

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Московский государственный университет экономики,
статистики и информатики

Гусынин А.Б
Минашкин В.Г.

**ТЕОРИЯ ВЫБОРОЧНЫХ
ОБСЛЕДОВАНИЙ**

Москва 2003

УДК 311
ББК 60.6
М 96

Гусынин А.Б, Минашкин В.Г. Теория выборочных обследований / Московский государственный университет экономики, статистики и информатики. – М., 2003 – 67с.

© Минашкин В.Г., 2003
© Гусынин А.Б., 2003
© Московский государственный университет экономики, статистики и информатики, 2003

Содержание

Введение.....	4
1. Сущность и задачи выборочного метода.....	6
1.1. Понятие о выборочном наблюдении	6
1.2. Выборочное обследование в системе методов несплошного статистического наблюдения	7
1.3. Основные причины применения выборочных обследований	12
1.4. Теоретические основы применения выборочного метода.....	14
2. Проектирование и проведение выборочного наблюдения.....	20
2.1. Программно-методологические вопросы и проектирование инструментария	20
2.2. Оценка исходной информации.....	34
2.3. Формирование выборочной совокупности	40
3. Обработка выборочных данных.....	48
3.1. Оценка пределов генеральных характеристик.....	48
3.2. Определение необходимого объема выборки.....	52
3.3. Распространение результатов выборочного наблюдения на генеральную совокупность	57
Список рекомендуемой литературы.....	67

Введение

Анализ социально-экономического развития, разработка стратегии бизнеса, формирование маркетинговой политики невозможны без соответствующего информационного обеспечения. Среди известных методов получения статистических данных особое место принадлежит выборочному методу, как весьма эффективному, оперативному и наименее трудоемкому. Выборочное исследование служит в настоящее время одним из важнейших источников статистической информации. Это во многом обусловлено переходом российской статистики на международные стандарты учета и отчетности.

Теория выборочных обследований базируется на общенаучных принципах исследования социально-экономических явлений, статистической методологии и компьютерных технологиях. Наиболее тесно теория выборочных обследований связана с основными положениями теории статистики, высшей математики, теории вероятностей и математической статистики. Она создает методологические основы информационного обеспечения в таких отраслях статистической науки и практики, как макроэкономическая, международная статистика, национальное счетоводство, статистика фирм, рынка товаров и услуг, прикладная социология и др. Это в равной степени можно отнести к макро- и микроэкономике, оценке бизнеса и недвижимости, финансовому анализу, техническому анализу товарных и финансовых рынков.

Необходимо отметить, что в настоящее время в Российской Федерации, в отличие от развитых зарубежных стран, возможности выборочного метода используются далеко не в полной степени. Это положение объясняется нехваткой специалистов в данной области, отсутствием детально проработанных методик (в особенности – для работы на микроуровне), недоверием региональных руководителей к результатам выборочных обследований. Улучшить ситуацию в данной области были призваны проведенные в апреле и сентябре 2000 года Российские региональные семинары в рамках проекта TACIS «Профессиональное обучение специалистов системы российской статистики» на тему «Несплошные методы статистического наблюдения», подготовленные с участием специалистов TES (Институт подготовки европейских статистиков, Люксембург) и INSEE

(центральный орган государственной статистики и институт подготовки кадров, Франция).

Данное учебно-практическое пособие дает возможность получить знания по теоретическим основам выборочного наблюдения, ознакомиться с методологическими проблемами подготовки и организации выборочных обследований и подходами к их решению, основными сферами практического применения выборочных обследований в современных условиях. Наряду с этим оно призвано помочь в приобретении умений и навыков решения основных вопросов проектирования, проведения, обработки и анализа результатов выборочных обследований на макроэкономическом, отраслевом и региональном уровнях, а также на уровне отдельных предприятий, учреждений и организаций всех форм собственности.

Учебное пособие предназначено для студентов специальностей «Статистика» и «Маркетинг», изучающих курс «Теория выборочных обследований». Кроме того, оно может быть рекомендовано студентам, специализирующимся в области менеджмента и права, экономики и финансов, а также слушателям факультета повышения квалификации преподавателей и специалистов.

1. Сущность и задачи выборочного метода

1.1. Понятие о выборочном наблюдении

Методология статистического исследования массовых социально–экономических явлений различает, как известно, два способа наблюдения в зависимости от полноты охвата объекта: сплошное и несплошное. Разновидностью несплошного наблюдения является выборочное, которое в условиях развития рыночных отношений в России находит все более широкое применение.

Под *выборочным* понимается такое несплошное наблюдение, при котором статистическому обследованию подвергаются единицы изучаемой совокупности, отобранные по специально разработанной схеме, базирующейся, как правило, на принципе случайности. Выборочное наблюдение ставит перед собой задачу – по обследуемой части дать характеристику всей совокупности единиц, при условии соблюдения всех правил и принципов поведения статистического наблюдения и научно организованной работы по отбору единиц.

Теория выборочных обследований базируется на принципе, исключающем субъективность и тенденциозность поведения отбора единиц для их последующего изучения по заранее разработанной программе. Она исходит из следующих положений:

- выбор той или иной конкретной единицы для обследования должен быть независим от воли, субъективного подхода лица, производящего отбор;
- выбор должен быть независим от значений изучаемых статистических характеристик (показателей), которыми обладают отдельные единицы совокупности;
- процесс отбора должен быть организован так, чтобы все единицы совокупности имели равные шансы, равные вероятности быть отобранными.¹

Предметом выборочного обследования является количественная сторона массовых социально–экономических явлений в неразрыв-

¹ В данном пособии не рассматриваются менее распространенные методы направленного отбора.

ной связи с качественной стороной в конкретных условиях места и времени, исследуемая на основе принципов выборочного метода.

Методология выборочного статистического исследования включает три элемента:

1. Выборочное наблюдение – на этом этапе собираются сведения о каждой единице совокупности, сформированные на основе того или иного способа отбора из общей совокупности единиц.
2. Сводка и группировка полученных на основе выборочного обследования материалов – на этом этапе сведения, относящиеся к каждой единице выборочной совокупности, объединяются в группы и подгруппы по заранее разработанной программе.
3. Собранные и сгруппированные данные анализируются и распространяются на всю изучаемую совокупность, при этом рассчитываются абсолютные, относительные, средние показатели и т.д.

Необходимым условием применения методологии выборочного наблюдения является решение таких вопросов, как предварительная оценка необходимого объема выборки и, когда это возможно, ее точности.

Такие вопросы применения статистической методологии, как формирование цели, выбор объекта, единицы наблюдения и др., решаются в соответствии с общими принципами планирования статистического наблюдения вне зависимости от того, сплошное оно или выборочное.

1.2. Выборочное обследование в системе методов несплошного статистического наблюдения

Сущность несплошного наблюдения состоит в том, что обследованию подлежит часть единиц статистической совокупности. Его важным преимуществом является возможность получения необходимой информации оперативно и с меньшими затратами всех видов ресурсов: финансовых, материальных, кадровых и др.

Основными видами несплошного статистического наблюдения являются: монографическое обследование, метод основного массива, выборочное наблюдение.

Монографическое обследование представляет собой детальное, глубокое изучение и описание отдельных, характерных в каком-либо отношении единиц изучаемой совокупности. Оно может при-

меняться, например, для изучения причин успешной деятельности отдельных предприятий в реализации своей продукции.

Метод основного массива заключается в том, что наблюдение охватывает наиболее существенные единицы, имеющие преобладающий удельный вес для данного признака (признаков). В частности, при изучении хода посевной кампании наблюдению подлежат категории хозяйств с наибольшим удельным весом посевных площадей, хотя их доля в общей численности хозяйств может быть сравнительно невелика.

Иногда к несплошному относят также анкетное, корреспондентское и цензовое наблюдения.

Анкетный способ предполагает сбор статистической информации с помощью специальных вопросников (анкет), рассылаемых определенному кругу лиц и распространяемых в СМИ на основе принципов добровольности и, как правило, анонимности. Его часто используют редакции газет и журналов для опроса читателей, он широко применяется в социологии, а также в ряде других наук и в практической деятельности.

При корреспондентском способе определенные лица (организации) на основе предварительной договоренности берут обязательство вести наблюдение за какими-либо явлениями, процессами и в заранее установленные сроки представлять результаты статистическим органам или иным организациям. До конца 20-х гг. нашего столетия таким методом исследовалась сельскохозяйственное производство в индивидуальных хозяйствах.

Цензовое наблюдение предписывает отбор единиц по некоторому определенному критерию (цензу). В качестве ценза может, например, использоваться некоторое заданное, критическое число работников, занятых на предприятии. Применение ценза при проведении несплошного обследования необходимо отличать от его использования при определении единицы отбора и объекта наблюдения. Также численность занятых может служить для определения малого предприятия в конкретной отрасли, например, в промышленности и строительстве – не более 200 человек. И если ставится задача обследования малых предприятий соответствующих отраслей, то предприятия с числом занятых более 200 человек не наблюдаются и не досчитываются, так как не относятся к объекту наблюдения.

Из рассмотрения приведенного выше перечня методов сплошного обследования видно, что он не претендует на классификацию, так как нельзя построить систему признаков по которому он составлен. Действительно, если отличительным признаком для основного массива и выборочного методов является принцип отбора единиц, то анкетный и корреспондентский методы различаются по способу сбора данных.

Разобраться в сложных взаимосвязях рассматриваемых методов и помочь получить целостное представление о них дает возможность вариант классификации, предложенной Г.И.Деевым. Он представляет собой иерархическую систему наиболее важных классифицирующих признаков. Если все методы в совокупности считать нулевым иерархическим уровнем, то к первому уровню классификации относятся следующие признаки:

- 1) *объектно-временные характеристики наблюдения*, определяемые особенностями исследуемого объекта и сроками наблюдения;
- 2) *тип отбора единиц наблюдения*, то есть метод формирования обследуемой части статистической совокупности;
- 3) *способ получения данных* – форма организации сбора первичной информации;
- 4) *метод распространения (досчета)* результатов сплошного наблюдения на всю обследуемую (генеральную) совокупность.

Рассмотрим дальнейшую детализацию первых четырех классифицирующих признаков. С позиции классификации важны следующие объектно-временные характеристики наблюдения:

- *количество наблюдаемых признаков* (число показателей, характеризующих объект наблюдения);
- *повторяемость наблюдения* во времени.

Имеет принципиальное значение – один или несколько признаков подлежат наблюдению. Большинство методологических разработок предполагает *одномерность наблюдения*, то есть относятся к случаю, когда наблюдается только один признак, на практике однако чаще приходится иметь дело с *многомерным наблюдением*, когда интерес представляет несколько признаков. Задача здесь усложняется – нет универсальных приемов формирования наблюдаемой части совокупности и приходится искать компромисс между точностями оценок интересующих показателей, что выливается в постановку

достаточно сложных задач многомерной оптимизации, решение которых возможно только с привлечением сложных вычислительных средств.

Большое значение при определении метода имеет также вопрос о повторяемости во времени. Дело в том, что методологические возможности несплошного наблюдения существенно зависят от наличия и полноты априорной (предшествующей моменту наблюдения) информации. Если речь идет об однократном наблюдении, то приходится обходиться минимумом такой информации, например, только численностью обследуемой совокупности или искать данные, каким-либо косвенным образом, связанные с объектом исследования и позволяющие дать хотя бы приблизительные оценки интересующих параметров. Если же ставится задача наблюдения, *повторяющегося во времени*, то возможно накопление и в течение определенного времени сохранение предшествующих данных, которые существенно расширяют арсенал методов, возможных для применения. В частности, возможны такие варианты наблюдения:

- *повторение отбора* единиц на каждом цикле сбора данных;
- *сохранение* их в течение достаточно долгого времени, то есть использование так называемой *статистической сети*.

Длительно существующие статистические сети, примером которых является сеть наблюдения за домашними хозяйствами, порождают проблему их постоянного обновления – *ротации*, которая представляет собой вариант, промежуточный между двумя последними.

Второй классифицирующий признак первого иерархического уровня – *тип отбора* – имеет два подуровня: *принцип отбора* и *структура отбора*. По принципу, на котором они основываются, отборы подразделяются на: *случайные*, *систематические (механические)* и *направленные (целевые)*. В первом случае требуется обеспечение определенной вероятности попадания в число наблюдаемых для каждой единицы совокупности.

Разновидностями случайного отбора является отбор по *JALES технологии* и *Пуассоновский отбор*, в основе которых лежит присвоение единицам наблюдения в качестве своеобразных номеров случайных чисел из интервала 0 – 1. Эти виды отбора применяются при периодических обследованиях с использованием регистров в

качестве источников базовой информации. С их помощью, в частности, более просто решается задача ротации.

Суть *систематического отбора* состоит в том, что составляется список единиц отбора и устанавливается шаг отбора, то есть интервал, через который единицы отбираются.

В основу *направленного (целевого) отбора* могут быть положены самые различные принципы. Примерами целевого отбора служат:

- *метод основного массива*;
- *цензовый метод*;
- *монографический метод*,

которые были рассмотрены ранее, а также:

- *юстированный отбор* – проводится таким образом, чтобы наблюдаемая часть совокупности обеспечивала наилучшее совпадение с базовыми данными (наблюдаемая часть «юстируется» – подгоняется под базовые данные);
- *экспертный отбор* – отбор, который проводит эксперт, полагаясь на свои знания и опыт, не всегда поддающиеся ясной формулировке.

Помимо используемых принципов виды отбора могут отличаться и своими *структурными характеристиками*, которые образуют следующие градации:

- *индивидуальный отбор*, когда единицами отбора являются единицы наблюдения;
- *групповой (серийный) отбор*, когда отбор проводится группами (сериями);
- *расслоенный отбор*, когда отбор проводится из предварительно расслоенной (разделенной на группы) совокупности;
- *одноступенчатый и многоступенчатый отбор* (в первом случае единицы отбора отбираются непосредственно для наблюдения, во втором – из серий, отобранных на предыдущей ступени, делаются последующие отборы, а наблюдению подвергаются единицы последнего отбора);
- *Однофазный и многофазный отбор*. Многофазный отбор отличается от многоступенчатого тем, что на каждой ступени наблюдаются все единицы, но по разным, обычно расширяющимся при переходах к последующим ступеням, программам.

Рассмотренный отбор подразделяется на отбор с *оптимальным* и *неоптимальным* расслоением. Под оптимальным расслоением или расслоением Чупрова–Неймана понимается такое размещение единиц наблюдения по группам, которое обеспечивает наименьшую ошибку репрезентативности при заданном объеме выборки и числе слоев. Особенно практически важным является оптимальное расслоение при многомерном наблюдении.

Наконец, последним классифицирующим признаком первого уровня является организационный *способ получения данных* от единиц наблюдения. Наиболее известными разновидностями его являются: *анкетный, корреспондентский, опросный, экспедиционный, дневниковых записей* и др.

В данном разделе подробно освещены аспекты классификации, общие для всех методов несплошного статистического наблюдения. Методологические и другие вопросы, относящиеся непосредственно к выборочным обследованиям, более детально будут рассмотрены в следующих разделах.

1.3. Основные причины применения выборочных обследований

К выборочному наблюдению статистика прибегает по различным причинам. На современном этапе появилось множество субъектов хозяйственной деятельности, которые характерны для рыночной экономики. Речь идет об акционерных обществах, малых и совместных предприятиях, фермерских хозяйствах и т.д. Сплошное обследование этих статистических совокупностей, состоящих из десятков и сотен тысяч единиц, потребовало бы огромных материальных, финансовых и иных затрат. Использование же выборочного исследования позволяет значительно сэкономить силы и средства, что имеет немаловажное значение.

Наряду с экономией ресурсов одной из причин превращения выборочного наблюдения в важнейший источник статистической информации является возможность получить необходимые данные значительно быстрее. Ведь при обследовании, скажем, 10% единиц совокупности будет затрачено гораздо меньше времени, а результаты могут быть представлены быстрее и будут более актуальными. Фактор времени важен для статистического исследования вообще, а

в условиях изменяющейся социально-экономической ситуации – в особенности.

Роль выборочного исследования в получении статистических данных возрастает в силу возможности и необходимости расширения программы наблюдения. Так как исследованию подвергается сравнительно небольшая часть всей совокупности, можно более широко и детально изучить отдельные единицы и их группы.

Проведение статистического наблюдения вообще требует соответствующего кадрового обеспечения. Сплошное обследование требует иногда слишком большого числа людей для его организации и проведения. Например, только сплошное обследование бюджетов домашних хозяйств потребовало бы привлечения, по различным оценкам, от 500 тыс. до 1 млн. специалистов-статистиков.

Обращение же к опыту выборочного наблюдения приводит к тому, что необходимый штат сотрудников становится меньше. Это позволяет привлекать более квалифицированных людей, снизить опасность появления субъективных ошибок, особенно при непосредственной регистрации фактов и достичь поставленных целей с помощью меньшего количества более компетентных специалистов-статистиков.

Следует также отметить, что на практике приходится сталкиваться со специфическими задачами изучения массовых процессов, которые могут быть решены только с использованием методологии выборки. К таким задачам относятся, например, исследование качества продукции, если продукция при этом уничтожается. На основе выборочного наблюдения изучается, например, качество электроламп, спичек, качество многих сплавов и так далее. Кроме того, в современных условиях развития внеэкономических связей России при наличии, в частности, большого числа импортируемых продуктов и непродовольственных товаров, контроль их качества обеспечивается также путем выборочного исследования.

Наконец, важным фактором превращения выборочного наблюдения в важнейший источник статистической информации является возможность его использования в целях уточнения и для разработки данных сплошного обследования. Выборочная разработка данных сплошного наблюдения связана с потребностью представления оперативных предварительных итогов обследования. Кроме того, при обобщении данных сплошного учета невозможно вести сплошную

разработку по всем сочетаниям рассматриваемых признаков. Она является сложной и дорогостоящей. В этих условиях выборочный метод позволяет получить необходимые сведения приемлемой точности, когда факторы времени и стоимости делают сплошную разработку нецелесообразной.

1.4. Теоретические основы применения выборочного метода

Основные положения выборочного метода, составляющие основу его практического применения, созданы научными исследованиями выдающихся ученых П.Л.Чебышев, А.М.Ляпунова, А.А.Маркова, Бернулли и др. В их работах содержится, в частности, доказательство теорем, в которых устанавливается факт приближения средних размеров большого числа наблюдения к некоторым постоянным величинам. Эти теоремы составляют фундамент методологии выборочных обследований.

Средняя и предельная ошибки выборки. Теоремы Чебышева–Ляпунова. В процессе проведения выборочного наблюдения, как и вообще при анализе данных любого обследования, статистика выделяет два вида ошибок: регистрации и репрезентативности. Ошибки регистрации могут иметь случайный (непреднамеренный) или систематический (тенденциозный) характер. Их можно избежать при правильной организации и проведении наблюдения. Ошибки репрезентативности органически присущи выборочному наблюдению и возникают в силу того, что выборочная совокупность не полностью воспроизводит генеральную. Избежать ошибок репрезентативности нельзя, однако, пользуясь методами теории вероятностей, основанными на использовании предельных теорем закона больших чисел, эти ошибки можно свести к минимальным значениям, границы которых устанавливаются с достаточно большой точностью.

Ошибка выборочного наблюдения – это разность между величиной параметра в генеральной совокупности и его величиной, вычисленной по результатам выборочного наблюдения. Для среднего значения ошибка будет определяться так:

$$\Delta_{\bar{x}} = |\bar{x} - \tilde{x}|,$$

$$\text{где } \bar{x} = \frac{\sum x_i}{N} \text{ – генеральная средняя;}$$

$$\tilde{x} = \frac{\sum x_i}{n} - \text{выборочная средняя.}$$

Величина $\Delta_{\tilde{x}}$ называется предельной ошибкой выборки.

Предельная ошибка выборки величина случайная. Исследованию закономерностей случайных ошибок посвящены предельные теоремы закона больших чисел. Наиболее популярно эти закономерности раскрыты в теоремах П.Л.Чебышева и А.М.Ляпунова.

Теорему П.Л.Чебышева применительно к рассматриваемому методу можно сформулировать следующим образом: при достаточно большом числе независимых наблюдений можно с вероятностью, близкой к единице (т.е. почти с достоверностью) утверждать что отклонение выборочной средней от генеральной будет сколь угодно малым. В теореме П.Л.Чебышева доказано, что величина ошибки не должна превышать $t\mu$. В свою очередь, величина μ , выражающее среднее квадратическое отклонение выборочной средней от генеральной средней, зависит от колеблемости признака в генеральной совокупности σ и числа отобранных единиц n . Эта зависимость выражается формулой

$$\mu = \frac{\sigma}{\sqrt{n}}$$

где μ зависит также и от способа формирования выборочной совокупности, о чем сказано ниже.

Величину $\sqrt{\frac{\sigma^2}{n}}$ называют **средней ошибкой выборки** и обозначают μ .

В этом выражении σ^2 – генеральная дисперсия, а n – объем выборочной совокупности.

Рассмотрим, как влияет на величину средней ошибки число отбираемых единиц n . Логически нетрудно убедиться, что при отборе большого числа единиц расхождения между средними будут меньше, т.е. существует обратная связь между средней ошибкой выборки и числом отобранных единиц. При этом здесь образуется не только просто обратная математическая зависимость, а такая зависимость,

которая показывает, что квадрат расхождения между средними обратно пропорционален числу отобранных единиц.

Далее посмотрим, как влияет колеблемость признака в генеральной совокупности на величину ошибки. Нетрудно доказать, что увеличение колеблемости признака влечет за собой увеличение среднего квадратического отклонения, а, следовательно, и ошибки. Если предположить, что все единицы будут иметь одинаковую величину признака, то среднее квадратичное станет равным нулю и ошибка выборки также исчезнет. Тогда нет необходимости применять выборку. Однако следует иметь в виду, что величина колеблемости признака в генеральной совокупности нам бывает не известна, поскольку неизвестны размеры единиц в ней. Мы можем рассчитывать лишь колеблемость признака в выборочной совокупности.

Соотношение между дисперсиями генеральной и выборочной совокупности выражается формулой

$$\sigma_x^2 = \sigma_{\bar{x}}^2 \cdot \frac{n}{n-1}$$

Поскольку величина $\frac{n}{n-1}$ при достаточно больших n близка к 1, можно приближенно считать, что выборочная дисперсия равна генеральной, т.е. $\sigma_{ген}^2 \approx \sigma_{выб}^2$.

Следовательно, средняя ошибка выборки показывает, какие возможны отклонения характеристик выборочной совокупности от соответствующих характеристик генеральной совокупности. Однако о величине этой ошибки можно судить с определенной вероятностью. На величину вероятности указывает множитель t .

А.М.Ляпунов доказал, что распределение выборочных средних (а, следовательно, и их отклонений от генеральной средней) при достаточно большом числе независимых наблюдений приближенно нормально при условии, что генеральная совокупность обладает конечной средней и ограниченной дисперсией.

Математически теорему Ляпунова можно записать так:

$$P\{|\bar{x} - \tilde{x}| \leq \Delta_x\} = \frac{1}{\sqrt{2\pi}} \cdot \int_{-t}^t e^{-\frac{t^2}{2}} dt = \Phi(t),$$

где $\Delta_x = t \sqrt{\frac{\sigma^2}{n}}$; $\pi \approx 3,14$ (математическая постоянная);

Δ_x – **предельная ошибка выборки**, которая дает возможность выяснить, в каких пределах находится величина генеральной средней.

Значения этого интеграла для различных значений коэффициента t вычислены и приводятся в специальных математических таблицах. В частности, при

$$t = 1 \quad \Phi(t) = 0,683; t = 1,5 \quad \Phi(t) = 0,866;$$

$$t = 2 \quad \Phi(t) = 0,954; t = 2,5 \quad \Phi(t) = 0,988;$$

$$t = 3 \quad \Phi(t) = 0,997; t = 3,5 \quad \Phi(t) = 0,999.$$

Поскольку t указывает на вероятность расхождения $|\tilde{x} - \bar{x}|$, т.е. на вероятность того, на какую величину генеральная средняя будет отличаться от выборочной средней, то это может быть прочитано так: с вероятностью 0,683 можно утверждать, что разность между выборочной и генеральной средними не превышает одной величины средней ошибки выборки. Другими словами, в 68,3% случаев ошибка репрезентативности не выйдет за пределы $\pm \mu$. С вероятностью 0,954 можно утверждать, что ошибка репрезентативности не превышает $\pm 2\mu$ (т.е. в 95% случаев). С вероятностью 0,997, т.е. довольно близкой к единице, можно ожидать, что разность между выборочной и генеральной средней не превзойдет трехкратной средней ошибки выборки и т.д.

Логическая связь здесь выглядит довольно ясно: чем больше пределы, в которых допускается возможная ошибка, тем с большей вероятностью судят о ее величине.

Для различных способов отбора предельная ошибка рассчитывается при проведении выборки по-разному.

Зная выборочную среднюю величину признака (\tilde{x}) и предельную ошибку выборки (Δ_x), можно определить границы (пределы), в которых заключена генеральная средняя:

$$\tilde{x} - \Delta_x \leq \bar{x} \leq \tilde{x} + \Delta_x \quad \text{или} \quad \tilde{x} - \bar{x} = \pm \Delta_x.$$

Теорема Бернулли. Теорема Бернулли была доказана раньше теоремы Чебышева–Ляпунова, но является лишь частным случаем последней. Она рассматривает ошибку выборки для альтернативного признака, т.е. признака, у которого возможны только два исхода: наличие признака (1) и отсутствие его (0).

Теорема Бернулли утверждает, что при достаточно большом объеме выборки вероятность расхождения между долей признака в выборочной совокупности (w) и долей признака в генеральной совокупности (p) будет стремиться к единице.

В математических символах выражение теоремы Бернулли будет иметь вид:

$$P[|w - p| \leq t\mu] \rightarrow 1,$$

т.е. с вероятностью, сколь угодно близкой к единице, можно утверждать, что при достаточно большом объеме выборки частость признака (выборочная доля) сколько угодно мало будет отличаться от доли признака (в генеральной совокупности).

В виду того,

В виду того, что вероятность расхождения между частостью и долей следует закону нормального распределения, эту вероятность можно найти по функции $F(t)$ в зависимости от задаваемой величины t .

Из теоремы Бернулли следует, что величина расхождения между долей признака в выборочной совокупности (частостью) и долей этого признака в генеральной совокупности зависит, так же как и в расхождениях средних, от средней ошибки выборки.

Поскольку $\mu = \frac{\sigma}{\sqrt{n}}$, а среднее квадратическое отклонение в генеральной совокупности для альтернативного признака равно \sqrt{pq} ,

где $q = 1 - p$, то средняя ошибка выборки для альтернативного признака должна быть найдена по формуле:

$$\mu = \sqrt{\frac{p \cdot q}{n}}.$$

Однако доля признака в выборочной совокупности нам не известна, мы вынуждены заменить ее через долю того же признака в

генеральной совокупности, т.е. принять $w \approx p$, а дисперсию альтернативного признака принять за $w(1-w)$, тогда средняя ошибка выборки выразится формулой

$$\mu = \sqrt{\frac{w(1-w)}{n}}.$$

Предельная величина разности между частостью и долей называется *предельной ошибкой выборки*. О величине предельной ошибки можно судить с некоторой вероятностью, которая зависит от множителя t , поскольку $\Delta_w = t\mu$.

Зная выборочную долю признака (w) и предельную ошибку выборки (Δ_w), можно определить границы, в которых заключена генеральная доля (p):

$$w - \Delta_w \leq p \leq w + \Delta_w$$

Уточнение формулы средней ошибки выборки. Если отбор единиц из генеральной совокупности произведен бесповторным способом, то в формулы средней ошибки выборки вносится поправка

$$\sqrt{1 - \frac{n}{N}}$$

где n – объем выборочной совокупности;
 N – объем генеральной совокупности.

2. Проектирование и проведение выборочного наблюдения

2.1. Программно-методологические вопросы и проектирование инструментария

Проектирование любого статистического наблюдения, в том числе и выборочного, начинается с формулировки его **цели**. При этом следует обратить внимание на два главных требования, предъявляемых к цели наблюдения. Во-первых, она должна быть *актуальной*. Проводить выборочные исследования по всем вопросам экономического и социального развития одновременно (не говоря о сплошных) просто невозможно. Поэтому из многообразия явлений и процессов необходимо выбрать те из них, от изучения которых в первую очередь зависит проводимых в настоящее время реформ.

Второе требование, которому должна удовлетворять цель выборочного наблюдения, заключается в ее *четкости и конкретности*. Сложность и многогранность социально-экономических явлений часто препятствует четкому и точному выражению целей и задач. В связи с этим организаторам выборочного обследования необходимо глубоко вникнуть в сущность подлежащих изучению проблем, чтобы затем возможно более лаконично и вместе с тем понятно определить цель наблюдения.

Одним из принципов проектирования статистического наблюдения является научно обоснованное определение того, что предполагается изучить, т.е. **объекта** наблюдения. Неточность в определении объекта неизбежно приведет к тому, что в процессе отбора, а затем собирания сведений отдельные части объекта могут быть не учтены, пропущены. В то же время обследованию могут подвергнуться части других объектов, не подлежащих наблюдению.

Определение объекта включает в себя указание его отличительных черт, позволяющих отделить данный объект от других, близких к нему по характеру объектов. Границы объекта могут устанавливаться на основе действующих классификаций, например, классификаций отраслей экономики, профессий рабочих и др.

Критериями по ограничению объекта могут быть и положения законов и постановлений руководящих органов. При этом наиболее удобны количественные критерии, с помощью которых удается ус-

тановить границы объекта более точно. Так, если объектом выборочного наблюдения являются малые предприятия, отделить их от других предприятий можно достаточно легко, зная ограничения по численности занятых. Например, в науке и научном обслуживании к малым относятся предприятия с численностью занятых до 100 чел., а в розничной торговле – до 15 чел.

Однако применение количественных критериев в целях определения границ возможно не всегда. Иногда ограничение объекта может быть достигнуто простым перечислением его составных частей, подлежащих учету. Такой способ использовался, в частности, при проектировании в 1994 г. выборочного обследования перевозочной деятельности предпринимателей, проводимого и в настоящее время. Оно проводится отделом статистики транспорта и связи Госкомстата с целью получения информации о развитии малого предпринимательства на автомобильном транспорте (см. пример 1.). Объект наблюдения отграничивался на основе выделения типа автомобилей (в данном случае обследовались только владельцы грузовых автомобилей).

Иногда для более четкого и единообразного понимания объекта в плане необходимо привести дополнительные разъяснения. Так, недостаточно узнать, что выборочному обследованию подлежат *научные учреждения*. Следует установить главные характерные особенности этой категории учреждений. При установлении типов и численности научных учреждений к ним относят те, которые систематически ведут научно-исследовательскую работу в той или иной области научных знаний по тематическому плану научных работ, имеют средства на проведение научных исследований и обеспечены кадрами научных работников. Наличие этих трех признаков является обязательным условием отнесения учреждений к научным.

С целью правильной организации статистического наблюдения важно обоснованно определить **единицу наблюдения** и **единицу отбора**. Решение данного вопроса тесно связано с выделением изучаемого объекта. Четкое отделение единицы наблюдения позволяет отграничить единицы друг от друга, исключить из наблюдения те единицы, которые не должны входить в состав обследуемой совокупности согласно определению объекта. В то же время четкость в

определении единицы дает возможность охватить наблюдением все отобранные единицы объекта.

Чем сложнее исследуемое с помощью выборочного метода явление, тем труднее определить объект и единицу наблюдения, но вместе с тем и более необходимо. Это касается в частности, изучения совокупностей, единицы которых отличаются высокой подвижностью, например, совокупностью посетителей выставок, пассажиров на транспорте и т.п. В этом и ряде других случаев бывает недостаточно общего краткого определения единицы наблюдения. В таких случаях целесообразно уточнять понятие единицы наблюдения с учетом соображений времени проведения наблюдения, территориальных ограничений и других факторов.

Иногда при планировании выборочных обследований заранее известно, что отдельные единицы совокупности существенно различаются между собой по различным признакам. При исследовании социально-экономических явлений, характеризующихся обычно высокой разнородностью, с этим приходится сталкиваться довольно часто. В таких случаях полезно наряду с общим определением единицы наблюдения давать понятие отдельных видов и разновидностей этих единиц.

Рассматривая принципы определения единицы наблюдения в социальной статистике, надо отметить, что в одном наблюдении может быть не одна, а несколько единиц. Широкий круг вопросов, ответы на которые планируется получить в ходе одного обследования, часто требует обращения к разным единицам наблюдения. При проектировании таких обследований необходимо дать четкое понятие каждой единицы наблюдения.

Наконец, важное требование к определению единицы наблюдения связано со спецификой организации выборочного обследования. В процессе его подготовки следует стремиться к обеспечению определенного соотношения между единицей отбора и единицей наблюдения. Оно заключается в том, что единица отбора не была меньше, чем единица наблюдения. В противном случае возникает угроза возникновения систематической ошибки. Соблюдение указанного принципа в значительной степени зависит от исходной информации о генеральной совокупности, которая находится в распоряжении организаторов и служит для составления основы выборки.

Одним из главных вопросов подготовки наблюдения является вопрос о его **программе**. От правильного решения этого вопроса зависит успех всего наблюдения. Содержание программы определяется прежде всего сущностью изучаемого объекта. Особенности изучаемых явлений и процессов предъявляют повышенные требования к предварительному анализу сущности и основных свойств объекта выборочного исследования. На содержание программы влияют также цель конкретного обследования, потребность в определенных данных о социальных явлениях для государственного управления, хозяйственного руководства и научных исследований.

Правильно составленная программа призвана обеспечить получение таких сведений, которые бы по возможности полно и всесторонне характеризовали изучаемое явление или процесс. Для этого программа должна быть достаточно подробной. Использование принципов выборочного метода в исследовании социальных явлений открывает широкие перспективы сбора информации по более обширной программе, чем при сплошном наблюдении.

В то же время имеется целый ряд факторов, ограничивающих ее объем. Наиболее важными из них являются: размер средств, выделенных на проведение обследования; требуемая срочность получения данных и др. В случае, организация выборочного наблюдения не дает возможность получения достоверных сведений по широкой программе, то лучше ограничить круг исследуемых вопросов, обеспечив сбор небольшого по объему, зато хорошего с точки зрения качества материала.

Для проектирования выборочного наблюдения социальных явлений это особенно важно, потому что программы многих проводимых в данной области обследований содержат большое число вопросов, от нескольких десятков до сотни и более вопросов. Так, программа одного из выборочных обследований социальной активности жителей микрорайона г. Москвы, включала 138 вопросов.

Наличие большого количества вопросов еще не гарантирует получения подобной характеристики изучаемого объекта. Оно может стать причиной отказа части опрашиваемых от участия в наблюдении или получения непродуманных, а, следовательно, недостаточно объективных ответов из-за отсутствия времени на осмысление каждого вопроса.

При разработке программы выборочных обследований необходимо иметь в виду, что она должна включать *наиболее существенные признаки*. Эти признаки должны по возможности непосредственно отражать изучаемое явление, его тип, основные черты, свойства. Второстепенные вопросы осложняют проведение наблюдения, а в дальнейшем – обработку и анализ данных.

В программу не следует также включать вопросы, по которым вряд ли можно рассчитывать на достоверные ответы. Это касается, наряду с другими, вопросов, содержание которых связано с прошедшими событиями из жизни опрашиваемых. Поэтому при составлении программы следует учитывать, что существует обратная зависимость между периодом времени, данные за который опрашиваемому необходимо восстановить в памяти, и вероятностью получения точных ответов.

Программа наблюдения находит свое отражение в его **формуляре** (см. пример 1). Основным принципом построения формуляра является как можно более краткая, ясная и четкая формулировка *вопросов*. Их понимание не должно вызывать затруднений и должно быть одинаковым как у опрашиваемых, так и у проводящих выборочное обследование работников.

В бланках выборочных обследований помимо собственно вопроса, часто необходимо приводить перечень *статистических подсказок*. Они, с одной стороны, уточняют вопрос, а с другой – ограничивают возможную интерпретацию ответов. Информация, полученная на основе вопросов с подсказками, требует меньших затрат времени, она проще в обработке.

Иногда в формуляр наблюдения нужно включать так называемые *контрольные вопросы*. Они помогают на основе арифметического и логического контроля выявить ошибки наблюдения и тем самым способствуют повышению достоверности его данных. Контрольный вопрос ставится с целью проверки правильности ответа на один из основных вопросов. В функции контрольных вопросов не входит получение какой-либо дополнительной информации.

При использовании контрольных вопросов необходимо иметь в виду, что в формуляре их не должно быть много по крайней мере по двум причинам. Во-первых, обилие контрольных вопросов «перегружает» формуляр сведениями, не несущими дополнительной ин-

формации, и повышает затраты на проведение обследования. Поэтому контроль целесообразно организовывать только по важнейшим для результатов наблюдения вопросам. Во-вторых, злоупотребление контрольными вопросами часто вызывает раздражение у опрашиваемых, которые, со своей точки зрения, вполне справедливо полагают, что не имеет смысла неоднократно спрашивать то, что не имеет особого смысла или не является необходимым.

Что касается *внешнего вида* формуляра, то стоит по возможности избегать больших по размеру бланков. С ними сложно работать, представляет трудность их хранение. Желательно также не пользоваться бланками, состоящими из нескольких не связанных между собой листов. Обычно вместо двух и более листов лучше воспользоваться обеими сторонами одного и того же листа или карточки (см. пример 2).

Составление *инструкции* для выборочного наблюдения социальных явлений в основном базируется на тех же принципах, что и для статистического наблюдения вообще. Она должна быть как можно более краткой, указания – ясными и четкими. Во многом сходны и подходы к решению других методологических и организационных вопросов статистического наблюдения (вопросы времени, сроков критического момента наблюдения; вопросы составления организационного плана, определения органов наблюдения и др.). Основываясь на общеметодологических принципах статистического наблюдения, эти вопросы в каждом конкретном случае должны решаться с учетом цели социального исследования, его объекта, форм, видов и способов собирания сведений, имеющихся ресурсов и т.п. (см. пример 3).

В заключение отметим, что как бы хорошо с точки зрения организаторов не был спроектирован инструментарий выборочного наблюдения, фактическая работа по проведению обследования не всегда будет совпадать с составленным заранее планом. Отсюда вытекает необходимость проверки статистического инструментария на основе *пробного обследования*, подобно тому, как это делается, например, перед проведением переписей населения. Пробное обследование поможет выяснить, составлен ли формуляр наилучшим образом для получения необходимой информации, установить, понятны ли вопросы и относящиеся к ним инструкции населению и работни-

кам, осуществляющим наблюдение, и в состоянии ли они следовать этим инструкциям.

Пример 1

ГОСУДАРСТВЕННЫЙ КОМИТЕТ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ПО СТАТИСТИКЕ
Отдел статистики транспорта и связи

Утверждена поста-
новлением Госком-
стата России
от 22.12.94 №265

Район _____
Республика в составе Российской федерации,
край, область _____

АНКЕТА

обследования перевозочной деятельности предпринимателей-
владельцев грузовых автомобилей

Уважаемый предприниматель!

В целях получения информации о развитии малого предприни-
мательства на автомобильном транспорте просим Вас ответить на
предлагаемые вопросы. Ответы на вопросы носят анонимный харак-
тер. Госкомстат России гарантирует использование полученной от
Вас информации только в сводном виде для оценки общего объема
автотранспортных перевозок в стране.

Период, за который сообщаются данные:

число, месяц начала обследования

число, месяц окончания обследования

(заполняются органами статистики)

1. Наличие грузовых автомобилей на дату начала обследования, штук

Наименование показателей	№ строки	Количество автомобилей (сумма граф с 9 по 4)	В том числе по грузоподъемности, кг:					
			до 1499	1500-4999	5000-6999	7000-9999	10000-14999	15000 и более
1	2	3	4	5	6	7	8	9
Всего в том числе технически исправных	10							
	11		×	×	×	×	×	×

2. Перевозки грузов за обследуемую неделю

№ ездки с грузом по порядку	Пункт погрузки	Пункт разгрузки	Расстояние перевозки (пробег автомобиля с грузом от пункта погрузки до пункта разгрузки), км	Перевезено грузов, тонн
1	2	3	4	5

3. Общий пробег автомобилей за обследуемую неделю (по показаниям спидометра), км _____

ЗАРАНЕЕ БЛАГОДАРИМ ВАС ЗА УЧАСТИЕ В ОБСЛЕДОВАНИИ

Краткие пояснения к заполнению

Данные по содержащимся в анкете показателям заполняются за обследуемую неделю. Период, за который сообщаются данные, указывает статорган.

1. Наличие автомобилей показывается на дату начала обследования.

2. В анкете отражается фактический вес перевезенных грузов (с учетом тары) по каждой езде автомобиля с грузом.

Пункт погрузки, разгрузки автомобиля – указывается город или другой населенный пункт, где грузы погрузались (выгружались). Если пункт погрузки (разгрузки) находится не на территории России, то кроме города или другого населенного пункта, где осуществлялась погрузка или разгрузка автомобиля, необходимо указать страну, например, Украина, Азербайджан, Эстония, Польша, Монголия, Германия и т.д.

Пример 2.

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ
--

Конфиденциальность гарантируется получателем информации

Код формы по ОКУД	Код отчитывающейся организации по ОКПО	Наименование отчитывающейся организации
1615003		Почтовый адрес

× **Линия отрыва**

СВЕДЕНИЯ О ГРУЗОВЫХ ПЕРЕВОЗКАХ ПРЕДПРИЯТИЯМИ АВТОМОБИЛЬНОГО ТРАНСПОРТА ЗА _____ 19__ г. (обследуемая неделя указывается органом статистики)
--

Представляют:	Сроки представления	Форма №2 – автотранспорт (выборочное обследование)
предприятия автомобильного транспорта, независимо от формы собственности, по списку, составленному органами государственной статистики – статистическому органу по месту, установленному статистическим органом республики, края, области, автономного округа	сроки определяются органами статистики	Утверждена постановлением Госкомстата России от 22.11.95. №184

1615003					
1	2	3	4	5	6
формы документа по ОКУД	Отчитывающейся организации по ОКПО	отрасли по ОКОНХ	территории по СОАТО	формы собственности по КФС	контрольной суммы (гр. 1–5)
Коды проставляет отчитывающаяся организация					

Пример 3

ГОСУДАРСТВЕННЫЙ КОМИТЕТ РОССИЙСКОЙ ФЕДЕРАЦИИ ПО СТАТИСТИКЕ (ГОСКОМСТАТ РОССИИ)

Утверждена постановлением
Госкомстата России
от 22.11.95 № 184

ИНСТРУКЦИЯ

по заполнению формы федерального государственного статистического наблюдения за грузовыми перевозками предприятиями автомобильного транспорта №2-автотранспорт (выборочное обследование)

Порядок заполнения формы

В форме выборочного обследования отражаются сведения о грузовых перевозках за обследуемую неделю, которая устанавливается территориальным комитетом государственной статистики.

1. Наличие автомобилей на начало обследуемой недели показывается независимо от их технического состояния (в исправном состоянии, в ремонте, в ожидании ремонта и т.п.), включая арендованные у других хозяйств.

2. Пребывание автомобилей в работе (в автомобиле-днях) определяется суммированием количества автомобилей, выпущенных из гаража на линию, за каждый день обследуемой недели, независимо от количества отработанных смен в течение суток. Автомобиль, воз-

вратившийся с линии в гараж ранее установленного заданием времени, независимо от причин возврата, учитывается бывшим в работе один день. При двухсменной работе – за день работы автомобиля считается работа обеих смен; при трехсменной работе, даже если, согласно суточному графику, последняя смена заканчивается после 24-х часов, по всем трем сменам учитывается один автомобиле-день работы. При выпуске автомобиля на линию в календарные сутки более одного раза при подсчете автомобиле-дней в работе такой автомобиль должен учитываться только один раз.

Если грузовой автомобиль, выехавший на линию, не выполнил никакой работы по причине отсутствия груза, то этот день также учитывается как автомобиле-день в работе.

При междугородных и международных перевозках и перевозках грузов по заказам, когда в связи с расстоянием маршрута и характером перевозки водитель выполняет задание более чем в течение одних суток, автомобиле-дни в работе определяются как количество дней нахождения в командировке, начиная со дня выхода на линию, включая день возвращения в гараж, за исключением целодневных простоев (по технической неисправности, бездорожью, ожиданию обратного груза, отдыха водителя и др.).

3. Перевозки грузов в тоннах и грузооборот в тонно-километрах, выполненные за неделю, отражаются по данным ежедневного учета перевозок грузов, выполненных на коммерческой основе (т.е. за плату), для предприятий, организаций, физических лиц – заказчиков автотранспорта. При этом отсутствие в отчитываемом предприятии учета перевозок грузов по товарно-транспортным накладным не может служить основанием для незаполнения II раздела формы. На период проведения обследования могут быть использованы другие способы учета перевезенных грузов, например, записи водителем в путевом листе о количестве перевезенных грузов по каждой езде автомобиля. Грузооборот определяется путем умножения фактически перевезенного за отдельные езды груза на расстояние перевозки и суммированием полученных произведений.

Пример расчета:

За день было выполнено 3 ездки.

1-ая ездка: 3 тонны на расстояние 20 км.

2-ая ездка: 4 тонны на расстояние 30 км.

3-я ездка: 3 тонны на расстояние 10 км.

В этом случае перевозки грузов составят 10 тонн.

Грузооборот: $(3т \times 20км) + (4т \times 30км) +$

$+(3т \times 10км) = 210$ тонно-километров.

4. В группе каменный уголь и кокс отражается уголь каменный, бурый, кокс всякий; нефть и нефтепродукты – нефть сырая, бензин, керосин, масла и смазки, топливо дизельное, мазут, асфальт, битум и гудрон, прочие нефтепродукты светлые и темные, газы сжиженные; металлы черные и цветные – чугун, сталь в слитках, ферросплавы, заготовки стальные всякие, прокат и лом черных и цветных металлов, прочие черные и цветные металлы; строительные материалы и цемент – минерально-строительные материалы природные, зола, шлаки строительные, стеновые и кровельные материалы, строительные материалы и изделия промышленного производства; лесные грузы – лес круглый, пиломатериалы, дрова; зерновые грузы – пшеница, рожь, овес, ячмень, кукуруза в зерне и початках, рис в зерне, прочие зерновые.

К опасным грузам относятся любые вещества, материалы, изделия, отходы производственной или иной деятельности, которые в силу присущих им свойств и особенностей могут при перевозке создавать угрозу для жизни и здоровья людей, нанести вред окружающей природной среде, привести к повреждению или уничтожению материальных ценностей.

Деятельность, связанная с перевозками опасных грузов автомобильным транспортом осуществляется по лицензии «ОГ», полученной в органах Российской транспортной инспекции.

5. Общий пробег грузовых автомобилей определяется путем суммирования ежедневных пробегов всех автомобилей, работающих на линии в течение обследуемой недели.

Величина пробега за день определяется по показаниям спидометра.

Пробег с грузом определяется суммированием пробега грузовых автомобилей по каждой езде между пунктами погрузки и выгрузки, независимо от количества перевезенного груза.

6. К городским перевозкам относятся перевозки, осуществляемые на маршрутах в пределах черты города (другого населенного пункта).

К пригородным – перевозки, выполняемые за пределы черты города (другого населенного пункта) на расстояние до 50 км включительно.

К междугородным – перевозки, выполняемые за пределы черты города (другого населенного пункта) на расстояние более 50 километров. В междугородные перевозки в пределах Российской Федерации включаются внутриобластные, которые осуществляются на маршрутах, проходящих по территории двух и более областей (краев, республик в составе России).

Между Россией и государствами-участниками СНГ – перевозки, выполняемые на маршрутах, проходящих между пунктами отправления и назначения по территории России и государств-участников СНГ, к которым относятся Азербайджан, Армения, Белоруссия, Грузия, Казахстан, Кыргызстан, Молдова, Туркменистан, Таджикистан, Узбекистан, Украина. Здесь же отражаются перевозки грузов автотранспортом России между пунктами назначения по территории государств-участников СНГ без захода на территорию России (например, Украина – Белоруссия).

Между Россией и странами дальнего зарубежья – перевозки, выполняемые между пунктами отправления и назначения по территории России и стран дальнего зарубежья (Польша, Германия, Монголия и т.д.). К странам дальнего зарубежья относятся также государства Прибалтики: Латвия, Литва, Эстония. Здесь же отражаются перевозки грузов автотранспортом России между пунктами отправления и назначения по территории стран дальнего зарубежья без захода на территорию России (например, Латвия – Литва, или Польша – Германия).

Управление статистики
транспорта и связи

2.2. Оценка исходной информации

Качество результатов статистического исследования в значительной степени зависит от уровня подготовки его первой стадии – процесса наблюдения. Среди проблем, связанных с подготовкой выборочного наблюдения, важное место занимает проблема получения **исходной информации** об объекте исследования.

Выделение объекта исследования и сбор необходимой информации о нем являются предварительными условиями его статистического изучения. Без возможно более подробной характеристики объекта на основе собранной о нем информации трудно правильно спроектировать выборочное наблюдение и обеспечить репрезентативность его данных.

Сущность проблемы исходной информации заключается в том, что при планировании выборочного наблюдения такая информация либо совсем отсутствует, либо в случае исследования нового объекта является недостаточно точной и полной, либо является настолько разнородной, что не позволяет составить целостного представления об исследуемом объекте.

Причиной отсутствия информации может быть исследование какого-либо нового объекта, ранее никогда не подвергавшегося статистическому изучению. Основной **причиной неточности и неполноты исходной информации** являются погрешности, возникающие в процессе ее получения. Под этими погрешностями подразумеваются ошибки, характерные для любого статистического наблюдения, в том числе и выборочного. Неполнота данных об изучаемом объекте может быть связана также с изменением методики отбора, в частности с образованием выборочной совокупности по признаку, который отсутствует в исходной информации. Такая информация может не удовлетворять организаторов наблюдения и потому, что расширение задач социально-экономических исследований требует более подробной характеристики совокупности, из которой затем предстоит произвести отбор.

Причины разнородности исходной информации следует искать, прежде всего в разнообразии ее источников. Источниками первичной информации для подготовки выборочного наблюдения могут служить как материалы органов государственной статистики (отчетность, данные единовременного наблюдения и др.), так и иные источники (результаты социологических исследований, кадровая ин-

формация, сведения, полученные из избирательных комиссий и т.п.). Обращение организаторов обследований к дополнительным данным, степень достоверности которых уступает точности данных государственной статистики, вызвано прежде всего тем, что в ряде случаев у органов государственной статистики нет необходимых данных для формирования представительной выборочной совокупности.

Наличие исходной информации определяет работу по образованию выборочной совокупности и, как предварительный этап, – построение **основы**, то есть перечня единиц отбора. Составления такого перечня на практике часто бывает одной из самых главных задач подготовки выборочного наблюдения.

К настоящему времени сложились представления об основе, которая могла бы удовлетворить принципам научной организации выборки. Они нашли свое отражение в ряде требований, предъявляемых к основе. В общем виде эти требования сводятся к следующему: основа выборки должна быть

1. полной;
2. точной;
3. не содержащей дублирования;
4. соответствующей задачам конкретного обследования;
5. как можно более оперативной;
6. удобной в работе.

Под **полнотой** подразумевается представленность всех типов или групп данной генеральной совокупности в основе выборки. Если в качестве исходной информации для формирования выборочной совокупности используется список, то в нем обязательно присутствие всех единиц, образующих те или иные типы (группы). В противном случае исходная информация является неполной, что может привести к серьезным ошибкам, особенно когда не включенные в выборочную совокупность единицы представляют достаточно многочисленные группы в генеральной совокупности.

Известным примером в этом отношении служит неверный прогноз исхода президентских выборов в США 1936 года, данный “Литэри дайджест” на базе почтового опроса более 2 млн. американцев. Неправильное мнение о возможном победителе объяснялось наличием исходной информации, не обеспечившей полного представительства всей генеральной совокупности. В телефонных книгах, которые использовались для организации отбора, были представлены лишь

наиболее обеспеченные слои американского населения, в частности домовладельцы. Поскольку обеспеченные слои составляют меньшую часть всего населения США, то распространение мнения этой части на страну в целом оказалось ошибочным¹.

Соблюдение второго из указанных выше общих требований предполагает, что сведения о перечисляемых в основе единицах и используемые при организации процесса отбора, не должны содержать каких-либо **неточностей**. Подобные неточности встречаются в списках избирателей. В них иногда отсутствуют лица, вновь прибывшие в данный населенный пункт, или, наоборот, остаются лица, изменившие свое место жительства, умершие и т.п.

Если те или иные единицы генеральной совокупности включаются в основу более одного раза, это может привести к их повторению в выборке и, как следствие, – к двойному счету или **дублированию**. Он возникает, например, в том случае, когда человек переезжает из одного района в другой и заносится в новый список раньше, чем исключается из старого. Причиной погрешностей такого рода являются недостатки в организации учета механического движения населения. Поэтому к использованию данных такого учета как исходной информации при подготовке выборочного наблюдения следует подходить весьма осторожно.

Опасность дублирования существует и при работе по составлению основы на базе материалов переписей населения. Величина двойного счета населения, находящегося во время переписи вне постоянного места жительства, в масштабах страны может достигать значительных размеров. Однако большой опыт, накопленный статистикой в решении данной проблемы, позволяет рассматривать материалы переписей в качестве вполне приемлемой основы при планировании как общесоюзных, так и региональных выборочных исследований.

Соответствие исходной информации задачам проводимого исследования означает, что она должна быть пригодна для реализации конкретных целей изучения ограниченного во времени и пространстве объекта. Одна и та же информация, адекватная для решения одних задач, может быть неадекватной для других.

¹ Рабочая книга социолога. М., 1983, с.204

Так, полный список работающих на предприятии может стать хорошей основой для выборочного исследования производительности труда на данном предприятии. Но если целью является производительность труда не всех работников, а их части (скажем, работников цеха, бригады и т.п.), то полный список в «чистом виде» здесь не пригоден. Он может служить лишь для образования новой основы, относящейся к указанному подразделению или группе.

В то же время если имеющаяся информация не удовлетворяет целям выборочного исследования по причине неполноты (но не вследствие ошибок, о чем говорилось выше), то она может применяться только для получения выборки из той части генеральной совокупности, которая представлена полностью соответственно и результаты наблюдения будут правомерно распространять лишь на эту часть совокупности. Список сотрудников какого-либо подразделения коммерческого банка не может быть основой для исследования, цель которого – составить суждение обо всех сотрудниках. При этом не имеет значения даже незначительный удельный вес остальных частей, по которым сведения отсутствуют. В данном случае цель выборочного наблюдения достигается путем расширения основы с помощью дополнительной информации.

Существенное значение для формирования выборочной совокупности имеет **оперативность** информации, содержащейся в основе. Основа выборки, которая в момент ее построения удовлетворяет требованиям полноты, точности, целям и задачам конкретного исследования, может уже не быть таковой ко времени, когда ею нужно будет пользоваться. Использование устаревшей информации о составе, структуре, основных характеристиках объекта, ведет к тому, что результаты выборочного наблюдения характеризуют совокупность, существовавшую в прошлом. Понятно, что чем больше период, отделяющий время применения данных о генеральной совокупности от времени, на которое эти данные составлялись, тем меньше вероятность получить характеристику современного состояния изучаемого объекта. После истечения определенного периода использование информации, относящейся к прошлому подлежащего изучению явления, вообще теряет всякий смысл. С учетом динамичности различных социально-экономических явлений такой период составляет часто от нескольких дней до нескольких месяцев.

Наконец, одним из условий, влияющих на процесс отбора, является **удобство работы с исходной информацией**. Оно предполагает возможность быстро получить сведения о генеральной совокупности; с полной определенностью идентифицировать их; систематизировать собранные данные (пронумеровать, ранжировать, объединить в группы и т.д.). Удобство работы предполагает также наличие определенного порядка хранения и обращения информации.

Говоря о требованиях, предъявляемых к исходной социальной информации, нельзя не отметить некоторую условность разграничения этих требований. Если теоретически существует специфика, позволяющая отделить одно требование от другого, то в практической деятельности сделать это сложнее.

В реальной действительности требование, например, точности является достаточно общим, так как включает в себя и полноту, и оперативность, и ряд других требований, рассмотренных ранее. Данные о генеральной совокупности, удовлетворяющие одним и не отвечающие другим, используются сравнительно редко. Это связано со значением исходной информации для подготовки выборочного наблюдения. Ведь даже при самом высоком уровне организации процесса наблюдения и самой совершенной обработке его материалов практическая ценность результатов будет зависеть от того, насколько точно выборочная совокупность отражает генеральную. Решить эту задачу можно только при условии, что для отбора будет использована исходная информация, отвечающая не одному или нескольким, а по возможности, всем требованиям, о которых шла речь выше. При этом трудно переоценить роль тщательного предварительного анализа данных, которые могут быть положены в основу выборки.

Несмотря на существование четких теоретических представлений об «идеальной» основе, в практической деятельности исследователи сталкиваются с проблемами получения такой информации для построения такой основы. Сложность социально-экономических явлений и задач их статистического изучения приводит к использованию в ряде случаев исходных данных, имеющих отклонения от рассмотренных требований. При этом следует иметь в виду неодинаковое влияние недостатков основы на последующие этапы выборочного наблюдения.

Неточность основы, если она будет характерна для отобранных единиц, может быть автоматически обнаружена при анализе выборочной совокупности и в ходе самого обследования. В этом случае отдельные неточности в сведениях о попавших в выборку единицах генеральной совокупности могут быть исправлены. И хотя погрешности исходной информации, безусловно, отразятся на точности результатов наблюдения, их влияние поддается контролю. В процессе корректировки данных оно может быть уменьшено.

В отличие от неточности неполнота основы при проведении обследования не обнаруживается. Следовательно, в той мере, в которой основа является неполной, население или какой-либо другой изучаемый объект не будут подвергаться обследованию. Данный недостаток является тем более серьезным, что часто причина неполноты заключается в сознательном исключении из основы единиц, обладающих особенностями по сравнению с остальными. Поэтому такие единицы будут представлены недостаточно в выборочной совокупности. Это ставит под сомнение правомерность распространения полученных результатов на всю совокупность социальных явлений, которые, как правило, характеризуются неоднородностью.

К недостаточной репрезентативности выборки приводит и повторный счет, содержащийся в исходной информации. Единицы отбора, дважды учтенные в ней, имеют соответственно больше шансов попасть в выборочную совокупность. Однако в то время как при анализе основы определить или исправить ее неполноту нельзя, повторный счет исправить и устранить можно, хотя это требует дополнительных расчетов и проверки качества исходной информации.

Несоответствие основы целям и задачам обследования обычно выясняется еще до его начала. Предварительный анализ исходной информации позволяет сопоставить ее с задачами, которые планируется решить. Выявленные расхождения могут быть устранены путем создания дополнительной основы. При этом вспомогательная основа либо создается из имеющейся выделением частной основы, либо расширением существующей за счет дополнительной информации строится новая основа.

Поскольку на практике каждая основа в той или иной мере обладает перечисленными выше недостатками, при подготовке обследования важно тщательно рассмотреть любую основу, которую возможно использовать, а при необходимости – изучить возможность

уточнения имеющихся данных. Только после этого можно приступить к процессу формирования выборочной совокупности.

2.3. Формирование выборочной совокупности

Достоверность и обоснованность рассчитанных по выборочным данным характеристик в значительной степени определяются репрезентативностью выборочной совокупности, которая, в свою очередь, зависит от способа отбора единиц из генеральной совокупности. В каждом конкретном случае в зависимости от целого ряда условий, а именно сущности исследуемого явления, объема совокупности, вариации и распределения наблюдаемых признаков, материальных и трудовых ресурсов, выбирают наиболее предпочтительную систему организации отбора, которая определяется видом, методом и способом отбора.

По виду различают индивидуальный, групповой и комбинированный отбор. При индивидуальном отборе в выборочную совокупность отбираются отдельные единицы генеральной совокупности, при групповом отборе - группы единиц, а комбинированный отбор предполагает сочетание группового и индивидуального отбора.

Метод отбора определяет возможность продолжения участия отобранной единицы в процедуре отбора.

Бесповторным или отбором без возвращения называется такой отбор, при котором попавшая в выборку единица не возвращается в совокупность, из которой осуществляется дальнейший отбор.

При повторном отборе или отборе с возвращением попавшая в выборку единица после регистрации наблюдаемых признаков возвращается в исходную (генеральную) совокупность для участия в дальнейшей процедуре отбора. При этом методе отбора объем генеральной совокупности на всем протяжении процедуры выборки остается неизменным, что обуславливает постоянную вероятность попадания в выборку для всех единиц совокупности.

Повторный метод отбора применяется в тех случаях, когда характер исследуемого явления предполагает возможность повторной регистрации единиц. Такая возможность в практике работы органов государственной статистики, прежде всего, может иметь место в изучении масштабов теневой экономики, а также в различных вы-

борочных обследованиях населения. К повторному также приравнивается отбор из совокупности, границы которой не определены. В подобных случаях значения отобранных единиц рассматриваются как гипотетические величины, не исключающие возможности многократного повторения. Однако, следует иметь в виду, что ошибки при таком отборе всегда получаются значительно выше.

Способ отбора определяет конкретный механизм или процедуру выборки единиц из генеральной совокупности. В мировой практике выборочных обследований наибольшее распространение получили следующие выборки:

- собственно-случайная (простая случайная);
- механическая (систематический отбор);
- типическая (расслоенная, стратифицированная);
- серийная (гнездовая);
- комбинированная;

Собственно-случайная или простая случайная выборка заключается в отборе единиц из генеральной совокупности в случайном порядке без каких-либо элементов системности. Однако прежде чем производить собственно-случайный отбор, необходимо убедиться, что все без исключения единицы генеральной совокупности имеют абсолютно равные шансы попадания в выборку, в списках или перечне отсутствуют пропуски, игнорирования отдельных единиц и т.п. Следует также установить четкие границы генеральной совокупности таким образом, чтобы включение или не включение в нее отдельных единиц не вызывало сомнений.

Для реализации простой случайной выборки могут применяться различные алгоритмы отбора. Рассмотрим те из них, которые имеют наибольшее распространение в зарубежной практике статистических обследований.

Алгоритм случайной сортировки.

- каждой единице совокупности присваивается случайное число u на равномерно распределенном отрезке от 0 до 1 (например, с помощью генератора случайных чисел);
- единицы совокупности ранжируются в порядке убывания (возрастания) u ;
- отбираются n первых единиц.

Данный алгоритм, в частности, использовался при разработке Госкомстатом России Методологии координации и ротации вы-

борочных статистических наблюдений за малыми предприятиями и реализован в соответствующем программном комплексе. Недостаток данного алгоритма заключается в том, что при его реализации требуется относительно долгая процедура сортировки в случае большого объема N .

- возможно получение нескольких выборок без перекрытия или с контролируемым перекрытием.

Алгоритм прямой реализации.

- все единицы генеральной совокупности, расположенные в случайном порядке или ранжированные по какому-либо признаку нумеруются от 1 до N ;
- образуют n чисел в интервале от 1 до N в соответствии с законом равномерного распределения;
- отбираются единицы, по номеру соответствующие образованным числам.

Преимущество данного алгоритма заключается в том, что он требует только одного считывания файла (в случае ранжирования n случайных чисел).

Данный алгоритм можно реализовать на основе таблиц случайных чисел, приведенных в приложениях многих учебников по теории статистики.

Алгоритм метода отбора-отказа.

- последовательно образуют случайные числа $u_1, u_2 \dots$ в соответствии с законом равномерного распределения в интервале от 0 до 1;
- если $u_1 < n/N$, то первая из списка единица извлекается в выборку, в противном случае - отбрасывается;

- для последующих единиц: если $u_{k+1} < \frac{n - n_k}{N - k}$, то единица $(k+1)$ извлекается, в противном случае - отбрасывается;
- где k - число просмотренных единиц;

n_k - число отобранных единиц среди первых k просмотренных единиц.

- процедура заканчивается, когда $n_k = n$.

Данный алгоритм является наиболее экономичным, так как при его реализации происходит только одно считывание файла без

каких-либо предварительных сортировок единиц генеральной совокупности или случайных чисел.

После проведения отбора для определения возможных границ генеральных характеристик рассчитываются средняя и предельная ошибки выборки. Средняя ошибка повторной собственно-случайной выборки определяется по формуле

$$\mu = \frac{\sigma}{\sqrt{n}}$$

При расчете средней ошибки собственно-случайной бесповторной выборки необходимо учитывать поправку на бесповторность отбора:

$$\mu = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$$

Соответственно уменьшится и предельная ошибка выборки, что вызовет сужение границ генеральной средней. Особенно ощутимо влияние поправки на бесповторность отбора при большом проценте выборки.

Механическая выборка (систематический отбор) применяется в случаях, когда генеральная совокупность каким-либо образом упорядочена, т.е. имеется определенная последовательность в расположении единиц.

Для проведения механической выборки устанавливается пропорция отбора, которая определяется соотношением объемов выборочной и генеральной совокупностей.

Генеральную совокупность при механическом отборе целесообразно ранжировать или упорядочить по величине изучаемого или коррелирующего с ним признака, что позволит повысить репрезентативность выборки. Однако в этом случае возрастает опасность систематической ошибки, связанной с занижением значения изучаемого признака (если из каждого интервала регистрируется первое значение) или его завышением (если из каждого интервала регистрируется последнее значение). Поэтому целесообразно отбор начинать с середины первого интервала, например при 5%-ной выборке

отобрать 10, 30, 50, 70 и с таким же интервалом последующие единицы.

Опасность систематической ошибки при механической выборке также может появиться вследствие случайного совпадения выбранного интервала и циклических закономерностей в расположении единиц генеральной совокупности. Именно по этой причине, а также в связи с трудоемкостью предварительной сортировки, данный способ отбора не имеет большого распространения в практике статистических работ.

Для определения средней ошибки механической выборки используется формула средней ошибки при собственно-случайном бесповторном отборе.

Типический отбор или стратифицированный отбор используется в тех случаях, когда все единицы генеральной совокупности можно разбить на несколько типических групп. При обследованиях населения такими группами могут быть, например, районы, социальные, возрастные или образовательные группы, при обследовании предприятий - отрасль и подотрасль, форма собственности и т.п. Типический отбор предполагает выборку единиц из каждой типической группы собственно-случайным или механическим способом. Поскольку в выборочную совокупность в той или иной пропорции обязательно попадают представители всех групп, типизация генеральной совокупности позволяет исключить влияние межгрупповой дисперсии на среднюю ошибку выборки, которая в этом случае определяется только внутригрупповой вариацией.

Данный способ отбора при переводе реального сектора на несплошные методы наблюдения, безусловно, должен быть основным, так как он позволяет учесть региональные и отраслевые классификации субъектов финансово-хозяйственной деятельности.

Отбор единиц в типическую выборку может быть организован либо пропорционально объему типических групп, либо пропорционально внутригрупповой дифференциации признака (оптимальное размещение Чупрова - Неймана).

При выборке, пропорциональной объему типических групп, число единиц, подлежащих отбору из каждой группы, определяются следующим образом:

$$n_i = n \frac{N_i}{N},$$

где N_i - объем i -й группы;

n_i - объем выборки из i -й группы.

Средняя ошибка такой выборки находится по формулам:

$$\mu = \sqrt{\frac{\overline{\sigma}^2}{n}} \quad (\text{повторный отбор}),$$

$$\mu = \sqrt{\frac{\overline{\sigma}^2}{n} \left(1 - \frac{n}{N}\right)} \quad (\text{бесповторный отбор}),$$

где $\overline{\sigma}^2$ - средняя из внутригрупповых дисперсий.

При выборке, пропорциональной дифференциации признака, число наблюдений по каждой группе рассчитывается по формуле

$$n_i = n \frac{\sigma_i N_i}{\sum \sigma_i N_i},$$

где σ_i - среднее квадратическое отклонение признака в i -й группе.

Средняя ошибка такого отбора определяется следующим образом:

$$\mu = \frac{1}{N} \sqrt{\sum \frac{\sigma_i^2 N_i^2}{n_i}} \quad (\text{повторный отбор}),$$

$$\mu = \frac{1}{N} \sqrt{\sum \frac{\sigma_i^2 N_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right)} \quad (\text{бесповторный отбор}).$$

бор).

Отбор, пропорциональный дифференциации признака, дает лучшие результаты, однако на практике его применение затруднено вследствие трудности получения сведений о вариации до проведения выборочного наблюдения. Возможные варианты решения этой проблемы сводятся к использованию оценки вариации изучаемого признака, основанной на:

- результатах предшествующих обследований данной совокупности;
- результатах обследования переменной Z , коррелированной с изучаемой переменной;
- экспертных оценках;
- результатах обследования аналогичных страт по другим регионам.

Серийный отбор или гнездовой способ отбора удобен в тех случаях, когда единицы совокупности объединены в небольшие группы или серии. Сущность серийной выборки заключается в собственно-случайном либо механическом отборе серий, внутри которых производится сплошное обследование единиц.

Поскольку внутри групп (серий) обследуются все без исключения единицы, средняя ошибка серийной выборки (при отборе равновеликих серий) зависит от величины только межгрупповой (межсерийной) дисперсии и определяется по следующим формулам:

$$\mu = \sqrt{\frac{\delta^2}{r}} \quad (\text{повторный отбор}),$$

$$\mu = \sqrt{\frac{\delta^2}{r} \left(1 - \frac{r}{R}\right)} \quad (\text{бесповторный отбор}),$$

где r - число отобранных серий;

R - общее число серий.

Межгрупповую дисперсию вычисляют следующим образом:

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2}{r},$$

где \bar{x}_i - средняя i-й серии;

\bar{x} - общая средняя по всей выборочной совокупности.

Основное применение серийного отбора - технологическая статистика и контроль качества продукции (упаковки готовой продукции, упаковки товара и т.п.). В практике работы органов государственной статистики применение гнездового отбора может быть оправдано при выборочном обследовании инвестиций физических лиц в индивидуальное строительство дачных и садовых домов, где в качестве серий будут рассматриваться садовые товарищества и дачные кооперативы.

В практике статистических обследований помимо рассмотренных выше способов отбора применяется и их комбинация. Так, например, можно комбинировать типическую и серийную выборки, когда серии отбираются в установленном порядке из нескольких типических групп. Возможна также комбинация серийного и собственно-случайного отборов, при которой отдельные единицы отбираются внутри серии в собственно-случайном порядке. Ошибка такой выборки определяется ступенчатостью отбора.

Многоступенчатым называется отбор, при котором из генеральной совокупности сначала извлекаются укрупненные группы, потом - более мелкие и так до тех пор, пока не будут отобраны те единицы, которые подвергаются обследованию.

В отличие от многоступенчатой **многофазная выборка** предполагает сохранение одной и той же единицы отбора на всех этапах его проведения, при этом отобранные на каждой стадии единицы подвергаются обследованию (на каждой последующей стадии отбора программа обследования расширяется).

3. Обработка выборочных данных

3.1. Оценка пределов генеральных характеристик

При оценке точности выборочных обследований используются показатели средней и предельной ошибок выборки. **Средняя ошибка** выборки (μ) показывает, какие возможные отклонения характеристик выборочной совокупности от соответствующих характеристик генеральной совокупности. **Предельная ошибка** выборки (Δ) гарантирует с вероятностью p и коэффициентом доверия t , что эти отклонения не превысят величины $t\mu$.

Эти два вида ошибок связаны следующим отношением:

$$\Delta = t\mu$$

где Δ – предельная ошибки выборки;

μ – средняя ошибки выборки;

t – коэффициент доверия, определяемых в зависимости от уровня вероятности p .

Ниже приведены некоторые значения t :

Вероятность, p_i	0,683	0,866	0,954	0,988	0,997	0,999
Значение t	1,0	1,5	2,0	2,5	3,0	3,5

Предельные ошибки выборки рассчитываются по-разному в зависимости от способа формирования выборочной совокупности. Формулы для определения Δ для наиболее часто используемых на практике способов отбора содержатся в таблице:

Метод отбора Вид выборки	Повторный		Бесповторный	
	для средней	для доли	для средней	для доли
1. Собственно-случайная и механическая	$t\sqrt{\frac{\sigma^2}{n}}$	$t\sqrt{\frac{w(1-w)}{n}}$	$t\sqrt{\frac{\sigma^2}{n}(1-\frac{n}{N})}$	$t\sqrt{\frac{w(1-w)}{n}(1-\frac{n}{N})}$
2. Типическая (при пропорциональном объему групп отборе)	$t\sqrt{\frac{\bar{\sigma}_i^2}{n}}$	$t\sqrt{\frac{w_i(1-w_i)}{n}}$	$t\sqrt{\frac{\bar{\sigma}_i^2}{n}(1-\frac{n}{N})}$	$t\sqrt{\frac{w_i(1-w_i)}{n}(1-\frac{n}{N})}$
3. Серийная (гнездовая)	$t\sqrt{\frac{\delta_x^2}{r}}$	$t\sqrt{\frac{\delta_w^2}{r}}$	$t\sqrt{\frac{\delta_x^2}{r}(\frac{R-r}{R-1})}$	$t\sqrt{\frac{\delta_w^2}{r}(\frac{R-r}{R-1})}$

Расчет средней и предельной ошибок выборки позволяет определить возможные пределы, в которых будут находиться характеристики генеральной совокупности. Например, для выборочной средней такие пределы устанавливаются на основе следующих соотношений:

$$\tilde{\chi} - \Delta_{\tilde{\chi}} \leq \bar{\chi} \leq \tilde{\chi} + \Delta_{\tilde{\chi}}$$

где $\tilde{\chi}$ и $\bar{\chi}$ – выборочная и генеральная средняя соответственно;

$\Delta_{\tilde{\chi}}$ – предельная ошибка выборочной средней.

Рассмотрим примеры определения пределов выборочной средней для различных способов отбора.

Пример. При проверке веса импортируемого груза на таможне методом случайной повторной выборки было отобрано 200 изделий. В результате был установлен средний вес изделия 30 г. при среднем квадратическом отклонении 4 г. С вероятностью 0,997 определите пределы, в которых находится средний вес изделий в генеральной совокупности.

Решение. Рассчитаем сначала предельную ошибку выборки. Так как при $p = 0,997$ $t = 3$, она равна:

$$\Delta_{\tilde{\chi}} = t \cdot \frac{\sigma_{\tilde{\chi}}}{\sqrt{n}} = 3 \cdot \frac{4}{\sqrt{200}} = 0,84$$

Определим пределы генеральной средней:

$$30 - 0,84 \leq \bar{X} \leq 30 + 0,84$$

или

$$29,16 \leq \bar{X} \leq 30,84$$

Следовательно, с вероятностью 0,997 можно утверждать, что средний вес изделий в генеральной совокупности находится в пределах от 29,16 г. до 30,84 г.

Пример. В городе проживает 250 тыс. семей. Для определения среднего числа детей в семье была организована 2%-ная случайная бесповторная выборка семей. По ее результатам было получено следующее распределение семей по числу детей:

Число семей	0	1	2	3	4	5
Количество семей	1000	2000	12000	400	200	200

С вероятностью 0,954 найдите пределы, в которых будет находиться среднее число детей в генеральной совокупности.

Решение. Вначале на основе имеющегося распределения семей определим выборочные среднюю и дисперсию:

Число детей в семье, X_i	Количество семей, f_i	$X_i f_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^2 f_i$
0	1000	0	-1,5	2,25	2250
1	2000	2000	-0,5	0,25	500
2	1200	2400	0,5	0,25	300
3	400	1200	1,5	2,25	900
4	200	800	2,5	6,25	1250
5	200	1000	3,5	12,25	2450
Итого	5000	7400	-	-	7650

$$\tilde{\bar{X}} = \frac{7400}{5000} \approx 1,5 \text{ (чел.)}; \quad \sigma_{\tilde{\bar{X}}}^2 = \frac{7650}{5000} = 1,53.$$

Вычислим теперь предельную ошибку выборки (с учетом того, что при $p = 0,954$ $t = 2$).

$$\Delta_{\tilde{\bar{X}}} = t \sqrt{\frac{\sigma_{\tilde{\bar{X}}}^2}{n} \left(1 - \frac{n}{N}\right)} = 2 \sqrt{\frac{1,53}{5000} \left(1 - \frac{5000}{250000}\right)} \approx 0,006$$

Следовательно, пределы генеральной средней:

$$\bar{X} = \tilde{\bar{X}} \pm \Delta_{\tilde{\bar{X}}} = 1,5 \pm 0,0006$$

Таким образом, с вероятностью 0,954 можно утверждать, что среднее число детей в семьях города практически не отличается от 1,5, т.е. в среднем на каждые две семьи приходится три ребенка.

Наряду с определением ошибок, выборки и пределов для генеральной средней эти же признаки. В этом случае особенности расчета связаны с определением дисперсии доли, которая вычисляется так:

$$\sigma_w^2 = w(1 - w)$$

где $w = \frac{m}{n}$ - доля единиц, обладающих данным признаком в выборочной совокупности, определяемая как отношение количества соответствующих единиц к объему выборки.

Тогда, например, при собственно-случайном повторном отборе для определения предельной ошибки выборки используется следующая формула:

$$\Delta_w = t \sqrt{\frac{\sigma_w^2}{n}} = t \sqrt{\frac{w(1-w)}{n}}$$

Соответственно, при бесповторном отборе

$$\Delta_w = t \sqrt{\frac{\sigma_w^2}{n} \left(1 - \frac{n}{N}\right)} = t \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$$

Пределы доли признака в генеральной совокупности p выглядят следующим образом:

$$W - \Delta_w \leq P \leq W + \Delta_w$$

Пример. С целью определения средней фактической продолжительности рабочего дня в государственном учреждении с численностью служащих 480 человек, в январе 1999 г. была проведена 25 %-ная механическая выборка. По результатам наблюдения оказалось, что у 10 % обследованных потери времени достигали более 45 мин. в день. С вероятностью 0,683 установите пределы, в которых находится генеральная доля служащих с потерями рабочего времени более 45 мин. в день.

Решение. Определим объем выборочной совокупности: $n = 480 \cdot 0,25 = 120$ чел.

Выборочная доля W равна по условию 10 %.

Учитывая, что показатели точности механической и собственно-случайной бесповторной выборки определяются одинаково, а также то, что при $p = 0,683$ $t = 1$, вычислим предельную ошибку выборочной доли:

$$\begin{aligned} \Delta_w &= t \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)} = 1 \cdot \sqrt{\frac{0,1(1-0,1)}{120} \cdot 120 \left(1 - \frac{120}{480}\right)} = \\ &= 0,0237 \approx 0,024 \text{ или } 2,4\% \end{aligned}$$

Пределы доли признака в генеральной совокупности:

$$1000 - 2,4 \leq p \leq 10 + 2,4 \text{ или } 7,6 \leq p \leq 12,4$$

Таким образом, с вероятностью 0,683 можно утверждать, что доля работников учреждения с потерями рабочего времени более 45 мин. в день находится в пределах от 7,6% до 12,4%.

Ошибки выборки и пределы генеральных характеристик при других способах формирования выборочной совокупности определяются на основе соответствующих формул, отражающих особенности этих видов выборки (см. предшествующую главу). Например, в случае типической выборки показателем вариации является средняя из внутригрупповых дисперсий $\overline{\sigma}_i^2$, при серийной выборке – межгрупповая (межсерийная) дисперсия δ^2 и т.д. Кроме того, в последнем случае вместо объема выборочной совокупности n используется показатель числа серий g .

Пример. В области, состоящей из 20 районов, проводилось выборочное обследование урожайности на основе отбора серий (районов). Выборочные средние по районам составили соответственно 14,5 ц/га; 16 ц/га; 15,5 ц/га; 15 ц/га и 14 ц/га. С вероятностью 0,954 найдите пределы урожайности во всей области.

Решение. Рассчитаем общую среднюю:

$$\tilde{x} = \frac{14,5 + 16 + \dots + 14}{5} = 15 \text{ ц/га.}$$

Межгрупповая (межсерийная) дисперсия равна:

$$\delta^2 = \frac{(14,5 - 15)^2 + (16 - 15)^2 + \dots + (14 - 15)^2}{5} = 0,5$$

Определим теперь предельную ошибку серийной бесповторной выборки ($t = 2$ при $p = 0,954$):

$$\Delta = 2 \sqrt{\frac{0,5}{5} \left(1 - \frac{5}{20}\right)} \approx 1,7$$

Следовательно, урожайность в области будет с вероятностью 0,954 находиться в пределах:

$$15 - 1,7 \leq \bar{x} \leq 15 + 1,7 \text{ или } 13,3 \text{ ц/га} \leq \bar{x} \leq 16,7 \text{ ц/га}$$

3.2. Определение необходимого объема выборки

Полученные в результате выборочного обследования границы генеральных характеристик могут быть слишком широки для формулировки обоснованных выводов об изучаемых процессах и принятия каких-либо управленческих решений. В этом случае встает вопрос определения такой необходимой численности выборки, которая обеспечила бы удовлетворительные результаты. Эта числен-

ность может быть определена на базе допустимой ошибки при выборочном наблюдении исходя из вероятности, на основе которой можно гарантировать величину устанавливаемой ошибки, а также с учетом способа отбора. Рассмотрев вначале величину необходимой численности в общем виде, мы исследуем в дальнейшем особые условия, создающиеся в процессе ее вычисления при разных способах отбора.

Для определения необходимой численности выборки исследователь должен задать уровень точности выборочной совокупности с определенной вероятностью. В частности, необходимая численность случайной повторной выборки определяется по формуле

$$n = \frac{t^2 \cdot \sigma^2}{\Delta^2},$$

которая вытекает из формулы предельной ошибки: $\Delta = t \sqrt{\frac{\sigma^2}{n}}$.

Эта формула показывает, что с увеличением предполагаемой ошибки выборки значительно уменьшается необходимый объем выборки. Так, увеличение допустимой ошибки выборки в 2 раза уменьшает необходимый ее объем в 4 раза. Необходимая численность выборки прямо пропорциональна дисперсии признака и величине t^2 . Формула необходимой численности выборки для разных способов отбора выводится из формулы предельной ошибки выборки.

Необходимая численность выборки рассчитывается по-разному для выборочного наблюдения, в котором устанавливается средний размер признака в совокупности, и для наблюдения, в котором определяется доля единиц, обладающих данным признаком, в силу различных методов вычисления меры колеблемости для варьирующего и альтернативного признаков.

На практике определение необходимого объема выборки часто составляет серьезную проблему. Она связана, в частности, с недостаточной разработанностью таких вопросов, как оценка вариации изучаемых признаков, обоснование численности выборки при изучении нескольких признаков, зависимость объема выборочной совокупности от программы разработки материалов наблюдения и др.

Трудности порождаются и тем, что кроме чисто статистических в определении необходимой численности выборочной совокупности

больше значение принадлежит факторам организационного порядка, которые должны быть обязательно учтены. К ним относятся, например, обеспеченность обследования ресурсами, длительность обработки и срочность представления результатов. Согласование объема выборки с материальными, финансовыми, кадровыми ресурсами и тому подобное вызывает определенную сложность.

Одним из наиболее важных и в то же время сложных вопросов определения необходимого объема выборки в исследованиях является расчет показателя вариации изучаемого признака (σ). При подготовке выборочного наблюдения у его организаторов часто отсутствуют необходимые для этих вычислений данные. Основой оценки степени колеблемости изучаемого признака служат, как правило, материалы предыдущих обследований. Обращение к ним при отсутствии какой-либо другой информации вполне оправдано. Однако следует иметь в виду, что использование данных прошлых обследований имеет смысл только тогда, когда за прошедший до нового обследования период в генеральной совокупности не произошло значительных изменений.

Во многих случаях более точное представление об изучаемой совокупности, в том числе о вариации интересующих исследователя признаков, может дать пробное обследование. По его данным возможно рассчитать среднее квадратическое отклонение и дисперсию для последующего обоснования необходимого объема выборки. Если же мера колеблемости признака неизвестна, то ее можно найти приближенно по величине предполагаемого размаха или среднего линейного отклонения по следующим формулам:

$$\sigma = \frac{R}{6} \quad \text{и} \quad \sigma = 1,25\bar{d}$$

где σ – среднее квадратическое отклонение;

R – размах вариации;

\bar{d} – среднее линейное отклонение.

Важным условием практического использования этих формул является близость фактического распределения к нормальному. Исчисление среднего квадратического отклонения для явно несимметричных распределений не имеет смысла.

При статистическом исследовании социально-экономических явлений очень часто приходится сталкиваться с качественными признаками, причем именно по ним нередко проводится расчет необхо-

димого объема выборочной совокупности. Способ выражения качественных признаков не позволяет рассчитать по ним средние значения, поэтому оценка колеблемости производится, как правило, исходя из долей единиц, обладающих значениями этих признаков, т. е. **выборочных долей**. Выборочная доля также называется **частотью**.

Если расчет проводится по качественному альтернативному признаку и не известна его доля в генеральной совокупности (хотя бы приблизительно), рекомендуется принять ее равной 0,5, так как дисперсия доли достигает максимума: $\sigma_w^2 = 0,25$ при $w=0,5$.

Преимущество такого приема заключается в том, что он позволяет определить численность выборочной совокупности, не располагая данными предыдущих обследований, и не проводить пробных обследований. Возможность экономии времени и ресурсов часто оказывается решающим фактором при обращении к данному методу.

Если же качественный признак, по которому определяется необходимая численность выборочной совокупности, не является альтернативным, то использовать формулу $\sigma_w^2 = w(1 - w)$ нельзя.

В ряде случаев приближенная оценка колеблемости может быть осуществлена с помощью превращения изучаемого признака в альтернативный. Например, все категории работников предприятия можно условно разделить в зависимости от принадлежности работающих к рабочим и служащим. Однако при этом следует учитывать, что такое деление неизбежно приведет к потере некоторой части информации. Ведь существуют отдельные категории работников (МОП, охрана и др.), которые выделяются в самостоятельные группы. Поэтому применять описанный выше прием можно лишь при условии, что существует уверенность в незначительной доле неучетных единиц во всей совокупности.

Приведем формулы необходимого объема выборки для наиболее часто используемых на практике способов формирования выборочной совокупности (см. приведенную ниже табл.).

Виды выборочного наблюдения	Повторный отбор	Бесповторный отбор
<i>Собственно-случайная выборка</i>		
а) при определении среднего размера признака	$n = \frac{t^2 \cdot \sigma_{\bar{x}}^2}{\Delta_{\bar{x}}^2}$	$n = \frac{t^2 \cdot \sigma_{\bar{x}}^2 \cdot N}{\Delta_{\bar{x}}^2 \cdot N + t^2 \cdot \sigma_{\bar{x}}^2}$
б) при определении доли признака	$n = \frac{t^2 \cdot W(1 - W)}{\Delta_W^2}$	$n = \frac{t^2 \cdot W(1 - W) \cdot N}{\Delta_W^2 \cdot N + t^2 \cdot W(1 - W)}$
<i>Механическая выборка:</i>	то же	то же
<i>Типическая выборка:</i>		
а) при определении среднего размера признака	$n = \frac{t^2 \cdot \sigma_{\bar{x}}^2}{\Delta_{\bar{x}}^2}$	$n = \frac{t^2 \cdot \sigma_{\bar{x}}^2 \cdot N}{\Delta_{\bar{x}}^2 \cdot N + t^2 \cdot \sigma_{\bar{x}}^2}$
б) при определении доли признака	$n = \frac{t^2 \cdot \overline{W(1 - W)}}{\Delta_W^2}$	$n = \frac{t^2 \cdot \overline{W(1 - W)} \cdot N}{\Delta_W^2 \cdot N + t^2 \cdot \overline{W(1 - W)}}$
<i>Серийная выборка:</i>		
а) при определении среднего размера признака	$r = \frac{t^2 \cdot \delta_{\bar{x}}^2}{\Delta_{\bar{x}}^2}$	$r = \frac{t^2 \cdot \delta_{\bar{x}}^2 \cdot R}{\Delta_{\bar{x}}^2 \cdot R + t^2 \cdot \delta_{\bar{x}}^2}$
б) при определении доли признака	$r = \frac{t^2 \cdot W_r(1 - W_r)}{\Delta_W^2}$	$r = \frac{t^2 \cdot W_r(1 - W_r) \cdot R}{\Delta_W^2 \cdot R + t^2 \cdot W_r(1 - W_r)}$

Рассмотрим несколько примеров расчета объема выборки при различных способах отбора.

Пример. В микрорайоне проживает 5000 семей. В порядке случайной бесповторной выборки предполагается определить средний размер семьи при условии, что ошибка выборочной средней не должна превышать 0,8 человека с вероятностью $P = 0,954$ и при среднем квадратическом отклонении 3,0 человека (ошибка и среднее квадратическое отклонение определены на основе пробного обследования).

Так как при $P = 0,954$ $t = 2$, то в этом случае необходимая численность выборки равна:

$$n = \frac{2^2 \cdot 3^2 \cdot 5000}{5000 \cdot 0,64 + 2^2 \cdot 3^1} = \frac{4 \cdot 9 \cdot 5000}{5000 \cdot 0,64 + 4 \cdot 9} = \frac{180000}{3236} = 56 \text{ семей}$$

Пример. Для определения средней длины детали следует провести выборочное обследование методом случайного повторного отбора. Какое количество деталей надо отобрать, чтобы ошибка выборки не превышала 3 мм с вероятностью 0,997 при среднем квадратическом отклонении 6 мм?

При $t = 3$ и $P = 0,997$ объем выборки рассчитывается следующим образом.

$$n = \frac{3^2 \cdot 6^2}{3^2} = 36 \text{ деталей}$$

3.3. Распространение результатов выборочного наблюдения на генеральную совокупность

Заключительным этапом выборочного наблюдения является распространение его результатов на генеральную совокупность. Однако часто при статистическом изучении социально-экономических явлений этому процессу предшествует оценка результатов наблюдения с точки зрения самой возможности распространения.

Вывод о возможности распространения в значительной степени зависит от качества основы выборки, прежде всего от ее полноты. *Под полнотой* подразумевается наличие или представленность всех типов или групп данной генеральной совокупности в основе выборки. Неполнота основы может привести к нарушению представительности выборки и, как следствие, к неправильным выводам при анализе данных наблюдения.

Однако не следует обосновывать возможность распространения выборочных данных только анализом качества исходной информации для отбора. Более точной основой суждения о возможности распространения представляется расчет относительной ошибки:

$$\text{для средней } \Delta_{\%} = \frac{\Delta_{\tilde{x}}}{\bar{x}} \cdot 100\%,$$

$$\text{для доли } \Delta_{\%} = \frac{\Delta_w}{\bar{p}} \cdot 100\%,$$

где $\Delta_{\%}$ – относительная предельная ошибка выборки;
 $\Delta_{\bar{x}}$ и Δ_w – предельная ошибка для среднего значения или доли признака соответственно;
 \bar{x} и \bar{p} – генеральная средняя и доля соответственно

Суждение о возможности распространения выборочных данных можно составить, если в формулах заменить \bar{x} и \bar{p} соответствующими выборочными характеристиками. Необходимым условием при этом является соответствие плановой и фактической численности и структуры выборочной совокупности. При больших расхождениях использование этого приема может привести к ошибочным суждениям.

Если величина относительной ошибки не превышает заранее установленного для данного обследования предельного значения, то данные выборочного наблюдения являются представительными и могут быть распространены на генеральную совокупность. В противном случае следует попытаться восстановить исходные пропорции генеральной совокупности. Процесс восстановления пропорций выборки на основе исходной информации о таких пропорциях в генеральной совокупности принято называть *корректировкой*¹ выборки.

При обработке данных выборочного наблюдения целесообразно использовать два наиболее часто применяемых способа корректировки. Первый способ ориентирован на группу единиц, которые оказались недостаточно представлены в выборочной совокупности после наблюдения. Формуляры с данными об этих единицах, пригодные для обработки, следует сохранять в полном объеме. На основе сведений о количестве таких формуляров проводятся дополнительные расчеты. Их целью является определение числа хорошо представленных в фактической выборке формуляров остальных групп, часть которых необходимо исключить из обработки для сохранения пропорций генеральной совокупности. Данный способ корректировки, называемый *методом «отсечения»*, поясним на следующем условном примере.

¹ В отечественной литературе данный процесс называют также коррекцией, «ремонт» выборки. В зарубежной литературе получил распространение термин «постстратификация».

С целью изучения общественного мнения из генеральной совокупности численностью 1000 человек было отобрано в порядке типической пропорциональной выборки 100 человек, принадлежащих к различным социальным группам: рабочие, служащие, студенты. При этом в генеральной совокупности было 50% рабочих, 35 служащих и 15% студентов, т. е. пропорция по группам населения составила примерно 3,3 : 2,3 : 1. Следовательно, для обеспечения представительности выборки по признаку социального положения требовалось бы получить данные о 50 рабочих, 35 служащих и 15 студентах. Однако по тем или иным причинам часть анкет не была получена, а другая часть была забракована. В результате пригодными для дальнейшей обработки оказались 40 анкет, заполненных рабочими, 30 – служащими и 10 – студентами. Таким образом, пропорции по различным группам в массиве для обработки составили 4 : 3 : 1, что свидетельствует о нарушении структуры генеральной совокупности.

Для проведения корректировки необходимо определить, анкеты какой социальной группы респондентов должны быть сохранены в процессе обработки полностью. Это можно сделать с помощью относительных величин, вычисляемых для каждой социальной категории как отношение числа пригодных для обработки анкет к общему количеству анкет по данной группе. Результат выражается в долях или процентах. Расчеты показывают, что наименьшая относительная величина получается по студентам (приблизительно 66,7 %). Следовательно, формуляры, относящиеся к данной группе населения, необходимо сохранить полностью. Чтобы восстановить реальные пропорции генеральной совокупности, нужно вновь обратиться к ее структуре, выраженной соотношением 3,3 : 2,3 : 1.

Несложные вычисления показывают, что для сохранения представительности выборки в массиве анкет для дальнейшей обработки должны быть 33 анкеты, заполненные рабочими, 23 – служащими и 10 – студентами. Таким образом, из дальнейшей обработки следует «отсечь» по 7 анкет, относящихся к рабочим и служащим. Для «отсечения» возможно пользоваться процедурой случайной выборки.

После «отсечения» следует проверить, как исключение некоторого числа формуляров повлияло на обобщающие показатели фактической выборки. Для этого вначале следует найти средние по важнейшим показателям в совокупности пригодных для обработки

формуляров, включая те, которые затем предполагается «отсечь». Затем те же средние рассчитать по совокупности формуляров, оставшихся после «отсечения», и сравнить полученные результаты. Для оценки различий средних можно воспользоваться принципами оценки точности выборки. Если расхождения между средними, рассчитанными до и после «отсечения», не превышают $\pm 5\%$, итоги корректировки считаются вполне удовлетворительными. В противном случае ее целесообразно повторить, исключив из обработки другие формуляры.

Для проверки степени влияния «отсечения» на обобщающие характеристики выборки используют и другой способ. Он заключается в сравнении средних, рассчитанных для различных вариантов «отсекаемых» формуляров, со средними, вычисленными до и после «отсечения».

Приведенный пример позволяет рассмотреть достоинства данного метода корректировки и его недостатки. Основным достоинством метода является то, что он дает возможность сохранить пропорции генеральной совокупности в массиве данных, на основе которого будут делаться обобщения. Это позволяет формулировать выводы на базе представительных данных.

В то же время корректировка способом «отсечения» имеет существенные недостатки. Во-первых, «отсечение» приводит к еще большему, если учитывать невозвращенные и забракованные формуляры, уменьшению объема выборки. Во-вторых, из обработки и анализа исключаются вполне пригодные для исследования формуляры. В нашем примере «отсекаются» 17,5% собранных формуляров по группе рабочих и примерно 23,3% – по группе служащих. Сами по себе эти показатели весьма значительны. Однако и они могут возрасти, если для обработки окажутся пригодными не 10, а, скажем, 7 анкет студентов. Тогда из обработки потребуется исключить 17 из 40 анкет рабочих (42,5%) и 14 из 30 анкет служащих (около 46,7%). В таких случаях более целесообразно пользоваться другим способом корректировки – с помощью «взвешивания».

В отличие от первого способа корректировки «взвешивание» дает возможность сохранить в обрабатываемом массиве все или почти все полученные формуляры. Достигается это путем многократного пользования при обработке части формуляров. При этом несколько раз используются, как правило, те формуляры, число ко-

торых настолько мало, что вызывает необходимость исключения из дальнейшей обработки большого числа для исследования формуляров, относящихся к другим группам. Многократное применение формуляров проводится на основе специально рассчитанных для этой цели «весов».

Метод «взвешивания» наиболее удобно применять при обработке материалов выборочных обследований в случаях высокого процента невозвращенных или забракованных формуляров. Это характерно прежде всего для почтовых опросов.

Наиболее широко применяемые в социальных исследованиях методы корректировки выборки могут использоваться и при выборочных обследованиях промышленности, аграрного сектора и других отраслей экономики. Одной из наиболее проработанных является методика обработки так называемых неответов при обследовании малых предприятий.

Для обработки полных неответов респондентов совокупность неответивших организаций разделяется на три группы, к каждой из которых для обработки неответов применяется свой метод коррекции и восстановления данных.

Первая группа объединяет малые предприятия, данные по которым не восстанавливаются. К ним относятся организации, прекратившие или приостановившие свою деятельность в силу различных причин.

Вторая группа объединяет предприятия, о которых достоверно известно, что они, несмотря на отсутствие отчета, активны и ведут финансово-хозяйственную деятельность.

Для восстановления данных по предприятиям данной группы применяется метод перевзвешивания (или заполнения по среднему в слое). Неответившему малому предприятию присваивается среднее значение ответивших в данном слое. Если не удастся восстановить значения показателей по среднему в слое (например, если в слое нет ни одного ответившего предприятия), тогда этим предприятиям приписываются средние значения показателей, рассчитанные по действующим предприятиям в отрасли.

Третья группа - малые предприятия, по которым не получена достоверная информация о том, действующие они или нет.

Для восстановления данных по этой группе применяется метод заполнения случайным подбором в классах замещения (random

hot deck within classes). Для этого выборочная совокупность малых предприятий (как функционирующих и представивших отчет, так и ликвидированных) делится на классы по отраслям экономики. Неответившей организации приписывается значение признака выбранного случайным образом предприятия-донора из соответствующего класса замещения.

Собранные в результате выборочного наблюдения и при необходимости откорректированные данные распространяются на генеральную совокупность. Существуют два основных метода распространения – **прямой пересчет и способ коэффициентов**.

Сущность **способа прямого пересчета** заключается в умножении среднего значения признака, найденного в результате выборочного наблюдения, на объем генеральной совокупности. Практические расчеты при этом не вызывают серьезных затруднений. Например, на основании выборочного обследования 1000 молодых семей требуется оценить потребность в местах в детских яслях. С помощью метода прямого пересчета это можно сделать следующим образом. Известно, что ясли могут посещать дети в возрасте до трех лет. По материалам выборочного обследования следует вычислить среднее число детей этого возраста. Предположим, что оно составляет 1,3 человека. Умножив это число на численность генеральной совокупности, получим, что в детских яслях потребуется выделить 1300 мест.

Производя такие расчеты, мы считаем, что были обследованы все единицы, попавшие в выборочную совокупность. Однако в социальных исследованиях объемы фактической и запланированной выборки часто не совпадают, что всегда следует учитывать. Как правило, несоответствие фактической и запланированной выборки приводит, естественно, к неадекватному отражению в выборочных характеристиках, полученных по фактическим данным соответствующих характеристик генеральной совокупности.

Предположим, в нашем примере некоторое число семей по тем или иным причинам не было обследовано. Это привело к снижению объема фактической выборки по сравнению с запланированной. Среднее число детей, вычисленное по этой «неполной» выборке, составило не 1,3, а 1,2. Тогда прямой пересчет выборочной характеристики на объем генеральной совокупности даст результат 1200 мест. Абсолютное отклонение от необходимого количества мест при

условии охвата обследованием всей выборочной совокупности составит 100 мест, а относительное – приблизительно 7,7 %. Если же объем генеральной совокупности был бы в 10 раз больше, т.е. 10 000 семей, то абсолютное отклонение также увеличится в 10 раз и при сохранении тех же различий среднего числа детей составит 1000 мест, относительное отклонение при этом не изменится. Таким образом, размер абсолютного отклонения находится в прямой зависимости от объема генеральной совокупности.

Данный пример показывает: недоучет обстоятельства, при котором на практике объемы фактической и запланированной выборок часто не совпадают, приводит к серьезным ошибкам при использовании распространенных на генеральную совокупность результатов таких исследований. Руководствуясь данными, рассчитанными на условном примере, пришлось бы принимать ошибочное решение о строительстве дополнительного числа детских учреждений, мест в которых не хватило бы на 100 (или на 1000 – в зависимости от объема генеральной совокупности) детей. Но могла возникнуть и обратная ситуация, когда вычисленное по «неполной» выборочной совокупности среднее число детей оказалось бы больше «истинного». В этом случае появились бы «лишние» места. Данный пример показывает, что результатами выборочного наблюдения необходимо пользоваться осторожно, особенно в случаях, когда их использование связано с большими материальными затратами.

В условиях существования большого числа факторов, влияющих на точность данных выборочного наблюдения, использование точечной оценки при распространении выборочных характеристик на генеральную совокупность в социальных исследованиях часто нецелесообразно. Во всех случаях, когда это возможно, правильнее пользоваться интервальной оценкой, позволяющей учесть размер предельной ошибки выборки, рассчитанной для средней или для доли признака. Так, если в нашем примере среднее число детей в возрасте до трех лет по выборочным данным составило 1,3 человека, а предельная ошибка – $\pm 0,1$ человека, то требуемое количество мест в детских учреждениях будет находиться в пределах от 1200 до 1400.

Наряду со способом прямого пересчета при распространении данных выборочного наблюдения на генеральную совокупность применяется так называемый **способ коэффициентов**. Данный спо-

соб целесообразно использовать в случаях, когда выборочное наблюдение проводится с целью проверки и уточнения данных сплошного наблюдения, в частности численности учетных единиц совокупности. При этом следует использовать следующую формулу:

$$Y_1 = Y_0 \cdot \frac{y_1}{y_0},$$

где Y_1 - численность совокупности с поправкой на недоучет;
 Y_0 - численность совокупности без этой поправки,
 y_0 - численность совокупности в контрольных точках по первоначальным данным,
 y_1 - численность совокупности в тех же точках по данным контрольных мероприятий/

Отметим, что цели исследования многих явлений могут быть достигнуты только путем сплошного наблюдения. Поэтому способ проверки результатов сплошного наблюдения на основе коэффициентов успешно применяется в социальной и экономической статистике. До сих пор возможности выборки при уточнении данных сплошного наблюдения используются недостаточно. В то же время в современных условиях данный способ может быть, например, инструментом в контроле за деятельностью коммерческих структур со стороны финансовых органов.

При уточнении данных сплошного наблюдения на основе контрольных выборочных мероприятий определяется так называемая *поправка на недоучет*. Метод ее расчета наиболее широко применяется в обследованиях относительно небольших совокупностей, когда их объем не превышает нескольких сотен или тысяч единиц.

Пример. При проведении учета коммерческих палаток в городе было зарегистрировано следующее их количество в районах: А – 2000, Б – 1500; В – 750. С целью проверки данных сплошного учета проведены контрольные обходы части обследованных районов. Их результаты содержатся в нижеприведенной таблице.

Количество коммерческих палаток в районах города до и после контрольных обходов

Районы	Зарегистрировано при сплошном учете	Установлено при контрольном обходе	Коэффициент недоучета
А	400	420	1,050
Б	300	310	1,033

В	150	160	1,067
---	-----	-----	-------

Рассчитанный по каждой категории работников коэффициент недоучета является основой уточнения при распространении данных, полученных при выборочных контрольных мероприятиях, на всю совокупность.

В нашем примере количество коммерческих палаток (по данным сплошного учета) следует умножить на рассчитанный для каждого района коэффициент недоучета. В итоге получим результаты, представленные в таблице:

Уточненные данные учета коммерческих палаток в районах города

	Количество коммерческих палаток в районах города		
	А	Б	В
Данные сплошного наблюдения	2000	1500	750
Численность с поправкой на недоучет*	2100	1550	800

Применение метода коэффициентов связано, как правило, с использованием выборочного наблюдения с целью проверки данных сплошного наблюдения. Однако это приводит к сознательному ограничению области применения данного метода. Метод коэффициентов можно использовать для проверки данных выборочного наблюдения, когда необходима очень высокая точность результатов и выборочная совокупность имеет большой объем – порядка нескольких тысяч или десятков тысяч единиц. В таких случаях списки единиц обследованной выборочной совокупности служат основой для отбора единиц в «контрольную» выборку, т. е. производится выборка из выборки. Контроль и, если это необходимо, уточнение данных основного обследования проводятся по методике, описанной выше на условном примере. Способ поправочных коэффициентов целесообразно использовать для распространения данных выборочного наблюдения в случаях, если его результаты значительно уступают в точности данным статистической отчетности или точность собран-

ного статистического материала вызывает сомнение. При этом, естественно, данные способы уточнения результатов выборочного наблюдения могут быть использованы при наличии времени и средств для проведения контрольных мероприятий.

Завершая рассмотрение вопросов оценки точности и распространения данных выборочного наблюдения на генеральную совокупность, отметим: публикуя где бы то ни было статистические данные, рассчитанные на основе выборки, обязательно следует пояснять, что соответствующие показатели получены расчетным путем на базе материалов выборочного наблюдения. Одновременно было бы правильно указать пределы ошибок, которые могли быть допущены в процессе выборки, а также вероятность этих пределов.

Список рекомендуемой литературы

1. Венецкий И.Г. Теоретические и практические основы выборочного метода. Учебное пособие. - М.: МЭСИ, 1975.
2. Голубков Е.П. Маркетинговые исследования: теория, методология, практика. – М.: Финпресс, 1998.
3. Джессен Р. Методы статистических исследований / Под ред. Е.М.Четыркина. - М.: Финансы и статистика, 1985.
4. Дружинин Н.К. Выборочное наблюдение и эксперимент. - М.: Статистика, 1977.
5. Елисеева И.И., Терехов А.А. Статистические методы в аудите. – М.: Финансы и статистика, 1998.
6. Кокрен У. Методы выборочного обследования / Под ред. А.Г.Волкова. - М.: Статистика, 1976.
7. Методологические положения по статистике. Вып. 3 / Госкомстат РФ. – М., 2000.
8. Практикум по теории статистики / Под ред. проф.Шмойловой Р.А. - М.: Финансы и статистика, 2000.
9. Теория статистики. Учебник / Под ред. проф.Шмойловой Р.А. - М.: Финансы и статистика, 1999.
10. Теория статистики. Учебно-практическое пособие / Под ред. Ми-нашкина В.Г. - М.: МЭСИ, 2000.
11. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере / Под ред. В.Э. Фигурнова. - М.: Финансы и статистика, 1998.
12. Шварц Г. Выборочный метод / Под ред. И.Г.Венецкого и В.М.Ивановой. - М.: Статистика, 1978.