кластерный анализ \ распределение Бенфорда \ аудиторская выборка

Статистические методы анализа данных позволяют повысить эффективность работы внутреннего аудитора по обнаружению нестандартных действий (в том числе мошеннических) в больших совокупностях учетной информации. В статье представлены результаты применения на практике двух методов выявления нетипичного поведения, относящихся к статистическим: кластерного анализа и выявления элементов совокупности на основе распределения Бенфорда.

Евгений ЗВЕРЕВ, *CIA, член Института внутренних аудиторов*
Андрей НИКИФОРОВ, *риск-менеджер*

Учетные данные с нестандартным поведением: практические способы выявления признаков мошенничества

Предполагается, что цифровизация позволит аудитору анализировать всю совокупность учетной информации (иначе говоря, выборка будет соответствовать совокупности). В анализе будет задействована разнообразная, полученная из различных источников, но в то же время взаимоувязанная информация. Выгрузка значительного объема оцифрованной финансово-экономической информации из ERP-систем с целью ее последующего анализа уже осуществляется. Но возникает парадоксальная ситуация: учетная информация для анализа подготовлена, но как проанализировать значительный объем информации в разумные сроки и с ограниченным ресурсом? В связи с этим актуален вопрос об определении критерия для отбора элементов совокупности с целью детальной проверки. Таким критерием может быть нестандартное (нетипичное) поведение элементов совокупности, выпадающее из общих тенденций, свойственных всей совокупности.

Кластерный анализ

Кластерный анализ[1] (далее — КА) позволяет разделить по заданным параметрам совокупность данных на отдельные группы (кластеры),

[1] Зверев Е., Никифоров А. Кластерный анализ: формирование индикатора риска для больших совокупностей учетной информации // Внутренний контроль в кредитной организации. 2018. № 3. С. 28–39.

Евгений ЗВЕРЕВ
Андрей НИКИФОРОВ

при этом в каждый кластер включаются элементы, которые, с одной стороны, максимально отличаются от элементов, попавших в прочие кластеры, а с другой — внутри кластера максимально схожи между собой. Потенциальный интерес представляют кластеры с относительно небольшим количеством элементов, максимально отличающиеся от остальных, то есть проявившие нестандартное поведение.

В результате проведения КА совокупность из 6932 номенклатурных единиц ТМЦ была разделена на 190 кластеров. При проведении КА был выбран алгоритм, который формирует кластеры на основе проверки нормальности распределения элементов совокупности. Данный подход методологически соответствует предположению, что отклонение от нормального распределения связано с нетипичным поведением.

Номенклатуры ТМЦ распределились по кластерам неравномерно (табл. 1). Основное количество элементов сосредоточено в небольшом количестве кластеров, содержащих более 30 элементов.

Интерес представляют кластеры с профилем (соотношением параметров, по которым проводилась кластеризация), существенно отличающимся от профилей остальных кластеров. То есть элементы, входящие в данные кластеры, проявляют нестандартное (нетипичное) поведение (рис. 1). Общее количество единиц номенклатуры, попавших в данные кластеры, равно 57.

Исходя из своего опыта, авторы хотели бы отметить, что обычно нестандартным поведением обладает небольшое количество кластеров, на которое и имеет смысл обратить внимание.

В результате последующего детального анализа учетной информации по номенклатурам ТМЦ, попавшим в кластеры с нестандартным поведением, выявлен ряд нарушений (табл. 2).

Таблица 1

Результат кластеризации совокупности ТМЦ по четырем параметрам

Диапазон количества номенклатур ТМЦ в одном кластере	Количество кластеров	Количество элементов
Менее 5	23	57
От 6 до 10	50	313
От 11 до 30	63	1161
От 31 до 100	40	2343
Более 100	14	3058

Рисунок 1

Профиль кластеров с нетипичным поведением

Таблица 2

Описание кластеров с нетипичным поведением

Кластер	Кол-во типов ТМЦ	Характеристика выявленного нарушения
1	14	Приобретение ТМЦ при отсутствии потребности. Личная заинтересованность
3	7	Несоответствие документального и фактического остатков ТМЦ. Недостачи. Пересортица
4	5	Срочное приобретение у «единственного поставщика». Личная заинтересованность
5	12	Аналогично кластеру № 4, но на другом ценовом уровне
6	14	Аналогично кластеру № 1, но на другом ценовом уровне
28	12	Аналогично кластеру № 3, но на другом ценовом уровне

При интерпретации результатов следует учитывать, что в большинстве случаев метод не выявляет нарушение (риск) в чистом виде, но позволяет сконцентрировать внимание на небольших выборках (кластерах).

Рекомендации по применению КА

1. Метод не ограничен рамками априорных предположений о структуре выборки и виде распределений значений анализируемых показателей.

Евгений ЗВЕРЕВ
Андрей НИКИФОРОВ

2. Применение метода полезно в тех случаях, когда из большой совокупности данных необходимо сделать выборку, содержащую признаки нестандартного поведения.

3. Результаты применения метода (кластеры с распределенными по ним элементами) необходимо профессионально интерпретировать, поскольку он является не более чем индикатором риска и необходим более детальный анализ, в котором возрастающее значение будет иметь профессиональное суждение.

4. При выборе параметров кластеризации необходимо заранее определить вид нарушения (риска), который предполагается обнаружить.

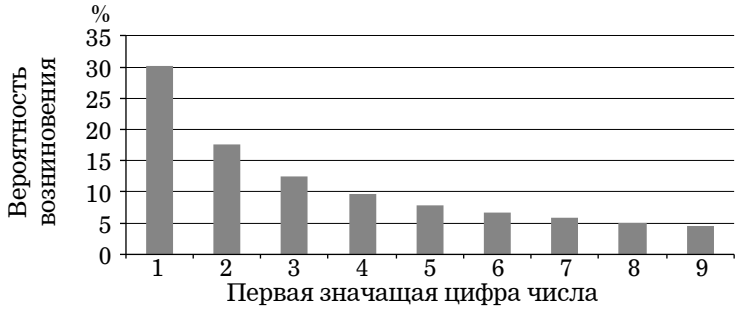
Распределение Бенфорда

Элементы совокупности с нестандартным поведением могут быть выявлены на основе анализа соответствия их поведения распределению Бенфорда[1] (далее — метод Бенфорда, МБ). Авторы применяли данный метод, чтобы определить выборку, представляющую интерес для аудита, из общей совокупности номенклатуры ТМЦ.

МБ заключается в определении частот появления первых значащих цифр (например: первой, второй, первой плюс второй и др.) данных, относящихся к анализируемой совокупности, и в сравнении этих частот с теоретическим распределением (распределением Бенфорда), отражающим вероятность возникновения первых значащих цифр. Пример распределения Бенфорда для первой значащей цифры показан на рис. 2.

Рисунок 2

Распределение Бенфорда для первой значащей цифры



[1] Зверев Е., Никифоров А. Распределение Бенфорда: выявление нестандартных элементов в больших совокупностях финансовой информации // Внутренний контроль в кредитной организации. 2018. № 4. С. 42-56.

Учетные данные с нестандартным поведением: практические способы выявления признаков мошенничества

Распределению Бенфорда, как правило[1], соответствуют так называемые *естественные финансовые учетные данные*. К таким данным относятся:

— остатки на счетах кредитных карт и (или) транзакции по ним;
— суммы бухгалтерских проводок;
— суммы страховых выплат;
— суммы гарантийного ремонта;
— суммы выставленных счетов к оплате;
— объемы поставок;
— суммы в налоговых декларациях;
— списки стоимости покупок или денежных поступлений и пр.

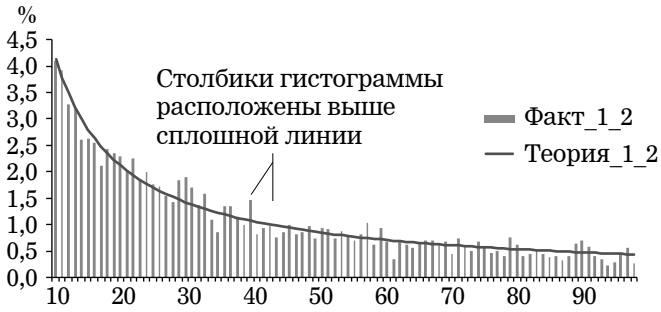
Отклонение от распределения Бенфорда может быть связано с нетипичным (нестандартным) поведением элементов совокупности.

Авторами были проанализированы данные на забалансовом учете с оборотом в календарный год. Исследуемая совокупность имела 3416 номенклатурных единиц давальческих ТМЦ. В качестве анализируемого параметра использован кредитовый оборот (транзакция списания) номенклатурных единиц. Рассчитывалась частота «выпадения» первой пары цифр (от 10 до 99) у каждого элемента (числа) совокупности. Полученное распределение частот сравнивалось с распределением Бенфорда.

В качестве элементов, представляющих интерес, были выбраны транзакции, у которых частота появления первых двух значащих цифр превышает теоретическое распределение (рис. 3). Всего выявлено 53 таких элемента.

Рисунок 3

Тест на соответствие распределению Бенфорда по 1-й + 2-й значащим цифрам



[1] См., например: Nigrini M.J. Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection. John Wiley & Sons, 2011. P. 330.

Евгений ЗВЕРЕВ
Андрей НИКИФОРОВ

Детальная аудиторская проверка (инвентаризация), проведенная по номенклатурным элементам, выбранным на основе отклонения транзакций списания от МБ, выявила недостачи (табл. 3).

Таблица 3

Описание результата применения МБ

Совокупность, ед.	Отклонения МБ, ед.	Нарушения, ед.	Характеристика выявленных нарушений
3416	53[1]	7	Несоответствие документального и фактического остатков ТМЦ

Рекомендации по применению МБ

1. Метод ограничен рамками априорных предположений о виде распределения значений анализируемых показателей. Авторы рекомендуют предварительно исследовать совокупность на соответствие распределению Бенфорда или иному распределению.

2–4. Аналогично рекомендациям по применению КА.

Представленные результаты подтверждают полезность применения статистических методов, в частности кластерного анализа и метода Бенфорда, для формирования выборки из большой совокупности учетных данных.

Применение данных методов направлено не на получение репрезентативной выборки, то есть выборки, максимально отражающей свойства генеральной совокупности, а на выявление элементов с нетипичным (нестандартным) поведением, которые с большей вероятностью будут представлять интерес для детальной аудиторской проверки[2].

Для получения выборки с нетипичным поведением элементов могут быть применены и другие статистические методы, например корреляционный анализ, анализ соответствия нормальному распределению и пр. Включение таких методов в алгоритмы обработки информации, выгружаемой из учетных систем, в рамках реализации проектов по цифровизации учетных данных было бы очень полезным для повышения эффективности и результативности аудита.

[1] В действительности подобных элементов больше, но объем выборки был ограничен уровнем статистической значимости отклонений при доверительной вероятности 95%.

[2] Зверев Е., Никифоров А. Экспертная выборка: формирование для большой совокупности // МСФО и МСА в кредитной организации. 2019. № 2. С. 67–73.