Е.М. Четыркин, Н.Е. Васильева

Выборочные методы в аудите

Издательство "Дело"

Е.М. Четыркин, Н.Е. Васильева

Выборочные методы в АУДИТЕ

Академия народного хозяйства
при Правительстве Российской Федерации

Москва
Издательство "ДЕЛО"
2003

УДК 519.2+657.6
ББК 22.172+65.053
Ч-52

Четыркин Е.М., Васильева Н.Е.

Ч-52 Выборочные методы в аудите. – М.: Дело, 2003. – 144 с.

ISBN 5-7749-0321-4

Выборочный метод, один из классических методов статистического анализа, сравнительно недавно получил новую область практического приложения. Как показал опыт крупнейших аудиторских фирм, выборка эффективно применяется в различных видах финансового контроля. В аудите используются как стандартные, так и специальные приемы извлечения выборок при проверках "на соблюдение" (контроле правильности оформления документов) и проверках "по существу" (контроле правильности количественных показателей).

Книга содержит описание выборочных методов, применяемых в аудите в увязке с рядом общетеоретических положений статистики. Приводится много примеров.

Книга предназначена работникам службы внешнего и особенно внутреннего аудита, студентам, специализирующимся в области финансов и кредита.

УДК 519.2+657.6
ББК 22.172+65.053

ISBN 5-7749-0321-4

ОГЛАВЛЕНИЕ

Глава 1. АУДИТ И ВЫБОРОЧНЫЕ МЕТОДЫ

Я пригласил вас, господа, с тем, чтобы сообщить вам пренеприятное известие: к нам едет ревизор.

Н.В. Гоголь

§ 1.1. Внешний и внутренний аудит

Аудиторская деятельность, которой в настоящее время в России уделяется большое внимание, как известно, осуществляется в виде *внешнего* и *внутреннего* аудита. Первый из названных выполняется аудиторскими фирмами или независимыми аудиторами, второй осуществляется специализированным подразделением той организации, чья деятельность подвергается аудиторской проверке. Конечная общая цель обоих видов аудита во многом совпадает – это контроль. Однако имеются и существенные различия в содержании внешнего и внутреннего аудита и характере деятельности специалистов, осуществляющих эти виды деятельности.

Внешний аудит предполагает оценку системы отчетности, проверку и оценку активов и пассивов организации, тестирование существующей системы внутреннего контроля. Главная задача внешнего аудита – установить, соответствуют ли реальности представляемые отчеты и балансы проверяемой организации, оценить ее финансовое положение и результат деятельности за определенный период. Внешний аудит осуществляется периодически, причем часто разными специализированными фирмами или аудиторами, приглашаемыми со стороны. Отчет, подготовленный внешним аудитором, предназначен главным образом вышестоящим органам, акционерам, кредиторам и т.д.

Внутренний аудит имеет другую природу, смысл, назначение и организацию по сравнению с внешним. По общепринятому мнению общей задачей внутреннего аудита является помощь организации в эффективном выполнении своих обязанностей в соответствии с ее назначением. Внутренний аудит – это орудие управления, предназначенное для обеспечения (гарантии) достижения

целей управления организацией. Для этого служба внутреннего аудита снабжает администрацию информацией о результатах выполненного анализа, оценках деятельности того или иного подразделения, различными рекомендациями. Важной задачей внутреннего аудита является сотрудничество с внешними аудиторами.

Выбор объектов для внутреннего аудита зависит от специфики экономического субъекта. Это система бухгалтерского учета, ее мониторинг, бухгалтерская и оперативная информация и т.п. Значительный объем работы внутреннего аудита приходится на тестирование массовой текущей первичной документации. Данное обстоятельство является одним из обоснований для применения выборочных методов контроля. (В большинстве случаев нет необходимости проверять все первичные документы – счета, накладные и т.д.) Во внешнем аудите с такого рода документами сталкиваются редко.

Служба внутреннего аудита, являясь независимым подразделением, созданным в рамках организации, выполняет систематическую, каждодневную работу по проверке и оцениванию деятельности этой организации. В России функции внутреннего аудита могут выполнять специальные аудиторские службы, ревизионные комиссии (ревизоры), или службы внутреннего контроля.

Как уже было отмечено выше, задачи, которые выполняются при внутреннем аудите, разнообразны и зависят от его целей, вида проверяемой организации и характера ее деятельности. Очевидно, что аудит государственных организаций, коммерческих банков, страховых компаний и пенсионных фондов имеет каждый свою специфику. Более того, многие задачи можно решать различными способами, с разной глубиной и детальностью и ориентируясь на различные критерии. В связи со сказанным неизбежно возникает проблема регламентирования аудиторской деятельности. Эта проблема в какой-то мере решается с помощью разработки соответствующих правил или стандартов, а также конкретных методик проведения аудита. Термин "стандарт" (калька с английского standard) здесь следует понимать в специальном смысле, который отличается от принятого у нас. Вместо Standards за рубежом применяют и другое название – Code of Ethics for Auditors, которое можно перевести как "стандарты поведения аудиторов", "этический кодекс" или, наконец, "правила для аудиторов". Заметим, что такие стандарты весьма расплывчаты, хотя и могут служить первичными ориентирами при разработке конкретных методик.

В странах с развитым государственным аппаратом проблемам аудита, в том числе внутреннего, уделяется много внимания. Достаточно упомянуть о существовании Международной организации высших институтов контроля (The International Organization of Supreme Audit Institutions), в которую входят более десятка стран, включая США, Англию и Японию. Международным является и Институт внутренних аудиторов (The Institution of Internal Auditors)[1]. В названных и других аналогичных зарубежных организациях имеются подразделения, ответственные за разработку соответствующих стандартов внутреннего аудита[2]. В России также разрабатываются стандарты аудиторской деятельности. Они рассматриваются Комиссией по аудиторской деятельности при Президенте Российской Федерации. Такие стандарты предназначены для общей регламентации аудиторской деятельности, в том числе и для внутреннего аудита[3].

Обратимся к двум стандартам внутреннего аудита, которые, как нам представляется, имеют непосредственное отношение к выборочным методам в аудите. Это стандарт разумной гарантии (Reasonable Assurance) и стандарт методов тестирования (Control Techniques)[4].

Разумная гарантия. Под этим понимается разумное (т.е. целенаправленное и без излишеств) достижение целей проверки. Иначе говоря, затраты на проведение проверки должны быть сопоставимы с ее эффективностью. При выполнении этого правила принимаются во внимание риски, свойственные операциям проверяемой организации, критерии для определения уровня риска, приемлемые размеры рисков для разных условий.

Методы тестирования. Основное назначение этого стандарта заключается в том, что выбранные методы внутреннего контроля должны быть эффективны и продуктивны для достижения постав-

[1] The Institution of Internal Auditors. Standards for Professional Practice of Internal Auditing. 1998. (Последний вариант стандарта опубликован в 2001 г.)

[2] Краткая характеристика международных стандартов внутреннего аудита приведена в статье: *Васильева Н.Е.* Стандарты и методы внутреннего аудита // Деньги и кредит. 2001. № 4. С. 35–39.

[3] Аудиторские ведомости. 1997. № 6; 1998. № 3. Текст стандарта приводится также в книге: *Богомолов А.М., Голощапов Н.А.* Внутренний аудит. Организация и проведение. М.: Экзамен, 1999. Гл. 5.

[4] The Institution of Internal Auditors. Op. cit.

ленных целей. Под методами тестирования (контроля) понимают различные процедуры, позволяющие достичь поставленных целей. Стандарт требует, чтобы эти методы обеспечивали высокую степень надежности проверки. Они включают планы организации аудита (в том числе разграничение обязанностей) и меры безопасности контроля.

Как будет показано ниже, оба указанных стандарта, дополняя друг друга, создают логическую и экономическую основу для применения выборочных методов в аудите (audit sampling). Действительно, аудиторская проверка, охватывающая всю совокупность документов, дает если не стопроцентную, то близкую к ней надежность обнаружения ошибок и различного рода погрешностей (разумеется, сплошной контроль не устраняет ошибки самого аудитора). Однако при большом количестве каких-либо однородных операций, учетных записей и т.п. сплошной контроль выливается в громадный объем работы со всеми вытекающими последствиями. Сказанное позволяет рассматривать выборочный метод как полезный инструмент внутреннего аудита. Выборочному методу посвящен международный стандарт SAS 39. Российский стандарт "Аудиторская выборка" (принят 25 декабря 1996 г.) дает "правовое" обоснование для применения выборок в аудите.

Практика аудита (о чем свидетельствуют инструкции и руководства по проведению аудита, в том числе разработанные членами большой пятерки[1]) показывает, что выборочный контроль позволяет получить необходимую аудитору информацию, в том числе для проверки представленных ему данных. При этом резко сокращаются издержки на проведение проверки и обеспечиваются приемлемые результаты в отношении точности и надежности получаемых сведений.

Выборочные методы при правильном их применении соответствуют упомянутым выше стандартам разумной гарантии и методов тестирования. В то же время выборочный контроль позволяет заметно сократить сроки проведения проверки и затраты на ее проведение, что часто имеет решающее значение. Сокращение объема работы позволяет для выполнения контроля привлечь более квалифицированный персонал.

Необходимость применения выборочных методов заставила крупные западные фирмы разрабатывать специальные программные продукты для выборочных проверок на основе хранящихся файлов.

Естественно, что выборочный контроль не может распространяться на любые объекты. Не могут быть такими объектами особо важные по своим финансовым последствиям сделки, бухгалтерские проводки. Иначе говоря, выборочный анализ не распространяется на так называемые *ключевые позиции* (key items), т.е. такие операции, сделки или элементы балансовых счетов, которые имеют особую важность в силу своей природы или своих размеров.

Кратко остановимся на структуре аудиторской проверки. Процесс аудита, как известно, распадается на несколько основных этапов. Типичными для аудита отдельного объекта являются следующие этапы:

- предварительное планирование;
- разработка плана и программы аудита;
- проведение аудита;
- оценка и документальное оформление результатов аудита в виде заключения.

На этапе предварительного планирования определяют общее направление аудита, его стратегию. На втором этапе эта стратегия конкретизируется. При этом устанавливаются цели, объекты и позиции для проверки, фиксируются процедуры, которые необходимо выполнить, их масштабы и сроки. Особо выделяются позиции, нуждающиеся в усиленной проверке с учетом прошлого опыта работы аудитора с проверяемым подразделением или клиентом, где существует высокая вероятность появления неверных сведений в представляемых документах. На этом этапе выделяются позиции, которые целесообразно проверить, используя выборочные методы.

При проведении аудита применяются два типа тестирования – *проверки на соответствие*, или *на соблюдение* (compliance tests)

[1] Можно указать на два фирменных руководства по применению выборочных методов в аудите: Touche Ross International. The Touche Ross Audit Sampling Manual. U.K. 1982; DRT International. Audit Sampling. A Guide to Applying the DRT International Audit Approach. N.Y., 1991.

Большое внимание уделено выборочному методу в известном руководстве по внутреннему банковскому аудиту: *Giliberti A.* Bank Internal Auditing Manual. N.Y.: Warren, Gorham & Lamont, 1997. Vol. 2. См. также: Ernst &Young Global Audit Methodology. 2001. April.

Общее описание условий применения выборочного метода в аудите (но не конкретных методик) содержится в пособии: Стандарты аудиторской деятельности. М.: ИНФРА-М, 2000.

и *проверки по существу* (substantive tests). Первый тип контроля нацелен на тестирование качества системы внутреннего контроля. Речь идет о проверке соответствия фактических данных принятым экономическим нормативам, соблюдения принятого порядка оформления документов. Например, своевременности и полноты оплаты уставного капитала, соответствия фактических данных установленным резервам, правильности оформления документов, наличия необходимых реквизитов и т.д.

В качестве примера приведем список позиций проверки на соответствие при аудите кассовых операций в банке. Проверяются:

а) правильность оформления объявлений на взнос наличными, а именно: наличие и тождественность подписи операционного работника образцу, полнота заполнения необходимых реквизитов, отсутствие исправлений в представленных документах, наконец, соответствие суммы цифрами сумме прописью;

б) своевременность зачисления средств на счет клиента;

в) правильность документального оформления поступивших в банк денег.

Приведенный краткий список позиций может быть значительно расширен. Однако и так ясно, что проверка на соответствие может быть распространена практически на любые действия оперативного работника. Причем выборка позволяет осуществить проверку при сравнительно небольших затратах.

Проверка по существу более информативна, более основательна. Она позволяет контролировать правильность стоимостных и других количественных показателей документов, например данных балансов, остатков на расчетных счетах и других записей в финансовых документах. В частности, особое внимание уделяется контролю дебиторских счетов[1]. В ходе тестирования выявляются ошибки. Выборочная оценка ошибок позволяет аудитору получить представление о правильности полученных им сведений и сделать соответствующие коррективы. Выборочный метод применяется и для получения отсутствующих стоимостных и других количественных данных, необходимых аудитору.

Разумеется, одни и те же документы, попавшие в выборку, могут быть тестированы и на соответствие, и по существу. Например, оплаченные организацией счета-фактуры могут быть выборочно проверены на полноту оформления. Вместе с тем эта же выборка может быть проверена и на правильность указанных в счетах сумм.

[1] *Loebbecke J., Neter J.* Statistical sampling in confirming receivables // The Journal of Accountancy. 1973. June.

Ошибки, обнаруженные при проверке на соответствие, сигнализируют о возможности наличия ошибок по существу. Короче говоря, если выявлено значительное количество ошибок на соответствие, то следует ждать и ошибки в стоимостных характеристиках. Предварительный контроль на соответствие сокращает затраты на контроль по существу и позволяет выявить наиболее важные участки для тестирования.

В заключении по аудиту фиксируется мнение аудитора о результатах, полученных в процессе работы, в частности:

- о финансовой устойчивости объекта проверки;
- достаточности данных для аудита и качестве представленной информации;
- применяемых принципах учета и отчетности;
- лояльности к аудиту администрации.

При применении выборочных методов общая схема процесса аудита сохраняется. Однако на этапах разработки детального плана проверки и при непосредственном проведении аудита учитываются специфические условия, зависящие от вида выборочного обследования. На этом мы остановимся в следующих главах.

§ 1.2. Существенность (материальность)

Одним из важных понятий в аудите является *существенность*, или *материальность* (materiality). Определение этого понятия довольно расплывчато. В различных руководствах[1] и стандартах аудита существенность определяется как размер ошибочных сведений (включая пропуски) в финансовой информации, которая с учетом конкретных условий проверки и сопутствующих обстоятельств влияет на суждение пользователя информации и в связи с этим делает вероятным изменение в принятых им решениях.

Задаваемый или планируемый в аудиторской проверке уровень существенности (planning materiality) представляет собой предварительную оценку ошибки в денежном измерении, которая рас-

[1] DRT International. The DRT International Audit Approach. 1990. P. 262.

сматривается как существенная для обоснования аудиторского заключения. Проще говоря, концепция существенности означает, что аудитор должен сосредоточить внимание на важной финансовой информации (документах) и меньше тратить усилий на менее значимые счета и операции. Таким образом, уровень существенности является некоторой условной границей, отделяющей важное от менее важного.

Операции или документы, стоимостные характеристики которых превышают принятый уровень существенности, подлежат сплошной проверке. Остальные можно тестировать выборочно. Важно ведь обнаружить те ошибки, которые приводят к существенным искажениям (material errors) и/или свидетельствуют о недобросовестности персонала, некачественном управлении, наличии мошенничества и хищений.

Понятие существенности может применяться не только к единичным наблюдениям, но и к суммам соответствующих показателей, например, установить существенность суммы задолженности отдельных хозяйствующих объектов и т.д. Выводы аудитора с учетом уровня существенности оказываются более надежными, чем в случае, когда он в равной мере распределяет свои усилия на все счета и операции.

При оценивании уровня существенности необходимы высокий профессиональный опыт аудитора, знание особенностей хозяйственных и финансовых операций проверяемого объекта, его положения и т.д. Сказанное дает лишь общее представление о сути планируемого уровня существенности. В различных западных руководствах приводятся конкретные рекомендации для практического определения этого уровня. Например, часто он связывается с объемом прибыли финансовой организации. Причем за базу может быть принята: прибыль до выплаты налогов (в этом случае уровень существенности может составлять от 0,5 до 5%), чистая прибыль (5–10%). Если за базу принята стоимость активов, то уровень существенности составляет 0,5–1% этой стоимости. В одном из руководств находим следующие рекомендации по определению уровня существенности: 5% прибыли от текущих операций или 1–2,5% стоимости акций[1]. Очевидно, что подобного рода рекомендации имеют весьма общий характер, так как и объемы текущих операций, и стоимость акций можно оценивать различными методами.

[1] *Giliberti A.* Op. cit. P. 65-3.

Оценивание уровня существенности производится на этапе разработки плана аудиторской проверки. Однако этот уровень при необходимости может быть изменен в ходе тестирования.

Конкретный уровень существенности рассматривается как база для определения другой, не менее важной для аудита величины – приемлемой ошибки (tolerable error, monetary precision). Под этим параметром понимается размер ошибки, при ориентации на который аудитор имеет некоторую разумную гарантию того, что он обнаружит значительную погрешность, если она существует.

Естественно, приемлемая ошибка должна быть меньше уровня существенности. Проблема в том – насколько меньше. Если эта ошибка слишком велика, то существует высокий уровень риска не достигнуть целей тестирования. Если этот уровень чрезмерно низкий, то на проверку придется затратить лишние усилия. В западных практических руководствах приводятся различные рекомендации о величине ошибки. Так "Делойт & Туш" предлагает брать 70–90% уровня существенности[1]. В руководстве по банковскому внутреннему аудиту[2] рекомендуют ограничиться 50% этого уровня. Как видим, размах довольно широкий.

§ 1.3. Аудиторский риск

Подготовка финансовой отчетности и прочей документации, регистрация различных событий, имеющих отношение к хозяйственной и коммерческой деятельности, и тому подобные действия осуществляются людьми, которым, как свидетельствует вся история человечества, увы, свойственно ошибаться. Ошибки могут быть случайными (пропуски при заполнении документов, неправильное их оформление, арифметические ошибки и т.д.) и, следовательно, не иметь явной тенденции в сторону преуменьшения или преувеличения каких-либо параметров. Значительно хуже, если речь идет не о случайных погрешностях, а о сознательных (тенденциозных) искажениях информации, направленных на маскировку хищений, мошенничества, плохого управления. Однако и сам аудитор может не заметить случайную или тенденциозную ошибку в документе, пропустить важную информацию, наконец, сделать неправильный вывод на основе полученной правильной

[1] DRT International. Op. cit. P. 33.
[2] *Giliberti A.* Op. cit P. 65-3, 65-4.

информации и т.д. Аудитор также может допускать ошибки. Вероятность того, что аудитор сделал неверное заключение (по любой из названных выше причин), обычно называют аудиторским риском (audit risk). Считается, что незначительный уровень аудиторского риска допустим.

В западной литературе для описания механизма формирования величины аудиторского риска (*AR*) последний представляют в виде произведения трех вероятностей, относящихся к различным элементам процесса тестирования:

$$AR = IR \cdot CR \cdot DR,$$

где *IR* – риск содержания ошибок в представленных сведениях; такой риск обычно называют *внешним риском* (inherent risk), он отражает вероятность того, что внешние условия (не деятельность аудитора) могут привести к существенным ошибкам;

CR – *риск контроля* (control risk), т.е. вероятность того, что система внутреннего контроля не предупреждает появление существенных ошибок;

DR – *риск обнаружения* (detection risk) измеряется вероятностью того, что аудитор не обнаружит имеющуюся существенную ошибку.

Наряду с аудиторским риском применяют и противоположное понятие – *аудиторская надежность* (audit assurance)[1]. Очевидно, что между аудиторским риском (*AR*) и аудиторской надежностью (*AS*) существует простая зависимость: $AR = 1 - AS$.

Рассмотренная выше модель аудиторского риска приводится в ряде западных руководств по аудиту. Однако она не является единственно возможной. Существуют и другие варианты такой модели[2]. Например, компоненты рассмотренной модели разлагаются на множители, отражающие влияние выделенных факторов. В отечественных публикациях также предлагаются мультипликативные модели риска, причем выделяются десятки (!) учитываемых факторов. Приведем некоторые из них. Например, риск, вызванный отсутствием в фирме четкого разделения обязанностей и ответственности, риск утраты данных, риск, вызванный использованием нелегальных копий системы электронной обработки данных, и т.п.

[1] DRT International. DRT International Audit Approach.1990. P. 256.

[2] Один из вариантов модели представлен в обширном руководстве по банковскому внутреннему аудиту: *Giliberti A.* Op. cit. V. 2. P. 65-6. Другой вариант обсуждается в учебном пособии: ACCA. Audit Framework. Examination Text. AT Foulks Lynch Ltd. 1993.

Приведенные выше и другие модели такого вида, на наш взгляд, являются чисто умозрительными и не могут быть применены на практике, и прежде всего потому, что используемые здесь вероятности (а каждому фактору соответствует некоторая вероятность) не могут быть однозначно и объективно оценены. Ну как, например, может быть определен уровень вероятности, характеризующей внешний риск, или утраты данных? Только интуитивно. Сказанное относится и к другим элементам модели. Да и само произведение указанных вероятностей, строго говоря, не имеет теоретического обоснования. Скорее подобного рода модели надо рассматривать как указание на то, что перечисленные факторы совместно (мультипликативно) влияют на величину риска, а изменение влияния любого из них увеличивает или сокращает величину аудиторского риска. Рекомендации по применению балльных оценок для оценивания субъективных вероятностей риска нисколько не устраняют практическую бесполезность указанных моделей.

Можно полагать, что такого рода модели приводятся скорее для придания некоторой "научности" обсуждению проблемы, чем для практического применения. Вместе с тем руководства по аудиту обычно содержат практические рекомендации, нацеленные на сокращение выделенных элементов аудиторского риска. Однако в таких руководствах не демонстрируется, да и не может быть показана сама возможность объективного их измерения.

Выше мы не останавливались на ошибках и соответствующем риске, связанных с применением выборочного контроля вместо сплошного. Как уже говорилось, выборочное тестирование резко сокращает издержки на проведение проверки, позволяет заметно сократить сроки его проведения и т.д., вместе с тем оно неизбежно связано с некоторыми погрешностями в получаемых результатах. Дело в том, что при выборочном наблюдении получают не точные значения необходимых параметров для тестируемой совокупности, а их *оценки*. Однако погрешности (ошибка выборки) в определенных условиях имеют объективное измерение, и их можно корректировать, связав с приемлемыми уровнями вероятностей. Эти вероятности можно рассматривать как один из элементов модели аудиторского риска.

§ 1.4. Цели и методы проведения выборочного обследования

Теория выборочных обследований и методики их применения разработаны весьма полно, в основном применительно к контролю качества продукции[1] (оценивание доли брака или, наоборот, доли продукции высшего качества), оценке видов на урожай и т.п. Широкое распространение получили выборочные методы в различных медико-биологических исследованиях и социологических обследованиях. Словом там, где обследуемая совокупность слишком велика для того, чтобы полностью охватить ее, или там, где контроль связан с разрушением проверяемого объекта (в некоторых видах технического контроля). Вместе с тем для практических или научных целей необходимы некоторые параметры, характеризующие совокупность в целом. Заметим, что во всех указанных областях объекты проверки, как правило, не измеряются по стоимости[2]. Аудитор же в основном имеет дело со стоимостными характеристиками. Соответственно в аудите в зависимости от специфики изучаемых объектов и целей проверки применяются как стандартные выборочные методы, так и свои, специальные методы выборочного контроля, которые нигде более не встречаются.

До характеристики конкретных целей и методов выборочного обследования необходимо ввести несколько специальных терминов, относящихся к теории выборочного метода, с учетом того, что эти методы применяются в аудите.

Единицы наблюдения в аудите – письменные свидетельства выполненных сделок и операций, договоры, учетные регистры, счета, платежные поручения, накладные и прочее (далее для краткости – документы), которые содержат сведения о проверяемых аудитором фактах и событиях.

[1] *Шторм Р.* Теория вероятностей. Математическая статистика. Статистический контроль качества. М.: Мир, 1970.

[2] По теории и методике применения выборок имеется множество публикаций на русском языке. Однако рассматриваемые в них методы не распространяются на стоимостные единицы наблюдения. Назовем некоторые из книг: *Дружинин Н.К.* Выборочный метод и его применение в социально-экономических исследованиях. М.: Статистика, 1970; *Шварц Г.* Выборочный метод. М.: Статистика, 1978. К более фундаментальным относятся книги: *Кокрен У.* Методы выборочных исследований. М.: Статистика, 1976; *Джессен Р.* Методы выборочных обследований. М.: Финансы и статистика, 1985.

Исследуемая (проверяемая, тестируемая) совокупность, или, кратко, совокупность, – все множество единиц наблюдения (в теории статистики такое множество обычно называют *генеральной совокупностью*), относительно которого необходимо, основываясь на выборке, получить обобщение в виде выборочной оценки.

Выборочная совокупность, или *выборка*, – отобранные для тестирования единицы наблюдения.

Отбор – процедура извлечения единиц наблюдения (документов) из тестируемой совокупности.

Оценка (выборочная оценка) – размер некоторого параметра (средней суммы, доли и т.д.), полученный по данным выборки. Выборочная оценка параметра для тестируемой совокупности несколько отличается от неизвестного истинного его значения.

Процедура оценивания (оценивание) – процесс вычисления оценки по выборочным данным.

Объем выборки – количество единиц наблюдения, попавших в выборку.

Доверительная вероятность (уровень надежности) – вероятность, достаточная, по мнению аудитора, для суждения о достоверности выборочной оценки параметра.

Точность оценки (или, наоборот, *погрешность оценки*) – близость оценки к истинному значению параметра. Оценка может быть весьма точной, но иметь низкую надежность, и наоборот.

Остальные термины будут объяснены по ходу рассмотрения соответствующих проблем.

Четкое определение единиц наблюдения и границ совокупности – важный шаг в применении выборочного метода. Тестируемая совокупность должна охватывать все единицы наблюдения, которые связаны с целью проверки и в одинаковой степени доступны для аудитора. Приведем пример. Пусть целью аудита является проверка законности совершаемых банком операций с позиций выполнения функций агента валютного контроля. Тогда единицами наблюдения являются досье клиентов-экспортеров или паспорта сделок и т.п., а все их множество представляет тестируемую совокупность.

В том случае, когда аудитор считает, что аудиторский риск зависит, например, от географического размещения единиц наблю-

дения, отрасли кредитования и т.п., тестируемая совокупность может быть разбита на соответствующие подмножества (слои, группы, страты). Каждое такое подмножество рассматривается как самостоятельный объект тестирования.

По целям и условиям проведения выборочные обследования можно разделить на две группы – репрезентативные и нерепрезентативные. Оба вида применяются в аудите.

Целью *репрезентативного* выборочного исследования (representative sampling) является получение (оценивание) характеристик генеральной совокупности по выборочным данным. Например, оценивание по выборке общей суммы на счетах, оценивание суммарной стоимостной ошибки и т.п. Иначе говоря, *полученная в ходе выборочного тестирования информация может быть распространена без больших погрешностей на всю проверяемую совокупность.* В таких случаях говорят о *репрезентативной (представительной) выборке.* Далеко не всегда выборка является репрезентативной. Выборка будет репрезентативной только тогда, когда она получена специальными методами при соблюдении определенных условий, основным из которых является равная вероятность попадания в выборку для каждой единицы наблюдения совокупности. Подробнее об этом будет сказано в гл. 2.

В аудите применяют и *нерепрезентативный отбор* (nonrepresentative sampling). Он также заключается в отборе некоторой части единиц наблюдения из всей исследуемой совокупности. Однако цели здесь иные, и отбор производится по другим принципам, чем при репрезентативном отборе. Нерепрезентативный отбор обычно производится на основе *субъективных суждений* аудитора. Основанием для отбора могут служить:

- интуиция, или "чутье", аудитора в отношении "подозрительности" документов или их источников, опыт прошлых проверок, дополнительная информация со стороны и т.п.;
- предположение аудитора о том, что отобранные единицы наблюдения наиболее характерны для проверяемой совокупности;
- ориентированность аудитора на проверку наиболее крупных по стоимости документов.

Не исключены случаи, когда отбор производится произвольно – без соблюдения требования равной вероятности отбора для каждой единицы совокупности.

Сформировав тем или иным путем выборочную совокупность, аудитор при нерепрезентативном отборе использует полученные данные для непосредственного суждения о соответствующих процессах или характеристиках только в рамках выборки. Он не имеет основания для их распространения на исходную генеральную совокупность.

Часто это выглядит следующим образом. Аудитор из всей совокупности счетов отбирает, например, счета, поступившие только от фирмы D. Обнаруживает, что d% документов в выборке оказалось с ошибками. Полученный результат относится только к выборочной совокупности (в данном случае к счетам фирмы D). Как видим, речь не идет об оценивании данных генеральной совокупности. Такой отбор называют направленным. Он решает определенные задачи, поставленные аудитором, однако, повторяем, он ограничен по своим возможностям, так как здесь нет оценки параметров тестируемой совокупности в целом.

В связи со сказанным уместно отметить, если так можно сказать, некоторую несимметричность получаемых результатов при нерепрезентативном отборе. Действительно, если выборка обнаруживает отрицательные факты, то для аудитора это уже данные для заключения. Однако если направленный отбор, да еще при небольшом объеме выборки не обнаружил отрицательные факты (например, ошибки или приписки в тестируемых документах), то такой результат мало что дает для аудитора.

В практике отечественного аудита в основном применяют направленный отбор, который, как было показано выше, является нерепрезентативным. В какой-то мере это можно объяснить тем, что такой отбор не имеет статистико-теоретического обоснования и поэтому очень прост при реализации. Не менее важным является и то, что применяемые здесь понятия не выходят за рамки бухгалтерского учета. В связи с этим он не вызывает вопросов у бухгалтера, и им можно пользоваться в любой ситуации. Однако за простоту, как известно, надо платить. Платой является получение крайне ограниченной информации и, главное, невозможность ее распространения на совокупность. Иное дело – репрезентативная выборка, которая дает аудитору обобщенные данные, относящиеся к тестируемой совокупности. Вместе с тем для осуществления подобного рода выборок аудитору надо быть знакомым с рядом статистичес-

ких понятий и методов анализа, которые никак не связаны с бухгалтерией. Ну как можно увязать с бухучетом (и соответствующим менталитетом) такие понятия, как "доверительная вероятность", "оценка интервала" и т.п. Последнее обстоятельство может вызвать естественное внутреннее сопротивление аудитора. Далее в книге обсуждаются только методы репрезентативных выборок.

Итак, выборочный метод в аудите кратко сводится к следующим моментам:

- извлечению из тестируемой или генеральной совокупности некоторой ее части (выборки);
- проверке по заданной программе единиц наблюдения (документов), попавших в выборку;
- обобщению результатов проверки и распространению выборочных результатов на совокупность.

Специфика исследуемых совокупностей и конкретных целей аудита определяет необходимость применения различных методов извлечения выборок и их последующего анализа. В теории и практике аудита обычно выделяют следующие виды выборочных исследований, различающихся по целям и соответственно по способам формирования выборок и методикам анализа:

- атрибутивные выборки;
- оценивание средней и суммы стоимостных показателей для совокупности;
- оценивание размера средней абсолютной ошибки для совокупности;
- оценивание размера средней относительной ошибки для совокупности.

Для проверок на соответствие, о которых говорилось выше, применяют атрибутивные выборки (attribute sampling). Этот метод необходим в ситуациях, когда аудитора интересует частота наступления некоторых случайных событий. Речь идет о фиксации в выборке наличия (или отсутствия) некоторого свойства, атрибута. Примерами могут служить: проверка наличия на документах разрешительной подписи, полноты комплекта документов по сделке, выявление ошибок в оформлении и т.д. Таким образом, речь идет о случаях, когда имеется два исхода – есть событие или

его нет (да, нет). Пока ограничимся этим кратким замечанием. Развернутому обсуждению проблем, относящихся к атрибутивным выборкам, посвящена гл. 4.

В свою очередь, при проверках по существу внимание аудитора нацелено на содержательную сторону документа – его стоимостные характеристики. Здесь применяются следующие виды выборочного контроля.

Оценивание средней и суммы стоимостных показателей исследуемой совокупности (mean-per-unit estimation). Необходимость в оценивании средней в основном возникает в тех случаях, когда общий размер исследуемой совокупности в денежном измерении неизвестен, поскольку стоимостные характеристики единиц наблюдения не регистрировались или они существуют, но имеются сомнения в истинности размера этой суммы, представленной аудитору. Аудитору требуется оценка этой суммы с заданной надежностью и приемлемой погрешностью.

Для решения данных задач применяют классический выборочный метод и некоторые специальные методы, разработанные в аудите для проверки стоимостных показателей. Соответствующие методики обсуждаются в гл. 5.

Оценивание абсолютных и относительных ошибок (difference estimation, ratio estimation) *в стоимостных показателях документов.* Необходимые оценки получают на основе выборочных данных с помощью специальных методов. Этому посвящена гл. 6.

Методы формирования выборочной совокупности для проведения аудита рассматриваются в гл. 2. Поскольку определение объема выборки является исключительно важным этапом выборочного обследования, соответствующие методики обсуждаются в главах, посвященных конкретным видам таких обследований. В связи с тем, что анализ выборочных данных и получение соответствующих оценок для проверяемой совокупности основан на распределении вероятностей случайных величин, в гл. 3 приводятся краткие характеристики нормального, биномиального, гипергеометрического распределений и распределения Пуассона.

Несомненно, что рассмотренные в книге методы можно конкретизировать и развивать с учетом особенностей проведения аудита в той или иной области (аудит банковской деятельности, ком-

мерческого предприятия, снабженческой организации и т.д.) и целей проверки. Очевидно, что такая конкретизация и соответствующие уточнения должны базироваться на накопленном опыте применения выборочных обследований.

В практике аудита могут применяться и другие, более сложные статистические методы. В частности, в последнее десятилетие аудиторы обратили внимание и на один из классических методов статистического анализа – проверку гипотез. Кратко суть такой проверки заключается в формулировании аудитором некоторой гипотезы о величине параметра проверяемой совокупности документов (сумме или средней стоимостных характеристик или размере ошибки). Сформулированная гипотеза проверяется по определенным правилам на основе выборочных данных. Она или принимается, или отвергается[1].

В последнее время в зарубежных публикациях по аудиту упоминаются случаи оценивания неизвестных параметров совокупности на основе регрессий[2] и байесовских методов[3]. Однако сомнительно, что в российской практике в ближайшее время упомянутые методы привлекут внимание отечественных аудиторов. Они и не рассматриваются в настоящей работе. Это дело будущего.

В связи с применением в аудите методов математической статистики возникла одна проблема методологического порядка. Дело в том, что аудитор обычно принадлежит, так сказать, к бухгалтерскому цеху. Как известно, бухгалтер имеет дело с величинами "с точностью до копейки" и не использует вероятности и случайные события. В статистическом же анализе, наоборот, вероятности и интервалы значений исследуемых параметров – обычные характеристики. В связи со сказанным при применении в аудите методов математической статистики приходится совмещать понятия из обеих областей знания.

В зарубежной аудиторской практике ориентируются на различные руководства по аудиту, разработанные в крупных аудиторских компаниях. В этих руководствах, как правило, уделяется определенное внимание конкретным методикам выборочного анализа. Характерной чертой таких руководств, по крайней мере тех, с которыми удалось ознакомиться авторам настоящей работы, является почти полное отсутствие теоретических обоснований предлагаемых методов. Обычно в этих пособиях указываются конкретные правила, похожие на рецепты в поваренной книге. В настоящей работе сделана попытка увязать методы выборочных обследований, применяемые в аудите, с рядом положений математической статистики и теории выборок.

[1] *Четыркин Е.М., Калихман И.Л.* Вероятность и статистика. М.: Финансы и статистика, 1982. Гл. 10.

[2] Подробную характеристику соответствующих методов можно найти в монографиях: *Кокрен У.* Указ. соч.; *Джессен Р.* Указ. соч.

[3] DRT International. Audit Sampling. Op. cit. P. 65–73.

Глава 2. СПОСОБЫ ИЗВЛЕЧЕНИЯ ВЫБОРОК В АУДИТЕ

Необычные случаи обычно повторяются.

К. Чапек

§ 2.1. Стандартные способы формирования выборки. Таблица случайных чисел

Отбор единиц наблюдения в выборку производится различными способами. Сгруппируем виды или типы отбора по принципам, на которых они основаны. В этом отношении прежде всего следует различать *случайный* и *направленный* отбор, о чем уже упоминалось в § 1.4.

Случайный, или вероятностный, отбор (в западной литературе по аудиту такой отбор обычно называют *статистическим* – statistical sampling) свободен от субъективных суждений аудитора. Такой отбор предполагает равную вероятность для каждой единицы наблюдения (документа), из которых состоит совокупность попадания в выборку. Ни одна единица наблюдения не должна иметь каких-либо преимуществ при извлечении из тестируемой совокупности. Полученные по выборке параметры – доли (например, доля документов с обнаруженными ошибками) или средние стоимостные характеристики (средняя сумма остатков на дебиторских счетах) – распространяются на проверяемую совокупность. Иначе говоря, выборочные оценки являются представителями (оценками) аналогичных по содержанию показателей исследуемой совокупности. Такие выборки, как уже упоминалось выше, называют репрезентативными. Именно репрезентативные выборки рассматриваются в данной и следующих главах книги.

Там, где целесообразно и возможно применение случайной выборки, необходимо наличие некоторых условий, а именно:

- проверяемая совокупность должна быть более или менее однородной по содержанию;

- совокупность должна быть достаточно большой, так как из небольшой нет смысла извлекать выборку;
- ожидания ошибок должны быть незначительны;
- единицы наблюдения легко различимы и одинаково доступны для контроля.

Рассматривая типы отбора единиц наблюдения в выборочную совокупность, нельзя не остановиться на так называемых *схемах повторного* и *бесповторного отбора*. Пусть отобранная единица наблюдения после обследования возвращается в исследуемую совокупность и, следовательно, может быть вновь отобрана (ситуация маловероятная, точнее, невозможная в аудите). Такую схему называют повторным отбором или отбором по схеме возвращенного шара. Происхождение последнего названия связано с отбором шаров из урны – классической, умозрительной модели, принятой в теории вероятностей. Бесповторный отбор, или отбор по схеме невозвращенного шара, не предусматривает возврата отобранной единицы наблюдения. Эта схема и применяется в аудите. Использование более простых формул, выведенных для повторного отбора, в выборках с бесповторным отбором, естественно, приводит к некоторым смещениям или искажениям получаемых оценок. Однако такими смещениями в большинстве случаев (при достаточно большой выборке) можно пренебречь.

Основное требование случайного отбора – равенство вероятностей попадания в выборку для каждой единицы исследуемой совокупности – обеспечивается несколькими методами извлечения выборок. К числу стандартных относятся следующие методы отбора:

- собственно-случайный;
- механический, или систематический;
- расслоенный, или стратифицированный;
- серийный, или групповой.

Специально для аудита разработаны такие методы, как:

- отбор с расслоением проверяемой совокупности на две группы;
- монетарная выборка – отбор с вероятностями попадания в выборку, пропорциональными стоимостным размерам единиц наблюдения.

Заметим, что при любом методе формирования выборки документы с наиболее крупными стоимостными характеристиками,

превышающими порог существенности, обычно подлежат сплошной проверке. Таким образом, выборочная проверка распространяется только на единицы наблюдения, которые не превышают этот порог.

Для начала кратко охарактеризуем особенности перечисленных стандартных способов отбора.

Собственно-случайный отбор (random selection). Такой отбор представляет собой наиболее простой вид выборки. Он не требует предварительного упорядочения единиц наблюдения. Единственное требование – обеспечение равенства вероятностей попадания в выборку для каждой единицы наблюдения.

Подлежащая тестированию совокупность сведений фиксируется на различного вида носителях информации – документах, регистрах, карточках и т.д. (далее – документах) или на электронных носителях. Хранение этих документов обычно каким-то образом упорядочено. Как правило, документы нумеруются в порядке поступления или регистрации. Если нумерации нет, то ее придется выполнить специально для собственно-случайного или механического отбора документов.

Для обеспечения случайности попадания документа в выборку и соответственно для исключения субъективности в отборе применяют *таблицы случайных чисел*. Названные таблицы публикуются в различных статистических справочниках. Они содержат n-значные ($n = 2, ..., 5$) случайные числа. Ниже помещена таблица, содержащая шесть столбцов равномерно распределенных пятизначных случайных чисел, полученная на основе электронной таблицы Excel. Ее продолжение (столбцы 7–36) помещено в Приложениях в конце книги.

Равномерно распределенные случайные числа могут быть получены и непосредственным их генерированием с помощью электронной таблицы Excel[1].

Для применения таблиц случайных чисел необходимо, чтобы носители информации (документы) имели сквозную нумерацию. Причем нумерация не должна быть связана с размерами признака, который представляет интерес для аудитора.

При отборе по таблице случайных чисел установленного заранее количества документов (объема выборки) поступают следующим образом:

[1] См. Дополнение к гл. 2.

- определяют "направление" отбора (по столбцу, строке, а иногда и по диагонали);
- произвольно выбирают столбец, строку и стартовое случайное число;
- из выбранного столбца (строки, диагонали) последовательно отбирают числа с нужным количеством цифр[1].

Если какое-либо случайное число встретится дважды, то второй раз оно не учитывается.

ПРИМЕР 2.1. Необходимо случайным образом отобрать 50 документов из 500. Порядковые номера документов от 1 до 500. По таблице случайных чисел (табл. 2.1) определим 50 случайных чисел в пределах от 1 до 500. Эти числа отберем, например, из шестого столбца таблицы случайных чисел (первая строка, первые три цифры). Приведем только десять отобранных случайных чисел этого столбца:

69024	60084
13763	98399
05590	50830
76436	45338
48644	71788

Таким образом, в выборку попадут документы с номерами: 137, 055, 486, 453 и т.д. В свою очередь, если отбирать последние три цифры в этом столбце, то получим следующую последовательность: 024, 436, 084, 399, 338 и т.д.

Таблица 2.1

Таблица случайных чисел

Строки	Столбцы					
	1	2	3	4	5	6
1	25004	81784	75721	16766	08532	69024
2	53095	42087	40548	29755	88902	13763
3	98625	40930	42859	91541	18468	05590
4	01872	16290	94369	06025	98919	76436
5	84254	37701	48958	12558	58293	48644
6	84689	83969	82126	14658	07056	60084
7	56811	61586	01603	84277	93561	98399

[1] Например, если отбирают двузначные числа из таблицы с четырехзначными случайными числами, то выбираются первые или последние две цифры.

Окончание табл. 2.1

Строки	Столбцы					
	1	2	3	4	5	6
8	05394	41420	64914	75710	26602	50830
9	68458	45774	57369	65537	85729	45338
10	39387	54019	66331	00972	52426	71788
11	73908	09180	12442	50513	93486	13244
12	91864	14843	71306	80781	69979	59001
13	45233	15541	94888	40441	86236	75488
14	11399	46145	33075	36026	62022	35489
15	06610	93248	92707	59095	16173	14224
16	43457	66930	06918	88421	63626	86849
17	85191	24892	23389	77671	60448	58990
18	84568	03132	17277	88453	74429	70171
19	12030	12391	23856	43764	50598	04860
20	95019	01117	89806	63045	96766	05618
21	72180	99252	92982	25860	23522	16485
22	37267	34153	07414	44536	58163	19085
23	91192	79276	46427	37158	83201	24556
24	27971	85002	48067	42809	91836	10312
25	52519	27159	07240	78046	32864	71399
26	71770	20237	09133	56989	61311	69936
27	98325	26034	72305	72656	78500	68807
28	51710	79863	81423	84173	22774	78811
29	24372	19377	40902	66014	74610	85363
30	80966	11563	52855	13087	09571	43019
31	84607	34443	30273	64261	37159	47319
32	77001	47506	92804	34505	59079	81513
33	98550	92463	15932	12035	66652	79498
34	81454	31174	92782	00089	74145	52638
35	41109	67765	83347	35227	18442	99908
36	52559	70811	26112	53559	94535	25147
37	25334	85918	49879	20836	71031	48096
38	31751	52979	70230	75537	07347	71672
39	17610	83851	62189	20949	13705	56816
40	01928	86771	77657	76546	33831	13747
41	92179	30442	29285	13920	68872	81738
42	62250	48299	12252	58486	04504	60379
43	35387	15998	86722	19994	38356	10988
44	16810	49516	99575	27816	86835	42670
45	80147	28241	46931	71075	07346	58170
46	70434	91808	84887	40542	56106	58511
47	97754	97523	34739	90043	01310	54432
48	83060	34616	65280	35636	85838	83023
49	36948	62775	51710	83314	14883	64636
50	78601	33455	73232	02907	94652	48170

Механический, или систематический, отбор (systematic sampling). По этому методу совокупность делится на столько групп или интервалов, сколько единиц наблюдения необходимо отобрать в выборку. Из каждой группы отбирают по одной единице. Например, если исследуемая совокупность охватывает 1000 единиц и из нее необходимо отобрать 40 документов – по 4 единицы из каждых 100, то в выборку попадают единицы наблюдения с интервалом, или шагом отбора, равным 25. Первый номер такой последовательности (в пределах 1–25) выбирают из первого интервала случайным образом (механический отбор со случайным началом). Если таким номером оказалось 4, то последовательно отбирают документы с номерами 29, 54, 79 и т.д.

Механический отбор часто легче реализовать, чем собственно-случайный. Его можно осуществить вручную или на компьютере. Если документы систематизированы в алфавитном порядке, то это не окажет влияния на случайность отбора. Однако равенство вероятностей попадания в выборку для любых единиц наблюдения может быть обеспечено только в случае, когда в их последовательности не наблюдается цикличности. Так, если в условиях нашего примера каждая 25-я единица обладает каким-то особым качеством (например, относится к одному объекту, к одному специфическому виду сделок и т.д.), то существует возможность для нарушения требования репрезентативности. Если же единицы наблюдения ранжированы по величине изучаемого признака, то в выборку попадут документы с систематически увеличивающимися показателями, что неминуемо приведет к смещениям в размерах выборочных оценок.

ПРИМЕР 2.2. Имеются данные в рублях о счетах-фактурах (всего 120 счетов), приведенные в табл. 2.2. Условно будем считать, что это генеральная совокупность. Пусть из нее для тестирования необходимо отобрать 18 счетов. Разумеется на практике из совокупности такого объема нет смысла извлекать выборку. Проще тестировать ее полностью. Однако в иллюстративных целях все же извлечем случайную выборку.

Для отбора методом собственно-случайной выборки (бесповторный отбор) пронумеруем счета (от 1 до 120) и воспользуемся приведенной выше таблицей случайных чисел и ее продолжением в Приложениях. В качестве случайного числа будем рассматривать три последние цифры пятизначного числа столбца в пределах от 1 до 120. Пусть начальным будет первое число столбца 6 этой таблицы. Нахо-

дим следующие случайные числа в этом столбце: 024, 084, 001, 085, 019, 024. Так как их всего шесть, а нам нужно восемнадцать, продолжим выбор случайных чисел. В столбцах 7 и 8 (см. Приложения) последовательно находим числа 030, 077, 106 и т.д. В выборку попадут счета с указанными номерами. Отобранные счета помечены в табл. 2.2 значком *.

Таблица 2.2

Суммы на счетах, руб.

Номер по порядку	Сумма	Номер по порядку	Сумма	Номер по порядку	Сумма
1	*1430	41	*8457	81	1587
2	684	42	1254	82	600
3	*1354	43	558	83	1525
4	4420	44	*64	84	*2300
5	4800	45	12548	85	*476
6	8742	46	1658	86	4123
7	7506	47	987	87	6324
8	876	48	12357	88	8450
9	46353	49	147	89	985
10	9250	50	6984	90	4750
11	3225	51	785	91	3655
12	48	52	7564	92	11450
13	3254	53	4531	93	174
14	1235	54	957	94	8666
15	854	55	875	95	35411
16	4457	56	1457	96	*99
17	658	57	3654	97	23300
18	5751	58	4800	98	1250
19	*24311	59	12657	99	4213
20	68	60	97	100	3564
21	126	61	5874	101	*78
22	14235	62	9666	102	542
23	*4568	63	33546	103	1358
24	*1750	64	1587	104	*9634
25	2316	65	*5478	105	8542
26	2547	66	132	106	*4450

Окончание табл. 2.2

Номер по порядку	Сумма	Номер по порядку	Сумма	Номер по порядку	Сумма
27	1247	67	4458	107	1352
28	6322	68	1324	108	775
29	4758	69	9854	109	3100
30	*1875	70	64	110	4521
31	6549	71	11202	111	3659
32	124	72	4357	112	875
33	27432	73	3256	113	14531
34	*11654	74	1524	114	112
35	5789	75	474	115	957
36	857	76	*15249	116	5488
37	3214	77	*684	117	3698
38	87	78	5478	118	1478
39	9480	79	2549	119	454
40	358	80	3562	120	17245

Средняя сумма счета при собственно-случайном отборе составляет 5214 руб. Во всей совокупности (120 счетов) средняя равна 5418 руб. При других вариантах определения случайных чисел (по другим столбцам таблицы случайных чисел) в выборке каждый раз окажутся другие счета, и, естественно, выборочные средние будут иными. Однако все они будут более или менее близки к средней, полученной для совокупности.

Механический отбор 18 счетов осуществим еще проще. Для получения шага отбора разделим объем совокупности на размер выборки: 120/18=6,7. Шаг равен 6. В выборку сначала случайным образом отберем первый номер счета. Им оказался счет 7. Далее в выборку отбираются счета с номерами, увеличивающимися на 6 , т.е. 13, 19, 25 и т.д. Средняя, полученная при механическом отборе, равна 5428 руб.

Расслоенный, или стратифицированный, отбор (stratified selection)[1]. Два рассмотренных выше метода случайного отбора, строго говоря, применимы в случаях, когда предполагается, что

[1] В отечественной статистической литературе такой отбор обычно называют типологическим или районированным.

проверяемые стоимостные характеристики единиц наблюдения не слишком разнятся между собой. Если же аудитор ожидает, что стоимостные характеристики имеют большой разброс или распределение единиц наблюдения по стоимости имеет заметную асимметрию (например, ожидается, что в совокупности преобладают документы с небольшими суммами), а также есть другие основания для сомнений в однородности совокупности, то для повышения точности тестирования применяют более сложные виды формирования выборки. К ним, в частности, относятся расслоенный отбор, согласно которому обследуемая совокупность первоначально делится на слои (группы, страты). Затем из каждой группы в случайном порядке (собственно-случайный или механический отбор) производится извлечение выборочных единиц наблюдения. В аудите в основном имеют дело с денежными показателями. Соответственно совокупность делится на интервалы или слои по величине стоимостных показателей.

Число слоев, границы между ними и распределение общего объема выборки между слоями устанавливаются аудитором. Подробнее об этом см. § 5.2. Переход от собственно-случайного к расслоенному отбору позволяет учитывать наиболее значимые для аудита слои документов.

На практике применяют несколько видов расслоения. Назовем некоторые из них.

Пропорциональное расслоение (размещение) выборки по слоям. Иначе говоря, доля отбора одинакова в каждом слое. Например, по 10% от количества документов в слое. Таким образом, количество документов, извлеченных в выборку из каждого слоя совокупности, пропорционально размеру слоя.

Непропорциональное расслоение по слоям. Доли отбора здесь изменяются от слоя к слою. Непропорциональное расслоение реализуется различными способами. Наиболее простым является отбор равных долей выборки из каждого слоя. Например, при 5 слоях из каждого слоя извлекается 1/5 установленного общего объема выборки.

Важным частным случаем непропорционального расслоения является так называемое *оптимальное расслоение* выборочных единиц наблюдения по слоям. Такой метод позволяет во многих случаях получить наибольший объем информации при фиксированных затратах труда на извлечение выборки. Оптимальным является расслоение объема выборки пропорционально взвешенным стандартным отклонениям по слоям. В качестве весов здесь берут доли слоев в совокупности. Методику оптимального размещения объема выборки по слоям целесообразно рассмотреть при определении общего объема выборки (см. § 5.2).

Расслоение, естественно, связано с некоторыми усложнениями как при извлечении выборки, так и при ее анализе. К тому же здесь необходима некоторая дополнительная информация. Однако почти всегда расслоение приводит к повышению точности полученных результатов по сравнению с оценками, полученными без расслоения.

Продемонстрируем распределение объемов выборки по слоям (стоимостным интервалам) на примере.

ПРИМЕР 2.3. Пусть единицы наблюдения характеризуются большим разбросом стоимостных параметров – от нескольких тысяч рублей до сумм, превышающих 1 млн руб. Аудитор принял решение проверить все документы со стоимостью 1 млн руб. (порог существенности) и выше. Остальные документы проверяются на основе выборки. Поскольку диапазон стоимостных показателей документов велик, документы распределены по четырем стоимостным интервалам (слоям) (см. первый столбец табл. 2.3).

Допустим, по каким-то соображениям необходимо отобрать 200 документов со стоимостными параметрами ниже порога существенности из общего их количества, равного 3700. Для этого есть два варианта – в случайном порядке отобрать это количество документов без учета стоимостных характеристик в каждом слое (пропорциональное размещение) и с их учетом. Во втором случае имеем непропорциональное числу документов размещение.

В столбце А графы "Объемы выборки" табл. 2.3 показаны размеры выборки, пропорциональные количеству документов в каждом интервале (5,4% от количества документов в слое).

При распределении объема выборки по интервалам с учетом стоимостных характеристик примем следующие условия. Пусть суммы стоимостных показателей по интервалам неизвестны (правдоподобная ситуация). Поэтому предварительно необходимо приближенно оценить эти величины. Для этого центральные значения каждого стоимостного интервала (показаны в скобках) умножим на количество документов, которые он включает.

Оценка общей суммы стоимостных показателей равна 345 млн руб. Следовательно, на каждый миллион рублей этой суммы приходится 200/345 ≈ 0,58 единицы выборки. В этом случае объем выборки для первого интервала составит 0,58 · 20 ≡ 12, для второго интервала 0,58 · 60 ≡ 35 и т.д. Соответствующие данные приведены в столбце B табл. 2.3. В столбце C показаны объемы выборки по слоям, полученные при равномерном размещении.

Таблица 2.3

Распределение объема выборки по интервалам

Стоимостные показатели, тыс. руб.	Количество документов N_k	Оценка суммы в интервале, млн руб.	Объемы выборки		
			A	B	C
0,1–20 (10)	2000	20	108	12	50
20–100 (60)	1000	60	54	35	50
100–400 (250)	500	125	27	72	50
400–1000 (700)	200	140	11	81	50
Всего	3700	345	200	200	200

Как видим, даже весьма приближенный учет стоимостного фактора (столбец B) заметно изменил распределение объемов выборки по слоям в пользу единиц наблюдения с более крупными стоимостными характеристиками. Объемы выборки по слоям, полученные для равномерного размещения (столбец C), занимают промежуточное место по своим размерам между вариантами A и B.

Серийный, или гнездовой, отбор с равновеликими сериями (serial selection, block sampling). Исследуемая совокупность делится на равновеликие "серии", или "гнезда". Выборка состоит из случайным образом отобранных серий, каждая из которых подвергается сплошной проверке. В практике аудита такой отбор пока не нашел применения. Сказанное вовсе не означает, что этот метод не может применяться в аудите. Вполне возможны ситуации, когда он целесообразен[1].

[1] С методикой серийного отбора можно ознакомиться по книгам: *Джессен Р.* Указ. соч. Гл.11; *Кокрен У.* Указ. соч. Гл. 9.

§ 2.2. Специальные способы отбора единиц наблюдения

Двухслойная выборка (two strata selection, TS). Это упрощенный вариант расслоенного отбора (отбор с двумя слоями), применяемый, вероятно, только в аудите. Порядок формирования выборки сводится к следующему. Устанавливается некоторый порог существенности в денежном измерении. Документы совокупности, не превышающие этот порог, тестируются в выборочном порядке. Документы, подлежащие выборочной проверке, распределяются по двум слоям. Учитывая то, что распределение единиц наблюдения по стоимостным показателям обычно является заметно асимметричным (преобладание документов с относительно небольшими стоимостными характеристиками), граница между слоями определяется как удвоенная средняя стоимостная характеристика документа[1]. Выборка документов из каждого слоя производится по таблице случайных чисел или с помощью механического отбора.

ПРИМЕР 2.4. Вернемся к примеру 2.3 и распределим тот же объем выборки (200) между двумя слоями. Пусть средняя стоимостная характеристика документа известна из предыдущих проверок и равна примерно 110 тыс. руб. Применяя приведенное выше правило, получим, что граница между двумя слоями составляет 110 · 2 = 220 тыс. руб. Соответствующие размещения количества документов по слоям показаны в табл. 2.4. Как и в примере 2.3, в графе A показаны объемы выборок без учета стоимостных характеристик документов, а в графе B – с учетом этого фактора.

Для первого слоя документов в графе A показан объем выборки 170. Эта величина получена элементарно:

$$\frac{3200}{3700} \cdot 200 \approx 170.$$

В свою очередь, для этого же слоя в графе B имеем:

$$\frac{122}{345} \cdot 200 \approx 70.$$

[1] Этот метод, в частности, рекомендуется аудиторам "Делойт энд Туш", см.: DRT International. DRT. Audit Sampling. 1991. P. 42–43.

Таблица 2.4

Распределения объема выборки по слоям

Слои, тыс. руб.	Количество документов	Общая сумма в интервале, тыс. руб.	Объем выборки	
			A	B
0,1–220	3200	122	170	70
220–1000	500	223	30	130
Всего	3700	345	200	200

Как показано в таблице, основное внимание аудитор, если он учитывает стоимостной фактор, должен уделить второй группе документов.

Монетарная выборка (monetary unit sampling, в американской литературе – dollar unit sampling). В отечественном Стандарте аудиторской деятельности этот способ отбора называют довольно громоздко – выборка по стоимостной оценке элементов совокупности. Этот способ в последнее время все больше привлекает внимание аудиторов. Суть его заключается в *отборе, пропорциональном стоимостным характеристикам документа*. За рубежом этот метод отбора иногда обозначают как PPS (probability proportional to size)[1].

Обсуждаемый метод базируется на принципе – чем больше стоимостная величина, содержащаяся в документе, тем больше должна быть вероятность для него оказаться в выборке. Например, если речь идет об отборе лицевых счетов, то счет с остатком, скажем, 20 тыс. руб. должен иметь в 10 раз большую вероятность оказаться в выборке, чем счет с остатком 2 тыс. На первый взгляд представляется, что здесь не соблюдается принцип случайности отбора. Однако это не так. Основополагающий принцип равенства вероятностей при отборе здесь не нарушается, он просто переносится с единиц наблюдения (балансовых счетов, кредитных дел и т.д.) на денежную единицу – каждый рубль в документах тестируемой совокупности должен иметь равную вероятность попадания в выборку. Каждый случайным образом отобранный рубль соответствует

[1] *Giliberti A.* Op.cit. P. 65-17.

конкретному документу совокупности, который и проверяется аудитором. На наш взгляд, монетарная выборка наиболее пригодна при оценивании относительной ошибки для совокупности (см. § 6.3).

Монетарная выборка по своим результатам близка расслоенному отбору. Однако в расслоенном отборе шанс попасть в выборку зависит от того, в какой слой (интервал) по стоимости попадает документ. В монетарной выборке этот шанс непосредственно связан со стоимостной характеристикой каждого документа.

При монетарном отборе единиц наблюдения в выборочную совокупность обычно применяют механический (систематический) отбор. Напомним, что при стандартном механическом отборе документы отбирают по присвоенным им порядковым номерам (например, каждый двадцатый документ). При монетарной механической выборке отбирают те документы совокупности, на которые указывают порядковые номера денежных единиц (рубля, тысячи рублей), например, отбирается каждый десятитысячный рубль в кумулятивной (последовательно накопленной) сумме. Разумеется, при определении кумулятивной суммы документы не ранжируют по величине стоимостной характеристики. В связи с той ролью, которую играют в рассмотренном способе отбора кумулятивные суммы, его иногда называют *методом кумулятивных сумм* (cumulative monetary amounts, CMA)[1].

Кумулятивные суммы определяются элементарно:

$$K_t = K_{t-1} + S_t,$$

где K_t – кумулятивная сумма после включения в нее денежной суммы t-го документа;

S_t – денежная сумма t-го документа.

Шаг отбора (стоимостной интервал отбора) определяется делением общей денежной суммы документов совокупности на объем выборки. Если на один документ совокупности пришлось более одного рубля при принятом шаге отбора, то документ попадает в выборку, разумеется, один раз.

[1] DRT International. DRT. Audit Sampling. 1991. P. 36–38.

ПРИМЕР 2.5. Пусть проверяются лицевые счета, имеющие порядковые номера, которые не связаны со стоимостными характеристиками документов. Количество счетов 1000, их номера – от 001 до 1000. Объем выборки – 100 единиц. Выборка осуществляется на основе монетарного механического отбора.

Общая сумма остатков на фиксированную дату – 2,5 млн руб. Соответственно шаг отбора равен 25 тыс. руб. Номер первого счета отбирается случайным образом из первого стоимостного интервала. Пусть это десятитысячный рубль из первых 25 тыс. руб. кумулятивной суммы.

В табл. 2.5 кумулятивные суммы показаны в графе 3. Первый счет не попадает в выборку, так как его сумма меньше 10 тыс. руб. На второй документ приходится кумулятивная сумма 5 + 3 = 8 тыс. руб., что также меньше 10 тыс. руб. Кумулятивная сумма на счетах, включая третий счет, составляет 8 + 4 = 12 тыс.руб. Таким образом, десятитысячным в кумулятивной сумме оказывается рубль, приходящийся на счет 003 с остатком 4 тыс. руб. (счет охватывает тысячи рублей с порядковыми номерами свыше 8 и до 12). Этот счет и попадает в выборку.

Следующие кумулятивные тысячи рублей, на основе которых производится отбор счетов, равны 35, 60 и т.д. (графа 4). Этим величинам соответствуют счета 005, 008, 057, 096 и т.д. В итоге в выборку отбирают счета с суммами 4, 20, 3, 13, 5 и т.д. тысяч рублей (см. графу 5). В таблице показан порядок отбора только первых семи документов.

Таблица 2.5

Пример монетарного отбора

Номер счета	Сумма на счете, тыс. руб.	Кумулятивная сумма, тыс. руб.	Порядковый номер тысячи рублей в выборке	Сумма на тестируемом счете
001	5	5		
002	3	8		
003	4	12	10	4
004	4	16		
005	20	36	35	20
006	1	37		
…	…	…		
08	3	61	60	3
…	…	…		
057	13	90	85	13
…	…	…		

Окончание табл. 2.5

Номер счета	Сумма на счете, тыс. руб.	Кумулятивная сумма, тыс. руб.	Порядковый номер тысячи рублей в выборке	Сумма на тестируемом счете
096	5	111	110	5
…	…	…		
124	11	141	135	11
…	…	…		
138	3	161	160	3
…	…	…		
1000	2	25000		
Итого	25000	–	–	–

При отборе счетов из совокупности можно было применить и таблицу случайных чисел. Однако механический отбор проще реализовать.

После того как тем или иным способом из совокупности отобраны единицы наблюдения, приступают к анализу полученных выборочных данных. Конечный результат их обработки при любом виде извлечения выборки заключается в точечном и интервальном оценивании неизвестных параметров тестируемой совокупности (средней, суммы, доли, средней абсолютной или относительной ошибки). Соответствующие методики обсуждаются в гл. 5 и 6. Поскольку обсуждаемые там методики основываются на распределениях случайных величин – важнейшем инструменте математической статистики, то следующая глава посвящена дискретным и непрерывным распределениям.

Дополнение к главе

Генерирование случайных чисел с помощью электронной таблицы Excel 5.

Последовательность действий.

1. Выделить в таблице Excel ячейку для первого случайного числа.

2. Последовательно вызвать: $f_x \rightarrow$ Статистические[1] $\rightarrow$ СЛЧИС( ).

3. В диалоговое окно вводят параметры, характеризующие диапазон последовательных номеров, из которого делается отбор:

– если случайные числа отбираются в диапазоне от нуля до, скажем, 500, то в окно вводится СЛЧИС( )*500 ;
– если диапазон начинается не с нуля, а с какой-либо положительной величины, допустим, 100, то в окне указывается СЛЧИС( )*100+500 .

4. Первое случайное число автоматически показывается в выделенной ячейке таблицы. Это число следует округлить до целочисленного значения.

5. Остальные числа столбца или строки получают, копируя эту ячейку стандартным путем.

[1] В последней версии Excel (XP Professional) вместо "Статистические" надо указать "Математические".

Глава 3. РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Что стало бы с нашим обществом, если бы доискивались до причин богатства каждого?

О. Бальзак

§ 3.1. Случайные величины. Нормальное распределение

Аудитор на основе случайной выборки получает в свое распоряжение некоторую совокупность единиц наблюдения (документов). Каждый документ, оказавшийся в выборке, содержит сведения, ради которых, собственно, и осуществлена выборка. Аудитор, таким образом, располагает необходимой ему количественной информацией или данными о наличии или отсутствии тех или иных свойств проверяемых документов (атрибутивные выборки). Очевидно, что выборочные данные, которые попадают в руки аудитора, являются *случайными величинами* в том смысле, что они в зависимости от сложившихся обстоятельств могут принимать те или иные значения в пределах определенной шкалы. Например, суммы на счетах у наугад отобранных клиентов банка, суммы приходных ордеров, накладных и других финансовых документов в общем случае заранее неизвестны и являются случайными величинами. В теории вероятностей случайная величина обычно определяется как переменная, принимающая то или иное числовое значение в зависимости от случайного исхода испытаний.

Случайные величины делятся на два класса – *непрерывные* и *дискретные*. Стоимостные данные, с которыми в большинстве случаев имеет дело аудитор, являются практически непрерывными. В самом деле, суммы остатков на текущих счетах, размеры оплаченных чеков и т.д. могут быть любыми положительными величинами от нуля и выше (с точностью до копейки, рубля или другой денежной единицы). Случайные величины, которые имеют целочисленные значения (их можно последовательно пронумеровать), являются дискретными. Например, количество доку-

ментов с нарушениями при оформлении, обнаруженными аудитором в выборке, может быть 0, 1, 2 и т.д.

Полученные по выборочным данным параметры (средние, доли, дисперсии) также являются случайными величинами. Например, если из одной и той же совокупности последовательно отобрать несколько выборок, то полученные выборочные средние будут разными – они варьируют по размерам. Их можно рассматривать как случайные величины, которые более или менее отличаются от неизвестного среднего исследуемой совокупности.

Как известно, для описания случайной величины необходимо не только указать множество ее возможных значений, но и показать вероятности для каждого из этих значений. В статистике такое полное описание случайной величины называют *теоретическим распределением*.

Методы математической статистики, используемые в ряде экономических исследований, в том числе в выборочном анализе, применяют следующие виды теоретических распределений: нормальное, распределение Стьюдента, биномиальное и гипергеометрическое распределения, распределение Пуассона. Первые два из перечисленных распределений относятся к группе *непрерывных*, остальные – к *дискретным распределениям*.

Нормальное распределение играет чрезвычайно важную роль в теории и практике выборочных обследований. Приведем краткие сведения о нем. Нормальное распределение характеризует распределение непрерывной случайной величины X. Параметрами нормального распределения являются среднее значение ($\bar{x}$) и дисперсия (σ^2). Обычно это распределение обозначается как $N(\bar{x}, \sigma)$ или $x \sim N(\bar{x}, \sigma)$. Функция симметрична относительно $\bar{x}$.

Кривая нормального распределения $N(\bar{x}, \sigma)$ показана на рис. 3.1. Площадь под кривой в пределах $\bar{x} \pm 3\sigma$ чуть меньше единицы.

На практике обычно применяют *нормированное (стандартизированное) нормальное распределение* N(0,1), которое не зависит от масштабов измерения случайной величины x. Вместо x здесь фигурируют нормированные по σ отклонения x от среднего $\bar{x}$, которые обозначим z:

$$z = \frac{x - \bar{x}}{\sigma}. \qquad (3.1)$$

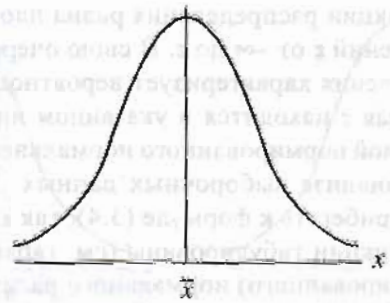

Рис. 3.1

Величина z также является случайной. Заменим в формуле (3.1) значения средней и стандартного отклонения в совокупности на их выборочные оценки. Получим

$$z = \frac{x - \tilde{x}}{s}, \qquad (3.2)$$

где $\tilde{x}$ – выборочная средняя;

s – стандартное отклонение в выборке.

Из (3.2) легко найти абсолютный размер отклонения x от средней как функцию z:

$$x - \tilde{x} = zs. \qquad (3.3)$$

ПРИМЕР 3.1. Для $x = 25$ при условии, что $\tilde{x} = 40$ и $s = 10$, нормированное отклонение составит

$$z = \frac{25 - 40}{10} = -1{,}5.$$

Кривая нормированного нормального распределения N(0,1) показана на рис. 3.2. Кривая симметрична относительно $z = 0$. Для z в пределах от $-\infty$ до $+\infty$ площадь под кривой распределения равна единице, а для z в пределах от –3 до +3 эта площадь чуть меньше единицы (0,997).

Функция нормированного нормального распределения определяется как

$$F_z = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \qquad (3.4)$$

Величина функции распределения равна площади под кривой в интервале значений z от $-\infty$ до z. В свою очередь, площадь под кривой распределения характеризует вероятность того, что случайная переменная z находится в указанном интервале. Именно это свойство кривой нормированного нормального распределения используется в анализе выборочных данных. Практически нет необходимости прибегать к формуле (3.4), так как соответствующие значения функции табулированы (см. таблицы нормированного (стандартизированного) нормального распределения[1]).

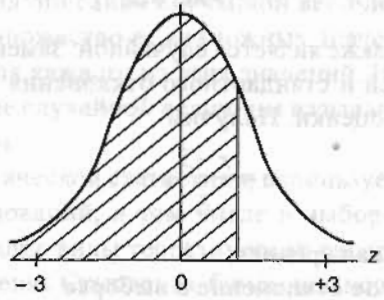

Рис. 3.2

В практических расчетах возможны следующие варианты определения вероятностей на основе указанных таблиц в зависимости от заданных значений случайной переменной z:

а) не превышает заданного значения z, т.е. z находится в пределах от $-\infty$ до z (рис. 3.2);

б) находится в пределах $\pm z$, т.е. интервал симметричен относительно начала координат (рис. 3.3);

в) находится в пределах от z_1 до z_2 (рис. 3.4).

Площади под кривой нормированного нормального распределения для варианта а приведены в графе F_z табл. 2 Приложений, а для варианта б – в графе F_{zz} этой же таблицы. Значение функции F_{zz} получены следующим образом:

$$F_{zz} = F_z - F_{-z} = 2(F_z - 0{,}5).$$

[1] В табл. 2 Приложений приводятся значения функции для z с шагом 0,05. Более полные таблицы значений функции распределения (3.4) приведены в справочнике: *Лукеш И., Ляга Й.* Основные таблицы математической статистики. М.: Финансы и статистика, 1985.

Значения функции нормального распределения можно получать и по электронной таблице Excel (см. функцию НОРМРАСП).

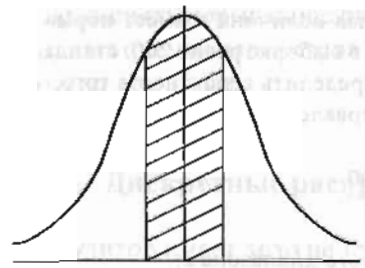

Рис. 3.3

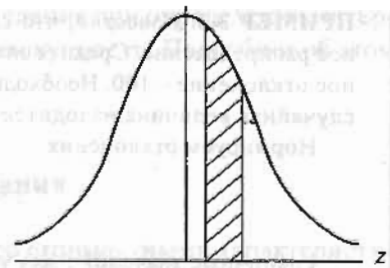

Рис.3.4

Что касается варианта в, то соответствующее значение получим как разность

$$F_{zz} = F_{z_2} - F_{z_1}, \quad -\infty < z_1 < z_2 < +\infty. \tag{3.5}$$

В силу симметричности распределения:

$$F_{-z} = 1 - F_z. \tag{3.6}$$

ПРИМЕР 3.2. Определим вероятности (площади под кривой нормального распределения N(0,1)) для следующих вариантов значений нормированного отклонения z:

- не превышает 0,65;
- находится в пределах ± 0,25;
- находится в пределах от 0,1 до 0,25.

По табл. 2 Приложений находим ответ для первого варианта: $F_{0,65} = 0{,}742$. Иначе говоря, вероятность того, что нормированное отклонение случайной переменной не превысит величину $z = 0{,}65$, немного превышает 74%.

Для второго варианта получим

$$F_{0,25;-0,25} = 2(F_{0,25} - 0{,}5) = 2(0{,}5987 - 0{,}5) = 0{,}1974$$

или в этой же таблице находим $F_{0,25;\,0,25} = 0{,}1974$ (т.е. вероятность указанного события порядка 20%).

Наконец, для третьего варианта имеем вероятность чуть ниже 6%:

$$F_{0,1;0,25} = F_{0,25} - F_{0,1} = 0{,}5987 - 0{,}5398 = 0{,}0589.$$

ПРИМЕР 3.3. Известно, что случайная величина x имеет нормальное распределение. Среднее значение в выборке равно 200, стандартное отклонение – 100. Необходимо определить вероятность того, что случайная величина находится в интервале от 150 до 400.

Нормируем отклонения:

$$z = \frac{x - 200}{100}.$$

Граничные значения z для указанного диапазона x:

$$z_1 = \frac{150 - 200}{100} = -0{,}5,\ z_2 = \frac{400 - 200}{100} = 2.$$

Искомая вероятность равна 69%:

$$F_{-0.5;2} = F_2 - F_{-0.5} = F_2 - \left(1 - F_{0.5}\right) = 0{,}9972 - \left(1 - 0{,}6915\right) = 0{,}6887.$$

Как было показано выше, нормированное нормальное распределение позволяет определить вероятность в зависимости от заданного интервала значений z.

Аудитору часто приходится оценивать по выборочным данным неизвестную сумму стоимостных характеристик для тестируемой совокупности. Эта задача равнозначна оцениванию неизвестного размера средней. Для этого аудитор отбирает из совокупности n документов (единиц наблюдения) и определяет значение выборочной суммы и средней. Допустим, что он повторяет отбор. Тогда, как правило, он получит другие значения суммы и средней. Для тестируемой совокупности эти величины остаются неизвестными. Для их оценивания необходимо располагать соответствующими выборочными распределениями.

Как утверждает *центральная предельная теорема*, распределение сумм независимых случайных величин асимптотически приближается к нормальному распределению[1]. Иначе говоря, каким бы ни было распределение исходных данных, с увеличением объема выборки распределение выборочных сумм (а следовательно, и средних) стремится к нормальному[2]. Это свойство позволяет применять нормальное распределение при определении интервальных оценок параметров для совокупности. Подробнее об этом сказано в гл. 5 и 6.

[1] Более строгое определение центральной предельной теоремы и ее доказательства для разных постановок задачи можно найти в классических курсах математической статистики.

[2] Многочисленные эксперименты показали, что если объем выборки равен 25 или больше единиц наблюдения, то распределение выборочной средней всегда оказывалось нормальным при любом виде исходного статистического распределения. Любопытный пример этого приводится в книге: *Джессен Р.* Указ. соч. С.60.

§ 3.2. Дискретные распределения

Аудитор имеет дело не только со стоимостными характеристиками. Ему часто приходится оценивать по выборке *качество* представленного материала – частоту ошибок, неправильно оформленных документов и т.д. Назовем такие ошибки *отклонениями*, а выборки, применяемые для оценивания количества отклонений, – *атрибутивными*.

Очевидно, что число отклонений в выборке является случайной величиной, случайной переменной. В одной выборке может оказаться два отклонения, в другой выборке из этой же совокупности – три и т.д.

Пусть случайная величина X принимает дискретные значения k ($k = 0, 1, 2, \ldots, n$) с вероятностями

$$P(X = k) = p_k.$$

По определению $0 \le p_k \le 1$. Сумма вероятностей от p_0 до p_m характеризует вероятность того, что случайная дискретная переменная не превысит некоторое заданное значение m:

$$F_k = P\left(X \le m\right) = \sum_{k=0}^{m} p_k. \qquad (3.7)$$

Функцию F_k называют *функцией распределения дискретной случайной величины* или *интегральной вероятностью*. В табл. 3.1 во второй графе показаны вероятности наступления случайного события (k отклонений в выборке), в третьей графе приведены значения такой функции для конкретного распределения.

В табл. 3.1 приведен пример распределения вероятностей наступления случайного события при условии, что вероятность отклонения $p = 0{,}04$, объем выборки $n = 40$. В графе k таблицы показаны количества документов с отклонениями – от 0 до 9. В графе p_k содержатся соответствующие значения вероятностей (как получены эти вероятности – обсуждается ниже). Например, вероятность того, что случайная величина (количество отклонений в

Таблица 3.1

Распределение случайной величины k ($p = 0{,}04$, $n = 40$)

k	p_k	F_k
0	0,19537	0,19537
1	0,32561	0,52098
2	0,26456	0,78553
3	0,13963	0,92516
4	0,05381	0,97898
5	0,01614	0,99512
6	0,00392	0,99905
7	0,00079	0,99984
8	0,00014	0,99998
9	0,00002	1,00000

выборке) равна 4, составляет 0,05381. Показатели, характеризующие интегральные (суммарные) вероятности для данного распределения, приведены в графе F_k. Для $k = 4$ находим по таблице $F_4 = 0{,}97898$. Это вероятность наличия в выборке не более четырех отклонений. Вероятность обнаружения не более пяти отклонений немного выше, а именно 0,99512 и т.д.

При выборочных исследованиях в аудите для описания дискретных величин (числа отклонений в выборках) применяют два основных вида дискретных распределений: биномиальное (Deloitte&Touche) и Пуассона (Arthur Andersen & Co). В некоторых условиях, на наш взгляд, более обоснованно применение гипергеометрического распределения. Кратко охарактеризуем названные дискретные распределения.

Биномиальное распределение. Допустим, на основе прошлого опыта известно или ожидается, что вероятность случайного события (в аудиторской поверке – отклонения) равна p. Вероятность противоположного события составит $(1 - p) = q$. Объем выборки равен n. С вероятностью p_k в выборке оказывается k таких документов. Указанные вероятности определяются по формуле Бернулли.

При выводе формулы Бернулли исходят из следующей модели. Пусть для начала производятся два независимых испытания (отбор только двух документов). В этом случае получим следующие исходы:

- в обоих документах нет отклонения, $k = 0$;
- в одном документе есть отклонение, в другом нет, $k = 1$;
- в обоих документах есть отклонения, $k = 2$.

Вероятности перечисленных исходов p_k в этом случае определяются следующим образом: $p_0 = q^2$, $p_1 = pq$, $p_2 = p^2$. В свою очередь, для трех документов получим следующие вероятности:

$$p_0 = q^3,\ p_1 = pq^2,\ p_2 = p^2q,\ p_3 = p^3.$$

При обобщении полученного результата на n независимых испытаний получим *формулу Бернулли*, определяющую вероятность появления k случайных событий (отклонений) из выборки объемом n:

$$p_k = P(X = k) = C_k^n p^n q^{n-k}, \quad k = 0, 1, 2, \ldots n, \tag{3.8}$$

где n – объем выборки;

p – вероятность единичного отклонения[1].

Число сочетаний определяется как

$$C_k^n = \frac{n!}{k!(n-k)!},$$

где $n!$ – факториал n (напомним: $0! = 1$, $1! = 1$, $2! = 1 \cdot 2 = 2$ и т.д.).

Для крайних значений k ($k = 0$ и $k = n$) по (3.8) получим следующие вероятности:

$$p_0 = C_0^n q^n = q^n \quad \text{и} \quad p_n = C_n^n p^n = p^n.$$

Вместо (3.8) можно воспользоваться рекуррентной формулой

$$p_{k+1} = \frac{n-k}{k+1} \cdot \frac{p}{q} \cdot p_k. \tag{3.9}$$

При применении атрибутивных выборок необходимы только значения функции распределения. Они определяются путем суммирования полученных вероятностей:

[1] Более подробные сведения о биномиальном и других распределениях можно получить в книгах: *Кеналл М., Стьюарт А.* Теория распределений. М.: Наука, 1966; *Феллер В.* Введение в теорию вероятностей и ее приложения. М.: Мир, 1964.

$$F_k = P(X \le k) = \sum_{x=0}^{k} p_x. \qquad (3.10)$$

Математическое ожидание (среднее значение переменной k) определяется весьма просто: $\mu = np$.

В табл. 3.1 приведены значения вероятностей и функции биномиального распределения для условий $p = 4\%$, $n = 40$. Соответствующий график показан на рис. 3.5.

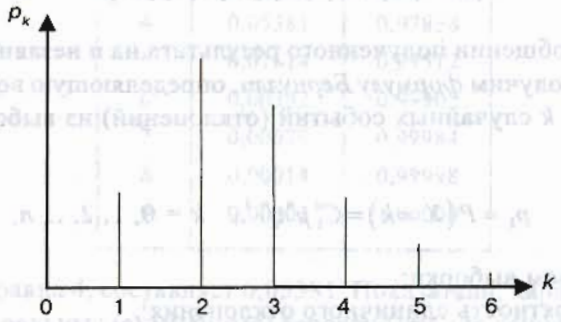

Рис. 3.5

Очевидно, что требование независимости испытаний при атрибутивном отборе на практике не соблюдается. Для соблюдения этого условия, строго говоря, пришлось бы возвращать каждый тестированный документ в обследуемую совокупность и заново осуществлять отбор, применяя таблицу случайных чисел. Причем это надо было бы сделать исключительно для того, чтобы не изменилась вероятность p, что лишено практического смысла. Несоблюдение требования независимости при безвозвратном отборе компенсируется большим объемом проверяемой совокупности, при котором можно считать вероятность p примерно постоянной.

Строго говоря, в расчетах значений p_k и F_k по приведенным выше формулам нет крайней необходимости, так как значения функций можно найти в специальных таблицах[1]. Хотя, как правило, соответствующие таблицы разработаны для небольших величин n. Эти же величины можно получить с помощью электронной

[1] Значения функции распределения F_k для биномиального распределения приводятся в ряде справочников по математической статистике. В Приложениях (табл. 3) помещены таблицы, в которых n равно 40, 80, 100, 150, 200, 250. Таблицы биномиального распределения для n от 2 до 40 приведены в справочнике: *Лике ш И., Ляга Й.* Основные таблицы математической статистики. М.: Финансы и статистика, 1985.

таблицы Excel[1]. Однако для лучшего понимания смысла получаемых величин целесообразно привести их расчет (см. пример 3.4).

ПРИМЕР 3.4. Определим вероятности наступления случайных событий (отклонений) в выборке для условий: объем выборки равен 25 единицам, ожидаемая вероятность одного отклонения в совокупности 5% ($n = 25$, $p = 0{,}05$). По формуле (3.8) последовательно получим показатели вероятности для следующих событий – отсутствия отклонений в выборке (p_0), наличия одного отклонения (p_1) и т.д.:

$$p_0 = \frac{25!}{0!\,25!} \cdot 0{,}05^0 \cdot 0{,}95^{25} = 0{,}2774,$$

$$p_2 = \frac{25!}{2!\,23!} \cdot 0{,}05^2 \cdot 0{,}95^{23} = 0{,}2305,$$

$$p_3 = \frac{25!}{3!\,22!} \cdot 0{,}05^3 \cdot 0{,}95^{22} = 0{,}093 \text{ и т.д.}$$

В свою очередь, применив рекуррентную формулу (3.9) и использовав полученную выше величину p_0, приходим к тем же результатам, но более простым путем:

$$p_1 = \frac{25-0}{0+1} \cdot \frac{0{,}05}{0{,}95} \cdot 0{,}2774 = 0{,}365,$$

$$p_2 = \frac{25-1}{1+1} \cdot \frac{0{,}05}{0{,}95} \cdot 0{,}365 = 0{,}2305,$$

$$p_3 = \frac{25-2}{2+1} \cdot \frac{0{,}05}{0{,}95} \cdot 0{,}2305 = 0{,}093 \text{ и т.д.}$$

Вероятность того, что число случайных событий окажется не более, допустим, трех, равна

$$F_3 = \sum_{k=0}^{3} p_k = 0{,}2774 + 0{,}365 + 0{,}2305 + 0{,}093 = 0{,}9659.$$

Если в данном примере речь идет об атрибутивной выборке, то с вероятностью порядка 97% аудитор может утверждать, что количество отклонений в проверяемой совокупности не превысит 3.

При обсуждении области применения биномиального распределения трудно удержаться от одной забавной иллюстрации. Аме-

[1] См. Дополнение к гл. 3.

риканский детский врач Б. Спок, чьи книги по уходу за детьми были широко известны в России, отличался активной общественной деятельностью. В связи с чем иногда привлекался к суду присяжных. На одном из таких процессов Спок отвел состав присяжных, поскольку, по его утверждению, они не были выбраны случайным образом, как это предусматривает закон.

Основанием для него послужили следующие рассуждения. Присяжные (по принятой выше терминологии – выборка) выбираются судом случайным образом из большого списка кандидатов (генеральная совокупность). В этом списке примерно поровну мужчин и женщин. Женщины в США знакомы с книгами Спока и, как правило, были его сторонницами. Наличие в составе присяжных одних мужчин явно указывало на направленный, а не случайный отбор.

В самом деле, объем выборки в данной ситуации 12 единиц, ожидаемая вероятность непопадания женщины в выборку (состав присяжных) 50%. Следовательно, вероятность отсутствия женщин среди двенадцати присяжных близка к нулю: $p_0 = 0{,}5^{12} = 0{,}00024$, а вероятность того, что среди них окажется, допустим, от одной до пяти женщин, составляет 0,39. Суд был вынужден согласиться со Споком.

Гипергеометрическое распределение. Это распределение также позволяет определить вероятности обнаружения в выборке ровно k случайных событий. Оно основано на той же модели, что и биномиальное распределение. Однако есть и одно отличие, которое заключается в отсутствии требования независимости испытаний и, следовательно, условия неизменности вероятности p. Иначе говоря, если в биномиальном распределении предполагается неизменность вероятности наступления случайного события (хотя практически это не принимают во внимание, см. пример 3.4), то в гипергеометрическом распределении она может изменяться в ходе отбора. Для расчета вероятностей p_k здесь необходима дополнительная информация – размер генеральной совокупности N.

Вероятность наступления ровно k случайных событий определяется следующим образом:

$$p_k = P(X = k) = \frac{C_k^M \cdot C_{n-k}^{N-M}}{C_n^N}, \qquad (3.11)$$

где M – ожидаемое число отклонений в совокупности;

p – вероятность случайного события до проведения испытаний.

По определению $M = pN$.

Найдем вероятность p_0 на основе формулы (3.11):

$$p_0 = \frac{C_0^M \cdot C_n^{N-M}}{C_n^N}.$$

Поскольку $C_0^M = 1$, то окончательно имеем

$$p_0 = \frac{C_n^{N-M}}{C_n^N} = \frac{(N-M)!\,(N-n)!}{(N-M-n)!\,N!}. \qquad (3.12)$$

Если вероятность p_0 определена, то расчет вероятностей p_k удобнее осуществить по рекуррентной формуле

$$p_{k+1} = \frac{(n-k)(M-k)}{(k+1)(N-M+k+1)} \cdot p_k. \qquad (3.13)$$

Функция распределения, как и в биномиальном распределении, определяется по общей формуле

$$F_k = \sum_{x=0}^{k} p_x.$$

ПРИМЕР 3.5. Определим вероятности наступления случайных событий и значения функции гипергеометрического распределения для условий: $N = 500$, $n = 40$, $p = M/N = 0{,}05$. Из последнего соотношения непосредственно следует $M = 0{,}05 \cdot 500 = 25$.

Вероятность нулевого события

$$p_0 = \frac{C_{40}^{490}}{C_{40}^{500}} = \frac{475!\,460!}{435!\,500!} = \frac{475 \cdot 474 \cdot \ldots \cdot 436}{500 \cdot 499 \cdot \ldots \cdot 461} = 0{,}11782.$$

По рекуррентной формуле (3.13) получим следующие значения вероятностей:

$$p_1 = \frac{40 \cdot 25}{500 - 25 + 1} \cdot 0{,}11782 = 0{,}27024,$$

$$p_2 = \frac{39 \cdot 24}{2 \cdot (500 - 25 + 1 + 1)} \cdot 0{,}27024 = 0{,}28941 \text{ и т.д.}$$

Значения p_k и F_k ($k = 0, \ldots, 8$) для условий данного примера приведены в табл. 3.2.

Таблица 3.2

Гипергеометрическое распределение ($N = 500$, $n = 40$, $p = 0{,}05$)

k	p_k	F_k
0	0,11782	0,11782
1	0,27024	0,38806
2	0,28941	0,67747
3	0,19250	0,86997
4	0,08923	0,95920
5	0,03066	0,98987
6	0,00811	0,99798
7	0,00169	0,99967
8	0,00028	0,99997

Подобного рода таблица позволяет определить вероятность нахождения в выборке количества отклонений в заданном диапазоне, например, от $k = 1$ до $k = 3$. В этом случае $F_k = \sum_{k=1}^{3} p_k = 0{,}75215$.

Разработать таблицы значений p_k и F_k для гипергеометрического распределения, аналогичные таблицам биномиального распределения, не удается, так как одним из аргументов здесь является объем генеральной совокупности[1]. Для определения необходимых значений вероятностей и функции распределения удобно воспользоваться электронной таблицей Excel[2].

Распределение Пуассона. При больших размерах тестируемой совокупности (N) гипергеометрическое распределение хорошо аппроксимируется *распределением Пуассона*. Распределение задается только одним параметром λ (среднее число случаев в выборке):

$$\lambda = pn, \tag{3.14}$$

где p – вероятность наступления случайного события;

n – объем выборки.

[1] В упоминаемом выше сборнике таблиц И. Ликеша и Й. Ляги помещены таблицы гипергеометрического распределения для N от 2 до 15.

[2] См. Дополнение к гл. 3.

Вероятность наступления ровно k случайных событий

$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \tag{3.15}$$

где e – основание натуральных логарифмов.

Вероятности p_k могут быть определены и по рекуррентной формуле:

$$p_{k+1} = \frac{\lambda}{k+1} \cdot p_k. \tag{3.16}$$

Функция распределения

$$F_k = P(X \le k) = \sum_{x=0}^{k} p_x. \tag{3.17}$$

Значения функций распределения Пуассона приводятся в соответствующих справочниках[1]. Их нетрудно определить и по электронной таблице Excel[2].

К распределению Пуассона в выборочных обследованиях обычно прибегают при невысоких значениях вероятности p.

ПРИМЕР 3.6. В иллюстративных целях определим вероятности для условия $\lambda = 1{,}2$ и функцию распределения Пуассона для $k = 7$. По формуле (3.15) находим:

$$p_0 = \frac{1{,}2^0}{0!} e^{-1{,}2} = 0{,}30119, \quad p_1 = \frac{1{,}2^1}{1!} e^{-1{,}2} = 0{,}36143,$$

$$p_2 = \frac{1{,}2^2}{2!} e^{-1{,}2} = 0{,}21686, \quad p_3 = \frac{1{,}2^3}{3!} e^{-1{,}2} = 0{,}08674,$$

$$p_4 = \frac{1{,}2^4}{4!} e^{-1{,}2} = 0{,}02602, \quad p_5 = \frac{1{,}2^5}{4!} e^{-1{,}2} = 0{,}00625,$$

$$p_6 = \frac{1{,}2^4}{5!} e^{-1{,}2} = 0{,}00125, \quad p_7 = \frac{1{,}2^5}{6!} e^{-1{,}2} = 0{,}00025.$$

[1] *Ликеш И., Ляга Й.* Указ. соч. С. 42.

[2] См. Дополнение к гл. 3.

Функция распределения для $k = 4$

$$F_4 = \sum_0^4 p_k = 0{,}99225.$$

В табл. 3.3 приведены значения вероятностей p_k и функции распределения F_k для k от 0 до 7.

Таблица 3.3

Распределение Пуассона ($\lambda = 1{,}2$)

k	p_k	F_k
0	0,30119	0,30119
1	0,36143	0,66263
2	0,21686	0,87949
3	0,08674	0,96623
4	0,02602	0,99225
5	0,00625	0,99850
6	0,00125	0,99975
7	0,00021	0,99996

Дополнение к главе

А. Биномиальное распределение в таблице Excel

Последовательность операций

1. Выделить ячейку в таблице Excel.
2. Вызвать: $f_x \rightarrow$ Статистические $\rightarrow$ БИНОМРАСП.
3. В диалоговое окно ввести:

Число успехов – количество случаев наступления события (k);
Число испытаний – объем выборки (n);
Вероятность успеха – вероятность наступления события (p);
Интегральная – если необходимо получить значение функции распределения F_k (интегральную вероятность), то вводится слово "ИСТИНА", если необходима вероятность наступления точно k событий p_k, то вводится "ЛОЖЬ".

Эти же данные после вызова функции можно ввести и непосредственно в формульное окно таблицы:

=БИНОМРАСП (k; n; p; ИСТИНА или ЛОЖЬ).

Например, для получения значения функции распределения при заданных параметрах $k = 8$, $n = 30$, $p = 0{,}1$ запишем в этом окне

= БИНОМРАСП (8; 30; 0,1; ИСТИНА).

В выделенной ячейке таблицы Excel получим значение функции $F_8 = 0{,}99798$.

Для варианта = БИНОМРАСП (8; 30; 0,1; ЛОЖЬ) получим величину вероятности наступления ровно восьми событий $p_8 = 0{,}00576$.

Б. Гипергеометрическое распределение в таблице Excel

Последовательность операций

1. Выделить ячейку в таблице Excel.
2. Вызвать: $f_x \rightarrow$ Статистические $\rightarrow$ ГИПЕРГЕОМЕТ.
3. В диалоговое окно ввести:

Число успехов в выборке – параметр k;
Размер выборки – параметр n;
Число успехов в совокупности – параметр M;
Размер генсовокупности – параметр N.

Эти же данные после вызова функции можно ввести и непосредственно в формульное окно таблицы:

= ГИПЕРГЕОМ (k; n; M; N).

Например, для параметров: $k = 1$, $n = 40$, $M = 10$, $N = 500$ запишем в формульном окне:

= ГИПЕРГЕОМ (1; 40; 10; 500).

В ячейке таблицы Excel получим значение вероятности $p_1 = 0{,}38222$.

В. Распределение Пуассона в таблице Excel

Последовательность операций

1. Выделить ячейку в таблице Excel.
2. Вызвать: $f_x \rightarrow$ Статистические $\rightarrow$ ПУАССОН.
3. В диалоговое окно ввести:

X – параметр k;

Среднее – параметр λ;

Интегральная – если необходимо получить значение функции распределения (интегральную вероятность), то вводится "ИСТИНА", если необходима вероятность наступления точно k событий, то вводится "ЛОЖЬ".

Эти же данные после вызова функции можно ввести непосредственно в формульное окно таблицы:

= ПУАССОН (k; λ; ИСТИНА или ЛОЖЬ) .

Например, для параметров: $k = 2$, $\lambda = 1{,}2$ запишем в формульном окне:

= ПУАССОН (2; 1,2; ИСТИНА) .

В ячейке таблицы Excel получим значение функции $F_4 = 0{,}21686$.

Глава 4. АТРИБУТИВНЫЕ ВЫБОРКИ

> Нет ничего тайного, что не сделалось бы явным, и ничего не бывает потаенного, что не вышло бы наружу.
>
> *Евангелие от Марка*

§ 4.1. Цель атрибутивного выборочного тестирования

Один из видов аудита заключается в проверке совокупности документов с целью определения их соответствия определенным требованиям или свойствам. Например, требованию наличия всех реквизитов, печатей и подписей, требованию отсутствия арифметических ошибок. В гл. 1 перечисленные и некоторые другие свойства были названы атрибутами. Атрибутом также может быть обязательный набор сопроводительных документов и тому подобное. Оговоренные перед проведением тестирования атрибуты должны присутствовать в каждом проверяемом носителе информации. Атрибуты или есть, или их нет. В аудите отсутствие атрибута называют *отклонением* или *ошибкой*. Так, если наличие разрешительной подписи рассматривается как атрибут, то ее отсутствие – как отклонение. В §1.2 подобного рода тестирование было названо "проверкой на соответствие".

Атрибутивное тестирование заключается в определении (оценивании) доли документов в совокупности с наличием отклонений. Возможна и обратная постановка задачи – определение доли документов с отсутствием соответствующего атрибута.

При большом массиве документов сплошные проверки на наличие атрибута (или его отсутствие) – весьма трудоемкие процедуры. Применение выборок в этих случаях является весьма продуктивным и во многих случаях целесообразным. Атрибутивная выборка позволяет оценить с определенной степенью надежности максимальное количество документов с наличием атрибута в проверяемой совокупности. В статистической теории подобного рода выборки обычно называют выборками для измерения каче-

ственных признаков. По своему назначению атрибутивные выборки напоминают выборки, применяемые при контроле качества продукции. Целью последних является оценка качества выпущенной партии изделий по выборочным данным[1].

Полученные в ходе атрибутивной выборки результаты говорят только о масштабах изучаемого явления – наличия атрибутов (например, при проверке обнаружено, что в 0,5% документов содержатся арифметические ошибки или 0,1% документов некомплектны и т.п.). Очевидно, что получаемые данные не позволяют определить обобщающую стоимостную или иную характеристику, связанную с обнаруженными отклонениями, да такая задача здесь и не ставится.

К атрибутивным выборкам обычно прибегают во внутреннем аудите при тестировании документов, сопровождающих массовые хозяйственные или деловые операции: счета-фактуры, накладные, платежные поручения и т.д. Необходимость в таком тестировании возникает как при проверке качества внутреннего контроля, так и по специальному запросу, например, клиента банка, наконец, при наличии сомнений у аудитора в качестве определенной группы документов, поступивших от ряда организаций.

При планировании атрибутивной выборки необходимо:

- сформулировать цель тестирования;
- определить единицу наблюдения;
- установить рамки исследуемой совокупности, причем каждая единица наблюдения (документ совокупности) должна быть одинаково доступна для отбора;
- четко определить, какое свойство является тестируемым атрибутом: атрибут может быть единичным или комплексным (документ содержит набор реквизитов), а в случае когда отсутствует даже одно свойство или реквизит, документ рассматривается как отклонение;
- принять уровень доверительной вероятности;
- определить приемлемый верхний уровень точности оценки количества отклонений в совокупности;
- определить размер (объем) выборки;
- извлечь выборку из совокупности, для этого обычно используют схемы случайного (в том числе механического) или монетарного отбора документов;
- тестировать единицы наблюдения, оказавшиеся в выборке;
- определить по выборке верхний уровень точности оценки количества отклонений;
- принять решение о качестве тестируемой системы.

[1] *Шторм Р.* Теория вероятностей. Математическая статистика. Статистический контроль качества. М.: Мир, 1970.

Обсудим некоторые из перечисленных этапов. Предварительно напомним (см. § 3.2), что количество документов с наличием атрибута (отклонения) в выборке является случайной величиной. Количество отклонений в выборке (k) может составлять 0, 1, 2 и т.д. Например, в одной выборке может быть обнаружено два отклонения, в другой выборке из той же совокупности – четыре и т.д. Каждое из значений k может быть сопоставлено с вероятностью соответствующего числа отклонений. Если задана величина ожидаемой вероятности отклонения в совокупности p, то применив правило определения таких вероятностей, получим соответствующее *дискретное распределение случайных величин*. Такое распределение дает возможность определения вероятности каждого возможного числа отклонений. В гл. 3 была приведена характеристика трех важнейших видов дискретных распределений.

§ 4.2. Максимально возможное количество отклонений в совокупности

Атрибутивные выборки на практике применяют при решении ряда вариантов тестирования. В каждом из них основным является оценивание неизвестного количества отклонений в совокупности. Получив такую оценку, аудитор должен вынести суждение о качестве работы проверяемой системы.

Остановимся на одной из наиболее распространенных постановок задачи, согласно которой аудитор по выборке определяет *максимально возможное количество отклонений в совокупности*. Эти величины в специальной литературе называют *расчетным верхним пределом точности* (computed upper precision limit). Обозначим этот предел как U.

Обсудим теперь смысл и общую методику определения U. Пусть для описания закона распределения случайных величин k (отклонений) принято биномиальное распределение. Каждому значению k в этом распределении соответствует определенная величина вероятности (см. формулу (3.8)). Напомним, что биномиальное распределение (см. § 3.2) зависит от двух параметров p и n.

Ожидаемая вероятность отклонения p (expected occurrence rate) устанавливается в виде доли документов в совокупности с наличием отклонения. Первоначальное определение параметра p может быть весьма приближенным – на основе прошлого опыта или специального выборочного обследования, даже по интуиции. Учитывая выборочные данные, получим следующую оценку:

$$p = \frac{b}{n} 100\%, \qquad (4.1)$$

где b – количество отклонений, обнаруженных в выборке;

n – объем выборки.

Что касается объема выборки n, то этот параметр, как известно, играет важную роль в тестировании и анализе полученных результатов. Методика его определения для атрибутивных выборок обсуждается в § 4.3. Здесь же ограничимся одним общим замечанием: если в ходе анализа выборочных данных обнаруживается недостаточность первоначального размера выборки, то следует осуществить дополнительный отбор единиц наблюдения.

Итак, на основе выборочной оценки вероятности случайного события (отклонения) p и принятого объема выборки n определяются вероятности появления k отклонений. Сумма вероятностей от p_0 до p_m биномиального распределения (функция распределения) составляет величину

$$F_m = P(X \le m) = \sum_{k=0}^{m} p_k. \qquad (4.2)$$

Функция характеризует вероятность того, что случайная дискретная переменная k не превысит некоторое заданное значение m. Чем больше m, тем больше слагаемых в формуле (4.2) и, следовательно, выше значение функции распределения.

Приведенные выше сведения о распределении величины k необходимы для пояснения смысла расчетного верхнего предела точности, названного выше параметром U. Для этого несколько изменим постановку задачи. Пусть значение функции распределения не вычисляется, а задается аудитором. В этом случае оно играет роль *доверительной вероятности*, или *уровня надежности* (confidence level, reliability level). Это понятие кратко было охарактеризовано в § 2.1. Если m в формуле (4.2) является приемлемой для аудитора величиной, то соответствующая величина вероятности F_m считается аудитором достаточной для обоснования своего заключения. Она измеряется в долях единицы, а в заключении аудитора обычно указывается в процентах. Например, "с вероятностью 90% можно утверждать, что количество отклонений в совокупности не превысит...". Если доверительная вероятность в долях единицы составляет α, то риск получить недостоверное заключение по выборке равен $1 - \alpha$.

Компания "Deloitte&Touche" рекомендует устанавливать доверительную вероятность на уровне от 90 до 95% в зависимости от важности проверяемых сведений[1]. В других зарубежных руководствах предлагается более широкий диапазон значений – от 80 до 99%[2].

Задавшись конкретным уровнем доверительной вероятности, аудитор может определить значение максимального числа отклонений m. Нетрудно понять, что величина U есть не что иное, как параметр m, соответствующий принятой доверительной вероятности F_m. В практических расчетах величину U удобнее измерять в процентах.

Заметим, что на практике нет необходимости определять U исходя из формулы (4.2). Для этого можно воспользоваться соответствующими таблицами. Методика табличного определения величины U демонстрируется в § 4.3.

Пусть в выборке, объем которой 200 единиц, обнаружено 12 отклонений. Соответственно

$$p = \frac{12}{200} 100\% = 6\%.$$

Допустим, аудитор установил доверительную вероятность F_m на уровне 95%. Табличное значение U для этого уровня вероятности и перечисленных выше условий примерно равно 8%. Таким образом, с вероятностью 95% он может утверждать, что наибольшее количество отклонений в совокупности не должно превышать 8%. Иначе говоря, вероятность того, что максимальное количество отклонений превысит 8%, составляет 100 – 95 = 5%.

[1] Touche Ross International. The Touche Ross Audit Sampling Manual. U.K. 1982. P. 18.

[2] *Giliberti A.* Op. cit. P. 65-13.

Подсчитанный по выборочным данным параметр U сопоставляют с "нормативом", который называют *приемлемым (допустимым) уровнем отклонений* или *планируемой точностью* (desired upper precision limit, tolerable rate of occurrence, planned precision). Это наибольшее приемлемое для аудитора количество отклонений в совокупности. Иначе говоря, это то количество отклонений, с которым аудитор готов примириться при заданном уровне доверительной вероятности. Обозначим приемлемый уровень отклонений в совокупности как T.

Очевидно, что выбор параметра T в известной степени субъективен. Естественные относительные его пределы: от 2 до 10–12%, так как считается, что вряд ли есть смысл применять выборочный контроль в том случае, когда в совокупности ожидается более 10–12% отклонений – система явно работает с большим "сбоем". С другой стороны, если предполагается, что ожидаемый уровень отклонений менее 2%, то для проверки выполнения этого условия потребуется очень большой объем выборки. Обычно этот параметр T устанавливается на уровне от 3 до 7%.

Итак, полученную выборочную оценку максимального количества отклонений в совокупности U сравнивают с приемлемым значением количества отклонений T. Если рассматриваемые параметры соотносятся как $U \le T$, то у аудитора есть основание считать, что проверяемая система контроля на соответствие функционирует нормально.

Ситуация, когда $U > T$, может служить указанием на то, что фактическая вероятность отклонения в проверяемой совокупности заметно превышает ожидаемую аудитором вероятность p. В этом случае аудитор может в определенных границах изменить принятые им первоначальные условия для расчета параметра U. Конкретно это сводится к снижению уровня доверительной вероятности и/или увеличению объема выборки. Если указанные поправки условий не достигают цели и параметр U остается больше T, то аудитор вынужден признать тестируемую систему контроля плохо функционирующей со всеми вытекающими отсюда последствиями.

§ 4.3. Табличное определение верхнего предела отклонений для совокупности

Как было показано в § 4.2, задача аудитора при применении атрибутивной выборки обычно заключается в оценивании величины верхнего предела точности U для совокупности. Задавшись уровнем доверительной вероятности F_m и зная объем выборки n и ожидаемую вероятность отклонения p, определяют искомую величину U. Для этого используют специальные таблицы.

Показатели таких таблиц получают на основе таблиц функции биномиального распределения. Для каждого уровня доверительной вероятности разрабатывается своя таблица. В табл. 4.1, 4.2 и 4.3 приведены значения U для доверительных вероятностей 85, 90 и 95%. Например, для $n = 100$, $p = 5\%$ и $F_m = 0{,}85$ табличное значение U получено следующим образом. Для $n = 100$ и $p = 0{,}05$ в таблице функции биномиального распределения находим наиболее близкие к 0,85 значения интегральной вероятности 0,76601 и 0,87024. Этим вероятностям соответствуют $m = 6\%$ и $m = 7\%$, где $m = k/n$. После интерполяции получим $U = 6{,}8\%$, что и показано в табл. 4.1.

Если известен размер совокупности N, то в табличное значение U можно внести небольшое уточнение, которое в статистике называют *поправкой на конечность совокупности (пкс)*. С учетом этой поправки верхний предел точности (максимально возможное количество отклонений) для совокупности оценивается следующим образом:

$$U' = p + (U - p)\sqrt{1 - \frac{n}{N}}. \qquad (4.3)$$

ПРИМЕР 4.1. Пусть выборочная оценка ожидаемой вероятности отклонений в совокупности равна 7%, объем выборки 150 единиц, доверительная вероятность принята на уровне 85%. Для $p = 7\%$, $n = 150$ и $F_m = 85\%$ по табл. 4.1 находим: $U = 9\%$. Таким образом, в выборке указанного объема максимальная доля документов с отклонениями при принятом уровне доверительной вероятности не должна превышать 9% . Иными словами, если в выборке обнаружено 7% отклонений, то предельная ошибка этого параметра не превышает 2% при принятом уровне доверительной вероятности: $U - p = 9 - 7 = 2$.

Немного изменим условия. Пусть теперь известен размер совокупности (750). Тогда с учетом пкс получим по формуле (4.3):

$$U' = 7 + (9 - 7)\sqrt{1 - \frac{150}{750}} = 8{,}8\%.$$

Таблица 4.1

Верхний предел точности для совокупности U, % (доверительная вероятность 85%)

Объем выборки n	Ожидаемая вероятность отклонения p, %										
	1	2	3	4	5	6	7	8	9	10	12
40	1,8	3,5	4,8	6,3	7,4	8,8	10,0	11,5	12,5	13,8	16,2
50	1,7	3,3	4,7	5,9	7,3	8,6	9,9	11,2	12,1	13,5	15,9
80	1,7	3,2	4,6	5,8	7,2	8,3	9,5	11,0	11,9	13,0	15,2
100	1,7	3,0	4,4	5,5	6,8	8,0	9,2	10,5	11,5	12,7	14,9
150	1,6	3,0	4,1	5,4	6,5	7,7	9,0	10,0	11,0	12,2	14,5
200	1,5	2,7	4,0	5,3	6,4	7,5	8,8	9,8	10,8	11,9	14,3
250	1,4	2,7	3,9	5,2	6,3	7,4	8,6	9,5	10,6	11,8	13,9
300	1,4	2,7	3,8	5,0	6,2	7,3	8,4	9,3	10,5	11,7	13,8
350	1,4	2,6	3,7	5,0	6,1	7,2	8,2	9,2	10,4	11,5	13,6
400	1,3	2,6	3,7	4,9	6,0	7,2	8,1	9,2	10,3	11,4	13,6

Таблица 4.2

Верхний предел точности для совокупности U, % (доверительная вероятность 90%)

Объем выборки n	Ожидаемая вероятность отклонения p, %										
	1	2	3	4	5	6	7	8	9	10	12
40	2,2	4,1	5,5	7,3	9,0	9,8	11,8	12,5	14,1	15,0	17,5
50	2,0	3,8	5,3	7,0	8,2	9,3	11,0	12,0	13,8	14,5	17,0
80	1,9	3,6	5,0	6,4	7,6	9,0	10,2	11,3	12,5	13,8	16,2
100	1,8	3,5	4,8	6,1	7,3	8,7	9,9	11,0	12,1	13,7	15,7
150	1,7	3,2	4,5	5,8	7,1	8,2	9,4	10,7	11,8	13,5	15,2
200	1,7	3,1	4,2	5,6	6,8	8,0	9,1	10,2	11,5	12,6	14,8
250	1,6	2,9	4,1	5,4	6,6	7,9	9,0	10,0	11,2	12,3	14,5
300	1,6	2,9	4,1	5,3	6,5	7,6	8,8	9,8	11,0	12,1	14,3
350	1,5	2,8	4,1	5,2	6,3	7,5	8,6	9,7	10,9	12,0	14,1
400	1,5	2,8	4,0	5,1	6,3	7,4	8,6	9,6	10,7	11,9	14,0

Таблица 4.3

Верхний предел точности для совокупности U, % (доверительная вероятность 95%)

Объем выборки n	Ожидаемая вероятность отклонения p, %											
	1	2	3	4	5	6	7	8	9	10	12
40	4,5	5,1	7,2	8,5	10,0	12,0	12,7	14,9	16,2	17,4	19,8
50	3,0	5,0	7,0	8,0	9,8	11,6	12,1	14,0	15,8	16,2	19,0
80	2,5	4,8	6,0	7,4	8,7	10,1	11,3	12,6	14,0	15,2	17,5
100	2,3	4,1	5,8	7,0	8,5	9,6	11,0	12,2	13,5	14,4	17,0
150	2,2	3,9	5,2	6,3	7,8	9,2	10,3	11,5	12,7	14,0	16,1
200	2,1	3,6	5,0	6,1	7,4	8,6	9,9	11,0	12,2	13,2	15,8
250	1,9	3,4	4,7	6,0	7,2	8,4	9,6	10,8	12,0	13,2	15,3
300	1,8	3,2	4,5	5,8	7,0	8,2	9,3	10,5	11,7	13,1	15,0
350	1,8	3,2	4,4	5,1	6,8	8,0	9,2	10,3	11,4	12,5	14,8
400	1,7	3,1	4,3	5,6	6,7	7,9	9,1	10,2	11,3	12,5	14,6

Повышение уровня доверительной вероятности увеличивает верхний предел точности. Так, для F_m = 90% и при сохранении остальных условий примера 4.1 по табл. 4.2 находим U = 9,4%. В свою очередь увеличение размера выборки снижает верхний предел точности при всех прочих равных условиях. Для варианта p = = 7%, n = 200 и F_m = 90% находим U = 9,1%.

Определить максимальное количество отклонений для совокупности можно и на основе гипергеометрического распределения (см. § 3.2).

ПРИМЕР 4.2. Определим максимальное количество отклонений для совокупности на основе гипергеометрического распределения. Воспользуемся при этом результатами выборочного обследования примера 4.1: p = 0,07, n = 150, N = 750. Для данных условий находим значения вероятностей p_k и функции распределения F_k, приведенные в табл. 4.4.

Таблица 4.4

Гипергеометрическое распределение

k	k/n, %	p_k	F_k
7	4,7	0,0728	0,1511
8	5,3	0,1053	0,2564
9	6,0	0,1312	0,3876
10	6,7	0,1426	0,5302
11	7,3	0,1364	0,6666
12	8,0	0,1156	0,7822
13	8,7	0,0876	0,8698
14	9,3	0,0594	0,9241
15	10,0	0,0364	0,9655

Если доверительная вероятность F_m = 85%, то ближайшее значение относительного максимального количества отклонений в совокупности примерно равно 8,5%. Это немного ниже, чем результат, полученный на основе биномиального распределения.

В некоторых руководствах по аудиту предлагают более простые методы оценивания максимального количества отклонений в совокупности. Например, "Deloitte&Touche" для быстрого и приближенного оценивания U рекомендует формулу

$$U = f_m \frac{100}{n}\%, \qquad (4.4)$$

где n – размер выборки;

f_m – коэффициент.

Впрочем, эта формула не предназначена для проверки качества контрольной системы[1].

Таблица 4.5

Значения коэффициентов f_m (доверительная вероятность 90%; 95%)

Приемлемое количество отклонений в выборке m	f_m	
	90%	95%
0	2,31	3,00
1	3,89	4,75
2	5,33	6,30
3	6,69	7,76
4	8,00	9,16
5	9,28	10,52

ПРИМЕР 4.3. В выборке размером 100 документов обнаружено 5 отклонений. Максимальное количество отклонений для совокупности при доверительной вероятности 90% составит по формуле (4.4)

$$U = 9{,}28 \cdot \frac{100}{100} = 9{,}28\%.$$

Для аналогичных условий по табл. 4.2 находим $U = 7{,}3\%$.

§ 4.4. Определение объема выборки

Для проведения выборочного тестирования необходимо определить объем выборки n. Разумеется, этот объем может быть установлен на основе опыта проведенных ранее случаев атрибутивного тестирования или по аналогии. Если же такого опыта нет, то

[1] Touche Ross International. Op. cit. P. 25.

приходится определять размер выборки исходя из принятого закона распределения случайной дискретной величины. Пусть в качестве закона формирования вероятностей появления k случайных событий принято биномиальное распределение. Напомним (см. § 3.2), что функция биномиального распределения определяется как

$$F_k = P(X \geq k) = \sum_k C_k^n p^k q^{n-k},$$

где p – ожидаемая вероятность отклонения в совокупности;

q – вероятность противоположного события ($q = 1 - p$);

k – количество отклонений;

C_k^n – число сочетаний.

Определить точное значение n, задаваясь размерами остальных параметров, на основе этой формулы довольно хлопотно. На практике необходимые объемы выборки приближенно (а в точном расчете здесь и нет необходимости) определяют по специальным таблицам. В зарубежных руководствах по аудиту приводятся несколько вариантов таких таблиц[1].

В табл. 4.6, 4.7 и 4.8 приводятся округленные значения необходимых объемов выборки. Данные таблиц определены для доверительных вероятностей 85, 90 и 95% в зависимости от приемлемых уровней отклонений T (в %) и ожидаемых показателей вероятностей отклонений в совокупности.

Символ (*) в ячейке таблицы означает, что объем выборки меньше 30, символ (***) – объем больше 500. Показатели таблиц получены путем подбора примерных округленных значений n, соответствующих значениям функции биномиального распределения, близким к 0,85, 0,90 и 0,95. Подбор производился по расширенным таблицам функции биномиального распределения, часть которых (до $n = 250$) приведена в Приложениях. Так, для доверительной вероятности 85%, ожидаемой вероятности отклонений в совокупности $p = 4\%$ и приемлемого уровня отклонений, равного 7%, находим ближайшее значение размера выборки. Им оказывается $n = 100$. Эта величина и показана в табл. 4.6 на пересечении строки $p = 4{,}0$ и столбца $T = 7$.

[1] Touche Ross International. Audit Sampling. Op. cit. P. 20; *Giliberti A.* Op.cit. P. 65. Что послужило основанием для получения таблиц – в источниках не указывается.

Таблица 4.6

Необходимые объемы выборки (доверительная вероятность 85%)

Ожидаемая вероятность отклонения p, %	Приемлемый уровень отклонений T, %											
	1	2	3	4	5	6	7	8	9	10	12
0,5	340	180	120	90	70	60	50	50	40	40	30
1,0	***	180	120	90	70	60	50	50	40	40	30
1,5		300	160	90	70	60	50	50	40	40	30
2,0		***	200	120	100	80	50	50	40	40	30
2,5			400	200	100	100	50	50	40	40	30
3,0			***	280	200	100	70	60	40	40	30
3,5				***	200	150	90	60	60	50	30
4,0					300	200	100	80	70	50	40
4,5					***	***	120	90	70	50	40
5,0							200	120	80	60	40
5,5							240	120	90	60	40
6,0							300	180	120	90	50
6,5							***	280	160	100	50
7,0								400	200	120	60
7,5								***	300	180	70
8,0									***	200	80
8,5										300	100
9,0										***	140
9,5											200
10,0											240
11,0											340
12,0											***

Если известен размер совокупности N, из которой извлекается выборка, то эта информация может быть использована для уточнения величины выборки (поправка на конечность совокупности). В этом случае

$$n = \frac{n_0}{1 + \frac{n_0}{N}}, \tag{4.5}$$

где n – размер выборки после поправки на конечность совокупности;

n_0 – размер выборки до поправки.

ПРИМЕР 4.4. Пусть допустимый уровень отклонений (приемлемая максимальная величина отклонений) равен 7%, ожидаемая вероятность отклонения (предполагаемая доля документов с отклонениями в тестируемой совокупности) составляет 3%. Если уровень надежности для данного вида проверки принят 85% (табл. 4.6), то из совокупности следует отобрать 70 документов.

Пусть теперь известен размер совокупности (N = 1000), тогда с учетом поправки получим по формуле (4.5)

$$n = \frac{70}{1 + \frac{70}{1000}} = 65.$$

Если ожидаемая доля документов с отклонениями установлена в виде интервала, скажем, 3–4%, то табличные значения размера выборки составят 70–120, примерно 100 единиц.

Как видно из приведенных таблиц, при увеличении ожидаемой вероятности отклонения и неизменности значений других параметров необходимый объем выборки увеличивается. Аналогичное по направлению влияние оказывает и рост доверительной вероятности. В свою очередь, рост приемлемого уровня отклонений сокращает этот объем.

Очевидно, что нет необходимости тратить большие усилия для получения точных значений размера выборки, тем более что его можно увеличить в ходе анализа. В связи с этим в руководствах по аудиту иногда предлагают более простые методы приближенного определения n. Например, "Deloitte&Touche" рекомендует следующую формулу[1]:

[1] Touche Ross International. Audit Sampling. Op. cit. P. 20.

Таблица 4.7

Необходимые объемы выборки (доверительная вероятность 90%)

Ожидаемая вероятность отклонения *p*, %	Приемлемый уровень отклонений *T*, %										
	1	2	3	4	5	6	7	8	9	10	12
0,5	400	200	140	100	80	70	60	50	50	40	40
1,0	***	400	180	180	80	70	60	50	50	40	40
1,5		***	320	200	120	70	60	50	50	40	40
2,0			***	320	140	90	80	50	50	40	40
2,5				***	160	90	80	70	60	40	40
3,0					250	120	100	90	60	60	50
3,5					400	160	140	100	80	70	50
4,0					***	200	150	100	90	70	50
4,5						300	220	160	120	80	60
5,0						***	320	160	120	80	60
5,5							***	280	160	120	70
6,0								380	200	160	80
6,5								***	260	180	90
7,0									400	200	100
7,5									***	280	120
8,0										460	160
8,5										***	200
9,0											260
9,5											380
10,0											***

Таблица 4.8

Необходимые объемы выборки (доверительная вероятность 95%)

Ожидаемая вероятность отклонения *p*, %	Приемлемый уровень отклонений *T*, %										
	1	2	3	4	5	6	7	8	9	10	12
0,5	***	240	160	120	80	70	60	50	50	40	40
1,0		320	160	120	80	70	60	50	50	40	40
1,5		***	260	160	120	70	60	50	50	40	40
2,0			400	200	140	90	80	50	50	40	40
2,5			***	300	160	90	80	70	60	40	40
3,0				***	260	120	100	90	60	60	50
3,5					400	160	140	100	80	70	50
4,0					***	200	200	100	90	70	50
4,5						300	220	160	120	80	60
5,0						***	320	160	120	80	60
5,5							***	280	160	120	70
6,0								380	200	160	80
6,5								***	260	180	90
7,0									400	200	100
7,5									***	280	120
8,0										460	160
8,5										***	200
9,0											260
9,5											380
10,0											***

$$n = f_m \frac{100}{T}, \qquad (4.6)$$

где T – приемлемое количество отклонения в совокупности (в %);
f_m – коэффициент из табл. 4.9.

Таблица 4.9

Коэффициенты f_k

Приемлемое количество отклонений в выборке k	f_k	
	90%	95%
0	2,31	3,00
1	3,89	4,75
2	5,33	6,30
3	6,69	7,76
4	8,00	9,16
5	9,28	10,52

ПРИМЕР 4.5. Необходимо определить размер выборки при проверке совокупности документов на соответствие. Условия: приемлемый уровень отклонений в совокупности 3%, в выборке ожидается не более 2 отклонений, доверительная вероятность 95%.

По табл. 4.9 находим $f_2 = 6{,}3$. Следовательно,

$$n = 6{,}3 \cdot \frac{100}{3} = 210.$$

Глава 5. СТАНДАРТНЫЕ МЕТОДЫ ОЦЕНИВАНИЯ СРЕДНЕЙ И СУММЫ СТОИМОСТНЫХ ХАРАКТЕРИСТИК СОВОКУПНОСТИ. ОЦЕНИВАНИЕ ДОЛЕЙ

Правда всегда торжествует. Но жизнь коротка.

Неизвестный автор

§ 5.1. Точечные и интервальные оценки. Нестратифицированная выборка

В предыдущей главе обсуждалась методика проверки на соответствие по данным атрибутивных выборок. Перейдем к проверкам по существу. В данной главе рассматривается один вид проверки по существу, который заключается в оценивании среднего размера какого-либо стоимостного показателя (mean per unit estimation), например среднего остатка на счете, средней суммы просроченного кредита и т.п. На основе выборочных средних легко оценить общие суммы соответствующих показателей в совокупности.

При оценивании средней и суммы возможны два варианта. Согласно первому правильность зарегистрированных стоимостных показателей документов не проверяется. По второму варианту выборочные стоимостные характеристики документов тестируются аудитором. Обнаруженные отклонения (ошибки) статистически обобщаются в виде соответствующих средних или сумм.

Необходимость оценивания по выборке средних и сумм стоимостных показателей для совокупности возникает в ряде ситуаций. Например, когда эти данные по каким-либо причинам отсутствуют, однако они нужны аудитору для анализа, для различных сопоставлений. Нельзя исключить случаи, когда необходимо выяснить правильность представленной для проверки итоговой суммы, не обращаясь ко всем документам совокупности.

Оценивание средней и суммы по выборке является стандартной и наиболее простой постановкой задачи определения неизвестных параметров совокупности. Обсуждение простой постановки целесообразно в методическом плане до рассмотрения более важной для аудита специальной постановки задачи. Последняя предполагает предварительное тестирование стоимостных характеристик, а также оценивание абсолютных и относительных ошибок регистрации и определение их влияния на суммарные стоимостные показатели в совокупности. Соответствующие методики обсуждаются в следующей главе.

При оценивании параметров совокупности по данным выборок получают два вида оценок: *точечные* и *интервальные*. В качестве точечной оценки средней для совокупности применяют выборочную среднюю арифметическую величину. Обозначим ее как $\tilde{x}$:

$$\tilde{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad (5.1)$$

где x_i – стоимостная характеристика документа i;

i – номер документа в выборке, $i = 1, 2, \ldots, n$;

n – объем выборки.

Оценивание среднего размера стоимостной характеристики документа является первым шагом при оценивании суммы. После того как выборочная средняя определена, нетрудно найти точечную оценку суммы для совокупности. Для этого распространим значение полученной средней на всю совокупность:

$$\tilde{Q} = \tilde{x}N, \qquad (5.2)$$

где $\tilde{Q}$ – точечная оценка суммы для совокупности;

N – общий размер тестируемой совокупности.

Точечную оценку суммы можно получить и на основе выборочного значения суммы $\sum_{1}^{n} x_i$. Естественно полагать, что точечная оценка суммы для совокупности в N/n раз больше, чем сумма, полученная по выборке, т.е.

$$\tilde{Q} = \sum_{i=1}^{n} x_i \frac{N}{n} = \tilde{x}N.$$

При получении точечной оценки средней не учитывается один важный момент – оценка содержит некоторую ошибку, связанную с тем, что при расчете была охвачена только часть документов, а не вся совокупность. В одной выборке средняя имеет одно значение, в другой из этой же совокупности – другое. В связи с этим следует ожидать, что действительная средняя тестируемой совокупности более или менее отличается от полученной выборочной средней, находится в ее окрестности. Таким образом, задача заключается в определении границ интервалов, в которых находятся средняя и сумма для совокупности. Основой для таких интервалов всегда являются соответствующие точечные оценки.

Границы интервала для средней совокупности находятся следующим образом:

$$\bar{x} = \tilde{x} \pm \Delta, \qquad (5.3)$$

ãääå Δ – предельная ошибка выборочной средней.

В свою очередь, границы интервала для суммы стоимостных показателей совокупности находят как

$$Q = N(\tilde{x} \pm \Delta) = \tilde{Q} \pm N\Delta. \qquad (5.4)$$

Как следует из (5.3) и (5.4), отклонения от точечных оценок (Δ и $N\Delta$) симметричны относительно центральных значений в соответствующих интервалах.

Если расчет $\tilde{x}$ не требует комментариев, то метод определения Δ, вероятно, нуждается в них. Начнем с того, что выборочная средняя является случайной величиной. Ее значение зависит от случая – от того, какие документы попадут в выборку. Изменчивость выборочной средней измеряется, как известно, с помощью *стандартной ошибки средней*. Обозначим эту величину как μ:

$$\mu = \frac{s}{\sqrt{n}}, \qquad (5.5)$$

где s – стандартное (среднее квадратическое) отклонение в выборке;

n – объем выборки.

Стандартное отклонение находится по выборочным данным следующим образом:

$$s = \sqrt{\frac{\sum_{i} (x_i - \tilde{x})^2}{n-1}}. \qquad (5.6)$$

Поскольку $\sum_{1}^{n}(x_i-\tilde{x})^2=\sum x_i^2-n\tilde{x}^2=n\sum_i x_i^2-(\sum_i x_i)^2$, то расчет s часто удобно осуществить, применив одну из следующих рабочих формул[1]:

$$s=\sqrt{\frac{\sum_i x_i^2-n\tilde{x}^2}{n-1}}=\sqrt{\frac{n\sum_i x_i^2-(\sum_i x_i)^2}{(n-1)n}}. \qquad (5.7)$$

Вернемся к предельной ошибке выборочной средней, или предельной ошибке выборки Δ, фигурирующей в (5.3) и (5.4). Эта величина характеризует наибольшее отклонение от выборочной средней, соответствующее принятому уровню надежности. Размер предельной ошибки в конечном счете зависит от двух факторов – степени изменчивости исходных данных и принятого уровня надежности оценки.

Влияние первого фактора измеряется стандартным отклонением средней μ. Чем больше изменчивость данных в выборке, тем, естественно, больше предельная ошибка. Аналогичное можно сказать и о влиянии второго фактора – уровня надежности, или доверительной вероятности, с которым мы уже познакомились в § 1.4. Уровень доверительной вероятности устанавливает аудитор в зависимости от важности оцениваемого параметра для аудиторского заключения.

Согласно центральной предельной теореме, о которой говорилось в § 3.1, распределение выборочной средней следует нормальному закону. Таким образом, предельная ошибка определяется с помощью нормированного нормального распределения в зависимости от заданного уровня доверительной вероятности.

С учетом влияния двух факторов, о которых говорилось выше, предельная ошибка выборки определяется следующим образом:

$$\Delta = z\mu, \qquad (5.8)$$

[1] Доказательство см. в Дополнении к гл. 5.
Для расчета стандартного отклонения s можно воспользоваться электронной таблицей Excel. Причем при большом объеме выборки, вероятно, удобнее с помощью электронной таблицы определить суммы Σx, $(\Sigma x)^2$, Σx^2, а далее воспользоваться формулой (5.7). Если же объем выборки невелик, то для получения величины s можно использовать функцию СТАНДОТКЛОНА в пакете Excel.
Сумму квадратов отклонений в (5.6) и (5.7) иногда делят не на $n-1$, а на n, что приводит к небольшому смещению величин s.

где z – нормированное отклонение от средней (см. формулу (3.1) из § 3.1), значения z для некоторых величин вероятности P приведены в табл. 5.1.

Окончательно для предельной ошибки выборочной средней имеем

$$\Delta = z\sqrt{\frac{\sum_i (x_i-\bar{x})^2}{(n-1)n}}. \qquad (5.9)$$

Корень всегда берется с положительным знаком.

Из формулы (5.9) следует, что чем больше объем выборки, тем меньше ошибка выборки при всех прочих равных условиях.

Таблица 5.1

Краткая таблица параметра z для нормального распределения[1]

P, %	80	85	90	95	99
z	1,28	1,44	1,64	1,6	2,57

На практике в выборочных обследованиях часто применяют $z = 2$, что соответствует доверительной вероятности 95,4% .

ПРИМЕР 5.1. Тестируется совокупность счетов-фактур. Необходимо выборочным путем определить их общую сумму. Количество счетов в совокупности 2 тыс., объем выборки 120 единиц. В табл. 5.2 приведены значения стоимостных показателей (в рублях) счетов-фактур (первые и последние семь счетов, оказавшихся в выборке) и суммы, которые требуются для расчета s (см. (5.7)).

Таблица 5.2

Расчет сумм, необходимых для определения стандартного отклонения

i	x_i	x_i^2
1	1205	1452025
2	92	8464
3	910	828100
4	3150	9922500

[1] Более полная таблица помещена в Приложениях (см. табл. 2).

Окончание табл. 5.2

i	x_i	x_i^2
5	2546	6482116
6	1293	1671849
7	315	99225
...	...	...
114	136	18496
115	1809	3272481
116	1521	2313441
117	837	700569
118	2100	4410000
119	3123	9753129
120	128	16384
Суммы	165445	350181362

Точечная оценка среднего значения для совокупности (выборочная средняя):

$$\bar{x} = \frac{165445}{120} = 1379 \text{ руб.}$$

Точечная оценка суммы для совокупности:

$$\tilde{Q} = 1379 \cdot 2 = 2758 \text{ тыс. руб.}$$

Перейдем к интервальному оцениванию. Определим стандартное отклонение и стандартную ошибку средней. По формулам (5.7) и (5.5) последовательно находим:

$$s = \sqrt{\frac{120 \cdot 350181362 - 165445^2}{119 \cdot 120}} = 1013 \text{ руб.},$$

$$\mu = \frac{1013}{\sqrt{120}} = 92 \text{ руб.}$$

Пусть доверительная вероятность 90% ($z = 1{,}64$). Тогда с вероятностью 90% можно утверждать, что средний размер счета в совокупности находится в пределах:

$$\bar{x} = 1379 \pm 1{,}64 \cdot 92 = 1379 \pm 152 \text{ руб.},$$

а общая сумма счетов составляет:

$$Q = 2 \cdot (1379 \pm 152) \text{ тыс. руб.}$$

Таким образом, с вероятностью 90% можно утверждать, что общая сумма счетов определяется пределами:

$$Q_н = 2 \cdot 1227 = 2454 \text{ и } Q_в = 2 \cdot 1531 = 3062 \text{ тыс. руб.}$$

Приведенные выше формулы (5.6) и (5.7) для расчета s, строго говоря, получены для повторного отбора (см. § 2.1), который в принципе вряд ли уместен в аудите. Применение указанных формул при бесповторном отборе приводит к некоторым смещениям в получаемых оценках. Впрочем, при небольшой величине соотношения n/N этими смещениями вполне можно пренебречь. Для бесповторного отбора оценку μ получим по формуле

$$\mu = \sqrt{\frac{s^2}{n}} \cdot \sqrt{1 - \frac{n}{N}} = \sqrt{\frac{s^2}{n}\left(1 - \frac{n}{N}\right)}. \qquad (5.10)$$

С множителем $\sqrt{1 - \frac{n}{N}}$ мы уже встречались в гл. 4. Там он был назван поправкой на конечность совокупности (имеется в виду ограниченность совокупности, из которой произведена выборка), или кратко пкс. При относительно небольшом относительном объеме выборки эта величина близка к единице.

ПРИМЕР 5.2. Продолжим пример 5.1 и оценим ошибку средней теперь по формуле для бесповторного отбора. По условиям примера $N = 2000$, $n = 120$, тогда пкс составит всего

$$\sqrt{1 - \frac{120}{2000}} = 0{,}97.$$

Таким образом, с учетом пкс: $s = 1013 \cdot 0{,}97 = 983$ руб., $\mu = 92 \cdot 0{,}97 = 89$ руб., $\bar{x} = 1379 \pm 1{,}64 \cdot 89 = 1379 \pm 146$ руб. и $Q = 2(1379 \pm 146)$, $Q_н = 2 \cdot 1233 = 2466$ и $Q_в = 2 \cdot 1525 = 3050$ тыс. руб. Как видим, учет пкс практически не изменил значения рассчитываемых параметров по сравнению с полученными в примере 5.1 ответами.

Выше внимание было сосредоточено на определении предельной ошибки выборочной средней. Для расчета последней необходимо иметь стандартную ошибку этой средней μ. Практически ничего не изменится, если расчет базировать на стандартной ошибке выборочной суммы μ_Σ:

$$\mu_\Sigma = \frac{NS}{\sqrt{n}}. \qquad (5.11)$$

Как видим, стандартная ошибка выборочной суммы в N раз больше стандартной ошибки выборочной средней.

§ 5.2. Расслоенный отбор

Расслоенный, или стратифицированный, отбор предусматривает деление обследуемой совокупности на группы, слои или страты (§ 2.1). Каждый слой содержит более или менее однородные в каком-либо отношении единицы наблюдения, например группы документов по источникам поступлений, по размеру сумм, по срокам задолженности и т.п. Из каждого слоя совокупности в случайном порядке производится отбор единиц наблюдения. В аудите наиболее часто критерием для разбивки документов по слоям является стоимостной показатель.

Применение расслоенного отбора, как правило, сокращает стандартное отклонение выборочной средней (суммы) и тем самым повышает точность оценивания без увеличения объема выборки и при одной той же надежности. Иначе говоря, расслоение сокращает интервалы оценок параметров совокупности. Заметим, что стандартное отклонение наиболее заметно уменьшается при увеличении количества слоев до 5. Далее эффект расслоения растет не так быстро. За пределами 20 слоев этот эффект пренебрежимо мал. Сказанное объясняет, почему в аудите часто применяют двухслойную выборку, о которой упоминалось в § 2.2, – затраты минимальные, а эффект ощутимый.

Аудитор, решив применить расслоение тестируемой совокупности, должен последовательно определить число слоев, границы между слоями, общий объем выборки и распределить этот объем между слоями.

Число слоев может быть установлено аудитором на основе собственного понимания структуры тестируемой совокупности, опыта или интуиции, иными словами, достаточно произвольно. Соответственно устанавливаются и границы интервалов. В то же время если единицы наблюдения ранжированы по величине стоимостных характеристик документов и установлено число слоев, то расслоение можно осуществить исходя из условия примерного равенства сумм в каждом слое[1].

ПРИМЕР 5.3. В примере 2.3 совокупность документов была разбита на четыре интервала произвольно. Общая оценка суммы стоимости совокупности документов равна 345 млн руб. Если исходить из условия равенства сумм стоимостных характеристик, то каждый слой должен содержать документов на сумму, примерно равную 345 : 4 ≈ 90 млн руб. Этому условию никак не отвечает расслоение, представленное в табл. 2.3.

После того как установлены число слоев и границы интервалов, необходимо общее количество единиц наблюдения в выборке (объем выборки) распределить по слоям. Как уже отмечалось в гл. 2, распределить этот объем можно *пропорционально* количеству документов в слое и *не пропорционально*. В первом варианте доля отобранных документов одинакова во всех слоях. Во втором – эта доля меняется от слоя к слою и обычно зависит от стоимостных характеристик каждого слоя.

Цели, порядок и методики оценивания в принципе остаются такими же, что и для выборки без расслоения, – получение точечных и интервальных оценок средней и суммы стоимостных характеристик.

Отбор единиц наблюдения, не пропорциональный размеру слоя в совокупности. Оценка средней для совокупности находится как средняя взвешенная из выборочных средних по сло-

[1] В литературе по выборочным методам в аудите иногда указывают на возможность применения еще одного метода определения границ интервалов – метода равных кумулятивных стандартных отклонений (equal cumulative square root). Для этого необходимо иметь данные о статистическом распределении совокупности и примерные оценки стандартных отклонений в каждом интервале распределения (см.: *Arens A.A., Loebbecke J.K.* Applications of Statistical Sampling to Auditing. N.Y.: Prentice-Hall, 1981). Однако если такая детальная информация о тестируемой совокупности уже имеется у аудитора, то возникает сомнение в необходимости применения выборок для оценивания параметров этой совокупности.

ям. Причем в качестве весов берутся количества документов в каждом из них.

Пусть общий размер выборки распределен по m слоям. Тогда средняя для слоя k составит

$$\tilde{x}_k = \frac{\sum_{i=1}^{n_k} x_i}{n_k}, \qquad (5.12)$$

где x_i – величина стоимостной характеристики выборочной единицы наблюдения i;

n_k – количество документов в выборке из слоя k, $k = 1,\ldots, m$.

Общая средняя для выборок по всем слоям

$$\tilde{x}_s = \frac{\sum_{k=1}^{m} \tilde{x}_k N_k}{N}, \qquad (5.13)$$

где N_k – количество документов в слое k;

N – размер совокупности.

По определению $N = \sum_k N_k$ и $n = \sum_k n_k$.

Параметр $\tilde{x}_s$ рассматривается как точечная оценка средней для совокупности. Заметим, что в общем случае средняя $\tilde{x}_s$ отличается от выборочной средней с весами, равными размерам выборок по слоям. Обозначим последнюю как $\tilde{x}_{sk}$:

$$\tilde{x}_{sk} = \frac{\sum_k \tilde{x}_k n_k}{n}. \qquad (5.14)$$

Известно, что дисперсия суммы независимых случайных величин (в нашем случае – выборочных средних по слоям) равна сумме дисперсий. Отсюда следует, что стандартная ошибка выборочной средней с учетом размеров слоев определяется в виде средней взвешенной стандартных ошибок по слоям:

$$\mu = \sqrt{\sum \frac{s_k^2}{n_k}\left(\frac{N_k}{N}\right)^2} = \frac{1}{N}\sqrt{\sum_k \frac{s_k^2}{n_k} N_k^2}. \qquad (5.15)$$

Стандартное отклонение выборочных данных из слоя k:

$$s_k = \sqrt{\frac{\sum_i (x_i - \tilde{x}_k)^2}{n-1}}.$$

Для расчета s_k можно воспользоваться рабочими формулами (5.7). В свою очередь, стандартную ошибку выборочной средней для бесповторного отбора (с учетом пкс) находим по формуле

$$\mu = \frac{1}{N}\sqrt{\sum_k \frac{s_k^2}{n_k} N_k^2 \left(1 - \frac{n_k}{N_k}\right)}. \qquad (5.16)$$

ПРИМЕР 5.4. Обратимся к данным примера 2.3 в гл. 2. Размер совокупности $N = 3700$. Массив данных разделен на 4 слоя по величине стоимостной характеристики. Необходимо по выборочным данным найти точечные и интервальные оценки суммы стоимостных показателей.

Допустим, что аудитор располагает некоторыми оценками стандартных отклонений по слоям. Из каждого слоя произведена выборка. Общий объем выборки n = 200. Распределение объема выборки по слоям (n_k) произведено с учетом стоимостного фактора.

Размеры слоев по числу документов, объемы выборки по слоям и другие данные, необходимые для расчета выборочной средней x_s и стандартной ошибки средней μ, показаны в табл. 5.3.

Таблица 5.3

Расчет данных для определения выборочной средней и суммы совокупности

Слои, тыс. руб.	N_k	n_k	$\tilde{x}_k$	$\tilde{x}_k N_k$	s_k^2	$\frac{s_k^2}{n_k} N_k^2$
0,1–20	2000	12	9,8	19600	11	$40{,}333 \cdot 10^6$
20–100	1000	35	68	68000	45	$57{,}857 \cdot 10^6$
100–400	500	72	236	11800	97	$32{,}67 \cdot 10^6$
400–1000	200	81	721	144200	152	$45{,}637 \cdot 10^6$
Всего	3700	200	–	243600	–	$176{,}797 \cdot 10^6$

Точечная оценка средней и суммы для совокупности:

$$\bar{x}_s = \frac{243600}{3700} = 65{,}8 \text{ тыс.руб.}, \quad \tilde{Q} = \sum_{k=1}^{m} \bar{x}_k N_k = 243600 \text{ тыс. руб.}$$

Перейдем к определению стандартной ошибки средней. Согласно (5.15) получим

$$\mu = \frac{1}{3700}\sqrt{176{,}797 \cdot 10^6} = 3{,}6 \text{ тыс. руб.}$$

Пусть доверительная вероятность равна 90%, тогда

$$\Delta = z\mu = 1{,}64 \cdot 3{,}6 = 5{,}9 \text{ и } Q = 243600 \pm 5{,}9 \cdot 3700 \text{ тыс.руб.}$$

Теперь есть основание утверждать, что оценка общей суммы для совокупности находится в пределах от 222 до 265 млн руб.

Как уже отмечалось выше, преимущество стратифицированного отбора заключается в более точной оценке предельной ошибки средней. Сказанное можно продемонстрировать даже на примере двухслойной выборки, о которой говорилось в § 2.2.

ПРИМЕР 5.5. Необходимо выборочным путем определить среднюю величину стоимостных показателей в совокупности документов. Общее количество документов равно 1500. Случайным образом (по таблице случайных чисел) отобрано без расслоения 150 документов (выборка 1).

В порядке эксперимента выборочное исследование было повторено. Теперь документы были разделены на две группы. В первую вошли документы со стоимостями до 100 тыс. руб., во вторую – превышающие эту границу. Из первой группы отобрано 50 единиц (выборка 2, слой 1), из второй – 100 единиц (выборка 2, слой 2). Таким образом, в первом варианте применен нестратифицированный отбор, во втором – двухслойная выборка. Сравним результаты оценивания, сопоставив только предельные ошибки выборочной средней, полученные по двум вариантам.

Выборочные средние и стандартные отклонения (без пкс) и другие данные, необходимые для получения соответствующих оценок для двух вариантов условий, показаны в табл. 5.4.

Таблица 5.4

Данные для расчетов точечных и интервальных оценок

	Выборка 1	Выборка 2, слой 1	Выборка 2, слой 2
N_k	–	950	550
n_k	150	50	100
$\bar{x}_k$	220,4	54,5	315,6
s, s_k	218	108	185

Определим интервал для средней по данным выборки 1 при условии, что доверительная вероятность равна 90% ($z = 1{,}64$):

$$\mu_1 = \frac{218}{\sqrt{150}} = 17{,}8; \quad \Delta_1 = z\mu_1 = 1{,}64 \cdot 17{,}8 = 29{,}2 \text{ тыс. руб.};$$

$$\bar{x} = 220{,}4 \pm 29{,}2 \text{ тыс. руб.}$$

Таким образом, средняя для совокупности находится в пределах от 191,2 до 249,6 тыс. руб.

Перейдем к двухслойной выборке. Общая средняя (точечная оценка для совокупности):

$$\bar{x}_s = \frac{\sum \bar{x}_k N_k}{N} = \frac{54{,}5 \cdot 50 + 315{,}6 \cdot 100}{150} = 228{,}6 \text{ тыс.руб.}$$

Стандартная ошибка средней при расслоении на два слоя:

$$\mu_2 = \frac{1}{1500}\sqrt{\frac{108^2 \cdot 950^2}{50} + \frac{185^2 \cdot 550^2}{100}} = 11{,}8 \text{ тыс.руб.}$$

Предельная ошибка:

$$\Delta_2 = z\mu_2 = 1{,}64 \cdot 11{,}8 = 19{,}3 \text{ тыс. руб.}$$

Средняя для совокупности лежит в интервале 228,6±19,3, а именно от 209,3 до 247,9 тыс.руб. Интервал здесь заметно уже, чем в предыдущем варианте.

Отбор, пропорциональный размеру слоя в совокупности. Оценку средней для совокупности можно получить по формуле

(5.13) или (5.14). Очевидно, что результат будет одинаковым, так как при пропорциональном отборе

$$\frac{n_k}{n} = \frac{N_k}{N}, \text{ отсюда } n_k = n\frac{N_k}{N}.$$

Не будем рассматривать во всех деталях порядок расчета границ интервала для средней и суммы. Вполне достаточно показать методику определения стандартной ошибки выборки без учета пкс. Для этого подставим значение n_k в формулу (5.15):

$$\mu = \frac{1}{N}\sqrt{\sum_k \frac{s_k^2}{n_k} N_k^2} = \sqrt{\frac{1}{n}\sum_k s_k^2 \frac{N_k}{N}} = \sqrt{\frac{\bar{s}^2}{n}}, \qquad (5.17)$$

где $\bar{s}^2$ – средняя взвешенная величина из выборочных дисперсий по слоям;

s_k – стандартное отклонение в слое k.

Аналогичная по содержанию величина, но для бесповторного отбора (с учетом пкс) находится следующим образом:

$$\mu = \sqrt{\frac{\bar{s}^2}{n}\left(1 - \frac{n}{N}\right)}. \qquad (5.18)$$

§ 5.3. Оценивание отношений (долей) для совокупности

К оцениванию отношений (долей) можно подойти, использовав методику, применяемую при оценивании средней по выборке для совокупности. Предварительно сделаем в методике оценивания некоторые уточнения. Разделим единицы наблюдения (документы) совокупности на две группы. Пусть в первой из них документы обладают некоторыми свойствами, например содержат приписки, ошибки, неправильно оформлены и т.д. Подобного рода случаи в гл. 4 были названы отклонениями. Во второй группе документов такие свойства не наблюдаются. Необходимо выборочным путем оценить для тестируемой совокупности долю документов с отклонениями. Иначе говоря, это обычная проверка на соответствие, с которой мы уже сталкивались в главе, посвященной атрибутивным выборкам.

Нестратифицированный отбор. Для начала положим, что оценивание производится без расслоения совокупности. Пусть количество документов первой группы в совокупности равно A. Величина A неизвестна, и ее необходимо оценить по выборке. Доля документов этой группы в совокупности соответствует отношению

$$P = \frac{A}{N},$$

где N – размер совокупности.

Для применения методики, обсуждавшейся в предыдущем параграфе, положим, что, если документ относится к первой группе (содержит отклонение), случайная переменная равна единице, $x_i = 1$, в противоположном случае $x_i = 0$. Таким образом, точечная оценка доли в совокупности, полученная по выборке, равна

$$p = \tilde{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{a}{n}, \qquad (5.19)$$

где a – количество отклонений в выборке;

n – объем выборки.

Итак, доля документов с отклонениями в выборке равна p, а без отклонений $q = 1 - p$. Точечная оценка количества документов с отклонениями в совокупности составит

$$\tilde{A} = pN. \qquad (5.20)$$

Для получения интервальной оценки доли в совокупности необходимо определить стандартную и предельную ошибки выборочной средней при условии, что x_i равно 1 или 0. Для этого воспользуемся формулой (5.6), в которой сумма квадратов отклонений определяется как $\sum_{i=1}^{n} x_i^2 - n\tilde{x}^2$. Рассмотрим элементы этой разности. Поскольку в оговоренных условиях $\sum_{i=1}^{n} x_i^2 = \sum x_i = a = np$ и $n\tilde{x}^2 = n\left(\frac{\sum x_i}{n}\right)^2 = np^2$, то окончательно

$$\sum_{i=1}^{n} x_i^2 - n\tilde{x}^2 = np - np^2 = np(1-p) = npq.$$

В итоге стандартное отклонение определяется как

$$s = \sqrt{\frac{npq}{n-1}}. \qquad (5.21)$$

В свою очередь, стандартная ошибка доли с учетом поправки на конечность совокупности составит

$$\mu = \sqrt{\frac{pq}{n} \cdot \frac{N-n}{N-1}} \approx \sqrt{\frac{pq}{n} \cdot \left(1 - \frac{n}{N}\right)}. \qquad (5.22)$$

Как и выше, при большом размере совокупности и относительно небольшом объеме выборки пкс можно пренебречь. Без пкс имеем

$$\mu = \sqrt{\frac{pq}{n}}. \qquad (5.23)$$

Предельную ошибку отношения находим стандартным путем: $\Delta = z\mu$, а границы интервалов оценок для отношения и количества документов с отклонениями в совокупности – как

$$\bar{p} = p \pm z\mu, \qquad (5.24)$$

$$A = N\left(p \pm z\mu\right) = \tilde{A} \pm z\mu N. \qquad (5.25)$$

ПРИМЕР 5.6. Из совокупности, содержащей 3000 документов, случайным образом отобрано 200 единиц. Оказалось, что 18 из них имеют отклонения. Итак: $N = 3000$, $n = 200$, $a = 18$. Необходимо по выборке оценить количество отклонений в тестируемой совокупности.

Точечные оценки доли и количества документов с отклонениями составят:

$$p = \frac{18}{200} = 0{,}09, \quad \tilde{A} = 0{,}09 \cdot 3000 = 270.$$

Для получения интервальных оценок определим стандартную ошибку доли (с учетом пкс):

$$\mu = \sqrt{\frac{0{,}09 \cdot 0{,}91}{200} \cdot \frac{3000-200}{3000-1}} = 0{,}0195.$$

Без учета пкс стандартная ошибка доли чуть больше 0,02.

Интервальную оценку доли найдем при уровне доверительной вероятности 85% ($z = 1{,}44$):

$$\bar{p} = 0{,}09 \pm 1{,}44 \cdot 0{,}0195 = 0{,}09 \pm 0{,}028.$$

В свою очередь, интервал для количества отклонений составит

$$A = 270 \pm 1{,}44 \cdot 0{,}0195 \cdot 3000 = 270 \pm 84.$$

Таким образом, аудитор с вероятностью 85% может утверждать, что в тестируемой совокупности содержится от 6,2 до 11,8% документов с отклонениями. Соответственно в совокупности ожидается от 186 до 354 таких документов.

Расслоенный отбор. Перейдем к оцениванию доли при расслоенном отборе единиц наблюдения. Не будем повторять то, что было сказано в § 5.2 о расслоении совокупности, и сразу приведем необходимые для расчетов формулы.

Пусть тестируемая совокупность разбита на m слоев. Тогда выборочная доля в слое k определяется как $p_k = a_k/n_k$, а точечной оценкой доли в совокупности является средняя

$$p_s = \frac{\sum_{k=1}^{m} p_k N_k}{N}, \qquad (5.26)$$

где a_k – количество документов с отклонениями в выборке из слоя k;

n_k – объем выборки из слоя k;

N_k – количество документов в слое k;

N – общее количество документов в совокупности, $N = \sum_{k=1}^{m} N_k$.

Приведем формулы расчета стандартной ошибки средней для разных условий отбора. Начнем с непропорционального расслоения. С учетом пкс имеем

$$\mu = \frac{1}{N}\sqrt{\sum_{k=1}^{m} \frac{p_k q_k}{n} N_k^2 \left(1 - \frac{n}{N}\right)}, \qquad (5.27)$$

где n – общий объем выборки, $n = \sum_{k=1}^{m} n_k$.

Для варианта без учета пкс получим

$$\mu = \frac{1}{N}\sqrt{\sum_{k=1}^{m}\frac{p_k q_k}{n}N_k^2}. \qquad (5.28)$$

В случае с пропорциональным отбором и учетом пкс

$$\mu = \frac{\sum_{k=1}^{m} p_k q_k n_k}{n\sqrt{n}}\sqrt{1-\frac{n}{N}}. \qquad (5.29)$$

Для этого варианта, но без учета пкс имеем

$$\mu = \frac{\sum_{k=1}^{m} p_k q_k n_k}{n\sqrt{n}}. \qquad (5.30)$$

ПРИМЕР 5.7. Несколько изменим условия примера 5.6. Пусть теперь совокупность разбита на два слоя: $N_1 = 1000$ и $N_2 = 2000$. Общий объем выборки 200 единиц наблюдения. Отбор пропорционально количеству документов в слое: $n_1 = 67$, $n_2 = 133$. Обнаружено, что в первом слое оказалось 8 документов с отклонениями, во втором – 10.

Выборочные доли по слоям: $p_1 = \frac{8}{67} = 0{,}119$; $p_2 = \frac{10}{133} = 0{,}075$.

Точечные оценки доли и количества документов для совокупности:

$$p_s = \frac{0{,}119 \cdot 67 + 0{,}075 \cdot 133}{200} = 0{,}0897; \quad \tilde{A} = 0{,}0897 \cdot 3000 = 269.$$

Стандартная ошибка доли

$$\mu = \frac{0{,}119 \cdot 0{,}801 \cdot 67 + 0{,}075 \cdot 0{,}825 \cdot 133}{200\sqrt{200}}\sqrt{1-\frac{200}{3000}} = 0{,}0048.$$

При доверительной вероятности 85% для доли документов с отклонениями в совокупности находим интервальные оценки

$$\tilde{p}_s = 0{,}0897 \pm 1{,}44 \cdot 0{,}0048 = 0{,}0897 \pm 0{,}0069.$$

Таким образом, аудитор вправе утверждать, что в проверенной совокупности документов содержится от 8,3 до 9,7% документов с отклонениями. В абсолютном выражении этому соответствует интервал от 249 до 291 документа.

§ 5.4. Определение необходимых объемов выборки (нестратифицированный отбор)

Все рассмотренные в предыдущих параграфах методики оценивания предполагали, что объемы выборок известны. Однако при проведении выборочных обследований прежде всего сталкиваются именно с проблемой определения объема выборки. При ее решении аудитор должен избежать двух крайностей. Слишком большой объем выборки потребует излишних денежных и трудовых затрат. В свою очередь, недостаточный объем обязательно скажется на надежности получаемых оценок параметров тестируемой совокупности.

На первый взгляд представляется, что размеры необходимого объема выборки легко получить на основе рассмотренных выше формул предельной ошибки средней, преобразовав их относительно n. Формально это так. Однако для расчета n необходимо иметь значение стандартного отклонения или оценки доли *до* отбора единиц наблюдения из совокупности. Возникает своего рода порочный круг. Выход состоит: а) в применении уже известной оценки стандартного отклонения или оценки доли, полученной ранее при оценивании параметров аналогичных совокупностей; б) в использовании ожидаемого или предполагаемого значения этого параметра; в) в определении значения n на основе данных специального (предназначенного именно для оценивания стандартного отклонения) выборочного обследования с последующей корректировкой полученного значения.

При определении объема выборки необходимо сформулировать некоторые требования к выборочным параметрам. Обычно это требования к точности оценок, которые можно связать с важностью документов тестируемой совокупности для аудиторского заключения. В качестве меры точности аудитор устанавливает (задает) допустимую, по его мнению, предельную ошибку средней, доли или суммарного значения тестируемых показателей для совокупности.

Начнем с нестратифицированной выборки. Для определения n при оценивании средней обратимся к формуле (5.8), в которой заменим Δ на d_x (приемлемую максимальную ошибку средней), а μ определим по (5.5). Если стандартная ошибка средней определена без учета пкс или при отсутствии данных о размерах тестируемой совокупности, то в итоге получим

$$n_0 = \left(\frac{z}{d_x} s\right)^2. \qquad (5.31)$$

С учетом пкс формулу (5.8) можно развернуть следующим образом:

$$d_x = z\mu = z\frac{s}{\sqrt{n}}\sqrt{1-\frac{n}{N}}.$$

Используя это выражение, получим предварительную оценку

$$n_0 = \frac{\left(\frac{z}{d_x} s\right)^2}{1+\frac{1}{N}\left(\frac{z}{d_x} s\right)^2} = \frac{1}{\left(\frac{z}{d_x} s\right)^{-2} + \frac{1}{N}}. \qquad (5.32)$$

Величина n_0 вычисляется по формулам (5.31) и (5.32) часто в качестве первого приближения. Если отношение n_0/N пренебрежимо мало, то полученные значения n_0 берутся в качестве окончательного результата. Если же этим отношением нельзя пренебречь, то окончательно имеем

$$n = \frac{n_0}{1+\frac{n_0}{N}}. \qquad (5.33)$$

ПРИМЕР 5.8. Необходимо определить объем выборки для оценки среднего размера остатков на счетах при простом случайном отборе. Всего счетов 30 000 единиц. Из предыдущих обследований аналогичных совокупностей известно, что стандартное отклонение составляет порядка 12,2 тыс. руб.

Аудитор полагает, что ошибка в оценке средней для совокупности не должна превышать 1,5 тыс. руб. при доверительной вероятности 90%.

Исходные данные: $N = 30000$, $s = 12{,}2$ тыс. руб., $d_x = 1{,}5$ тыс. руб., $z = 1{,}64$. Для этих условий объем выборки составит

$$n_0 = \left(\frac{1{,}64 \cdot 12{,}2}{1{,}5}\right)^2 = 178 \text{ счетов}.$$

Отношение n_0/N здесь пренебрежимо мало, около 0,5%, поэтому не будем корректировать полученный результат. После округления $n = 180$.

Изменим теперь одно условие задачи. Заметно сократим размер тестируемой совокупности. Пусть теперь $N = 1500$. В этом случае учет пкс обязательно скажется на размере выборки. После корректировки по формуле (5.33) получим

$$n = \frac{178}{1+\frac{178}{1500}} = 119 \text{ счетов}.$$

После округления $n = 120$.

Аналогичным путем определим n при оценивании доли в совокупности. В этом случае предварительная оценка

$$n_0 = \frac{\frac{z^2}{d_p^2} pq}{1+\frac{1}{N}\left(\frac{z^2}{d_p^2} pq - 1\right)}, \qquad (5.34)$$

где d_p – приемлемая максимальная ошибка доли (задается аудитором).

Без учета пкс или при отсутствии данных о размерах тестируемой совокупности имеем

$$n_0 = \frac{pq}{d_p^2}. \qquad (5.35)$$

Если отношение n_0/N достаточно велико и им нельзя пренебречь, то полученное значение n_0 корректируется (см. (5.33)).

ПРИМЕР 5.9. Определим объем выборки для оценивания доли документов с отклонениями в тестируемой совокупности. Размер совокупности 2500. Предварительная оценка для доли составляет 2–3%. Приемлемая для аудитора погрешность в оценке порядка 1% при доверительной вероятности 90%.

Исходные данные: $N = 2500$, $p = 0{,}03$, $q = 0{,}97$, $d_p = 0{,}01$, $z = 1{,}64$.

По формуле (5.35) находим

$$n_0 = \frac{0{,}03 \cdot 0{,}97}{0{,}01^2} = 291.$$

После корректировки с учетом размера совокупности получим

$$n = \frac{291}{1+\frac{291}{2500}} = 260 \text{ документов.}$$

Выборочное обследование может быть нацелено на оценивание не одной, а нескольких характеристик совокупности. Например, аудитору необходимы оценка суммы и одновременно доли документов с какими-либо признаками и т.п. В этом случае необходимый объем выборки определяют в отношении каждой оценки. Окончательно принимают наибольший из этих объемов.

§ 5.5. Определение необходимых объемов выборки (расслоенные выборки)

Перейдем к определению объема выборки при расслоенном отборе (размещении по слоям). Начнем с непропорционального размещения (непропорционально количеству документов в слое), когда доля каждого слоя в объеме выборки (f_k) установлена заранее на основе каких-либо правил или суждений аудитора. Например, доли выборки по слоям могут быть установлены пропорциональными суммам стоимостных характеристик слоев (см. пример 2.3) или с учетом каких-либо особенностей слоев и т.д.

Если пренебречь пкс, то в качестве первого приближения для определения n в общем случае имеем[1]

$$n_0 = \frac{z^2}{d_x^2}\sum_{k=1}^{m}\frac{F_k^2 s_k^2}{f_k}, \qquad (5.36)$$

$$f_k = \frac{n_k}{n}, \quad F_k = \frac{N_k}{N},$$

где N_k – количество документов в слое k;

N – размер тестируемой совокупности;

n_k – объем выборки из слоя k;

n – общий размер совокупности.

В свою очередь, для случая, когда пкс нельзя пренебречь, скорректированный объем можно получить по формуле

$$n = \frac{n_0}{1+\frac{z}{d_x N}\sum_{k=1}^{m} F_k s_k^2}. \qquad (5.37)$$

[1] Доказательство см. в Дополнении к гл. 5.

На первый взгляд представляется, что наиболее простым случаем непропорционального отбора является отбор одинаковых долей выборки из каждого слоя (равномерное размещение по слоям). По определению

$$f = 1/m = \text{const} \text{ и } n_k = \frac{n}{m}.$$

В итоге для равномерного размещения имеем

$$n_0 = \frac{z^2}{d_x^2 f}\sum_k F_k^2 s_k^2. \qquad (5.38)$$

Однако таким размещением можно воспользоваться лишь при относительно небольших различиях между слоями по количеству документов.

Перейдем к пропорциональному отбору (объем выборки из слоя пропорционален количеству документов в слое). Этот вид отбора относительно прост и поэтому обычно привлекателен для аудитора. По определению здесь

$$f_k = \frac{n_k}{n} = \frac{N_k}{N} = F_k.$$

Поэтому вместо (5.36) получим более простую формулу

$$n_0 = \frac{z^2}{d_k^2}\sum_{k=1}^{m} F_k s_k^2. \qquad (5.39)$$

Как и выше, корректировку первоначальной оценки объема выборки осуществим по формуле (5.33).

Расчет по формулам (5.36)–(5.39) предполагает, что стандартные отклонения по слоям (s_k) известны, например, исходя из прошлых проверок. Однако такое бывает весьма редко. Как правило, приходится обходиться примерными, правдоподобными оценками. Разумеется, можно, если возникнет необходимость, уточнить значение n на основе полученных по выборке величин стандартных отклонений.

Из сказанного выше следует, что метод отбора существенно влияет на объем выборки и его распределение по слоям.

ПРИМЕР 5.10. Необходимо определить объем выборки из тестируемой совокупности счетов-фактур. Число слоев 5. Размер тестируемой совокупности 1900 документов. Разбивка на слои, значения N_k, s_k и F_k представлены в табл. 5.5. Стандартные отклонения по слоям s_k (тыс. руб.) получены в предыдущих проверках аналогичных по содержанию документов.

Таблица 5.5

Исходные данные для расчета объема выборки

Номер слоя, k	Интервал m	N_k	s_k	F_k
1	>20	940	5	0,495
2	21–100	490	16	0,258
3	101–250	200	30	0,105
4	251–500	170	50	0,089
5	501–1000	100	60	0,053
Итого	–	1900	–	1,000

Данные, необходимые для расчета объема выборки из расслоенной совокупности при применении непропорционального и пропорционального отборов, показаны в табл. 5.6.

Доверительная вероятность установлена на уровне 90%, максимальная приемлемая ошибка в значении средней принята равной 2 тыс. руб. Таким образом, $z = 1{,}64$, $d_x = 2$ тыс. руб.

Начнем с непропорционального отбора. Зададимся некоторыми значениями f_k (см. табл. 5.6).

Таблица 5.6

Размещение объемов выборки по слоям (непропорциональное и пропорциональное)

k	f_k	$F_k^2 s_k^2$	$F_k^2 s_k^2 / f_k$	$F_k s_k^2$	n_k(НП)	n_k(ПР)
1	0,3	6,1	20,4	12,4	78	158
2	0,3	17,0	56,8	66,0	78	82
3	0,2	9,9	49,6	94,7	52	34
4	0,1	19,8	190,8	223,7	26	29
5	0,1	10,1	101,1	189,5	26	17
	1,0	62,9	418,7	586,3	260	320

По формуле (5.36) находим

$$n_0 = \frac{1{,}64^2}{2^2} \cdot 418{,}7 = 282 \text{ документа.}$$

Скорректируем полученную оценку с учетом пкс (см. (5.37)). Общий объем выборки теперь составит

$$n = \frac{282}{1 + \frac{1{,}64}{2 \cdot 1900} \cdot 586{,}3} = 255.$$

Округлим результат, пусть $n = 260$. Остается распределить этот объем выборки по слоям. Для этого воспользуемся заданными отношениями f_k. Объем выборки из первого и второго слоев $n_1 = n_2 = 0{,}3 \cdot 260 = 78$, из третьего слоя $n_3 = 0{,}2 \cdot 260 = 52$ и т.д. Полученные по всем слоям результаты округлены и показаны в столбце n_k(НП) табл. 5.6.

Перейдем к пропорциональному размещению по слоям. По формуле (5.39) находим:

$$n_0 = \frac{1{,}64^2}{2^2} \cdot 586{,}3 = 395.$$

После стандартной корректировки получим $n = 316$. Округлим до 320. Для размещения по слоям используем показатели F_k. Так, для первого интервала получим $n_1 = 0{,}495 \cdot 320 = 158$. Соответствующее размещение показано в столбце n_k (ПР) табл. 5.6.

Рассмотрим еще один метод определения объема выборки и размещения по слоям. Он предполагает отбор, пропорциональный взвешенным стандартным отклонениям по слоям. Такое размещение называют предположительно оптимальным (далее – оптимальным). В качестве весов здесь берутся размеры слоев. Таким образом, учитываются два фактора. Доля слоя в объеме выборки составляет:

$$f_k = \frac{n_k}{n} = \frac{N_k s_k}{\sum_k N_k s_k}. \qquad (5.40)$$

Объем выборки определяется по формуле

$$n = \frac{\left(\sum_k F_k s_k\right)^2}{\frac{z}{d_x} + \frac{1}{N}\sum F_k s_k^2}. \qquad (5.41)$$

ПРИМЕР 5.11. Продолжим пример 5.10. Определим объем выборки при оптимальном размещении. Данные для расчета и его результаты представлены в табл. 5.7.

Таблица 5.7

Определение объемов выборки по слоям (оптимальное размещение)

k	F_k	s_k	$F_k s_k^2$	N_k	$N_k s_k$	$(N_k/N)s_k$	f_k	n_k
1	0,495	5	12,4	940	4700	2,5	0,142	38
2	0,258	16	66,0	490	7840	4,1	0,237	64
3	0,105	30	94,7	200	6000	3,2	0,182	49
4	0,089	50	223,7	170	8500	4,5	0,257	70
5	0,053	60	189,5	100	6000	3,1	0,182	49
Итого	1,000		586,3	1900	33040	17,4	1,000	270

По формуле (5.41) находим

$$n = \frac{17{,}4^2}{\frac{1{,}64}{2} + \frac{1}{1900}\cdot 586{,}284} = 268 \text{ счетов}.$$

Округлим ответ до 270. Распределим объем между слоями. Для первого слоя получим $n_1 = 0{,}142 \cdot 270 = 38$, для второго – $n_2 = 0{,}237 \cdot 270 = 64$ и т.д.

Перейдем к расслоенному размещению выборки при оценивании доли для совокупности. Здесь также возможны пропорциональное и непропорциональное размещения по слоям. При пропорциональном размещении найденное по формуле (5.34) значение n делится по слоям в соответствии со структурой тестируемой совокупности:

$$n_k = f_k n,$$

где $f_k = N_k / N = F_k$.

Наиболее эффективным является размещение пропорционально стандартному отклонению (предположительно оптимальное размещение). В этом случае

$$f_k = \frac{N_k\sqrt{p_k q_k}}{\sum_k N_k\sqrt{p_k q_k}}. \qquad (5.42)$$

ПРИМЕР 5.12. Вернемся к тестируемой совокупности примера 5.9, где $N = 2500$, $p = 0{,}03$. Расчетный объем выборки 260 единиц наблюдения.

Совокупность разбита на 4 слоя по стоимостным показателям документов. Необходимо распределить объем выборки между слоями. Применим пропорциональное и оптимальное размещения. Расчет по первому варианту показан в табл. 5.8. Для первого интервала находим $n_1 = 0{,}4 \cdot 260 = 104$.

Таблица 5.8

Определение объемов выборки по слоям (пропорциональное размещение)

k	Интервал	N_k	$f_k = F_k$	n_k
1	>30	1000	0,40	104
2	31–200	800	0,32	83
3	201–500	500	0,20	52
4	501–1000	200	0,08	21
Итого	–	2500	1,00	260

Данные, необходимые для оптимального размещения по слоям, приведены в табл. 5.9. В столбце p_k показаны ожидаемые аудитором значения вероятностей отклонений по слоям. Доля выборки из слоя определяется по формуле (5.42). Объем выборки из первого слоя составляет $n_1 = 0{,}334 \cdot 260 = 87$.

Таблица 5.9

Определение объемов выборки по слоям (оптимальное размещение)

k	N_k	p_k	$\sqrt{p_k q_k}$	$N_k\sqrt{p_k q_k}$	f_k	n_k
1	1000	0,02	0,140	140	0,334	87
2	800	0,03	0,171	136	0,325	85
3	500	0,046	0,209	104	0,248	64
4	200	0,04	0,196	39	0,093	24
Итого	2500	0,03	–	419	1,000	260

§ 5.6. Малые выборки

Далеко не всегда аудитор может осуществить выборку в полном объеме. В силу разных причин, например из-за нехватки времени, он ограничивается небольшим объемом выборки – менее 20 единиц. Это так называемые *малые выборки*. Аудитор и при анализе малых выборок в ряде ситуаций может быстро получить полезную информацию.

Как уже отмечалось в § 5.1, согласно центральной предельной теореме выборочные средние имеют нормальное или близкое к нему распределение. Однако это утверждение, справедливое для больших выборок, не относится к выборочным средним, извлеченным из малых выборок. Выходом является применение другого распределения, а именно распределения нормированных отклонений выборочных средних от средней генеральной совокупности. Это распределение получило название *t-распределение Стьюдента*, или кратко *t-распределение*[1]. Распределение получено для случайной выборки из нормально распределенной генеральной совокупности. Оно зависит только от одного параметра – числа степеней свободы v:

$$v = n - 1,$$

где n – размер выборки.

[1] Характеристика распределения Стьюдента на интуитивном уровне, достаточном для понимания существа дела лицами со скромной математической подготовкой, приведена в книге: *Вайнберг Дж., Шумекер Дж.* Статистика. М.: Статистика, 1979, § 14.2.

Величина t находится как отношение разности выборочной и генеральной средних к стандартной ошибке выборки:

$$t = \frac{\tilde{x} - \bar{x}}{\mu}, \qquad (5.43)$$

где $\tilde{x}$ – выборочная средняя;

$\bar{x}$ – средняя для совокупности;

μ – стандартная ошибка выборки.

Стандартная ошибка выборки находится как

$$\mu = \frac{s}{\sqrt{n}}. \qquad (5.44)$$

Выборочное стандартное отклонение s можно определить по формуле (5.7).

Как видим, величина t характеризует отклонение выборочной средней от средней совокупности, нормированное по стандартной ошибке выборки.

Распределение Стьюдента симметрично относительно нуля, как и нормальное распределение, но более пологое. С ростом числа степеней свободы оно приближается к нормальному. Площадь под кривой распределения равна 1, или 100%. Площадь под кривой для значений, превышающих t (или меньше $-t$), характеризует вероятность превышения указанного значения. Обозначим эту вероятность как p.

Значения t-распределения Стьюдента для разных величин степеней свободы и уровней вероятностей превышения абсолютного значения t приведены в табл. 4 Приложений. Например, для $v = 7$ и вероятностей, равных 0,2 , 0,15 и 0,1, находим следующие табличные значения t: 1,415, 1,617 и 1,895. Значения t в зависимости от p и v можно определить и по электронной таблице Excel (см. Дополнение к главе).

Вернемся к формуле (5.43). Преобразуем ее относительно средней генеральной совокупности:

$$\bar{x} = \tilde{x} \pm t\mu. \qquad (5.45)$$

Как видим, приведенная формула позволяет оценить границы для неизвестной средней генеральной совокупности. Границы

определяются таким же способом, что и при оценивании интервалов на основе большой выборки. Соответствующие методики обсуждались в § 5.1. Различие состоит только в том, что здесь вместо квантиля z нормального распределения используется величина t.

С доверительной вероятностью, равной $1 - p$, можно утверждать, что неизвестная генеральная средняя находится в пределах, определяемых формулой (5.45). Произведение $t\mu$ в этой формуле характеризует *предельную ошибку выборочной средней*. Чем больше доверительная вероятность, тем больше t и, следовательно, предельная ошибка выборки. Уровень доверительной вероятности можно указывать при записи величины t. Например, для $p = 0{,}05$ запишем t_{95} и т.д.

ПРИМЕР 5.13. Необходимо на основе выборки быстро оценить пределы, в которых находится средняя величина для совокупности чеков.

Для оценивания случайным образом отобрано 18 чеков со следующими характеристиками (тыс. руб.):

40,9	14,6	48,9	35,6
29,3	41,4	38,7	56,1
50,5	18,5	28,9	65,7
45,9	20,6	43,7	
50,6	32,7	64,9	

По приведенным данным число степеней свободы равно 17, выборочная средняя 40,42 тыс. руб., стандартное отклонение 14,72 тыс. руб. Стандартная ошибка выборочной средней

$$\mu = \frac{14{,}72}{\sqrt{18}} = 3{,}47 \text{ тыс.руб.}$$

По таблице распределения Стьюдента (табл. 4 Приложений) для 17 степеней свободы получим следующие значения t для доверительных вероятностей 85, 90, 95 и 99% (с округлением до третьего знака): $t_{85} = 1{,}508$; $t_{90} = 1{,}740$; $t_{95} = 2{,}101$; $t_{99} = 2{,}898$.

Для принятых уровней доверительных вероятностей оценим величины интервалов, в которых следует ожидать среднее значение суммы чеков совокупности (тыс. руб.):

$$\bar{x} = 40{,}42 \pm 1{,}508 \cdot 3{,}47 \approx 35{,}2 \div 45{,}7,$$
$$\bar{x} = 40{,}42 \pm 1{,}740 \cdot 3{,}47 \approx 34{,}4 \div 46{,}5,$$
$$\bar{x} = 40{,}42 \pm 2{,}101 \cdot 3{,}47 \approx 33{,}1 \div 47{,}4,$$
$$\bar{x} = 40{,}42 \pm 2{,}898 \cdot 3{,}47 \approx 30{,}4 \div 50{,}5.$$

Как следует из полученных ответов, рост доверительной вероятности с 85 до 99% вдвое увеличил размеры интервалов. Решение этой же задачи, но на основе большой выборки значительно сузит искомые интервалы.

Дополнение к главе

А. Расчет стандартного отклонения (среднего квадратического отклонения) по таблице Excel

Последовательность операций.

1. Разместить исходные данные (выборку) в ячейках таблицы Excel (например, в ячейках от A1 до A18).
2. Выделить для результата ячейку в той же таблице.
3. Вызвать функцию: $f_x \rightarrow$ Статистические $\rightarrow$ СТАНДОТКЛ.
4. В диалоговое окно ввести координаты ячеек таблицы:

например, `= СТАНДОТКЛ (A1:A18)` .

Ответ показывается в выделенной ячейке таблицы.

Б. Определение значений t-распределения Стьюдента по таблице Excel

Последовательность операций

1. Выделить ячейку в таблице Excel.
2. Вызвать функцию: $f_x \rightarrow$ Статистические $\rightarrow$ СТЬЮДРАСПОБР.
3. В диалоговое окно ввести:

Вероятность – указывается принятая доверительная вероятность;

Степень свободы – показывается число степеней свободы.

Эти же данные после вызова функции можно ввести и непосредственно в формульное окно таблицы `=СТЬЮДРАСПОБР ( )`.

Например, для параметров $p = 0{,}2$ и $\nu = 20$ запишем в формульном окне:

`СТЬЮДРАСПОБР (0,2;16)` .

В выделенной ячейке таблицы Excel получим 1,336757.

В. Рабочие формулы для расчета суммы квадратов отклонений от выборочной средней

Запишем:

$$\sum_{i=1}^{n}(x_{ii}-\tilde{x})^2=\sum_{1}^{n}(x_i^2-2x_i\tilde{x}+\tilde{x}^2)=\sum_{1}^{n}x_i^2-\left(2\sum_{1}^{n}x_i\right)\tilde{x}+\sum_{1}^{n}\tilde{x}^2=$$

$$=\sum_{1}^{n}x_i^2-2\left(\sum_{1}x_i\right)\frac{\sum x_i}{n}+n\left(\frac{\sum x_i}{n}\right)^2=\frac{n\sum x_i^2-\left(\sum x_i\right)^2}{n}. \quad (1)$$

Эту же сумму можно получить и иначе. Поскольку

$$(2\sum_{1}^{n}x_i)\tilde{x}=2\left(\frac{\sum x_i}{n}n\right)\tilde{x}=2n\tilde{x}^2,$$

то получим другой вариант формулы:

$$\sum_{i=1}^{n}(x_{ii}-\tilde{x})^2=\sum_{1}^{n}x_i^2-\left(2\sum_{i}^{n}x_i\right)\tilde{x}+\sum_{1}^{n}\tilde{x}^2=$$

$$=\sum_{1}^{n}x_i^2-2n\tilde{x}^2+n\tilde{x}^2=\sum_{1}^{n}x_i^2-n\tilde{x}^2. \quad (2)$$

Г. Доказательство формулы (5.39)

Заменим в (5.8) Δ на d_x (приемлемую максимальную ошибку средней). Стандартная ошибка средней без учета пкс определяется по формуле (5.15). Поскольку $n_k=f_k n$, то справедливо следующее равенство:

$$d_x=z\mu=\sqrt{\sum\frac{s_k^2}{f_k n}F_k^2}.$$

Отсюда получим

$$n=\frac{z^2}{d_x^2}\sum\frac{s_k^2}{f_k}F_k^2.$$

Глава 6. ОЦЕНИВАНИЕ СУММ ОШИБОК ПО ВЫБОРКЕ

Если в одном месте что-то убудет, то в другом ровно столько же присовокупится.

М. Ломоносов

§ 6.1. Абсолютные ошибки

Атрибутивные выборки, которые обсуждались в гл. 3, позволяют оценивать *количество* документов в совокупности, содержащих те или иные отклонения, в том числе ошибки оформления. Напомним, что к отклонениям могут быть отнесены и различного рода погрешности в размерах стоимостных показателей (тест на соответствие). Причем размер самих ошибок во внимание не принимался. Во многих случаях подобного рода контроль является недостаточным для аудитора, поскольку для анализа ему необходимы оценки *обобщенного размера стоимостных ошибок* для совокупности (тест на содержание). Для решения такой задачи воспользуемся методикой точечного и интервального оценивания средних и сумм стоимостных показателей для совокупности, которая рассмотрена в гл. 5. В связи с этим несколько изменим постановку задачи и распространим методику оценивания не на сами показатели, а на обнаруженные в ходе тестирования отклонения от реальных или действительных данных. Такие данные получает аудитор при проверке и корректировке. Далее отклонения от реальных данных назовем *ошибками*.

Для аудитора представляет безусловный интерес оценка суммы ошибок для совокупности, а иногда и оценка среднего их размера. Эти параметры дают возможность характеризовать качество представленной для тестирования документации и определить действительные размеры совокупности по стоимости с учетом ошибок, обнаруженных аудитором. Для такого тестирования применяют два метода, которые условно назовем *оцениванием разности* (difference estimation) и *оцениванием отношения* (ratio estimation). В первом случае в качестве исходных данных рассмат-

риваются *абсолютные ошибки*, обнаруженные в представленных документах, во втором – *относительные ошибки*. В данном параграфе главы обсуждается первый метод, в следующем – второй.

К оцениванию размера ошибок на основе выборок аудитор прибегает при необходимости проверить большой массив первичных документов, например счетов-фактур, накладных, различных документов, подтверждающих оплату товаров с указанием цены и стоимости и т.п. Ожидается, что эти документы содержат незначительные ошибки.

Как и в гл. 5, обсудим методики получения точечных и интервальных оценок. Точечные оценки служат базой при определении интервальных оценок средней абсолютной ошибки и суммарной ошибки для совокупности. По-видимому, для аудитора наибольший интерес представляет последняя величина. Располагая такой оценкой, аудитор может с определенной степенью надежности скорректировать представленные ему данные, сформулировать определенные выводы для своего заключения.

Для записи формул, применяемых при точечном и интервальном оценивании, примем следующие обозначения:

x_i – величина стоимостной характеристики в i-й выборочной единице наблюдения (документе);

y_i – скорректированный после тестирования (реальный) стоимостной показатель i-й единицы наблюдения;

e_i – обнаруженная аудитором величина абсолютной ошибки в i-м документе;

$\tilde{e}$ – среднее значение ошибки в выборке;

n – объем выборки;

N – размер совокупности, из которой произведена выборка;

E – точечная оценка суммы ошибок для совокупности;

$\overline{e}$ – интервальная оценка средней ошибки для совокупности;

A – интервальная оценка сумм ошибок для совокупности.

Величину абсолютной ошибки стоимостной характеристики i-й единицы наблюдения в выборке получим как разность:

$$e_i = x_i - y_i. \tag{6.1}$$

Ошибка может быть положительной, отрицательной или отсутствовать (равна нулю), так как x_i и y_i только положительные величины.

Среднее значение абсолютной ошибки в расчете на единицу наблюдения в выборке:

$$\tilde{e} = \frac{\sum_{i=1}^{n} e_i}{n}. \tag{6.2}$$

Среднее значение абсолютной ошибки рассматривается как *точечная оценка средней ошибки для совокупности*. Экстраполируем полученное значение средней на весь объем совокупности. Получим *точечную оценку суммы ошибок для совокупности*:

$$E = \tilde{e}N. \tag{6.3}$$

Если ошибки случайны, то следует ожидать, что их сумма и среднее значение близки к нулю, поскольку положительные ошибки в какой-то мере компенсируются отрицательными. Заметное преобладание положительных (или, наоборот, отрицательных) ошибок может свидетельствовать о приписках и других преднамеренных искажениях в представленных документах.

При полном совпадении по суммам ошибок с противоположными знаками (разумеется, это крайний, но теоретически возможный случай) имеем: $\sum e_i = 0$ и $\tilde{e} = 0$. В этой или близкой ситуации точечная оценка средней абсолютной ошибки теряет свой смысл, но только в качестве индикатора наличия и размера ошибок. В самом деле, нулевая или близкая к нулевой точечная оценка вовсе не означает, что тестируемые документы не содержат ошибок. Однако сказанное не относится к интервальным оценкам.

Перейдем к методике интервального оценивания. Средняя ошибка для совокупности лежит в интервале

$$\overline{e} = \tilde{e} \pm \Delta, \tag{6.4}$$

где Δ – предельная ошибка выборочной средней для заданного уровня надежности, или *интервал точности* (precision interval). Если $\tilde{e} = 0$, то $\overline{e} = \pm\Delta$.

Как было показано в § 5.1, выборочные средние распределяются по нормальному закону или близко к нему. Соответственно для определения доверительных границ следует воспользоваться кван-

тилями нормального распределения z, о которых говорилось в гл. 5 (см. табл. 5.1). Предельную ошибку находим стандартным путем:

$$\Delta = z\mu = z\frac{s}{\sqrt{n}}, \qquad (6.5)$$

где s – стандартная ошибка в выборке;

μ – стандартная ошибка выборочной средней;

n – объем выборки.

С увеличением объема выборки размер предельной ошибки уменьшается.

Для определения s обратимся к одной из рабочих формул, приведенных в § 5.1. Перепишем ее в символах, принятых в данном параграфе. Без поправки на конечность совокупности формула (5.7) записывается следующим образом:

$$s = \sqrt{\frac{\sum_i e_i^2 - n\tilde{e}^2}{n-1}}, \qquad (6.6)$$

а с учетом пкс:

$$s = \sqrt{\frac{\sum_i e_i^2 - n\tilde{e}^2}{n-1}} \cdot \sqrt{1-\frac{n}{N}}. \qquad (6.7)$$

Границы интервальной оценки суммы ошибок для совокупности находятся элементарно:

$$A = (\tilde{e} \pm \Delta)N = E \pm \Delta N. \qquad (6.8)$$

Как видим, с увеличением размера совокупности растет интервал для суммы ошибок. Если стандартная ошибка средней определяется без учета пкс, то размер интервала строго пропорционален объему совокупности.

Остановимся на одном частном случае, который, впрочем, нельзя исключить. Пусть все ошибки одинаковы по размеру и имеют один знак (например, одинаковые приписки к цене). В этом случае стандартная ошибка равна нулю. Соответственно: $\Delta = s = \mu = 0$ и, следовательно, интервала для оценки средней нет: $\bar{e} = \tilde{e}$.

ПРИМЕР 6.1. Общий объем документов в совокупности равен 2000. Для проверки случайным образом отобрано 200 документов. Тестирование показало, что в 15 документах выборки обнаружены ошибки. Причем имели место как завышения, так и приуменьшения. Данные выборки и результаты корректировки в тыс. руб. показаны в табл. 6.1 (столбцы x_i, y_i и e_i).

Таблица 6.1

Абсолютные ошибки в стоимостных показателях выборки

i	x_i	y_i	e_i	e_i^2
1	3,0	2,8	0,2	0,04
2	5,4	5,6	–0,2	0,04
3	1,7	1,8	–0,1	0,01
4	10,9	10,8	0,1	0,01
5	9,9	9,5	0,4	0,16
6	3,7	3,4	0,3	0,09
7	16,0	16,1	–0,1	0,01
8	8,2	8	0,2	0,04
9	7,5	7,6	–0,1	0,01
10	12,3	12,6	–0,3	0,09
11	6,4	6	0,4	0,16
12	7,8	8	–0,2	0,04
13	2,1	2,3	–0,2	0,04
14	3,3	3,2	0,1	0,01
15	9,0	9,3	–0,3	0,09
Итого	107,2	107,0	0,2	0,84

Ошибки имеют разные знаки. Как показано в таблице, наибольшая по абсолютной величине отрицательная ошибка равна –0,3, а наибольшая положительная 0,4. Сумма положительных ошибок немного больше, чем отрицательных (соответственно 1,7 и –1,5), что может служить свидетельством об отсутствии преднамеренных ошибок. Имеет место частичное взаимное их погашение, что уменьшает их влияние на размер суммарной ошибки для совокупности.

По данным таблицы находим среднюю ошибку (точечную оценку абсолютной ошибки для совокупности):

$$\tilde{e} = \frac{0,2}{200} = 0,001 \text{ тыс. руб.}$$

Иначе говоря, на одну единицу наблюдения в выборке приходится ошибка, равная 1 руб. Экстраполируя полученную среднюю на все 2000 документов, получим точечную оценку суммы ошибок для совокупности:

$$E = 0{,}001 \cdot 2000 = 2 \text{ тыс. руб.}$$

Перейдем к определению более важных для аудитора интервальных оценок. Начнем с расчета стандартной и предельных ошибок. Для этого найдем сумму квадратов отклонений: $\sum e_i^2 = 0{,}84$. По формуле (6.6) определим стандартное отклонение

$$s = \sqrt{\frac{0{,}84 - 200 \cdot 0{,}001^2}{200 - 1}} = 0{,}065 \text{ тыс. руб.}$$

Примем доверительную вероятность на уровне 90%, тогда $z = 1{,}64$. В этом случае предельная ошибка средней составит

$$\Delta = 1{,}64 \frac{0{,}065}{\sqrt{200}} = 0{,}0075 \text{ тыс. руб.}$$

Интервал для средней ошибки в рублях в расчете на один документ составляет $\bar{e} = 1 \pm 7{,}5$ руб., иначе говоря, ошибка средней для совокупности при принятом уровне доверительной вероятности лежит в границах от –6,5 до +8,5 руб.

В свою очередь, интервал для суммы ошибок во всех документах тестируемой совокупности определяется как

$$A = 2 \pm 0{,}0075 \cdot 2000 \text{ тыс. руб.}$$

Аудитор с вероятностью 90% может утверждать, что суммарная ошибка находится в пределах от –13 до +17 тыс. руб. У аудитора теперь есть объективное основание для суждения о качестве представленных ему документов.

Если приведенные выше параметры определить с учетом пкс, то получим:

$$s = \sqrt{\frac{0{,}84 - 200 \cdot 0{,}001^2}{200 - 1}} \cdot \sqrt{1 - \frac{200}{2000}} = 0{,}065 \cdot 0{,}95 = 0{,}061 \text{ тыс. руб.},$$

$$\Delta = 1{,}64 \frac{0{,}061}{\sqrt{200}} = 0{,}007, \quad \bar{e} = 0{,}01 \pm 0{,}07.$$

Окончательно имеем

$$A = 2 \pm 0{,}007 \cdot 2000 \text{ тыс. руб.}$$

Эти интервалы немного уже, чем полученные ранее. Так, интервал для суммы ошибок с учетом пкс теперь составляет от –12 до +16 тыс. руб.

ПРИМЕР 6.2. В приведенном выше примере ошибки, обнаруженные аудитором в документах, попавших в выборку, имели разные знаки. Изменим одно условие этого примера. Пусть теперь имеют место только ошибки с положительным знаком (явное свидетельство приписок). Данные (в тыс. руб.) приведены в табл. 6.2.

Таблица 6.2

Абсолютные положительные ошибки в стоимостных показателях выборки

i	x_i	y_i	e_i	e_i^2
1	3	2,9	0,1	0,01
2	5,4	5,1	0,3	0,09
3	1,7	1,6	0,1	0,01
4	10,9	10,5	0,4	0,16
5	9,9	9,7	0,2	0,04
6	3,7	3,6	0,1	0,01
7	16	14,9	1,1	1,21
8	8,2	8	0,2	0,04
9	7,5	7,4	0,1	0,01
10	12,3	11,5	0,5	0,25
11	6,4	6,3	0,1	0,01
12	7,8	7,6	0,2	0,04
13	2,1	2	0,1	0,01
14	3,3	3,2	0,1	0,01
15	9	8,7	0,3	0,09
Итого	107,2	103	3,9	1,99

На основе этих данных последовательно получим (все результаты в тыс. руб.):

$$\tilde{e} = \frac{3,9}{200} = 0,0195; \quad E = 0,0195 \cdot 2000 = 39;$$

$$s = \sqrt{\frac{1,99 - 200 \cdot 0,0195^2}{200 - 1}} = 0,098;$$

$$\Delta = 1,64 \frac{0,098}{\sqrt{200}} = 0,0136;$$

$$A = 39 \pm 0,0136 \cdot 2000 = 39 \pm 27,2.$$

Как видим, интервал для суммы ошибок в совокупности тестируемых документов существенно расширился и составляет теперь от 11,8 до 66,2 тыс. руб.

§ 6.2. Относительные ошибки

В предыдущем параграфе обсуждались методики оценивания ошибок и различного рода искажений, установленных аудитором (разности показанных в документах данных и этих же показателей после их корректировки). Возможен и другой вариант измерения ошибок – в виде относительных величин. Обобщающей характеристикой таких ошибок является оценка доли суммарной ошибки в общей стоимости. Оценка доли, или оценка отношения, позволяет найти и сумму абсолютных ошибок для совокупности. В статистической литературе такого рода выборочные оценки относительных величин называют *оценками по отношению*.

Итак, вместо абсолютных разностей e_i определим их отношения для каждой единицы наблюдения в выборке:

$$a_i = \frac{e_i}{x_i}. \qquad (6.9)$$

Введем несколько новых обозначений:

$\tilde{a}$ – выборочная оценка относительной ошибки в величине стоимостного показателя;

$\tilde{x}$ – среднее значение стоимостной характеристики в выборке;

$\bar{X}$ – среднее значение стоимостной характеристики в совокупности;

Q – сумма стоимостных показателей совокупности;

E – точечная оценка суммы ошибок для совокупности;

$\bar{a}$ – интервальная оценка относительной ошибки для совокупности;

A – интервальная оценка суммы ошибок для совокупности.

Обобщенной выборочной характеристикой размера относительных ошибок является величина $\tilde{a}$, которую примем в качестве точечной оценки относительной ошибки для совокупности. Определим ее следующим образом:

$$\tilde{a} = \frac{\sum_1^n e_i}{\sum_1^n x_i} = \frac{(\sum e_i)/n}{(\sum x_i)/n} = \frac{\tilde{e}}{\tilde{x}}. \qquad (6.10)$$

Оценку по отношению $\tilde{a}$ можно применить для определения точечной оценки абсолютной суммы ошибок для совокупности. Для этого необходима общая сумма стоимостных показателей. Искомая величина составит:

$$E = \tilde{a}Q = \tilde{a}\bar{X}N. \qquad (6.11)$$

Перейдем к интервальным оценкам относительной ошибки и суммы ошибок для совокупности. Прежде всего необходимо определить стандартную ошибку оценки по отношению, а затем предельную ошибку выборки. Однако величина $\tilde{a}$ определяется как отношение двух выборочных средних $\tilde{e}$ и $\tilde{x}$, каждая из которых имеет свое распределение. В силу сказанного для расчета s нельзя воспользоваться рабочими формулами (5.7), которые получены для средней, а не отношения двух средних. Оказывается, что точной формулы для расчета стандартного отклонения по отношению нет. Приходится пользоваться приближенной формулой[1]:

$$s = \frac{1}{\bar{X}} \sqrt{\frac{\sum e_i^2 - 2\tilde{a}\sum e_i x_i + \tilde{a}^2 \sum x_i^2}{n-1}}, \qquad (6.12)$$

а с учетом пкс имеем

$$s = \frac{1}{\bar{X}} \sqrt{\frac{\sum e_i^2 - 2\tilde{a}\sum e_i x_i + \tilde{a}^2 \sum x_i^2}{n-1}} \cdot \sqrt{1 - \frac{n}{N}}. \qquad (6.13)$$

[1] Вывод формулы см. в Дополнении к главе.

Предельную ошибку выборки находим по стандартной формуле (6.5). В итоге получим границы интервальной оценки по отношению:

$$\bar{a} = \tilde{a} \pm \Delta. \qquad (6.14)$$

Экстраполируем эти границы на всю сумму стоимостных характеристик совокупности. Получим интервал для оценки абсолютной суммы ошибок в совокупности:

$$A = (\tilde{a} \pm \Delta)Q = E \pm \Delta \bar{X} N. \qquad (6.15)$$

Как уже отмечалось в комментариях к формуле (6.2), в определенных условиях $\sum e_i = 0$, соответственно $\tilde{a} = 0$ и $E = 0$. Нулевая точечная оценка вовсе не указывает на отсутствие ошибок в совокупности. Поскольку $\Delta \neq 0$, то интервальные оценки в этой ситуации отличаются от нуля: $\bar{a} = \pm\Delta$, $A = \pm\Delta Q$.

На первый взгляд представляется, что оценивание суммарной ошибки на основе абсолютных разностей дает близкие результаты, как и оценивание по отношению. Однако это не так. Дело в том, что методика, базирующаяся на абсолютных разностях, не принимает во внимание такой важный фактор, как стоимостные характеристики документов. Вместе с тем, как правило, абсолютные размеры ошибок находятся в корреляционной зависимости от этих характеристик.

ПРИМЕР 6.3. Воспользуемся условиями примера 6.1 и определим точечные и интервальные оценки по отношению и оценки абсолютной суммы ошибок в совокупности. Исходные и необходимые для расчетов данные представлены в табл. 6.3.

Пусть общая сумма стоимостных характеристик тестируемых документов совокупности равна 13 млн руб. Средняя для совокупности составит

$$\bar{X} = \frac{Q}{N} = \frac{13000}{2000} = 6{,}5 \text{ тыс. руб.}$$

В 15 из 200 документов выборки обнаружены различного рода искажения. Как показано в столбце a_i таблицы, относительные ошибки находятся в интервале от минус 0,095 до плюс 0,081. Относитель-

Таблица 6.3

Относительные ошибки в стоимостных показателях выборки

i	x_i	y_i	e_i	e_i^2	a_i	$e_i x_i$	x_i^2
1	3	2,8	0,2	0,04	0,067	0,60	9,0
2	5,4	5,6	-0,2	0,04	-0,037	-1,08	29,2
3	1,7	1,8	-0,1	0,01	-0,059	-0,17	2,9
4	10,9	10,8	0,1	0,01	0,009	1,09	118,8
5	9,9	9,5	0,4	0,16	0,040	3,96	98,0
6	3,7	3,4	0,3	0,09	0,081	1,11	13,7
7	16	16,1	-0,1	0,01	-0,006	-1,60	256,0
8	8,2	8	0,2	0,04	0,024	1,64	67,2
9	7,5	7,6	-0,1	0,01	-0,013	-0,75	56,3
10	12,3	12,6	-0,3	0,09	-0,024	-3,69	151,3
11	6,4	6	0,4	0,16	0,063	2,56	41,0
12	7,8	8	-0,2	0,04	-0,026	-1,56	60,8
13	2,1	2,3	-0,2	0,04	-0,095	-0,42	4,4
14	3,3	3,2	0,1	0,01	0,030	0,33	10,9
15	9	9,3	-0,3	0,09	-0,033	-2,70	81,0
Итого	107,2	107,0	0,2	0,84	–	-0,68	1000,4

ная ошибка в выборке, или точечная оценка по отношению для совокупности, составит:

$$\tilde{a} = \frac{0{,}2}{107{,}2} = 0{,}001866 \approx 0{,}0019.$$

Обобщенная относительная ошибка в величине стоимостного показателя равна 0,19%. Иначе говоря, на каждую тысячу рублей стоимостной характеристики приходится в среднем ошибка, которая чуть меньше двух рублей. Распространим эту оценку на всю суммарную стоимость документов в совокупности:

$$E = 0{,}0019 \cdot 13000 = 24{,}7 \text{ тыс. руб.}$$

Перейдем к определению интервальных оценок. По формуле (6.12) рассчитаем стандартное отклонение выборочной доли. По табл. 6.3 находим необходимые для расчета суммы:

$$\sum e_i^2 = 0{,}84; \quad \sum e_i x_i = -0{,}68; \quad \sum x_i^2 = 1000{,}4.$$

Стандартное отклонение без учета пкс:

$$s = \frac{1}{6{,}5}\sqrt{\frac{0{,}84 - 2 \cdot 0{,}0019 \cdot (-0{,}68) + 0{,}0019^2 \cdot 1000{,}4}{199}} = 0{,}01.$$

Предельная ошибка выборки при доверительной вероятности 90% равна

$$\Delta = z\frac{s}{\sqrt{200}} = 1{,}64\frac{0{,}01}{\sqrt{200}} = 0{,}0012.$$

Интервал оценки по отношению для совокупности составит

$$\bar{a} = \tilde{a} \pm \Delta = 0{,}0019 \pm 0{,}0012, \text{ т.е. от 0,07 до 0,31\%.}$$

По формуле (6.15) находим интервал для суммы ошибок в совокупности

$$A = 24{,}7 \pm 0{,}0012 \cdot 6{,}5 \cdot 2000, \text{ т.е. от 9,1 до 40,3 тыс. руб.}$$

С учетом пкс последовательно получим: $s = 0{,}0095$; $\Delta = 0{,}0011$ и

$$\bar{a} = 0{,}0019 \pm 0{,}0011; \; A = 24{,}7 \pm 0{,}0011 \cdot 6{,}5 \cdot 2000.$$

Сравним результаты точечного и интервального оценивания сумм ошибок для совокупности, полученные двумя методами (на основе абсолютных разностей и на основе отношений).

Для точечной оценки по первому методу (см пример 6.1) была получена оценка, равная 2 тыс. руб., по второму – 24,7 тыс. руб. Разница, как видим, громадная. Размеры интервалов сумм абсолютных ошибок для совокупности примерно одинаковы (30 и 31 тыс. руб.). Интервалы смещены относительно друг друга: от –13 до +17 тыс. руб. по первому методу и от 9,1 до 40,3 тыс. руб. по второму.

Рассмотренный выше метод оценивания ошибок по отношению не является единственно возможным. В практике зарубежного аудита[1] иногда вместо соотношения $a_i = e_i/x_i$ объектом анализа является выборочное соотношение $r_i = y_i/x_i$. Обобщающая

[1] *Arens A.A., Loebbecke J.K.* Applications of Statistical Sampling to Auditing. N.Y.: Prentice-Hall, 1981.

характеристика для выборки или точечная оценка для совокупности находится как отношение:

$$\tilde{r} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\tilde{y}}{\tilde{x}}. \tag{6.16}$$

Между двумя выборочными характеристиками относительных ошибок существует функциональная зависимость. Нетрудно доказать зависимость между $\tilde{r}$ и $\tilde{a}$:

$$\tilde{a} = \frac{\sum (x_i - y_i)}{\sum x_i} = 1 - \tilde{r}. \tag{6.17}$$

Стандартное отклонение отношения $\tilde{a}$ находим по приближенной формуле

$$s = \frac{1}{\bar{X}}\sqrt{\frac{\sum_i y_i^2 - 2\tilde{r}\sum_i x_i y_i + \tilde{r}^2 \sum_i x_i^2}{n-1}}. \tag{6.18}$$

Предельная ошибка выборочного отношения определяется стандартным путем, см. (6.5). Точечная и интегральные оценки суммы ошибок составляют:

$$E = (1 - \tilde{r})Q, \tag{6.19}$$

$$A = \left[1 - (\tilde{r} \pm \Delta)\right]Q. \tag{6.20}$$

Таким образом, принципиально нового здесь ничего нет. Немного меняется лишь порядок расчетов. Результат должен быть таким же, что и при применении $\tilde{a}$. Разница в результатах может быть только следствием округлений. В силу сказанного не будем больше останавливаться на рассмотренной выше методике.

Дальнейшее развитие методики оценивания по отношению заключается в применении расслоенного отбора. Однако для реализации такой методики необходима дополнительная информация, что в рамках аудиторской проверки, вероятно, создаст определенные трудности, а часто и просто представляется невозможным.

§ 6.3. Монетарная выборка. Оценивание суммарной ошибки для совокупности

Цель монетарной выборки остается такой же, что и при собственно-случайном и расслоенном отборах – расчет суммарной ошибки для совокупности (ошибка в сумме остатков на счетах, в суммарной стоимости товарных документов и т.д.) в виде точечной и интервальной оценок. Методика получения необходимых статистических параметров в принципе не изменяется.

Напомним, что при монетарной выборке документы отбирают случайным образом. Однако принцип равной вероятности здесь применяется не к документам (единицам наблюдения), а к отдельным денежным единицам (см. § 2.1). В результате: чем больше стоимостная характеристика документа, тем больше вероятность для него попасть в выборку. Например, если стоимость, указанная на одном документе совокупности, равна 200 тыс. руб., а на другом 20 тыс. руб., то вероятность оказаться в выборке у первого документа при монетарной выборке в 10 раз выше, чем у второго. При собственно-случайном отборе эти вероятности должны быть одинаковыми. Очевидно, что при монетарном отборе внимание аудитора в большей мере сосредоточивается на документах с более высокими стоимостными характеристиками. Соответственно уменьшается риск пропустить значительную погрешность. Более того, такой метод отбора увеличивает охват по стоимости тестируемой совокупности без увеличения размера выборки.

Перейдем к оцениванию суммарной ошибки для совокупности на основе данных монетарной выборки в виде относительных ошибок, обнаруженных аудитором. Что касается использования абсолютных ошибок, то в литературе в этих случаях иногда предлагается применять методику определения точечных и интервальных оценок для совокупности, которая была рассмотрена в § 6.1. С этим трудно согласиться. Дело в том, что выборочная средняя абсолютная ошибка при монетарном отборе оказывается смещенной (завышенной), так как она определяется в расчете на единицу наблюдения, а в выборке преобладают документы с высокими стоимостями. Причем следует ожидать, что имеет место корреляция размера абсолютной ошибки и стоимостной характеристики документа – чем выше стоимостная характеристика, тем больше абсолютная ошибка. В итоге получаемые оценки сумм ошибок для совокупности будут завышенными. Попутно заметим, что монетарный отбор приводит к смещению и при оценивании суммарной стоимости для совокупности. В связи со сказанным остановимся только на методике оценивания на основе относительных ошибок.

Пусть, как и выше, x_i и y_i означают стоимостные характеристики документа в выборке – стоимости, показанные в документе, и стоимости после корректировки аудитором. Обнаруженная абсолютная ошибка в документе равна $e_i = x_i - y_i$, обе переменные больше нуля. Ошибка может быть равна нулю, больше или меньше нуля. Ошибка в расчете на 1 руб. стоимостной характеристики документа составляет

$$h_i = \frac{e_i}{x_i}.$$

По выборочным данным находим значения h_i и затем их обобщающую оценку

$$\tilde{h} = \frac{\sum_{i=1}^{n} \frac{e_i}{x_i}}{n}. \qquad (6.21)$$

Эту величину можно трактовать как среднюю ошибку в расчете на один рубль стоимостной характеристики единицы наблюдения. В силу этого оценкой $\tilde{h}$ можно воспользоваться для получения точечной оценки суммы ошибок для совокупности:

$$E = \tilde{h}Q. \qquad (6.22)$$

Определим теперь стандартную ошибку выборочной средней без пкс. Для этого заменим в (6.6) e_i на h_i. Получим

$$s = \sqrt{\frac{\sum h_i^2 - n\tilde{h}^2}{n-1}}. \qquad (6.23)$$

Предполагается, что выборочная средняя имеет распределение, близкое к нормальному. С учетом этого условия предельную ошибку выборки для средней находим стандартным путем:

$$\Delta = z\frac{s}{\sqrt{n}},$$

Предельная ошибка для суммы составит

$$\Delta_s = z\frac{s}{\sqrt{n}}Q. \qquad (6.24)$$

Интервальная оценка суммарной ошибки для совокупности

$$A = E \pm \Delta Q = E \pm \Delta_s. \qquad (6.25)$$

ПРИМЕР 6.4. На основе монетарной выборки из совокупности отобрано 100 документов. В 15 из них обнаружены ошибки в стоимостных показателях. Суммарная стоимость всех документов совокупности равна 2 млн руб. Выборочные данные и показатели, необходимые для расчета предельной ошибки, приведены в табл. 6.4.

Таблица 6.4

Оценивание суммарной ошибки для совокупности. Монетарная выборка

i	x_i	y_i	e_i	h_i	h_i^2
1	13,2	12,4	0,8	0,0606	0,0037
2	25,4	23,6	1,8	0,0709	0,0050
3	11,7	12	–0,3	–0,0256	0,0007
4	10,9	10,3	0,6	0,0550	0,0030
5	2,9	2,8	0,1	0,0345	0,0012
6	3,7	3,4	0,3	0,0811	0,0066
7	17,9	16,1	1,8	0,1006	0,0101
8	18,2	17,4	0,8	0,0440	0,0019
9	7,5	7,6	–0,1	–0,0133	0,0002
10	12,3	12,6	–0,3	–0,0244	0,0006
11	6,4	6,2	0,2	0,0313	0,0010
12	17,8	17,1	0,7	0,0393	0,0015
13	2,1	2,3	–0,2	–0,0952	0,0091
14	3,3	3,2	0,1	0,0303	0,0009
15	9,5	9,3	0,2	0,0211	0,0004
Итого	162,8	156,3	6,5	0,4099	0,0459

По данным таблицы находим средний размер обнаруженных в выборке ошибок:

$$\bar{h} = \frac{0,4099}{100} = 0,0041.$$

Точечная оценка суммы ошибок для совокупности

$$E = 0,0041 \cdot 2000 = 8,2 \text{ тыс. руб.}$$

Стандартное отклонение в выборке

$$s = \sqrt{\frac{0,0459 - 100 \cdot 0,0041^2}{100-1}} = 0,0211.$$

Предельная ошибка средней и суммы ($z = 1,64$):

$$\Delta = 1,64\frac{0,0211}{\sqrt{100}} = 0,00346, \quad \Delta_s = 0,00346 \cdot 2000 = 6,92 \text{ тыс. руб.}$$

Интервальная оценка суммы ошибок для совокупности:

$$A = 8,2 \pm 6,92 \text{ тыс. руб.}$$

Таким образом, аудитор, основываясь на выборке, с доверительной вероятностью 90% может утверждать, что в проверенной им совокупности суммарная ошибка лежит в широких пределах от 1,28 до 15,12 тыс. руб. Сузить интервал можно за счет снижения уровня доверительной вероятности. Если последнюю принять на уровне 85%, то интервал будет находиться в пределах 2,12–14,28 тыс. руб.

§ 6.4. Необходимые объемы выборки

В предыдущих параграфах главы обсуждались методики оценивания размеров ошибок стоимостных характеристик для совокупности. Во всех рассмотренных случаях объемы выборок считались известными параметрами. С этой проблемой мы уже сталкивались в гл. 5 при оценивании средней и общей суммы стоимостных характеристик. Кратко напомним суть вопроса. При оценивании интервала для средней и суммарной ошибки необходимо определить предельную ошибку выборки. В свою очередь, для этого надо иметь значения n и s, т.е. знать объем выборки и выборочное стандартное отклонение до извлечения выборки. Создается своего рода порочный круг.

Для того чтобы выйти из этого круга, процедуру определения n, как и в гл. 5, разобьем на два этапа – предварительный расчет и, если в этом возникнет необходимость (обнаружится недостаточность объема выборки), расчет окончательного объема. Методика определения n при оценивании сумм и средних была рассмотрена в гл. 5. В принципе она остается такой же и в выборочном анализе абсолютных и относительных ошибок. Остановимся на определении объема выборки при оценивании абсолютных ошибок. Для этого обратимся к (6.5) и найдем n:

$$n = \left(z \frac{s}{\Delta} \right)^2. \qquad (6.26)$$

Поскольку значение s до осуществления выборки неизвестно, на предварительном этапе придется воспользоваться соответствующей информацией из прошлого опыта аудитора или других источников. В крайнем случае можно просто задаться некоторой правдоподобной величиной стандартного отклонения, применить метод "проб и ошибок".

Что касается фигурирующей в формуле (6.26) предельной ошибки, то вместо нее следует взять величину, которую назовем *приемлемой ошибкой* (обозначим этот параметр как d_x). Итак, расчет предварительного объема выборки осуществим по формуле

$$n_0 = \left(z \frac{s}{d_x} \right)^2. \qquad (6.27)$$

где n_0 – предварительный объем выборки;

s – предварительная оценка стандартной ошибки выборочной средней;

d_x – величина приемлемой погрешности для одного документа.

Формула (6.27) не учитывает пкс. Если отношение n_0/N пренебрежимо мало или нет сведений о величине N, то полученное значение объема не корректируется. В противном случае применим знакомую формулу

$$n = \frac{n_0}{1 + \frac{n_0}{N}}.$$

Следующим шагом после определения n_0 или n является отбор единиц наблюдения в выборку и оценка выборочных параметров $\tilde{e}_0$, s_0, Δ_0.

Если найденное по выборке значение предельной ошибки не удовлетворяет аудитора (интервал для суммы ошибок оказывается слишком широким), то можно сократить предельную ошибку путем уменьшения принятого уровня доверительной вероятности или заново определить, скорректировать объем выборки на основе выборочного значения s_0, доведя объем до n_1.

Очевидно, что при увеличении объема выборки нет необходимости отбирать все n_1 единиц наблюдения заново. Достаточно добавить к ранее отобранным документам некоторое дополнительное их количество $n_{доп} = n_1 - n_0$. После окончательного определения объема выборки можно приступить к определению выборочных параметров.

Средний размер абсолютной ошибки с учетом данных дополнительного отбора:

$$\tilde{e}_1 = \frac{\sum_1^{n_0} e_i + \sum_1^{n_{доп}} e_i}{n_1} = \frac{\sum_1^{n_1} x_i - \sum_1^{n_1} y_i}{n_1}. \qquad (6.28)$$

Средний размер стоимостной характеристики

$$\tilde{x} = \frac{\sum_1^{n_0} x_i + \sum_1^{n_{доп}} x_i}{n_1}. \qquad (6.29)$$

Стандартное отклонение

$$s_1 = \sqrt{\frac{\sum_1^{n_1} e_i^2 - n_1 \tilde{e}_1^2}{n_1 - 1}}. \qquad (6.30)$$

Предельная **ошибка**

$$\Delta_1 = z \frac{s_1}{\sqrt{n_1}}. \qquad (6.31)$$

Интервал суммарной ошибки для совокупности теперь равен

$$A = E \pm \Delta_1 N.$$

ПРИМЕР 6.5. Необходимо определить объем выборки для тестирования стоимостных показателей совокупности документов – оценивания суммарной ошибки. Размер совокупности $N = 3000$. Примерная средняя стоимостная характеристика одного документа порядка 2 тыс. руб. Приемлемая суммарная ошибка для совокупности равна 6 тыс. руб.

Для предварительного определения объема выборки воспользуемся формулой (6.27). Предположим, что тестирование данного вида документов производится впервые и нет никаких сведений о возможном или ожидаемом значении стандартного отклонения для абсолютных ошибок. Поэтому для начала примем, что оно заметно больше предполагаемой средней ошибки. Пусть допустимая ошибка в расчете на один документ равна 2 руб., $s_0 = 20$ руб. Доверительная вероятность 90%.

Предварительная оценка объема выборки

$$n_0 = \left(1{,}64\frac{20}{2}\right)^2 = 269.$$

В связи с неопределенностью исходных данных результат можно округлить до 250 единиц наблюдения.

После случайного отбора 250 единиц наблюдения обнаружено, что 15 документов содержат ошибки. По выборке получены следующие данные: $\tilde{e} = 1{,}9$ руб., $\sum_1^{n_0} e_i^2 = 908$. На основе этих данных получим: $s_0 = 0.15$; $\Delta_0 = 0{,}016$ тыс. руб.; $E = 0{,}0019 \cdot 3000 = 5{,}7$ тыс. руб.

Интервальная оценка суммы ошибок для совокупности

$$A = 5{,}7 \pm 0{,}015 \cdot 3000 \text{ тыс. руб.}$$

Если такая интервальная оценка удовлетворяет аудитора, то выборочное исследование на этом заканчивается, если же аудитор считает, что необходимо повысить точность оценки, то он может определить n_1 с учетом уже имеющейся оценки стандартного отклонения.

Аналогичный подход применим и для определения объема выборки при оценивании суммарной ошибки на основе относительных выборочных показателей ошибки.

Дополнение к главе

Вывод формулы (6.12)

В каждой единице наблюдения простой случайной выборки, в которой найдена ошибка, фигурируют переменные e_i (ошибка) и x_i (стоимостной показатель). Ошибка определяется как разность $e_i = x_i - y_i$, $x_i > 0$, $y_i > 0$. Значения переменной y_i задаются аудитором.

Относительная ошибка в выборке определяется как отношение двух выборочных средних:

$$\tilde{a} = \frac{\sum_1^n e_i / n}{\sum_1^n x_i / n} = \frac{\tilde{e}}{\tilde{x}}.$$

Перейдем к определению дисперсии и стандартной ошибки выборочной величины $\tilde{a}$. Для этого найдем разность

$$\tilde{a} - a = \frac{\tilde{e}}{\tilde{x}} - a,$$

где a – неизвестная величина соотношения для совокупности.

Выборочное среднее стоимостной характеристики является оценкой этой средней для совокупности $\tilde{x} \approx \bar{X}$. Таким образом,

$$\tilde{a} - a = \frac{\tilde{e} - a\tilde{x}}{\bar{X}}.$$

Дисперсия соотношения для совокупности составит:

$$S^2 = \frac{1}{\bar{X}^2} \cdot \frac{\sum_1^N (e_i - a x_i)^2}{n-1}.$$

Выборочная оценка этой величины:

$$s^2 = \frac{1}{\bar{X}^2} \cdot \frac{\sum_1^n (e_i - \tilde{a} x_i)^2}{n-1}.$$

После несложных преобразований окончательно получим искомую формулу:

$$s = \frac{1}{\bar{X}}\sqrt{\frac{\sum_i e_i^2 - 2\tilde{a}\sum e_i x_i + \tilde{a}^2\sum_i x_i^2}{n-1}}.$$

ПРИЛОЖЕНИЯ

Таблица 1

Случайные числа (продолжение[1])

Строки	Столбцы					
	7	8	9	10	11	12
1	07030	29945	74800	16928	21810	63181
2	93192	28233	13380	62885	06450	89714
3	59135	07396	25770	64179	86717	00638
4	96129	75428	52790	16417	77466	79030
5	52333	37301	96479	40370	29856	20070
6	97454	69400	06511	59504	88080	60296
7	12905	48721	18962	18236	75012	57953
8	02926	15760	56725	76075	63767	83000
9	47568	63143	95015	94967	57344	87296
10	77779	06950	51755	96185	98409	74471
11	89172	19862	67498	18017	13837	63343
12	21805	30292	18573	16526	04419	02585
13	68891	70497	63003	79538	96279	52365
14	86685	03886	89236	51865	47239	69018
15	91258	42253	48032	11835	07748	34384
16	57287	29593	34488	62989	54031	76951
17	38077	11722	23225	60737	70964	69580
18	41691	90207	71344	34238	16802	28907
19	85399	83399	12510	90133	88939	09843
20	41156	47003	03760	81752	68207	11511
21	06393	22104	73394	63706	77186	25746
22	32553	34173	65147	72237	45907	28380
23	32910	57643	15521	66693	98589	04040
24	41399	15041	28123	82172	39595	95858
25	30106	59506	39979	23543	42776	68590
26	98101	40044	80625	64572	17508	32837
27	20783	20578	18919	21820	99827	94235
28	77790	11675	16158	65034	33748	74790
29	80476	83266	67505	43320	89331	07379
30	25240	93076	02390	92983	44931	74893
31	83144	57475	25785	10325	12737	11515
32	55340	15195	09092	34690	71257	59134

[1] Начало табл. 1 см. в § 2.1, табл. 2.1.

Продолжение табл. 1

Строки	Столбцы					
	7	8	9	10	11	12
33	87571	95771	12247	41339	53428	20256
34	68303	97887	62578	10156	70419	45255
35	08973	86945	78400	42171	35518	48092
36	59905	66270	77308	24891	57052	32312
37	98718	14813	66510	49246	95971	10139
38	83860	11349	08094	00599	94673	42315
39	20978	44578	00287	32276	11527	87841
40	65153	48788	01031	91547	69389	01217
41	15373	85602	36317	79087	52913	31426
42	39243	48816	15908	46675	55384	32862
43	41474	06804	41379	03831	66488	34991
44	18065	98951	37265	11100	81162	47771
45	34796	85400	29250	50444	28605	41005
46	96153	84388	92186	00272	36298	92212
47	86933	85802	72148	99170	18132	21138
48	03146	60623	39796	73131	02520	11545
49	77476	35745	99258	09427	93808	08041
50	12034	52876	46404	89382	80314	89264
	13	14	15	16	17	18
1	75271	96289	21370	30425	87456	75977
2	37843	60372	95463	89866	91973	99956
3	54005	62441	99230	68830	80333	19892
4	43824	57278	06184	62049	18597	51563
5	38217	81053	47130	72806	50112	16732
6	38122	01773	20240	73658	23017	39283
7	10021	97799	30137	55881	38513	14277
8	72263	24463	67794	92989	49707	48257
9	03943	42764	54090	04453	00446	86588
10	22000	70573	74468	20513	30608	00215
11	68331	41915	14349	44618	75453	70772
12	18659	27003	61359	97621	78992	74622
13	98651	20434	36962	52267	73580	06817
14	47580	80829	33801	16395	53986	08862
15	24467	61583	73632	80751	64613	19147
16	17646	32939	57684	37546	39010	42706
17	14644	11587	54800	67868	45733	60332
18	24441	00163	90569	11890	32939	37177
19	34051	03412	45564	81980	03672	43263
20	28227	16903	62376	86248	11686	12075

Продолжение табл. 1

Строки	Столбцы					
	13	14	15	16	17	18
21	14438	07819	46786	89545	41602	83436
22	83267	98000	60979	28616	80166	05331
23	34364	31067	88149	39453	54599	02345
24	36838	03789	19701	37430	76857	26276
25	27350	77691	18873	33947	13401	78952
26	63645	18567	57825	59173	87245	00669
27	77933	24942	84544	84815	12215	25638
28	28063	23776	27065	32655	53550	77226
29	94792	15332	35230	48164	11101	26982
30	77192	23486	02037	35351	72285	30837
31	31220	81357	27651	08859	29680	46534
32	40650	87997	56243	32047	10407	87739
33	38628	07913	42662	48887	57899	59763
34	59336	93090	42009	77102	05657	78035
35	96574	09864	72547	70782	83337	05576
36	61112	93997	42345	48032	25173	38876
37	35406	24826	43764	93241	11819	88330
38	05750	67568	05910	04743	18001	03248
39	12451	17003	50239	81970	45511	20591
40	57377	38981	96998	93776	36256	50791
41	30784	17260	17381	98615	55002	07294
42	59066	82103	22519	81153	40888	59324
43	95106	09976	10334	94653	84520	63299
44	85892	16227	43641	37880	15021	61073
45	84008	87289	49340	69314	29574	66939
46	72350	43300	42564	14787	60889	24116
47	76897	51233	27689	99654	23560	09490
48	46878	33197	19185	18196	04256	26401
49	04888	72852	55203	60985	23047	92384
50	36185	27056	18329	96802	17827	50891
	19	20	21	22	23	24
1	06216	67466	33474	40659	62317	45987
2	96744	28732	77610	60234	77705	87767
3	14425	36411	82804	11909	11892	65278
4	89527	75437	82899	48657	10216	73270
5	24198	23043	36768	93017	14792	41675
6	53143	41903	87610	95252	09029	67354
7	42964	82051	97664	31390	90918	61956
8	17772	36849	50373	46393	54294	83814
9	46336	62837	51934	10695	03970	02544

Продолжение табл. 1

Строки	Столбцы					
	19	20	21	22	23	24
10	38782	86887	57669	04688	31757	99125
11	63555	85983	48271	58225	67431	24025
12	44314	30284	22758	73720	14250	27043
13	31331	22819	08360	55797	13281	68119
14	09444	91425	06727	63179	45244	06560
15	12961	01435	56357	87663	20147	98615
16	64604	72044	54531	45650	47849	17292
17	86824	90802	92319	40655	80371	60000
18	19014	57244	26793	45566	47238	71635
19	78816	49471	55608	62095	53400	73122
20	30817	79331	35766	06915	54158	63021
21	29696	15111	03838	57528	54224	15144
22	86070	31388	00653	61896	32180	70574
23	05565	33417	77081	55540	20928	53039
24	23697	57662	40580	42279	02751	10058
25	48517	57064	49286	10480	29497	47919
26	73376	18598	88769	45369	61259	63674
27	76585	88215	72464	96635	09854	91649
28	12175	95845	80478	85252	65142	62982
29	35439	60627	72522	07245	41074	43195
30	65790	22131	75583	98529	59381	50498
31	84831	72209	62799	67289	76698	73501
32	99333	78859	98193	06209	30617	06507
33	67548	23078	81336	09730	02137	43120
34	75718	29809	45953	67552	38300	13880
35	18012	39089	45079	85707	28415	96111
36	25087	56471	87958	84917	60326	50355
37	29277	64129	08048	76610	16806	39395
38	15907	76509	16990	34660	32961	50923
39	41058	13096	47622	06855	85468	94457
40	45777	78447	01523	58719	11327	07819
41	60787	84028	11166	34853	85787	94386
42	27713	93990	61724	32414	55270	90483
43	85721	52496	37413	41057	99778	96475
44	22592	29987	10417	38284	74505	47658
45	39734	69078	28831	61339	61698	36874
46	70453	57168	60140	07264	56805	91512
47	26787	31291	47257	15459	75796	86148
48	49350	85951	41764	66706	97785	32988
49	72092	76861	88164	99121	25708	03035
50	14773	34364	05754	27032	59862	01745

Продолжение табл. 1

Строки	Столбцы					
	25	26	27	28	29	30
1	32362	35507	46290	37331	75534	81484
2	46468	06183	96913	33426	10000	23153
3	12738	32389	05683	86782	52242	86400
4	85650	13600	82078	74306	06929	29945
5	49945	34338	46891	96612	17134	12463
6	19407	04704	19569	00356	58159	03211
7	46613	40956	73508	70792	25848	11781
8	32592	25162	83200	26959	30181	45373
9	76940	26062	09124	59249	21417	84747
10	20768	11923	50218	00268	86710	19770
11	61124	34084	65860	94465	28671	55728
12	06775	70531	71973	57250	51190	35162
13	91377	83756	93938	83034	51467	62105
14	59886	22924	15732	81386	96972	52599
15	65608	25267	24540	81550	85802	42772
16	59630	06022	09987	24742	61939	96766
17	19140	24682	90409	61640	27924	56940
18	15786	21148	89431	01219	88105	62078
19	86869	58097	54754	44999	61463	20532
20	22197	93035	74226	35521	94666	90609
21	66294	74529	40236	02242	21737	22421
22	68883	79813	62128	48085	50670	09136
23	39068	17342	85765	84540	45470	82662
24	37976	60902	75770	07775	79237	29341
25	50912	24970	99797	83900	89264	12857
26	46216	86531	00760	69879	85658	96779
27	18061	66052	82206	69612	01532	41613
28	46362	59762	71830	87980	24974	42756
29	78976	57841	41160	16171	96874	37285
30	51280	13871	69015	18292	22065	98727
31	54402	50131	44785	96276	40014	00684
32	58261	71066	63390	84212	23354	58650
33	76972	04191	47234	18463	96477	46683
34	65505	87643	06248	68047	01023	95660
35	26889	40750	22780	70013	83316	44289
36	46444	51830	43522	54936	15328	38997
37	06386	54392	51811	62695	56506	32078
38	17301	34762	17311	78983	97949	50777
39	27342	31199	36851	53122	97763	07919
40	44974	18758	41350	75605	15371	31534

Окончание табл. 1

Строки	Столбцы					
	25	26	27	28	29	30
41	44823	56351	17283	43101	61422	88030
42	49208	39765	17027	55198	70098	58594
43	34145	70697	33474	15488	80185	57414
44	55212	82643	32233	37818	95858	44504
45	03464	04845	66967	22711	56664	25113
46	46594	20734	97607	78269	77646	11797
47	84889	84611	96412	30870	35242	60952
48	12284	44974	52584	64121	80750	01197
49	60929	76004	04679	76620	43311	15184
50	75165	41468	24327	56929	73866	91116

Таблица 2

Функция стандартного нормального распределения N(0,1)

z	F_z	$F-0,5$	F_{II}	z	F_z	$F-0,5$	F_{II}
0,00	0,5000	0,0000	0,0000	1,55	0,9394	0,4394	0,8789
0,05	0,5199	0,0199	0,0399	1,60	0,9452	0,4452	0,8904
0,10	0,5398	0,0398	0,0797	1,65	0,9505	0,4505	0,9011
0,15	0,5596	0,0596	0,1192	1,70	0,9554	0,4554	0,9109
0,20	0,5793	0,0793	0,1585	1,75	0,9599	0,4599	0,9199
0,25	0,5987	0,0987	0,1974	1,80	0,9641	0,4641	0,9281
0,30	0,6179	0,1179	0,2358	1,85	0,9678	0,4678	0,9357
0,35	0,6368	0,1368	0,2737	1,90	0,9713	0,4713	0,9426
0,40	0,6554	0,1554	0,3108	1,95	0,9744	0,4744	0,9488
0,45	0,6736	0,1736	0,3473	2,00	0,9772	0,4772	0,9545
0,50	0,6915	0,1915	0,3829	2,05	0,9798	0,4798	0,9596
0,55	0,7088	0,2088	0,4177	2,10	0,9821	0,4821	0,9643
0,60	0,7257	0,2257	0,4515	2,15	0,9842	0,4842	0,9684
0,65	0,7422	0,2422	0,4843	2,20	0,9861	0,4861	0,9722
0,70	0,7580	0,2580	0,5161	2,25	0,9878	0,4878	0,9756
0,75	0,7734	0,2734	0,5467	2,30	0,9893	0,4893	0,9786
0,80	0,7881	0,2881	0,5763	2,35	0,9906	0,4906	0,9812
0,85	0,8023	0,3023	0,6047	2,40	0,9918	0,4918	0,9836
0,90	0,8159	0,3159	0,6319	2,45	0,9929	0,4929	0,9857
0,95	0,8289	0,3289	0,6579	2,50	0,9938	0,4938	0,9876
1,00	0,8413	0,3413	0,6827	2,55	0,9946	0,4946	0,9892
1,05	0,8531	0,3531	0,7063	2,60	0,9953	0,4953	0,9907
1,10	0,8643	0,3643	0,7287	2,65	0,9960	0,4960	0,9920
1,15	0,8749	0,3749	0,7499	2,70	0,9965	0,4965	0,9931
1,20	0,8849	0,3849	0,7699	2,75	0,9970	0,4970	0,9940
1,25	0,8944	0,3944	0,7887	2,80	0,9974	0,4974	0,9949
1,30	0,9032	0,4032	0,8064	2,85	0,9978	0,4978	0,9956
1,35	0,9115	0,4115	0,8230	2,90	0,9981	0,4981	0,9963
1,40	0,9192	0,4192	0,8385	2,95	0,9984	0,4984	0,9968
1,45	0,9265	0,4265	0,8529	3,00	0,9987	0,4987	0,9973
1,50	0,9332	0,4332	0,8664				

Таблица 3

Функция биномиального распределения

n	k	k/n	p				
			0,01	0,02	0,03	0,04	0,05
40	0	0,00	0.66897	0,44570	0,29571	0,19537	0,12851
	1	2,50	0,93926	0,80954	0,66154	0,52098	0,39906
	2	5,00	0,99250	0,95433	0,88217	0,78553	0,67674
	3	7,50			0.96860	0,92516	0,86185
	4	10,00				0,97898	0,95197
50	0	0,00	0,60501	0,36417	0,21807	0,12989	
	1	2,00	0,91056	0,73577	0,55528	0,40048	0,27943
	2	4,00	0,98618	0,92157	0,81080	0,67671	0,54053
	3	6,00		0,98224	0,93724	0,86087	0,76041
	4	8,00			0,98319	0,95103	0,89638
	5	10,00					0,96222
80	0	0	0,44752	0,19865	0,08745	0,03817	0,01652
	1	1,25	0,80916	0.52297	0,30381	0,16539	0,08605
	2	2,5	0,95345	0,78442	0,56812	0,37479	0,23062
	3	3,75		0,92315	0,78067	0,60163	0,42845
	4	5		0,97764	0,90721	0,78358	0,62888
	5	6,25			0,96669	0,89881	0,78922
	6	7,5				0,95883	0,89471
	7	8,75					0,95341
100	0	0	0,36603	0,13262	0,04755	0,01687	0,00592
	1	1	0,73576	0,40327	0,19462	0,08716	0,03708
	2	2	0,92063	0,67669	0,41978	0,23214	0,11826
	3	3	0,98163	0,85896	0,64725	0,60163	0,25784
	4	4		0,94917	0,81785	0,62886	0,43598
	5	5		0,98452	0,91916	0,78837	0,61600
	6	6			0,96877	0,89361	0,76601
	7	7				0,95249	0,87204
	8	8					0,93691
	9	9					0,97181

Продолжение табл. 3

n	k	k/n	p				
			0,01	0,02	0,03	0,04	0,05
150	0	0	0,2215	0,0483	0,0104	.	
	1	0,67	0,5570	0,1961	0,0585	0,0159	
	2	1,33	0,8095	0,4209	0,1693	0,0584	0,0182
	3	2,00	0,9353	0,6472	0,3384	0,1458	0,0548
	4	2,67	0,9820	0,8170	0,5307	0,2796	0,1256
	5	3,33		0,9181	0,7043	0,4424	0,2344
	6	4,00		0,9680	0,8340	0,6063	0,3729
	7	4,67			0,9166	0,7468	0,5228
	8	5,33			0,9622	0,8515	0,6638
	9	6.00				0,9203	0,7809
	10	6,67				0,9607	0,8678
	11	7,33					0,9260
	12	8,00					0,9615
200	1	0,5	0,4046	0,0894	0,0162		
	2	1	0,6767	0,2351	0,0593	0,0125	
	3	1,5	0,8580	0,4315	0,1472	0,0395	
	4	2	0,9483	0,6288	0,2810	0,0950	0,0264
	5	2,5	0,9840	0,7867	0,4432	0,1856	0,0623
	6	3		0,8914	0,6063	0,3084	0,1237
	7	3,5		0,9507	0,7461	0,4501	0,2133
	8	4			0,8504	0,5926	0,3270
	9	4,5			0,9192	0,7192	0,4547
	10	5			0,9599	0,8200	0,5831
	11	5,5				0,8925	0,6998
	12	6				0,9401	0,7965
	13	6,5				0,9688	0,8701
	14	7					0,9219
	15	7,5					0,9556
250	1	0,4	0,2858	0,0391			
	2	0,8	0,5432	0,1221	0,0190		
	3	1,2	0,7581	0,2622	0,0565		
	4	1,6	0,8922	0,4387	0,1282	0,0270	
	5	2	0,9588	0,6160	0,2373	0,0633	0,0623
	6	2,4		0,7637	0,3750	0,1250	0,0314
	7	2,8		0,8687	0,5235	0,2147	0,0650
	8	3,2		0,9339	0,6630	0,3282	0,1186
	9	3,6		0,9696	0,7790	0,4554	0,1946
	10	4			0,8655	0,5831	0,2909
	11	4,4			0,9239	0,6991	0,4016
	12	4,8			0,9598	0,7955	0,5175
	13	5,2				0,8690	0,6293
	14	5,6				0,9208	0,7288
	15	6				0,9548	0,8113
	16	6,4					0,8750
	17	6,8					0,9212
	18	7,2					0,9526

Продолжение табл. 3

n	k	k/n	p					
			0,06	0,07	0,08	0,09	0,10	0,12
40	0	0	0,08416	0,05487	0,03561	0,02300	0,01478	
	1	2,50	0,29904	0,22006	0,15945	0,11397	0,08047	0,03883
	2	5,00	0,56650	0,46252	0,36945	0,28942	0,22281	0,12609
	3	7,50	0,78274	0,69369	0,60075	0,50921	0,42313	0,27680
	4	10,00	0,91042	0,85463	0,78679	0,71029	0,62902	0,46691
	5	12,50	0,96909	0,94185	0,90327	0,85347	0,79373	0,65356
	6	15,00		0,98015	0,96236	0,93608	0,90048	0,80203
	7	17,50				0,97576	0,95810	0,90037
40	8	20,00						0,95569
50	0	0	0,04533	0,02656	0,01547			
	1	2	0,19000	0,12649	0,08271	0,05324	0,03379	0,01310
	2	4	0,41625	0,31079	0,22597	0,16054	0,11173	0,05126
	3	6	0,64730	0,53274	0,42530	0,33034	0,25029	0,13453
	4	8	0,82060	0,72903	0,62895	0,52766	0,43120	0,26795
	5	10	0,92236	0,86495	0,79187	0,70719	0,61612	0,43534
	6	12	0,97108	0,94169	0,89813	0,84037	0,77023	0,60652
	7	14		0,97799	0,95621	0,92316	0,87785	0,75325
	8	16				0,96717	0,94213	0,86080
	9	18					0,97546	0,92924
50	10	20						0,96750
80	1	1,25	0,04325	0,02114	0,01009			
	2	2,5	0,13445	0,07503	0,04038	0,02106	0,01068	
	3	3,75	0,28579	0,18051	0,10886	0,06309	0,03531	0,01010
	4	5	0,47174	0,33333	0,22350	0,14311	0,08797	0,02989
	5	6,25	0,65216	0,50817	0,37502	0,26341	0,17692	0,07091
	6	7,5	0,79611	0,67267	0,53972	0,41213	0,30045	0,14084
	7	8,75	0,89325	0,80357	0,69111	0,56761	0,44556	0,24165
	8	10	0,94982	0,89347	0,81125	0,70794	0,59268	0,36708
	9	11,25	0,97871	0,94760	0,89481	0,81896	0,72345	0,50392
	10	12,5		0,97653	0,94641	0,89693	0,82662	0,63640
	11	13,75			0,97496	0,94599	0,89956	0,75136
	12	15				0,97390	0,94617	0,84151
	13	16,25					0,97325	0,90580
	14	17,5						0,94776
80	15	18,75						0,97294
100	1	1	0,01517	0,00601				
	2	2	0,05661	0,02579	0,01127			
	3	3	0,14302	0,07441	0,03671	0,01730		
	4	4	0,27678	0,16316	0,09034	0,04739	0,02371	
	5	5	0,44069	0,29142	0,17988	0,10452	0,05758	0,01522
	6	6	0,60635	0,44428	0,30316	0,19398	0,11716	0,03674
	7	7	0,74835	0,59878	0,44711	0,31279	0,20605	0,07614
	8	8	0,85371	0,73397	0,59263	0,44940	0,32087	0,13859
	9	9	0,92246	0,83798	0,72198	0,58751	0,45129	0,22565
	10	10	0,96239	0,90922	0,82433	0,71180	0,58316	0,33369
	11	11	0,98325	0,95310	0,89715	0,81238	0,70303	0,45422

Продолжение табл. 3

n	k	k/n	p					
			0,06	0,07	0,08	0,09	0,10	0,12
	12	12		0,97759	0,94412	0,88616	0,80182	0,57612
	13	13			0,97176	0,93555	0,87612	0,68865
	14	14				0,96590	0,92743	0,78400
	15	15					0,96011	0,85855
	16	16						0,91256
	17	17						0,94895
100	18	18						0,97183
150	3	2,00	0,0186					
	4	2.67	0,0499	0,0180				
	5	3,33	0,1083	0,0448	0,0169			
	6	4,00	0,0052	0,0934	0,0397	0,0155		
	7	4,67	0,3166	0,1688	0,0806	0,0350	0,0140	
	8	5,33	0,4516	0,2702	0,1442	0,0696	0,0307	
	9	6,00	0,5875	0,3907	0,2315	0,1235	0,0600	0,0111
	10	6,67	0,7098	0,5185	0,3384	0,1987	0,1060	0,0234
	11	7,33	0,8091	0,6410	0,4568	0,2934	0,1709	0,0446
	12	8,00	0,8826	0,7477	0,5760	0,4018	0,2545	0,0781
	13	8,67	0,9323	0,8330	0,6861	0,5157	0,3530	0,1266
	14	9,33	0,9634	0,8959	0,7797	0,6259	0,4602	0,1914
	15	10,00		0,9387	0,8536	0,7247	0,5682	0,2715
	16	10,67		0,9660	0,9078	0,8071	0,6694	0,3636
	17	11,33			0,9449	0,8714	0,7581	0,4626
	18	12,00			0,9687	0,9184	0,8308	0,5624
	19	12,67				0,9507	0,8870	0,6569
	20	13,33					0,9279	0,7413
	21	14,00					0,9560	0,8126
	22	14,67						0,8696
	23	15,33						0,9128
	24	16,00						0,9440
150	25	16,67						0,9654
200	5	2,5	0,0177					
	6	3	0,0413	0,0119				
	7	3,5	0,0829	0,0274				
	8	4	0,1470	0,0556	0,0183			
	9	4,5	0,2343	0,1010	0,0374	0,0121		
	10	5	0,3407	0,1661	0,0691	0,0251		
	11	5.5	0,4580	0,2508	0,1168	0,0472	0,0168	
	12	6	0,5760	0,3513	0,1821	0,0816	0,0320	
	13	6,5	0,6849	0,4606	0,2643	0,1308	0,0566	
	14	7	0,7777	0,5705	0,3597	0,1958	0,0929	0,0146
	15	7,5	0,8512	0,6731	0,4626	0,2755	0,1431	0,0267
	16	8	0,9054	0,7623	0,5660	0,3666	0,2075	0,0457
	17	8,5	0,9429	0,8351	0,6634	0,4642	0,2849	0,0739
	18	9	0,9672	0,8907	0,7494	0,5623	0,3724	0,1129
	19	9,5		0,9308	0,8211	0,6553	0,4655	0,1638
	20	10		0,9582	0,8775	0,7385	0,5592	0,2266

Окончание табл. 3

n	k	k/n	p					
			0,06	0,07	0,08	0,09	0,10	0,12
	21	10,5			0,9196	0,8090	0,6484	0,3001
	22	11			0,9493	0,8657	0,7290	0,3816
	23	11,5			0,9694	0,9092	0,7983	0,4676
	24	12				0,9409	0,8551	0,5541
	25	12,5				0,9629	0,8995	0,6372
	26	13					0,9328	0,7134
	27	13,5					0,9566	0,7804
	28	14						0,8368
	29	14,5						0,8824
	30	15						0,9179
	31	15,5						0,9444
200	32	16						0,9635
250	7	2,8	0,0156					
	8	3,2	0,0334					
	9	3,6	0,0640	0,0170				
	10	4	0,1110	0,0339				
	11	4,4	0,1766	0,0616	0,0177			
	12	4,8	0,2599	0,1032	0,0334			
	13	5,2	0,3573	0,1604	0,0585	0,0178		
	14	5,6	0,4625	0,2334	0,0954	0,0323		
	15	6	0,5681	0,3198	0,1458	0,0550	0,0175	
	16	6,4	0,6672	0,4153	0,2103	0,0879	0,0309	
	17	6,8	0,7542	0,5143	0,2875	0,1327	0,0513	
	18	7,2	0,8261	0,6107	0,3743	0,1900	0,0808	
	19	7,6	0,8821	0,6993	0,4665	0,2592	0,1207	0,0162
	20	8	0,9235	0,7764	0,5591	0,3383	0,1719	0,0274
	21	8,4	0,9523	0,8399	0,6473	0,4239	0,2342	0,0441
	22	8,8		0,8896	0,7272	0,5121	0,3063	0,0678
	23	9,20		0,9267	0,7960	0,5986	0,3857	0,0998
	24	9,60		0,9532	0,8526	0,6794	0,4692	0,1411
	25	10,00			0,8971	0,7517	0,5530	0,1920
	26	10,40			0,9306	0,8136	0,6336	0,2521
	27	10,80			0,9547	0,8644	0,7079	0,3201
	28	11,20				0,9044	0,7736	0,3940
	29	11,60				0,9347	0,8296	0,4710
	30	12,00				0,9567	0,8753	0,5485
	31	12,40					0,9114	0,6234
	32	12,80					0,9389	0,6933
	33	13,20					0,9590	0,7563
	34	13,60						0,8111
	35	14,00						0,8573
	36	14,40						0,8948
	37	14,80						0,9245
	38	15,20						0,9471
	39	15,60						0,9639

Таблица 4

t-распределение Стьюдента

v	$p = 0,2$	$p = 0,15$	$p = 0,1$	$p = 0,05$	$p = 0,01$
5	1,476	1,699	2,015	2,571	4,032
6	1,440	1,650	1,943	2,447	3,707
7	1,415	1,617	1,895	2,365	3,499
8	1,397	1,592	1,860	2,306	3,355
9	1,383	1,574	1,833	2,262	1,833
10	1,372	1,559	1,812	2,228	3,169
11	1,363	1,548	1,796	2,201	3,106
12	1,356	1,538	1,782	2,179	3,055
13	1,350	1,530	1,771	2,160	3,012
14	1,345	1,523	1,761	2,145	2,977
15	1,341	1,517	1,753	2,131	2,947
16	1,337	1,512	1,746	2,120	2,921
17	1,333	1,508	1,740	2,110	2,898
18	1,330	1,504	1,734	2,101	2,878
19	1,328	1,500	1,729	2,093	2,861
20	1,325	1,497	1,725	2,086	2,845
21	1,323	1,494	1,721	2,080	2,831
22	1,321	1,492	1,717	2,074	2,819
23	1,319	1,489	1,714	2,069	2,807
24	1,318	1,487	1,711	2,064	2,797
25	1,316	1,485	1,708	2,060	2,787
26	1,315	1,483	1,706	2,056	2,779
27	1,314	1,482	1,703	2,052	2,771
28	1,313	1,480	1,701	2,048	2,763
29	1,311	1,479	1,699	2,045	2,756
30	1,310	1,477	1,697	2,042	2,750
31	1,309	1,476	1,696	2,040	2,744
32	1,309	1,475	1,694	2,037	2,738
33	1,308	1,474	1,692	2,035	2,733
34	1,307	1,473	1,691	2,032	2,728
35	1,306	1,472	1,690	2,030	2,724
40	1,303	1,468	1,684	2,021	2,704
50	1,299	1,462	1,676	2,009	2,678
100	1,290	1,451	1,660	1,984	2,626

Евгений Михайлович ЧЕТЫРКИН,
Наталия Евгеньевна ВАСИЛЬЕВА

Выборочные методы в аудите

Главный редактор *Ю.В. Луизо*
Зав. редакцией *Г.Г. Кобякова*
Редактор *М.Н. Глухова*
Художник *Н.В. Пьяных*
Компьютерная подготовка
оригинал-макета *С.П. Кильдишев*
Технический редактор *Л.А. Зотова*
Корректоры *Л.М. Филькова, Г.В. Платова*

Подписано в печать 2.07.2003. Формат 60×90 1/16. Бумага офсетная.
Гарнитура Таймс. Печать офсетная. Усл. печ. л. 9,0.
Тираж 3000 экз. Заказ № 747. Изд. № 409.

Издательство "Дело"
119571, Москва, пр-т Вернадского, 82

Коммерческий отдел – тел. 433-2510, 433-2502
E-mail: delo@ane.ru
Internet: http://www.delo.ane.ru

ФГУП "Московская типография № 6" Министерства Российской Федерации
по делам печати, телерадиовещания и средств массовых коммуникаций
115088, Москва, Южнопортовая ул., 24

ISBN 5-7749-0321-4