А.Б. Лось
канд. техн. наук, доцент
E-mail: alos@hse.ru
А.С. Кабанов
канд. техн. наук, доцент
E-mail: kabanov_as@mail.ru
(Московский институт электроники и математики Национального исследовательского университета «Высшая школа экономики»)
Москва, Российская Федерация

Особенности оценки рисков информационной безопасности с использованием регрессионного анализа в системе менеджмента информационной безопасности

В статье обсуждаются вопросы построения системы менеджмента информационной безопасности и методы оценки рисков ее нарушения. Рассматриваются особенности использования регрессионного анализа в системе менеджмента информационной безопасности. Проведен анализ преимуществ и недостатков методов регрессионного анализа в применении к процедуре оценки рисков нарушения информационной безопасности. Рассмотрены положения национального стандарта ГОСТ Р ИСО 10017-2005 в части использования методов регрессионного анализа. Особое внимание уделяется специфике использования регрессионного анализа с целью обеспечения адекватности применяемых моделей при оценке рисков нарушения информационной безопасности. Сформулированы условия для получения необходимых результатов при построении модели регрессии. Рассмотрены особенности прогнозирования состояний при использовании регрессионного анализа. Кратко рассмотрен алгоритм заполнения пропусков модели регрессии для получения дополнительной информации. Приведены рекомендации по использованию регрессионного анализа для построения оценок рисков нарушения информационной безопасности.

Ключевые слова: *риски нарушения информационной безопасности; методики оценки рисков; регрессионный анализ; ранговая корреляция; мультиколлинеарность.*

A.B. Los
Cand. of Techn. Sciences, Associate Professor
E-mail: alos@hse.ru
A.S. Kabanov
E-mail: kabanov_as@mail.ru
(Moscow Institute of Electronics and Mathematics National Research University «Higher School of Economics»)
Moscow, Russian Federation

Features Information Security Risk Assessment Using Regression Analysis in the Information Security Management System

The article discusses the design of information security management system and methods of risk assessment of its violation. Discusses the features of the use of regression analysis in the information safety management system. The analysis of advantages and disadvantages of the methods of regression analysis applied to the procedure of assessment of the risk of a security breach. The provisions of the national standard GOST R ISO 10017-2005 regarding the use of regression analysis. Particular attention is given to use regression analysis to ensure the adequacy of the applied models in risk assessment information security breaches. The conditions for obtaining the necessary results when constructing the regression model. Considered are the peculiarities of forecasting States when using regression analysis. Briefly describes an algorithm for filling gaps regression model for more information. Contain recommendations on the use of the regression analysis for the estimation of the risk of a security breach.

Keywords: *Information security risks; risk assessment methods; regression analysis; rank correlation, multicollinearity.*

Особенности системы менеджмента информационной безопасности

Использование IT-решений для поддержки основных бизнес-процессов компании или непосредственно для оказания услуг клиентам предъявляет высокие требования к их качеству – доступности, непрерывности и безопасности использования. Реализация угроз различного характера – от вирусной эпидемии во внутренней сети до отказа системы электропитания в масштабах города, может привести к нарушению деятельности организации, прямым финансовым потерям и ущербу репутации. Зрелая Система менеджмента информационной безопасности обеспечивает эффективное управление информационной безопасностью: отсутствие неприемлемых рисков со стороны IT-систем и поддержание баланса между рисками и затратами на обеспечение информационной безопасности [1].

Выгоды от реализации Системы менеджмента информационной безопасности в организации достигаются за счет [1]:

- обеспечения соответствия требованиям законодательства и бизнес-требованиям в области информационной безопасности;
- предупреждения возникновения инцидентов информационной безопасности и снижения ущерба в случае их возникновения;
- повышения культуры информационной безопасности в организации;
- повышения зрелости в области управления обеспечением информационной безопасности;
- оптимизации расходования средств на обеспечение информационной безопасности.

Современная Система менеджмента информационной безопасности представляет собой процессно-ориентированную систему управления, включающую организационный, документальный и программно-аппаратный компоненты. Можно выделить следующие уровни Системы менеджмента информационной безопасности: процессный, документальный и зрелостный.

Процессы Системы менеджмента информационной безопасности создаются в соответствии с требованиями стандарта ISO/IEC 27001:2005, в основе которого лежит цикл управления *Plan-Do-Check-Act*. В соответствии с ним жизненный цикл Системы менеджмента информационной безопасности состоит из четырех типов деятельности: Создание – Внедрение и эксплуатация – Мониторинг и анализ – Сопровождение и совершенствование. Процессы Системы менеджмента информационной безопасности интегрируются в существующую структуру бизнес-процессов организации для выполнения всех требований стандарта.

Для их автоматизации применяется специализированное программное обеспечение, использование которого позволяет существенно уменьшить трудоемкость эксплуатации Системы менеджмента информационной безопасности, повысить уровень зрелости процессов менеджмента и упростить процедуры внутреннего и внешнего сертификационного аудита.

Документация Системы менеджмента информационной безопасности состоит из политик, документированных процедур, стандартов и записей и делится на две части: документация менеджмента Системы менеджмента информационной безопасности и эксплуатационная документация Системы менеджмента информационной безопасности [1].

Документация менеджмента информационной безопасности представлена Политиками информационной безопасности и Системой менеджмента информационной безопасности, основной процедурой – «Менеджмент информационной безопасности» и сопутствующими формами записей, процедурами «Внутренний аудит», «Управление документацией» и «Управление Записями». При необходимости Система менеджмента информационной безопасности интегрируется с существующей в организации Системой менеджмента.

Модель зрелости Системы менеджмента информационной безопасности определяет состав и детализацию разрабатываемой документации, последовательность построения Системы менеджмента информационной безопасности, детальность разрабатываемой документации и степень автоматизации процессов менеджмента и эксплуатации Системы менеджмента информационной безопасности. При оценке и планировании используется модель зрелости CobiT. В Программе повышения зрелости Системы менеджмента информационной безопасности определяются состав и сроки мероприятий по совершенствованию процессов менеджмента информационной безопасностью и управления эксплуатацией средств безопасности [1].

Одним из ключевых элементов Системы менеджмента информационной безопасности является процедура оценки рисков нарушения информационной безопасности. Необходимость проведения анализа рисков в сфере информационной безопасности вызвана следующими причинами [2, 3]:

- выявление проблем в сфере безопасности (не только уязвимостей компонент системы, но и недостатков политик безопасности и т. д.);
- анализ рисков позволяет специалистам, не связанным с техническими вопросами (в частности, руководству организации), оценить выгоды от внедрения средств и механизмов защиты и принять участие в процессе определения требуемого уровня защищенности;
- проведение оценки рисков повышает обоснованность рекомендаций по безопасности;
- анализ возможных рисков позволяет выделить наиболее приоритетные направления для внедрения новых систем защиты, мер и процедур обеспечения информационной безопасности;
- подробно описанные методики анализа рисков позволяют ознакомиться с полученными результатами широкому кругу специалистов, что способствует повышению их обоснованности.

В настоящее время для анализа рисков и их оценки применяются специализированные программные средства, использующие следующие методики [4, 5]:

- оценки риска на качественном уровне (например, по шкале «высокий», «средний», «низкий»), к таким методикам, в частности, относится *FRAP*;
- количественных показателей (риск оценивается через числовое значение, например, размер ожидаемых годовых потерь), к этому классу относится методика *RiskWatch* [6];
- смешанных оценок (такой подход используется в *CRAMM* и *Microsoft*).

Оценка рисков на качественном уровне не позволяет однозначно сравнить затраты на обеспечение информационной безопасности и получаемую от них отдачу (в виде снижения суммарного риска). В связи с этим, более предпочтительными представляются количественные методики (например, *RiskWatch*). Но они требуют наличия оценок вероятности возникновения для каждой из рассматриваемых угроз безопасности [6]. Кроме того,

использование интегральных показателей, таких как ожидаемый годовой ущерб (*ALE*), опасно тем, что неправильная оценка вероятности угрозы в отношении очень дорогостоящего актива может кардинально изменить оцениваемое значение суммарной стоимости рисков [4, 5].

Конкретную методику проведения анализа рисков на предприятии и инструментальные средства, поддерживающие ее, нужно выбирать, учитывая следующие факторы [5]:

- наличие экспертов, способных дать достоверные оценки объема потерь от угроз информационной безопасности;
- наличие на предприятии достоверной статистки по инцидентам в сфере информационной безопасности;
- необходима ли точная количественная оценка последствий реализации угроз или достаточно оценки на качественном уровне.

Общим отличительным признаком всех факторов оценки рисков является использование в той или иной мере экспертных систем (например, определение ожидаемых годовых потерь, оценка риска по шкале и т. д.).

Применение методов регрессионного анализа для оценки рисков нарушения информационной безопасности

В силу широких возможностей, предоставляемых регрессионным анализом в вопросах прогнозирования различного рода показателей, целесообразно рассмотреть возможность его применения для оценки рисков нарушения информационной безопасности. Основной задачей при этом является выработка рекомендаций по применению регрессионного анализа для оценки рисков в Системе менеджмента информационной безопасности.

Регрессия (лат. *regressio* – обратное движение, переход от более сложных форм развития к менее сложным) – одно из основных понятий в теории вероятности и математической статистике, выражающее зависимость среднего значения одной случайной величины от значений другой случайной величины или нескольких случайных величин. Данное понятие введено Ф. Гальтоном в 1886 году.

Регрессионный анализ – это метод моделирования измеряемых данных и исследования их свойств. Данные состоят из пар значений зависимой переменной (переменной отклика) и независимой переменной (объясняющей переменной). Регрессионная модель есть функция независимой переменной и параметров с добавленной случайной переменной. Параметры модели настраиваются таким образом, чтобы модель наилучшим образом приближала данные. Критерием качества приближения обычно является среднеквадратичная ошибка: сумма квадратов разности значений модели и зависимой переменной для всех значений независимой переменной в качестве аргумента. Регрессионный анализ используется для прогноза, анализа временных рядов, тестирования гипотез и выявления скрытых взаимосвязей в данных.

Наиболее часто используемая множественная линейная модель регрессионного анализа имеет вид:

$$Y = b_0 + b_1 x_1 + \ldots + b_r x_i + \varepsilon,$$

где b_i – параметры регрессионной модели, ε – случайные ошибки наблюдения.

Регрессионный анализ показывает, во-первых, адекватность применяемой модели, то есть насколько данная совокупность переменных x_i объясняет Y. Во-вторых, регрессионный анализ вычисляет значения коэффициентов b_i, то есть определяет степень влияния аргументов x_i на функцию Y.

Построение уравнения регрессии осуществляется, как правило, методом наименьших квадратов, суть которого состоит в минимизации суммы квадратов отклонений фактических значений от его расчетных значений. Для того чтобы регрессионный анализ, основанный на обычном методе наименьших квадратов, давал наилучшие результаты, должны выполняться следующие условия, известные как условия Гаусса-Маркова [7].

Первое условие. Математическое ожидание случайной составляющей в любом наблюдении должно быть равно нулю. Иногда случайная составляющая является положительной, иногда отрицательной, но она не должна иметь систематического смещения ни в одном из направлений. Фактически, если уравнение регрессии включает постоянный член, то это условие выполняется автоматически, так как роль константы состоит в определении любой систематической тенденции, которую не учитывают переменные, включенные в уравнение регрессии [7].

Второе условие состоит в том, что дисперсия случайной составляющей должна быть постоянна для всех наблюдений. Иногда случайная составляющая будет больше, иногда меньше, однако не должно быть причины, порождающей ошибку в одних наблюдениях большей, чем в других. Данное условие носит название гомоскедастичности (постоянство дисперсии отклонений).

Третье условие предполагает отсутствие систематической связи между значениями случайной составляющей в любых двух наблюдениях. Например, если случайная составляющая велика и положительна в одном наблюдении, это не должно приводить к тому, что она будет большой и положительной в следующем наблюдении. Случайные составляющие должны быть независимы друг от друга.

Четвертое условие состоит в том, что в регрессионной модели зависимая переменная есть величина случайная, а объясняющая переменная – неслучайная. Если это условие выполнено, то ковариация между независимой переменной и случайным членом равна нулю.

Наряду с условиями Гаусса-Маркова, в регрессионной модели, обычно, предполагается нормальность распределения случайного члена.

Качество модели регрессии связывают с адекватностью модели наблюдаемым (эмпирическим) данным.

Проверка адекватности (соответствия) модели регрессии наблюдаемым данным проводится на основе анализа остатков. Анализ остатков позволяет получить представление, насколько хорошо подобрана сама модель и насколько правильно выбран метод оценки коэффициентов. Согласно общим предположениям регрессионного анализа, остатки должны вести себя как независимые (в действительности, как почти независимые) одинаково распределенные случайные величины.

Качество модели регрессии оценивается по следующим направлениям [7]:

- проверка качества всего уравнения регрессии;
- проверка значимости всего уравнения регрессии;
- проверка статистической значимости коэффициентов уравнения регрессии;
- проверка выполнения предпосылок метода наименьших квадратов.

При анализе качества модели регрессии, в первую очередь, используется коэффициент детерминации. Коэффициент детерминации показывает долю вариации результативного признака, находящегося под воздействием изучаемых факторов, то есть определяет, какая доля вариации Y учтена в модели и обусловлена влиянием на него факторов. Чем ближе коэффициент детерминации к 1, тем выше качество модели.

Для оценки качества регрессионных моделей целесообразно также использовать коэффициент множественной корреляции (индекс корреляции). Данный коэффициент является универсальным, так как он показывает точность модели, а также может использоваться при любой форме связи между переменными.

Важным моментом является проверка значимости построенного уравнения в целом и отдельных параметров. Оценить значимость уравнения регрессии означает установить, соответствует ли математическая модель, выражающая зависимость между Y и x_i, фактическим данным и достаточно ли включенных в уравнение объясняющих переменных x_i для описания зависимой переменной Y.

Оценка значимости уравнения регрессии производится для определения возможности его практического использования. Для проверки значимости модели регрессии используется F-критерий Фишера.

Важным этапом регрессионного анализа является определение функции, характеризующей зависимость между признаками. Основанием для этого должен служить содержательный анализ природы изучаемой зависимости и ее механизма. Вместе с тем, теоретически обосновать тип связи каждого из факторов с результативным показателем можно далеко не всегда, поскольку исследуемые явления сложны и, факторы, формирующие их уровень, тесно переплетаются и взаимодействуют друг с другом. В связи с этим, на основе теоретического анализа нередко могут быть сделаны только самые общие выводы относительно характера связи, возможности ее изменения в исследуемой совокупности, правомерности использования линейной зависимости, возможного наличия экстремальных значений и тому подобное. Необходимым дополнением такого рода предположений должен быть анализ конкретных фактических данных [7].

Определенное представление о характере связи можно получить на основе эмпирической линии регрессии. Эмпирическая линия регрессии, как правило, является ломаной линией. Объясняется это влиянием неучтенных факторов, оказывающих воздействие на вариацию результативного признака. В связи с этим, эмпирической линией связи для выбора и обоснования типа теоретической кривой можно воспользоваться лишь при условии, что число наблюдений будет достаточно велико.

Одним из элементов исследований в данном направлении является сопоставление различных уравнений зависимости, основанное на применении критериев качества аппроксимации эмпирических данных конкурирующими вариантами моделей. Наиболее часто для характеристики связей показателей используют следующие типы функций:

- линейная;
- гиперболическая;
- показательная;
- параболическая;
- степенная;
- логарифмическая.

Следует заметить, что регрессию, достаточно часто, разделяют на следующие виды:

- гиперболическая (регрессия равносторонней гиперболы);
- линейная (применяется в статистике в виде четкой интерпретации параметров);
- логарифмически линейная;
- множественная (несколько независимых, объясняющих переменных);
- нелинейная (регрессия, нелинейная по оцениваемым параметрам);
- парная (только x и Y);
- обратная (регрессия, приводимая к линейному виду).

Стандарты серии ISO 27000, определяющие порядок построения Системы менеджмента информационной безопасности и оценки рисков, неразрывно связаны со стандартами качества серии ISO 9000 [8]. В связи с этим, с точки зрения применения статистических методов, целесообразно рассмотреть национальный стандарт Российской Федерации ГОСТ Р ИСО/ТО 10017-2005 «Статистические методы. Руководство по применению в соответствии с ГОСТ Р ИСО 9001». Стандарт идентичен международному стандарту ИСО/ТО 10017:2003 «Руководство по статическим методам применительно к ISO 9001:2000» (ISO/TR 10017:2003 «Guidance on statistical techniques for ISO 9001:2000»).

Одна из задач стандарта состоит в оказании содействия при выборе статистических методов, используемых при разработке, внедрении, поддержке и улучшении системы менеджмента качества согласно требованиям ISO 9001:2000. Необходимость

применения статистических методов вызвана существенной изменчивостью фактически всех процессов даже в условиях очевидной стабильности. Такая изменчивость наблюдается для количественных характеристик изделий и процессов, а также для данных, используемых на различных стадиях жизненного цикла изделий. Стандарт предоставляет организациям руководство по выбору статистических методов. Критерии определения потребности в статистических методах и пригодности выбранных методов остаются прерогативой организаций. Значительное место в стандарте уделено регрессионному анализу.

Регрессионный анализ позволяет следующее [8]:

- проверять гипотезы относительно влияния независимых переменных на отклик и использовать эту информацию для оценок изменений в отклике при заданном изменении независимой переменной;
- предсказывать значения переменной отклика при заданных значениях независимых переменных;
- предсказывать (с заданным уровнем доверия) интервал значений, в котором будет находиться ожидаемое значение отклика при заданном значении независимой переменной;
- оценивать направление и характер связи между переменной отклика и независимой переменной (хотя такая связь не означает причинную зависимость). Данная информация может использоваться для определения влияния изменения одного фактора (например, стоимость актива) на выходные характеристики процесса, в то время как другие факторы остаются постоянными.

В стандарте отмечаются следующие достоинства регрессионного анализа:

- Регрессионный анализ может обеспечить понимание соотношений между различными факторами и наблюдаемым откликом. Такое понимание может помочь в принятии решений, связанных с изучаемым процессом, и будет способствовать улучшению процесса. Регрессионный анализ позволяет в сжатом виде представлять данные отклика, сравнивать различные, но связанные наборы данных, и анализировать потенциальные отношения «причина – следствие». Регрессионный анализ позволяет оценить относительные величины влияния независимых переменных, а также относительный вклад этих переменных. Эта информация очень важна при управлении или улучшении выходных характеристик процесса [8].
- Регрессионный анализ позволяет определить характер влияния на отклик, вызванного факторами, которые или не измерены, или не исследовались при анализе. Данная информация может использоваться для совершенствования системы измерения или управления процессом. Регрессионный анализ может использоваться для прогнозирования значений переменной отклика при заданных значениях одной или более независимых переменных, а также для прогнозирования влияния изменений независимых переменных на полученный или предсказанный отклик. При решении ряда задач проведение таких исследований может быть полезно для оценки эффективности предполагаемых действий.

Ограничения и предостережения, сформулированные в стандарте

При моделировании поведения какого-либо процесса требуется навык в построении модели регрессии (линейной, многомерной и т. д.) и использовании диагностики для улучшения модели. Наличие неучтенных переменных, погрешностей измерений и других источников необъясненных вариаций отклика может усложнить моделирование. Выбор подходящего метода оценки определяется предположениями, лежащими в основе рассматриваемой регрессионной модели, и характеристиками имеющихся данных [8].

Включение или не включение в анализ единичного наблюдения или их небольшой группы может оказать влияние на оценку отклика. В связи с этим, наблюдения, влияющие на результаты, должны быть освобождены от случайных выбросов, то есть от экстремальных значений, пригодность которых для анализа должна быть исследована. Важным моментом при моделировании является минимизация количества независимых переменных. Включение ненужных переменных может скрыть влияние независимых переменных и уменьшить точность прогнозов, сделанных с помощью модели. Однако, опустив существенную независимую переменную, можно серьезно ограничить модель и снизить достоверность результатов.

Учитывая рассмотренные характеристики регрессионного анализа (в том числе, сформулированные в национальном стандарте ГОСТ Р ИСО/ТО 10017-2005), представляется целесообразным провести анализ возможности его применения для оценки рисков при нарушения информационной безопасности.

Отметим возможности, предоставляемые регрессионным анализом [9]:

1. *Возможность оценки адекватности разрабатываемой модели*. Для практического использования моделей регрессии большое значение имеет их адекватность, т. е. соответствие фактическим статистическим данным. Анализ качества эмпирического уравнения регрессии начинается с построения эмпирического уравнения регрессии, которое является начальным этапом анализа. Первое же, построенное по выборке, уравнение регрессии редко является удовлетворительным по тем или иным характеристикам. Следующим шагом является проверка качества уравнения регрессии, которая проводится по следующим направлениям [10]:

- проверка статистической значимости коэффициентов уравнения регрессии;
- проверка общего качества уравнения регрессии;
- проверка свойств данных, выполнимость которых предполагалась при оценивании уравнения.

Прежде чем проводить анализ качества уравнения регрессии, необходимо определить дисперсии и стандартные ошибки коэффициентов, а также интервальные оценки коэффициентов. Регрессионный анализ, как правило, проводится для ограниченной по объему

совокупности, поэтому параметры уравнения регрессии (показатели регрессии и корреляции), коэффициент корреляции и коэффициент детерминации могут быть искажены действием случайных факторов. Для проверки характера влияния этих показателей на всю генеральную совокупность необходимо проверить адекватность построенных статистических моделей.

При анализе адекватности уравнения регрессии (модели) исследуемому процессу, возможны следующие варианты [11]:

- построенная на основе *F*-критерия Фишера модель в целом адекватна и все коэффициенты регрессии значимы. Такая модель может быть использована для принятия решений и осуществления прогнозов;
- модель, построенная по *F*-критерию Фишера адекватна, но часть ее коэффициентов не значима (модель пригодна для принятия некоторых решений, но для прогнозов непригодна);
- модель, построенная по *F*-критерию адекватна, но все коэффициенты регрессии не значимы. Модель полностью считается неадекватной. На ее основе не принимаются решения и не осуществляются прогнозы.

Проверить значимость (качество) уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным, достаточно ли включенных в уравнение объясняющих переменных для описания зависимой переменной. Для определения качества модели по каждому наблюдению из относительных отклонений определяют среднюю ошибку аппроксимации. Проверка адекватности уравнения регрессии (модели) осуществляется с помощью средней ошибки аппроксимации, величина которой, в соответствии с рекомендациями, не должна превышать 10…12 %.

Возможность прогнозирования состояния. Прогноз получают путем подстановки в регрессионное уравнение переменных. Результат представляет собой оценку среднего значения зависимой переменной при данных факторах. Для уравнения регрессии обычно определяют доверительные интервалы, которые можно использовать в прогнозировании. Расчет доверительных интервалов позволяет определить область, в которой следует ожидать значение прогнозируемой величины. Выход этой величины за границы интервала в силу случайных колебаний имеет незначительную вероятность – меньше, чем дополнение до единицы доверительной вероятности, т. е. меньше уровня значимости [12].

Возможность восполнения пропусков показателей. Заполнение пропусков позволяет не только получить дополнительную информацию (предсказанные значения), но и сохранить уже имеющуюся, часто очень важную и полученную ценой значительных усилий информацию, за счет сохранения наблюдений, изначально содержавших пропуски [13].

Помимо очевидных достоинств, импутирование, как способ решения проблемы недостающей информации, имеет несколько недостатков, которые нельзя не учитывать:

- использование для предсказания пропусков имеющихся полных данных искажает структуру результирующих данных (после импутирования), которая смещается в сторону структуры только полных наблюдений;
- искусственная подстановка пропусков вносит в массив определенную долю искусственных данных, которые в свою очередь приводят к смещению значимости получаемых на их основе результатов.

Модели, построенные по импутированным данным, как правило, менее точны по сравнению с идеальной моделью, построенной только на полных наблюдениях. Потери в их точности будут зависеть от качества предсказания отсутствующих значений. Однако, потеряв в точности, можно выиграть в репрезентативности результатов.

В большинстве случаев, импутирование при помощи регрессионных моделей осуществляется в два этапа [12]:

1. На первом этапе по совокупности полных наблюдений строится регрессионная модель и оцениваются коэффициенты в уравнении, где в качестве зависимой переменной выступает целевая переменная, пропущенные значения по которой необходимо восстановить.

2. По полученному на предыдущем этапе уравнению, в которое подставляются известные значения независимых переменных предикторов, для каждого целевого объекта рассчитывается отсутствующее значение по зависимой целевой переменной. В случае интервальных и абсолютных переменных рассчитывается конкретное значение, а для порядковых и номинальных переменных с некоторой вероятностью предсказывается категория, к которой должен быть отнесен объект.

Выбор регрессионной модели для расчета пропущенных значений переменной определяется уровнем измерения целевой зависимой переменной, значения которой необходимо восстановить, и независимых переменных, по которым будут предсказываться отсутствующие значения.

Следует заметить, что помимо указанных выше возможностей регрессионный анализ обладает рядом недостатков. Отметим факторы, которые необходимо учитывать при использовании регрессионного анализа для оценки рисков нарушения информационной безопасности.

Проблема недостаточности одного уравнения

Для разных групп показателей влияние на окончательный результат различно [15]. Существенной проблемой в этом случае является то обстоятельство, что итоговая зависимость ищется единой для всей совокупности показателей. Иными словами, мы предполагаем, что для всех показателей характер зависимости Y от x_i единый. В том случае, когда выборочная совокупность достаточно однородна, такого рода допущение имеет под собой определенные основания.

В то же время, в случае неоднородности выборочной совокупности, единая форма уравнения сильно огрубляет реальную зависимость, качество модели

неизбежно оказывается весьма низким, а роль регрессионных коэффициентов, определяющих степень влияния x_i на Y, сводится к оценке показателя «средняя информационная безопасность системы».

По-видимому, в данном случае, для существенно различающихся между собой групп показателей целесообразно строить отдельные модели. Однако такой подход имеет свои негативные стороны.

Следовательно, при оценке рисков нарушения информационной безопасности необходимо использовать формальные критерии, позволяющие определять границы показателей, для которых действуют одинаковые, либо различные механизмы.

Различное влияние величины приращения (возмущения)

Одно из главных условий эффективного применения регрессионного анализа состоит в том, что разброс точек вокруг линии регрессии должен быть достаточно равномерен по всей протяженности линии регрессии [15]. Возможна, например, ситуация, в которой при небольших значениях величин x_i отклонения кривой от линии регрессии относительно невелики, но с увеличением значения отклонения x_i возрастают и отклонения кривой. Такие регрессионные модели называются *моделями с гетероскедастичностью возмущений*.

В некоторых случаях является эффективным (для более реального отображения действительности) использование в регрессионном уравнении логарифма от зависимой переменной. Это связано с тем, что воздействие величины прироста (либо уменьшения) на показатели зависит не только от величины прироста, но и от самого значения приращения (уменьшения).

Рассмотрим пример, характерный для оценки рисков нарушения информационной безопасности. Изменение (увеличение/уменьшение) ожидаемого уровня ущерба от возможного нарушения информационной безопасности на 100 рублей при стоимости актива некоторой организации 1 000 рублей достаточно существенно. И такое же изменение ущерба при стоимости актива 10 000 рублей мало заметно.

Если не ограничиваться визуальной констатацией нарушения требования равномерности разброса по всей протяженности линии, то можно использовать статистические тесты, которые покажут наличие или отсутствие нарушения данного ограничения. Одним из возможных тестов является тест ранговой корреляции Спирмена – непараметрический метод, который используется с целью статистического изучения связи между явлениями. В этом случае определяется фактическая степень соответствия между двумя количественными рядами изучаемых признаков и дается оценка характеристики установленной связи с помощью количественно выраженного коэффициента [15]. Коэффициент ранговой корреляции целесообразно применять при наличии небольшого количества наблюдений. Данный метод может быть использован не только для количественно выраженных данных, но также и в случаях, когда регистрируемые значения определяются описательными признаками различной интенсивности. Для обнаружения нарушения требования равномерности разброса также используют тест Голдфельда-Квандта, тест Глейзера, двусторонний критерий Фишера и другие.

Мультиколлинеарность

Классический регрессионный анализ предполагает, что величины x_i независимы между собой. В реальных приложениях это бывает достаточно редко, поскольку, как правило, между ними есть корреляция. Само по себе это является нарушением регрессионной модели и носит название *мультиколлинеарности* [15]. Основной недостаток регрессионной модели в случае мультиколлинеарности – неустойчивые значения коэффициентов модели. Это влечет за собой неточности при получении конечного результата.

Под мультиколлинеарностью понимается высокая взаимная коррелированность объясняющих переменных, которая приводит к линейной зависимости нормальных уравнений. Мультиколлинеарность может возникать в силу разных причин. Например, несколько независимых переменных могут иметь общую временную динамику, относительно которой они совершают малые колебания. Существует несколько способов для определения наличия или отсутствия мультиколлинеарности. Один из подходов заключается в анализе матрицы коэффициентов парной корреляции. Явление мультиколлинеарности в исходных данных считается установленным, если коэффициент парной корреляции между двумя переменными больше 0,8. Другой подход состоит в исследовании матрицы. Если определитель матрицы близок к нулю, то это свидетельствует о наличии мультиколлинеарности [11].

Для устранения или уменьшения мультиколлинеарности используется ряд известных методов. Наиболее распространенные в таких случаях следующие приемы: исключение одного из двух сильно связанных факторов, переход от первоначальных факторов к их главным компонентам, число которых быть может меньше, затем возвращение к первоначальным факторам.

Самый простой из них (но не всегда самый эффективный) состоит в том, что из двух объясняющих переменных, имеющих высокий коэффициент корреляции (больше 0,8), одну переменную исключают из рассмотрения. При этом какую переменную оставить, а какую удалить из анализа, решают в первую очередь на основании соображений информационной безопасности. Если с точки зрения информационной безопасности ни одной из переменных нельзя отдать предпочтение, то оставляют ту, которая имеет больший коэффициент корреляции с зависимой переменной [11].

Одним из возможных методов устранения или уменьшения мультиколлинеарности является использование стратегии шагового отбора, реализованного в ряде алгоритмов пошаговой регрессии.

Наиболее широкое применение получили следующие схемы построения уравнения множественной регрессии: метод включения факторов и метод исключения – отсев факторов из полного его набора.

В соответствии с первой схемой признак включается в уравнение в том случае, если его включение существенно увеличивает значение множественного коэффициента корреляции, что позволяет последовательно отбирать факторы, оказывающие существенное влияние на результирующий признак даже в условиях мультиколлинеарности системы признаков, отобранных в качестве аргументов из содержательных соображений. При этом первым в уравнение включается фактор, наиболее тесно коррелирующий с Y, вторым в уравнение включается тот фактор, который в паре с первым из отобранных дает максимальное значение множественного коэффициента корреляции, и т. д. Существенно, что на каждом шаге получают новое значение множественного коэффициента (большее, чем на предыдущем шаге); тем самым определяется вклад каждого отобранного фактора в объясненную дисперсию Y.

Вторая схема пошаговой регрессии основана на последовательном исключении факторов с помощью tкритерия. Она заключается в том, что после построения уравнения регрессии и оценки значимости всех коэффициентов регрессии, из модели исключают тот фактор, коэффициент при котором незначим и имеет наименьший коэффициент t. После этого получают новое уравнение множественной регрессии и снова производят оценку значимости всех оставшихся коэффициентов регрессии. Если среди них опять окажутся незначимые, то опять исключают фактор с наименьшим значением tкритерия. Процесс исключения факторов останавливается на том шаге, при котором все регрессионные коэффициенты значимы.

Ни одна их этих процедур не гарантирует получения оптимального набора переменных. Однако при практическом применении они позволяют получить достаточно хорошие наборы существенно влияющих факторов.

При отборе факторов рекомендуется пользоваться следующим правилом: число включаемых факторов обычно в 6…7 раз меньше объема совокупности, по которой строится регрессия. Если это соотношение нарушено, то число степеней свободы остаточной дисперсии очень мало. Это приводит к тому, что параметры уравнения регрессии оказываются статистически незначимыми [7].

Особым случаем мультиколлинеарности при использовании временных выборок является наличие в составе переменных линейных или нелинейных тенденций. В этом случае рекомендуется сначала выделить и исключить тенденции, а затем определить параметры регрессии по остаткам.

Игнорирование наличия тенденций в зависимой и независимой переменных ведет к завышению степени влияния независимых переменных на результирующий признак, что получило название *ложной корреляции*.

Препятствием к применению регрессии является ограниченность исходной информации, при этом наряду с указанными выше затрудняющими обстоятельствами ценность информации может снижаться за счет ее «засоренности», т. е. проявления новых обстоятельств, которые ранее не были учтены.

Резко отклоняющиеся наблюдения могут быть результатом действия большого числа сравнительно малых случайных факторов, которые в достаточно редких случаях приводят к большим отклонениям, либо это действительно случайные один или несколько выбросов, которые можно исключить как аномальные. Однако при наличии не менее трех аномальных отклонений на несколько десятков наблюдений приписывают это наличию одного или нескольких неучтенных факторов, которые проявляются только для аномальных наблюдений [11].

Наиболее распространенны в таких случаях следующие приемы: исключение одного из двух сильно связанных факторов, переход от первоначальных факторов к их главным компонентам, число которых быть может меньше, затем возвращение к первоначальным факторам.

Приведенный анализ и рекомендации указывают на тот факт, что при поверхностном рассмотрении конкретной системы (без учета всех взаимосвязей и показателей, степени и характера их влияния) велика вероятность получения неточной регрессионной модели для оценки рисков информационной безопасности. Поэтому, универсальные программные продукты, используемые для оценки рисков нарушения информационной безопасности, предположительно могут иметь много неточностей, которые будут негативно отражаться на конечном результате. Создание максимально точных регрессионных моделей невозможно без учета замечаний, рассмотренных в данной статье.

Список литературы

1. Системы менеджмента информационной безопасности. *Электронный ресурс:* http://www.ot.ru/facilities_infsecurity_5.html
2. Астахов А. Как управлять рисками информационной безопасности? *Электронныйресурс:* http://iso27000.ru/chitalnyi-zai/upravlenie-riskami-informacionnoi-bezopasnosti/kak-upravlyat-riskami-informacionnoi-bezopasnosti/
3. Куканова Н. Практические аспекты применения международного стандарта безопасности информационных систем ISO 27001:2005. *Электронный ресурс:* http://www.dsec.ru/about/articles/practice_iso_27001/
4. Лопарев С., Шелупанов А. Анализ инструментальных средств оценки рисков утечки информации в компьютерной сети предприятия. *Электронный ресурс:* http://www.iso27000.ru/chitalnyi-zai/upravlenie-riskami-informacionnoi-bezopasnosti/analiz-instrumentalnyh-sredstv-ocenki-riskov-utechki-informacii-v-kompyuternoi-seti-predpriyatiya
5. Поликарпов А.К. Обзор существующих методов оценки рисков и управления информационной безопасностью. *Электронный ресурс:* http://is.isa.ru/PolikOtc.html
6. Сайт компании RiskWatch. *Электронный ресурс:* http://www.riskwatch.com
7. Орлова И.В., Половников В.А. *Экономико-математические методы и модели: компьютерное моделирование. Учебное пособие*. М.: Вузовский учебник, 2007. 365 с.

8. Национальный стандарт Российской Федерации ГОСТ Р ИСО/ТО 10017-2005 «Статистические методы. Руководство по применению в соответствии с ГОСТ Р ИСО 9001». *Электронный ресурс:* http://tehnorma.ru/normativbase/4/4787/index.htm
9. Радченко С.Г. *Методология регрессионного анализа: Монография*. К.: Издательство «Корнийчук», 2011. 45 с.
10. Дрейпер Н., Смит Г. *Прикладной регрессионный анализ. Множественная регрессия*. М.: Издательство «Диалектика», 2007. 35 с.
11. Орлов А.И. *Эконометрика. Учебник*. М.: Издательство «Экзамен», 2008. 66 с.
12. Паклин Н. Б., Орешков В.И. *Бизнес-аналитика: от данных к знаниям*. СПб.: Издательство «Питер». 2009. 25 с.
13. Радченко С.Г. *Устойчивые методы оценивания статистических моделей*: *Монография.* К.: Издательство «Санспарель». 2005. 18 с.
14. Крыштановский А.О. Ограничения метода регрессионного анализа. *Электронный ресурс:* http://socioline.ru/pages/ao-kryshtanovskij-ogranichenia-metoda-regressionnogo-analiza.
15. Глинский В.В., Иония В.Г. *Статистический анализ*. М.: Издательство «Филинъ». 1998. 28 с.

References

1. Sistemy menedzhmenta informatsionnoy bezopasnosti [Information security management systems]. *Available at:* http://www.ot.ru/facilities_infsecurity_5.html
2. Astakhov A. Kak upravlyat riskami informatsionnoy bezopasnosti? [How to manage information security risks?]. *Available at:* http://iso27000.ru/chitalnyi-zai/upravlenie-riskami-informacionnoi-bezopasnosti/kak-upravlyat-riskami-informacionnoi-bezopasnosti/
3. Kukanova N. Prakticheskie aspekty primeneniya mezhdunarodnogo standarta bezopasnosti informatsionnykh ystem ISO 27001:2005 [Practical aspects of international information systems security standard ISO 27001:2005]. *Available at:* http://www.dsec.ru/about/articles/practice_iso_27001/
4. Loparev S., Shelupanov A. Analiz instrumentalnykh sredstv otsenki riskov utechki informatsii v kompyuternoy seti predpriyatiya [Analysis of the tools of risk assessment information leakage in computer network]. *Available at:* http://www.iso27000.ru/chitalnyi-zai/upravlenie-riskami-informacionnoi-bezopasnosti/analiz-instrumentalnyh-sredstv-ocenki-riskov-utechki-informacii-v-kompyuternoi-seti-predpriyatiya
5. Polikarpov A.K. Obzor sushchestvuyushchikh metodov otsenki riskov i upravleniya informatsionnoy bezopasnostyu [A review of existing methods of risk assessment and management of information security]. *Available at:* http://is.isa.ru/PolikOtc.html
6. Saytkompanii RiskWatch [Web-site of RiskWatch]. *Available at:* http://www.riskwatch.com
7. Orlova I.V., Polovnikov V.A. *Ekonomiko-matematicheskie metody i modeli: kompyuternoe modelirovanie. Uchebnoe posobie* [Economic-mathematical methods and models: computer simulation. Textbook]. M.: Vuzovskiy uchebnik [Moscow: High school textbook]. 2007. 365 p.
8. Natsionalnyy standart Rossiyskoy Federatsii GOST R ISO/TO 10017-2005 «Statisticheskie metody. Rukovodstvo po primeneniyu v sootvetstvii s GOST R ISO 9001» [National Standard of the Russian Federation GOST R ISO / TR 10017-2005 «Statistical methods. Guidance on the application in accordance with ISO 9001»]. *Available at:* http://tehnorma.ru/normativbase/4/4787/index.htm
9. Radchenko S.G. *Metodologiya regressionnogo analiza: Monografiya* [The methodology of regression analysis: Monograph]. K.: Izdatelstvo «Korniychuk»[Kiev: Publishing House «Korniichuk»]. 2011. 45 c.
10. Dreyper N., Smit G. *Prikladnoy regressionnyy analiz. Mnozhestvennaya regressiya* [Applied Regression Analysis. Multiple regression]. M.: Izdatelstvo «Dialektika» [Moscow: Publishing House «Dialectic»]. 2007. 35 p.
11. Orlov A.I. *Ekonometrika. Uchebnik* [Econometrics. Textbook]. M.: Izdatelstvo «Ekzamen» [Moscow: Publishing House «Exam»]. 2008. 66 p.
12. Paklin N. B., Oreshkov V.I. *Biznes-analitika: ot dannykh k znaniyam* [Business Intelligence: from data to knowledge]. SPb.: Izdatelstvo «Priter» [St. Petersburg.: Publishing «Priter»]. 2009. 25 p.
13. Radchenko S.G. *Ustoychivye metody otsenivaniya statisticheskikh modeley: Monografiya* [Sustainable methods of estimation of statistical models: Monograph]. K.: Izdatelstvo «Sansparel» [Kiev: Publishing House «Sansparel»]. 2005. 18 p.
14. Kryshtanovskiy A.O. Ogranicheniya metoda regressionnogo analiza [Limitations of regression analysis]. *Available at:* http://socioline.ru/pages/ao-kryshtanovskij-ogranichenia-metoda-regressionnogo-analiza
15. Glinskiy V.V., Ioniya V.G. *Statisticheskiy analiz* [Statistical analysis]. M.: Izdatelstvo «Filin» [Moscow: Publishing House «Filin»]. 1998. 28 p.

Информация об авторах

Лось Алексей Борисович, канд. техн. наук, доцент
E-mail: alos@hse.ru
Кабанов Артем Сергеевич, канд. техн. наук, доцент
E-mail: kabanov_as@mail.ru
Московский институт электроники и математики
Национального исследовательского университета «Высшая школа экономики»
109028, Москва, Российская Федерация, Трехсвятительский пер., дом 3

Information about the authors

los Aleksey Borisovich, Cand. of Techn. Sciences, Associate Professor
E-mail: alos@hse.ru
Kabanov Artem Sergeevich, Cand. of Techn. Sciences, Associate Professor
E-mail: kabanov_as@mail.ru
Moscow Institute of Electronics and Mathematics
National Research University «Higher School of Economics»
109028, Moscow, Russian Federation, Trehsvyatitelsky per., 3