НЕКОТОРАЯ ОЦЕНКА ПРОГНОЗНОГО ЗНАЧЕНИЯ РИСКА ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ РЕГРЕССИОННОГО АНАЛИЗА

Козлов А.Д., Нога Н.Л.

Институт проблем управления им. В.А. Трапезникова РАН,
Россия, г. Москва, ул. Профсоюзная, д.65
alkozlov@ipu.ru, noga@ipu.ru

Аннотация: Предложена методика определения прогнозного значения риска и его доверительного интервала методами регрессионного анализа в условиях зависимости риска от различных факторов: ценность информационных ресурсов, уровень угроз, наличие уязвимостей, потенциальный ущерб, уровень затрат на создание и эксплуатацию системы, уровень контроля информационных ресурсов.

Ключевые слова: информационная безопасность, множественная регрессия, угроза, риск, нечеткая логика, доверительный интервал, взаимная корреляция.

Введение

Современная экономика не мыслима без обработки огромных массивов данных. Информация является как средством достижения целей в экономике, так и сама становится своего рода товаром, а ее предоставление – востребованным видом услуг.

Информационные системы сплетаются в причудливые, сложные сети, окружают практически каждого человека, внедряются в деятельность предприятий и корпораций. Но помогая в одном, облегчая решение отдельных задач и достижение поставленных целей, они, в свою очередь, привносят новые вызовы и угрозы. Эти вызовы и угрозы в первую очередь связаны с вопросами информационной безопасности, с возможностями несанкционированного доступа к информационным ресурсам, утечками данных, нарушениями нормального функционирования систем.

При этом необходимо понимать, что эти угрозы могут быть как внешние, так и внутренние. Так по оценке группы компаний InfoWatch [1-2] количество утечек с некоторыми колебаниями постоянно возрастает. Возрастает и объем этих утечек. Так за 9 месяцев 2020 года было скомпрометировано порядка 9,9 млрд. записей, а число инцидентов с утечкой более 1 млн. записей возросло по сравнению с 2019 годом с 6 до 15.

За период с 2013 по 2019 годы доля внутренних утечек (по вине внутреннего нарушителя) колебалась от 70% до 53%, но все время была выше доли внешних нарушителей. Внутренние нарушители это, как правило, сотрудники, имеющие легитимный доступ к данным. Утечки происходят или по халатности, или умышленно. Причем, если утечка персональных данных или финансовой информации чаще всего происходит по халатности или не досмотру, то утечки, связанные с высокотехнологической информацией, в основном, происходят умышленно.

Все вышеперечисленные факты необходимо учитывать при решении задачи правильной оценки рисков на всех этапах жизнедеятельности информационных систем, без которой выполнить требования по ускорению обработки огромных массивов данных в сложных сетях и при этом соблюсти требования по обеспечению доступности, целостности и конфиденциальности этих данных невозможно. Оценивать риск необходимо не только в начале какого-либо проекта, но и прогнозировать возможное его изменение в процессе реализации проекта.

Важность учета влияния на риск угроз от внутреннего нарушителя продиктована, в том числе, и все большим распространением использования облачных технологий. В этом случае внутренний нарушитель у провайдера облачных сервисов получает доступ к информационным ресурсам компании, которая использует эти сервисы, а сама компания частично утрачивает контроль над своими информационными ресурсами.

Необходимо отметить, что регрессионный анализ, как используемый в эконометрике метод для получения оценки уравнения (эконометрической модели), которое наилучшим способом соответствует множеству наблюдений зависимых и независимых переменных, дает возможность получать наилучшую оценку взаимосвязи между этими переменными.

Для правильной оценки риска информационной безопасности необходимо знать, как связаны между собой зависимые и независимые переменные. При этом сформулировать задачу по количественному измерению связей межу различными информационными, экономическими процессами и явлениями на основе данных, полученных методами нечеткой логики, и при помощи

статистических методов соответствующим образом интерпретировать и использовать получаемые результаты, в частности для вычисления оценок прогнозных значений рисков информационной безопасности.

Сегодня эконометрические методы применяются в качестве стандартных в различных отраслях прикладной экономики, изучающей все, начиная от расходов домашних хозяйств и предпринимательских инвестиций, заканчивая организацией производств, рынков труда и эффектами государственной политики.

В работах [3-4] авторами было предложено в силу наличия факторов неопределенности вычислять риски с помощью методов нечеткой логики. Надо отметить, что это позволяет определять риск информационной безопасности и его зависимость от различных параметров. Но этот метод не дает ответа на вопросы коррелированности параметров модели, возможной их избыточности, качества самой модели, а также границ нахождения прогнозных значений рисков. Использование же методов эконометрики дает возможность ответить на эти вопросы.

1 Краткий обзор методик оценки риска информационной безопасности

На текущий момент существует множество методик, основанных на оценке, анализе и управлении рисками информационной безопасности [5-6]. В понятие оценки риска вкладывается оценка угроз, уязвимостей, через которые реализуются угрозы, ущерб, возникающий при их реализации. Под анализом и управлением рисками понимается построение моделей, отображающих ситуации появления неблагоприятных условий с учетом различных факторов, которые характеризуют эти риски и принятие решений с целью уменьшения ущерба, наносимого в результате реализации угроз или в результате принятия мер по предотвращению различных атак на информационную систему. Чтобы получать более-менее точные оценки риска для рассматриваемых информационных систем необходимо строить модель угроз и нарушителя, используя данные из публикуемых банков данных угроз и уязвимостей [7-8].

Рассмотрим кратко недостатки и положительные стороны следующих методологий [5-6].

CORAS и OCTAVE:

- приводятся только качественные оценки рисков;
- оценки рисков и их обновление выполняются в CORAS единоразово, в OCTAVE регулярно;
- CORAS свободно распространяемый программный продукт, не требующий значимых ресурсов по использованию;
- пригодны для работы в частном облаке;
- в OCTAVE для снижения рисков реализованы только способы снижения и принятия затрат;

CRAMM и ГРИФ:

- приводятся как качественные, так и количественные оценки рисков;
- периодически выполняются оценки рисков и их обновление;
- для снижения рисков в CRAMM реализован только способ снижения, в ГРИФ - обход, снижение и принятие;
- приводится идентификация различных элементов риска: активов, угроз и т.п.;
- у CRAMM высокая стоимость;
- в ГРИФ учитывается описание бизнес-процессов организации;
- необходимость внесения поправок в оценки в случае использования публичного облака или использование оценок провайдера услуг.

RiskWatch:

- дается количественная оценка общего объема потерь за указанный период и мера возврата от инвестиций во внедрение средств защиты информации;
- периодически выполняются оценки рисков и их обновление;
- наличие удобных шаблонов, а также баз с данными о частотах реализации угроз;
- анализ рисков реализуется только на программно-техническом уровне без учета организационных мероприятий;
- высокая стоимость.

В работах [3-4] авторами была предложена методология по оценке рисков на основе методов нечеткой логики с учетом субъективных факторов риска, реализующая процедуру оценки рисков в условиях неопределенности. Для реализации методологии использовалась опция Fuzzy Logic Toolbox системы Matlab [9].

2 Задача оценки риска информационной безопасности

В работе [4] авторы сформулировали задачу оценки риска, используя методы нечеткой логики, с учетом влияния особенностей использования облачных технологий и затрат на создание информационных систем следующим образом.

Вводятся в рассмотрение пять лингвистических переменных: ценность активов, вероятность угрозы, объем ущерба, уровень контроля информационных ресурсов, затраты на создание и эксплуатацию системы. Пусть совокупность входных параметров A_j (j = 1,…,J) соответствует ценности активов, совокупность входных параметров P_i (i = 1,…,I) соответствует уровням угроз. Совокупность входных параметров D_l (l = 1,…,L) соответствует уровням ущерба; совокупность входных параметров уровня контроля информационных ресурсов – Km (m = 1,…,M), а совокупность входных параметров по затратам на создание и эксплуатацию системы – Z_k (k = 1,…,K). Значения всех этих параметров могут принимать как количественные значения в промежутке [0;1], так и качественные значения, например: низкий, средний, высокий, критический.

В данной постановке задачи требуется найти выходной параметр G, зависящий от перечисленных входных параметров, который и определяет уровень риска.

Границы термов задаются экспертным путем, например, для уровня контроля информационных ресурсов, либо для затрат, могут рассматриваться конкретные варианты построения систем.

Таким образом, авторами предлагалось уровень риска представлять в виде функции $G = G$ (ценность активов, уровень угрозы, уровень ущерба, уровень контроля информационных ресурсов, затраты на создание и эксплуатацию системы) или

$$G = G(A, P, D, K, Z), \quad (1)$$

где A – ценность активов, P - вероятность реализации угрозы через заданную уязвимость, D - величина значения ущерба от реализации данной угрозы, K - уровень контроля информационных ресурсов, Z – затраты.

Следует отметить, что вычислять вероятность реализации угрозы крайне сложно. По этой причине было предложено использовать методы нечеткой логики для вычисления риска.

В отличие от представленных выше методов предложенный метод позволяет осуществлять многофакторную оценку риска. При этом взаимовлияние различных параметров (факторов) может быть неочевидным. Точность оценки зависит от того, насколько качественно и подробно будут написаны продукционные правила, насколько подробно описаны термы для каждого параметра и правильно указаны диапазоны их значений.

Необходимо отметить, что данный метод имеет и ряд недостатков. Основной недостаток, по мнению авторов, заключается в том, что выбор пяти указанных параметров производился исходя из общих рассуждений о зависимости риска от этих параметров. При этом, например, нет ясности в вопросе взаимной корреляции указанных параметров и, соответственно, возможной их избыточности. В связи с этим авторы предлагают исследовать зависимость (1) методами регрессионного анализа, в частности, представив уравнение (1) в виде уравнения множественной регрессии в следующей спецификации:

$$G = a_0 + n_1 x_A + n_2 x_P + n_3 x_D + n_4 \frac{1}{x_K} + n_5 \frac{1}{x_Z} + \varepsilon, \quad (2)$$

где x_A - ценность активов; x_P - вероятность реализации угрозы, опосредовано связанная с уровнем опасности уязвимости; x_D - величина ущерба от реализации угрозы; x_K - уровень контроля информационных ресурсов; x_Z - затраты на создание и эксплуатацию системы; ε - возмущение, включающее в себя влияние неучтенных в данной модели факторов, случайные ошибки и особенности измерений. Также принимаем, что при повышении уровня контроля информационных ресурсов и увеличении затрат на создание и эксплуатацию системы риск уменьшается.

Уравнение (2) приводится к линейному виду путем замены: $x_A = y_1, x_P = y_2, x_D = y_3$, $\frac{1}{x_K} = y_4$ и $\frac{1}{x_Z} = y_5$. Получаем уравнение множественной регрессии в линейном виде:

$$G = a_0 + n_1 y_1 + n_2 y_2 + n_3 y_3 + n_4 y_4 + n_5 y_5 + \varepsilon. \quad (3)$$

Теперь к этому уравнению для определения коэффициентов $a_0, n_i, i = 1,\ldots,5$ применяем метод наименьших квадратов (МНК). При этом строится система уравнений (4), решая которую получаем оценки коэффициентов регрессии.

$$\begin{cases} \sum G = ka_0 + n_1 \sum y_1 + n_2 \sum y_2 + n_3 \sum y_3 + n_4 \sum y_4 + n_5 \sum y_5 \\ \sum G y_1 = a_0 \sum y_1 + n_2 \sum y_1 y_2 + n_3 \sum y_1 y_3 + n_4 \sum y_1 y_4 + n_5 \sum y_1 y_5 + n_1 \sum y_1^2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \sum G y_5 = a_0 \sum y_5 + n_1 \sum y_1 y_5 + n_2 \sum y_2 y_5 + n_3 \sum y_3 y_5 + n_4 \sum y_4 y_5 + n_5 \sum y_5^2 \end{cases}. \quad (4)$$

где k – объем исследуемой совокупности, а коэффициенты $a_0, n_i, i = 1,\dots,5$ вычисляются методом определителей [8].

При этом необходимо отметить, что полученные коэффициенты несравнимы. Для того чтобы можно было сравнивать коэффициенты уравнения регрессии и при этом ранжировать параметры по степени влияния на риск, можно воспользоваться матрицей парных коэффициентов корреляции и построить уравнение регрессии в стандартизованном масштабе [8]:

$$t_G = l_1 t_{y_1} + l_2 t_{y_2} + l_3 t_{y_3} + l_4 t_{y_4} + l_5 t_{y_5}, \quad (5)$$

где $l_i, i = 1,\dots,5$, – стандартизованные коэффициенты, а $t_G, t_{y_i}, i = 1,\dots,5$ - стандартизованные переменные, такие что

$$t_G = \frac{G-\bar{G}}{\sigma_G}, t_{y_i} = \frac{y_i - \bar{y}_i}{\sigma_{y_i}}, \bar{t}_G = \bar{t}_{y_i} = 0, \sigma_{t_G} = \sigma_{t_{y_i}} = 1. \quad (6)$$

Теперь, если будем применять МНК к уравнению (5) можно получить следующую систему уравнений, решая которую получим стандартизованные коэффициенты для (5):

$$\begin{cases} R_{Gy_1} = l_1 + l_2 R_{y_2 y_1} + l_3 R_{y_3 y_1} + l_4 R_{y_4 y_1} + l_5 R_{y_5 y_1} \\ R_{Gy_2} = l_1 R_{y_2 y_1} + l_2 + l_3 R_{y_3 y_2} + l_4 R_{y_4 y_2} + l_5 R_{y_5 y_2} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ R_{Gy_5} = l_1 R_{y_5 y_1} + l_2 R_{y_5 y_2} + l_3 R_{y_5 y_3} + l_4 R_{y_5 y_4} + l_5 \end{cases},$$

где $R_{y_i y_j}$ - парные коэффициенты корреляции.

Известно [10], что стандартизованные коэффициенты в уравнении регрессии показывают, на сколько единиц в среднем изменится, в данном случае риск, если, например, параметр y_i изменится на единицу при неизменном уровне остальных параметров. Из (6) следует, что переменные - центрированные и нормированные, т.е. стандартизованные коэффициенты сравнимы между собой, а значит при сравнении можно отранжировать параметры по их воздействию на риск. Это ранжирование позволяет определить избыточные параметры, после чего просто исключить из уравнения параметры с наименьшими значениями l_i.

Известно также [10], что коэффициенты регрессии из уравнения (3) связаны с коэффициентами из уравнения (5) следующим образом:

$$n_i = l_i \frac{\sigma_G}{\sigma_{y_i}}. \quad (7)$$

Таким образом, можно от уравнения (5) перейти к исходному уравнению регрессии (3).

Чтобы теперь получить оценку прогнозного значения риска рассмотрим следующие обозначения:

$$Y = \begin{pmatrix} 1 & y_{11} & \dots & y_{15} \\ 1 & y_{21} & \dots & y_{25} \\ \dots & \dots & \dots & \dots \\ 1 & y_{k1} & \dots & y_{k5} \end{pmatrix}, Y' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ y_{11} & y_{21} & \dots & y_{k1} \\ \dots & \dots & \dots & \dots \\ y_{15} & y_{25} & \dots & y_{k5} \end{pmatrix}, n = \begin{pmatrix} a_0 \\ n_1 \\ \dots \\ n_5 \end{pmatrix}, G = \begin{pmatrix} g_1 \\ g_2 \\ \dots \\ g_k \end{pmatrix},$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_k \end{pmatrix}.$$

Тогда уравнение (3) можно переписать следующим образом:

$$G = Yn + \varepsilon. \quad (8)$$

И пусть вектор $\boldsymbol{y}_p = (1, y_{1p}, y_{2p}, y_{3p}, y_{4p}, y_{5p})$ составлен из значений объясняющих переменных, для которых вычисляется прогнозное значение риска G. Тогда точечное прогнозное значение риска вычисляется по уравнению регрессии в (9) следующим образом:

$$\hat{G}_p = a_0 + n_1 y_{1p} + n_2 y_{2p} + \dots + n_5 y_{5p} = y_p n. \quad (9)$$

Чтобы найти доверительный интервал для прогнозного значения риска, необходимо вычислить стандартную ошибку прогнозного значения риска следующим образом [8]:

$$m_{\hat{G}_p} = S_{\text{ост}}\sqrt{1 + y_p(Y'Y)^{-1}y_p'},$$

где $S_{ост}$ – корень квадратный из остаточной дисперсии.

Тогда доверительный интервал для прогнозного значения риска определяется следующим образом:

$$\hat{G}_p - m_{\hat{G}_p}t_{\text{табл}} < G_p < \hat{G}_p + m_{\hat{G}_p}t_{\text{табл}}, \quad (10)$$

где $t_{табл}$ - критическое значение при заданном уровне значимости по таблице распределения Стьюдента, соответствующее k = 31-1-5 = 25 степеням свободы.

Примерный алгоритм получения оценки прогнозного значения риска, основанный на положениях регрессионного анализа и, в частности, множественной регрессии, используя при этом данные из совокупности, получаемой с помощью продукционных правил в работе [3], предлагается реализовать следующим образом.

Формулируются продукционные правила для пяти входных переменных: ценность актива, уровень реализации угрозы, связанный с уровнем опасности уязвимости, уровень потенциального ущерба, уровень контроля информационных ресурсов, уровень затрат на создание и эксплуатацию информационной системы.

Каждая из перечисленных входных переменных оценивается по своей шкале [3] как на качественном уровне, так и, соответственно, в количественном виде в некотором интервале. Например, входная переменная – уровень угроз, предварительно отобранных из БД угроз [7] и отфильтрованных с помощью деревьев атак. Далее в таблице продукционных правил качественные значения переменных заменяются на соответствующие им усредненные количественные значения. Таким образом, мы получаем совокупность данных для исследования методами регрессионного анализа. Т.е. строим уравнение множественной регрессии с пятью объясняющими переменными, с помощью МНК находим коэффициенты уравнения регрессии, исследуем переменные на зависимость. Возможно исключение некоторых переменных из рассмотрения. Получаем точечную оценку прогнозного значения риска, а также доверительный интервал для прогнозного значения риска информационной безопасности.

В качестве примера рассмотрим Таблицу 1 данных, полученных из таблиц с диапазонами изменений термов в [3] с заменой качественных значений на соответствующие усредненные количественные значения.

Таблица 1. Соответствие качественных значений усредненным количественным значениям переменных

Переменная	**Качественные значения**	**Границы терма**	**Усредненные значения**
Риск, G	Незначительный	0 – 0,21	0,1
	Допустимый	0,16 – 0,41	0,3
	Высокий	0,35 – 0,65	0,5
	Критический	0,60 – 1,0	0,8
Уровень контроля, y_4	Полный	0,95-1,00	0,97
	Высокий	0,60-0,90	0,7
	Средний	0,40-0,75	0,5
	Низкий	0,10-0,50	0,25
Уровень затрат, y_5	Низкий	0-0,30	0,15
	Средний	0,25-0,60	0,42
	Высокий	0,55-0,80	0,55
	Значительный	0,75-1,00	0,75
Ценность актива, y_1	Пренебрежимо малая	0-0,20	0,1
	Низкая	0,21-0,40	0,3
	Средняя	0,41-0,60	0,5
	Высокая	0,61-0,80	0,7
	Критически высокая	0,81-1,00	0,9
Уровень угрозы, y_2	…	…	…
Уровень воздействия, y_3	…	…	…

Используя усредненные значения из Таблицы 1 в продукционных правилах, приведенных в работе [3], получаем нечеткую базу знаний в виде Таблицы 2. Для нашего примера достаточно взять каждую восьмую строку из 240, начиная с первой строки, представленные в Таблице 2, так как, чтобы построить уравнения регрессии с пятью переменными достаточно объема исследуемой совокупности в размере n = 31. При необходимости повышения точности расчетов можно использовать все продукционные правила из работы [3].

Таблица 2. Нечеткая база знаний, продукционные правила

	Риск, G	**Ценность актива, y_1**	**Уровень угрозы, y_2**	**Уровень воздействия, y_3**	**Уровень контроля, y_4**	**Уровень затрат, y_5**
1	0,1	0,1	0,15	0,1	0,97	0,42
2	0,1	0,1	0,15	0,1	0,75	0,15
…						
240	0,8	0,9	0,8	0,9	0,25	0,42

Таким образом, получим данные, расположенные в Таблице 3, необходимые для построения множественной регрессии.

При вводе данных из Таблицы 2 в систему MATLAB находим значения стандартизованных коэффициентов и уравнение (5) принимает следующий вид:

$$t_G = 0{,}3361t_{y_1} + 0{,}4381t_{y_2} + 0{,}6289t_{y_3} - 0{,}0071t_{y_4} + 0{,}1835t_{y_5} \qquad (11)$$

Т.к. стандартизованные коэффициенты можно сравнивать между собой, то ранжируя коэффициенты перед переменными из (11) в порядке возрастания, можно определить, какую из объясняющих переменных можно исключить из рассмотрения. При этом исключаем из рассмотрения переменную с наименьшим коэффициентом.

Из полученного результата следует, что наименьшее влияние на общее значение риска оказывает уровень контроля собственных информационных ресурсов, и его можно исключить из рассмотрения.

Тем не менее в нашем примере при дальнейших вычислениях будем рассматривать все пять вышеперечисленных переменных.

Таблица 3. Совокупность данных

	G	**y1**	**y2**	**y3**	**y4**	**y5**		**G**	**y1**	**y2**	**y3**	**y4**	**y5**
1	0,1	0,1	0,15	0,1	0,97	0,42	128	0,3	0,5	0,48	0,65	0,25	0,15
8	0,3	0,1	0,48	0,1	0,25	0,15	136	0,5	0,5	0,15	0,9	0,25	0,15
16	0,3	0,1	0,48	0,35	0,25	0,15	144	0,8	0,5	0,8	0,9	0,25	0,15
24	0,5	0,1	0,8	0,35	0,45	0,25	152	0,3	0,7	0,48	0,1	0,45	0,42
32	0,3	0,1	0,48	0,65	0,25	0,15	160	0,3	0,7	0,15	0,35	0,25	0,42
40	0,5	0,1	0,15	0,9	0,25	0,15	168	0,5	0,7	0,8	0,35	0,25	0,42
48	0,5	0,1	0,8	0,9	0,25	0,15	176	0,5	0,7	0,48	0,65	0,25	0,55
56	0,3	0,3	0,48	0,1	0,45	0,25	184	0,5	0,7	0,15	0,9	0,25	0,42
64	0,3	0,3	0,15	0,35	0,25	0,15	192	0,8	0,7	0,8	0,9	0,45	0,42
72	0,3	0,3	0,8	0,35	0,25	0,15	200	0,5	0,9	0,48	0,1	0,25	0,42
80	0,3	0,3	0,48	0,65	0,25	0,15	208	0,5	0,9	0,15	0,35	0,25	0,55
88	0,5	0,3	0,15	0,9	0,45	0,25	216	0,5	0,9	0,8	0,35	0,25	0,42
96	0,5	0,3	0,8	0,9	0,25	0,15	224	0,5	0,9	0,48	0,65	0,7	0,42
104	0,3	0,5	0,48	0,1	0,25	0,15	232	0,5	0,9	0,15	0,9	0,25	0,42
112	0,3	0,5	0,15	0,35	0,25	0,15	240	0,8	0,9	0,8	0,9	0,25	0,55
120	0,5	0,5	0,8	0,35	0,45	0,25							

Чтобы оценить совместное влияние представленных пяти параметров на значение риска необходимо вычислить по данным из Таблицы 2 коэффициент множественной корреляции и коэффициент множественной детерминации.

$$R_{Gy_1y_2y_3y_4y_5} = \sqrt{1 - \frac{\sum(G - \hat{G})^2}{\sum(G - \bar{G})^2}} = 0{,}8439,\ R^2_{Gy_1y_2y_3y_4y_5} = 0{,}7122.$$

$\overline{G}$ - средний уровень риска.

Последний коэффициент показывает, что 71,2% изменения риска объясняется изменением параметров, включенных в уравнение регрессии, и дает оценку качества построенной модели. Кроме того, с помощью F-критерия Фишера и t-критерия Стьюдента можно убедиться в статистической значимости полученного уравнения регрессии и коэффициентов регрессии соответственно при уровне значимости $\alpha = 0{,}05$. Вопрос мультиколлинеарности снимается, поскольку максимальное значение парного коэффициента корреляции

$$\max_{i \neq j} \left| r_{y_i y_j} \right| = r_{y_4 y_5} = 0{,}502 < 0{,}7.$$

Пусть теперь $y_p = (0{,}3; 0{,}4; 0{,}8; 0{,}25; 0{,}4)$. Найдем доверительный интервал для прогнозного значения риска G_p. При указанных значениях параметров $\hat{G}_p = 0{,}4977$, а

$$S_{\text{ост}} = \sqrt{\frac{\sum (G - \hat{G})^2}{31 - 5 - 1}} = 0{,}0918.$$

Тогда стандартная ошибка $m_{G_p} \approx 0{,}167$ и доверительный интервал для прогнозного значения риска G_p из (10) при заданном уровне значимости $\alpha = 0{,}05$ будет следующий:

$$0{,}4977 - 0{,}167 \cdot 2{,}0595 < G_p < 0{,}4977 + 0{,}167 \cdot 2{,}0595 \Rightarrow 0{,}1538 < G_p < 0{,}8416.$$

Необходимо отметить, что если из рассмотрения была исключена какая-либо переменная, то можно вновь построить уравнение регрессии с меньшим количеством переменных и получить более точное прогнозное значение риска.

Анализируя полученный результат и сравнивая его с диапазоном значений риска, приведенного в Таблице 1, получаем, что в рассматриваемых условиях нашего примера будет допущен высокий риск. Из этого следует, что настоятельно требуется пересмотр исходных допущенных условий.

В случае изменений условий необходимо повторно произвести расчет прогнозируемого значения риска и его доверительного интервала. Таким образом, процесс оценки риска будет производиться за несколько итераций.

Из (11) следует, что чем больше значение коэффициентов у переменных, тем большее влияние на значение риска оказывает данный параметр. Изменения условий должны быть направлены на уменьшение, как прогнозируемого значения риска, так и доверительного интервала.

Заключение

Предлагаемая методика, основанная на совместном использовании методов нечеткой логики и регрессионного анализа, позволяет при оценке риска учитывать такие параметры как ценность активов, уровни угроз, опасности имеющихся уязвимостей, возможный ущерб от реализации угроз, уровень затрат на создание и эксплуатацию системы. Кроме влияния на риск перечисленных выше параметров, методика позволяет также оценить риски, связанные с частичной утратой контроля за собственными информационными ресурсами, возникающие при использовании облачных структур, и с рядом других субъективных факторов. Используя данную методику, можно как дополнять модель новыми факторами (параметрами), так и исключать факторы, которые не оказывают значительного влияния на риск информационной безопасности. Таким образом, становится возможным, используя аппарат регрессионного анализа:

- определять степень влияния различных параметров на уровень риска информационной безопасности;
- исключать из модели параметры, незначительно влияющие на уровень риска информационной безопасности;
- включать в модель дополнительные параметры, значительно влияющие на уровень риска информационной безопасности;
- вычислять прогнозное значение риска информационной безопасности, как точечное, так и в пределах доверительного интервала;
- оптимизировать расходы на эксплуатацию корпоративных информационных систем, включая системы со сложными сетевыми структурами, с возможностью контролировать в полной мере собственные информационные ресурсы.

Литература

1. Утечки информации ограниченного доступа: отчет за 9 месяцев 2020 г. URL: https://www.infowatch.ru/form-modal/report-download/30708
2. Утечки данных организаций по вине внутреннего нарушителя. Сравнительное исследование. 2013-2019 гг. URL: https://www.infowatch.ru/form-modal/report-download/24339
3. *Kozlov A., Noga N.* Some Method of Complex Structures Information Security Risk Assessment in Conditions of Uncertainty / Proceedings of the 13th International Conference "Management of Large-Scale System Development" (MLSD). M.: IEEE, 2020. P. 1-5, https://ieeexplore.ieee.org/document/9247662.
4. *Козлов А.Д., Нога Н.Л.* Риски информационной безопасности корпоративных информационных систем при использовании облачных технологий // Управление риском, 2019. - №3. – С. 31-46.
5. *Разумников С.В.* Анализ возможности применения методов OCTAVE, RiskWatch, CRAMM для оценки рисков ИТ для облачных сервисов // Современные проблемы науки и образования, 2014. - № 1. - С. 247-248.
6. *Баранова С.Ю.* Методики анализа и оценки рисков информационной безопасности // Вестник Московского университета им. С.Ю. Витте. Серия 3. Образовательные ресурсы и технологии, 2015. – № 1(9). – С. 73-79.
7. Банк данных угроз информационной безопасности. Список уязвимостей. URL: http://www.bdu.fstec.ru/vul
8. Банк данных угроз информационной безопасности. Список угроз. URL: http://www.bdu.fstec.ru/threat
9. *Matlab версия 9.6.0 R2019a [Электронный ресурс]*. – Режим доступа: https://1progs.ru/matlab/ - (Дата обращения: 05.09.2019)
10. *Елисеева И.И. и др.* Эконометрика // М.: Финансы и статистика, 2003. – С.344.