

PROPERTIES OF COVARIANCE
IN DEPENDENCES BETWEEN
INFORMATIONAL SIGNS.
APPLICATION OF COEFFICIENTS OF
CONTINGENCY SIMILARITY
– YULE’S AND FISCHER’S CORRELATION
COEFFICIENT IN THE ANALYSIS OF
MISTAKES IN THE TECHNOLOGICAL
INFORMATION PROCESSING CHAINS.
DETERMINATION OF FLUCTUATIONS OF
INFORMATION SIGNS IN TIME

T. Ivanova, associate professor, Candidate of Economic sciences,
Associate Professor
Moscow State Pedagogical University, Russia

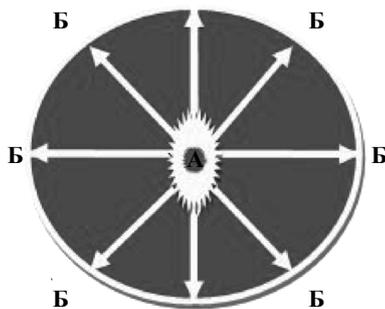
In this article, the author raises the question of the use of statistical, mathematical methods for the analysis of information and process, provides examples of such methods in information technology.

Keywords: Process chain, data processing, information workflow, system errors, mathematical methods.

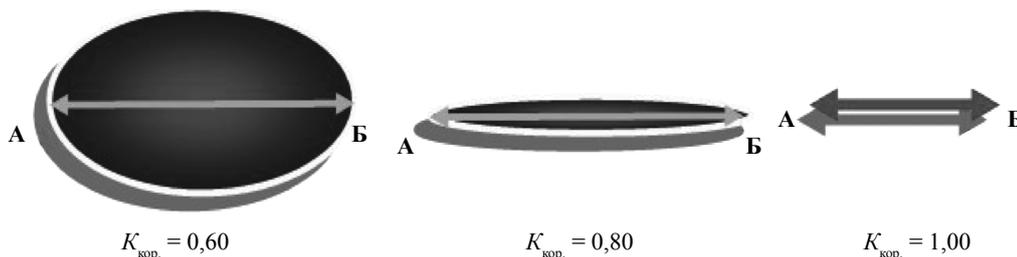
Conference participant, National championship
in scientific analytics, Open European and Asian
research analytics championship

При состоянии системы ковариации между точками А и Б, при реакции движения, вокруг точки А образуется окружность, а точка Б будет любая точка на линии окружности.

Система останавливается перед равновыбором (любой исход по любому направленному вектору). Коэффициент корреляции $K_{кор.} = 0$ (это и есть ковариация).



Что бы система вышла из равновесного состояния, необходимо хотя бы малая сплюснутость окружности по полюсам. Тогда коэффициент корреляции будет отличаться от 0. И начнется медленный процесс взаимного влияния (корреляции) двух систем.



СВОЙСТВА КОВАРИАЦИИ
В ЗАВИСИМОСТЯХ МЕЖДУ
ИНФОРМАЦИОННЫМИ ПРИЗНАКАМИ.
ПРИМЕНЕНИЕ КОЭФФИЦИЕНТОВ
СХОДСТВА КОНТИНГЕНЦИИ – ЮЛА
И АССОЦИИИ ФИШЕРА ПРИ АНАЛИЗЕ
ОШИБОК В ТЕХНОЛОГИЧЕСКИХ
ЦЕПОЧКАХ ОБРАБОТКИ ИНФОРМАЦИИ.
ОПРЕДЕЛЕНИЕ КОЛЕБАНИЙ
ИНФОРМАЦИОННЫХ ПРИЗНАКОВ
ВО ВРЕМЕНИ

Иванова Т.А., канд. экон. наук, доцент
Московский государственный педагогический университет,
Россия

В данной статье автор поднимает вопрос об использовании статистических, математических методов для анализа работы информационно-технологического процесса, приводит примеры применения таких методов в информационных технологиях.

Ключевые слова: Технологические цепочки, обработка данных, информационно технологический процесс, ошибки системы, математические методы.

Участник конференции, Национального первенства
по научной аналитике, Открытого Европейско-Азиатского
первенства по научной аналитике

Если рассматривать прохождение программ в алгоритме конкретного задания Прямая линия будет означать «четкое беспрепятственное прохождение программных команд», далее эллипс – «разно выраженная задержка и препятствия в прохождении программ». И наконец, ровная окружность – «выполнение программ невозможно» – прерывание цепи последовательных действий. Эти сообщения можно использовать для отслеживания разного рода ошибок в технологических процессах прохождения программ через электронные устройства обработки данных.

Коэффициенты сходства двух признаков (логическая проверка).

Таблица 2×2.

Ошибки		1 пр.		Σ
		+	-	
2 пр.	+	a	b	a + b
	-	c	d	c + d
Σ		a + c	b + d	n

1 – ошибки в **формате данных** на электронных носителях («+» – есть ошибки; «-» – нет ошибок)

2 – ошибки в **прохождении программы** технологической цепочки алгоритма обработки данных («+» – есть ошибки; «-» – нет ошибок).

Формулы ассоциации и контингенции применяются в позициях логического квадрата, таблицы 2×2. В таких таблицах существуют установки: да–да; нет–нет; если не да, то нет (нет–да); если не нет, то да (да–нет).

Коэффициенты покажут зависимость (логику) двух существенных признаков, в данном случае прерывание прохождения пакета прикладных программ и обработку статистического множества данных.

Коэффициент ассоциации Юла:

$$A = \frac{ad - bc}{ad + bc}$$

показывает долю разности диагональных значений совпавших значений признаков (да–да и нет–нет) и не совпавших (да–нет и нет–да) первого порядка к их сумме.

Коэффициент контингенции Фишера:

$$K = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$$

Коэффициент контингенции Фишера отличается от коэффициента ассоциации Юла тем, что в делители (базе сравнения) берем среднегеометрическое значение произведения двух строковых суммы по всей матрице (табл. 1).

$$A = \frac{60 \cdot 70 - 30 \cdot 40}{60 \cdot 70 + 30 \cdot 40} = \frac{3000}{5400} \approx 0,556;$$

$$K = \frac{60 \cdot 70 - 30 \cdot 40}{\sqrt{90 + 110 + 100 + 100}} = \frac{3000}{9949,9} = 0,302.$$

Коэффициенты сходства показывают среднестатистическую (не ярко выраженную) зависимость (сходство) признаков, то есть выстраивается логический квадрат. (Существуют общепринятые, пограничные критерии: если $A \geq 0,5$; $K \geq 0,3$ можно говорить о наличие прямой или обратной связи между двумя рассматриваемыми признаками).

Если собираются данные о «сбоях» в информационных системах, то есть образуются ряды динамики во времени (например, недели, месяцы и т.д.), используют метод сопоставлений средних значений признака (средние детализованных разбиений и общие, агрегатные средние).

Например, если разные группы системных инженеров осуществляют дежурства и контроль за отладкой системы по дням недели, можно отследить недельные индексы сезонности ошибок за 4 недели месяца и выстроить график – сезонную волну. При расчете средние индексы сезонности по неделям взятые за месяц сопоставляются со средним недельным уровнем ряда динамики, то есть общая средняя за период.

Формула расчета индекса сезонности такая:

$$I_{S_i} = \frac{\bar{X}_i}{\bar{X}} \cdot 100\%$$

где \bar{X}_i – средний уровень i -го ряда динамики;

\bar{X} – средняя величина из всех уровней рядов динамики.

Эта формула представляет долю отдельной частной средней в общем среднем уровне ряда динамики.

Совокупность рассчитанных индексов характеризует сезонную волну.

Число ошибок за месяц (4 недели) по дням недели (табл. 2).

Таблица 1.

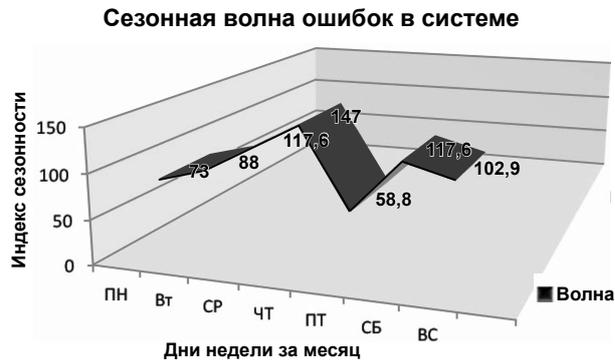
Данные коэффициенты сходства имеют значения в пределе от -1 до +1

Прохождение программ	Без прерывания	Прерывание обработки данных	Σ
Ошибки в информац. носителях			
Верное форматирование (без ошибок)	60	40	100
Неверные данные в инф. носителях Ошибки в форматировании	30	70	100
Σ	90	110	200

Таблица 2.

Дни недели	Число ошибок				Σ	\bar{X}_i	$I_{S_i} = \frac{\bar{X}_i}{\bar{X}} \cdot 100\%$
	1 нед.	2 нед.	3 нед.	4 нед.			
Пн.	2	1	1	1	5	1,25	73
Вт.	1	1	2	2	6	1,5	88
Ср.	2	1	2	3	8	2	117,6
Чт.	3	3	2	2	10	2,5	147,0
Пт.	1	1	1	1	4	1,0	58,8
Сб.	2	2	2	2	8	2,0	117,6
Вс.	1	2	2	2	7	1,75	102,9
					48	$\bar{X} = 1,7$	

Волна сезонности понедельных ошибок за месяц будет такой:



Из графика следует, что максимальный экстремум сезонной волны за месяц приходится на четверг, это значит, что группе системных инженеров необходимо серьезно проанализировать ошибки и осуществить необходимую коррекцию в проблемных областях технологической цепочки обработки данных.

Таким образом, возможен своеобразный симбиоз статистико-математических методов и информационных систем. Особенно это направление полезно в областях соединения информатики и математики. Так, например, при

изучении мобильных и имобильных элементов информационных систем можно использовать первый метод, описанный в данной статье (ковариация – она же имобильность, неподвижность) или тем же методом отслеживать состояния, когда система переходит в «спящий» режим; вторые методы хороши для анализа логической амбивалентности (то есть противоречивости) отдельных позиций системы, третий метод можно применять при анализе деструкции иерархических зависимостей в информационных системах, изучении дистрибуции и редистрибуции информационных сложенных.

В процессе образования данная корреляция математических методов и информатики призвана научить обучающихся апеллировать математическими знаниями в исследовании крупных информационных массивов и анализе алгоритмов обработки данных.

References:

1. Статистический словарь. – М.: Энциклопедия, 2007.
2. Статистико-математическая энциклопедия. – М.: Энциклопедия, 2004.
3. Математическая статистика; учебник / В.С. Мхитарян. – М.: Финансы и статистика, 2002.



The AICAC Secretariat

Tel: + 12 024700848
 Tel: + 44 2088168055
 e-mail: secretariat@court-inter.us
 skype: court-inter



AMERICAN INTERNATIONAL
 COMMERCIAL
 ARBITRATION COURT

The American International Commercial Arbitration Court LLC – international non-government independent permanent arbitration institution, which organizes and executes the arbitral and other alternative methods of resolution of international commercial civil legal disputes, and other disputes arising from agreements and contracts.

The Arbitration Court has the right to consider disputes arising from arbitration clauses included into economic and commercial agreements signed between states.

Upon request of interested parties, the Arbitration Court assists in the organization of ad hoc arbitration.

For additional information
 please visit:
court-inter.us