

РЕДАКЦИОННЫЙ КОММЕНТАРИЙ К СТАТЬЕ

В статье рассмотрена актуальная для образовательных организаций проблематика, связанная с управлением жизненным циклом электронных образовательных ресурсов (ЭОР). Изложение достаточно последовательное, охват материала соответствует названию работы, библиографический список адекватен ее содержанию.

Можно сделать следующие замечания. 1) В работе используются только диаграммы в нотациях UML и IDEF0, а изложение «привязывается» к ним. Однако целесообразность использования именно этих типов диаграмм не обосновывается. Существуют ведь и другие типы диаграмм, которые потенциально для решения рассматриваемых в работе задач. 2) Недостаточно внимания в статье уделено изменениям нормативных документов, определяющих условия деятельности образовательных организаций – в частности постоянным корректировкам требований в Федеральных государственных образовательных стандартах (ФГОС) высшего образования. При этом отметим, что прямые (явные) требования к номенклатуре и содержанию именно ЭОР в ФГОС не ются. 3) В работе фактически не рассматриваются вопросы оценки целесообразности создания новых ЭОР на основе оценки «полноты и качества» совокупности уже существующих разработок – причем не в масштабе конкретного вуза, а России в целом. Это отражает сложившуюся практику, когда почти каждый вуз самостоятельно разрабатывает ЭОР и поддерживает их использование – в т.ч. для обеспечения процессов дистанционного 4) В статье вообще не затрагиваются вопросы «аудита качества» ЭОР, их аттестации независимыми экспертами. Между тем, для «учебников на бумаге», учебных пособий и пр. получение грифов учебно-методических объединений о «допуске» к использованию в учебном процессе является важным показателем, подтверждающим их качество. 5) При характеристике жизненного цикла ЭОР стоило бы что-то сказать и о необходимых усилиях по их информационному продвижению в образовательном пространстве, в т.ч. регистрации в соответствующих фондах. Такая регистрация, кроме того, является наиболее рациональным средством фиксации авторских прав на выполненные разработки. 6) Также в рамках анализа ЖЦ ЭОР можно было бы рассмотреть вопросы обмена ими между вузами-разработчиками и/или о взаимном обеспечении доступа к ЭОР на серверах; об использовании ЭОР (включая тестирующие модули) не только для обучения, но и для сертификации (аттестации) специалистов и пр. 7) Вопросы «оптимизации решений», а также технологий проведения такой оптимизации, в работе представлены относительно слабо. Для расчетов в таблице 6 по формуле «1» используются только бинарные («да-нет») оценки показателей программных средств. При переходе к количественным оценкам этих показателей (по полноте/качеству реализации соответствующих функций в программных средствах), результаты выбора по таблице 6 могут измениться. Кроме того, для таблицы 6 будет целесообразно оценить «устойчивость» выбора оптимального решения по отношению к величинам экспертных оценок показателей. 9) В статье ничего не сказано о рисках, связанных с разработкой и использованием ЭОР; управлением этими рисками. Между тем, приведенные в статье диаграммы обеспечивают для этого достаточно удобные возможности – по крайней мере, на качественном уровне.

УДК 519.226+004.9

**КОМБИНАТОРНЫЕ МЕТОДЫ ОБНАРУЖЕНИЯ СОГЛАСОВАННОСТИ
СУЖДЕНИЙ ЭКСПЕРТОВ ПРИ ЧИСЛОВОЙ ШКАЛЕ ОЦЕНИВАНИЯ**

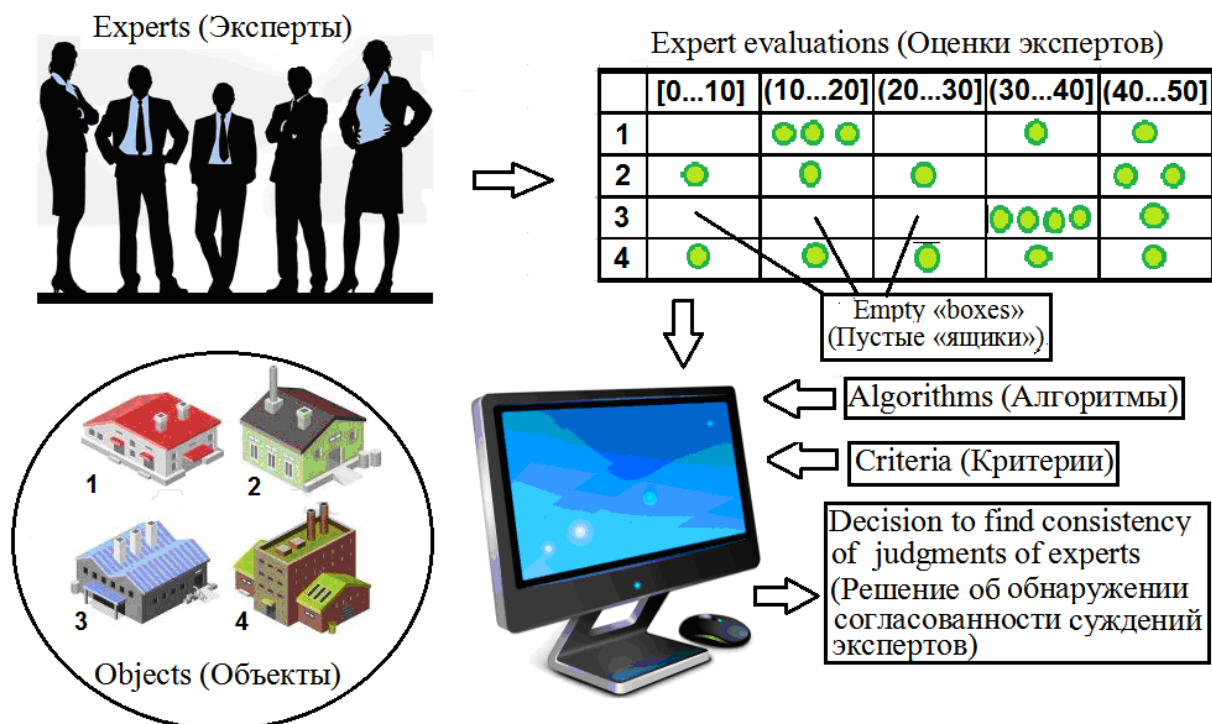
Статья поступила в редакцию 19.01.2018, в окончательном варианте – 19.02.2018.

Гolik Феликс Валентинович, Новгородский филиал Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации. 173003, Российская Федерация, Новгородская область, г. Великий Новгород, ул. Германа, 31,
доктор технических наук, профессор, ORCID <https://orcid.org/0000-0002-4208-456X>,
e-mail: felix.golik@mail.ru

Цель статьи – разработка процедур обнаружения согласованности суждений экспертов (ССЭ) в виде экспертных оценок, полученных при проведении групповых экспертиз в числовых шкалах; определение вероятностных характеристик предложенных процедур и результатов оценивания. Объект исследования – экспертные оценки, полученные в результате проведения групповой экспертизы объектов с оцениванием их на основе использования числовой шкалы. Предмет исследования – оценка ССЭ. В качестве мер согласованности оценок ССЭ при использовании числовой шкалы оценивания применяются среднее значение, медиана, коэффициент вариации величин экспертных оценок. При этом корректную проверку адекватности этих показателей выполнить не удастся, поскольку неизвестен закон распределения вероятности оценок. Причина этого – существует связь между показателями ССЭ и видом закона распределения оценок. При полной ССЭ распределение вырожденное, а при его отсутствии – равномерное. В связи с этим предложено вместо количественной оценки ССЭ проверять факт ее наличия. Процедура сводится к проверке гипотезы о законе распределения экспертных оценок. Если распределение незначимо отличается от равномерного, то принимается решение об отсутствии ССЭ. В противном случае считают, что мнения экспертов согласованы. При решении поставленной задачи используется аппарат комбинаторной теории размещений, теории вероятностей и математической статистики. Оценки вероятностных характеристик процедур обнаружения ССЭ получены посредством компьютерного имитационного моделирования. Основные результаты исследования: разработаны конструктивные методы обнаружения ССЭ при использовании числовой шкалы оценивания; предложены процедуры обнаружения ССЭ относительно конкретных объектов и относительно групповой экспертизы в целом; путем компьютерного имитационного моделирования найдены оценки вероятностных характеристик процедур оценивания. Результаты работы могут найти применение при принятии управленческих решений, прогнозировании и долгосрочном планировании в различных сферах деятельности.

Ключевые слова: экспертная оценка, шкала оценивания, согласованность экспертных оценок, критерий «пустых ящиков», статистическая гипотеза, уровень значимости, доверительная вероятность, кривые Пирсона, состояние полной вариации, компьютерное моделирование, принятие решений, информационные технологии

Графическая аннотация (Graphical annotation)



COMBINATORIAL METHODS FOR IDENTIFYING THE COHERENCE OF EXPERTS' JUDGMENTS WITH A NUMERICAL GRADING SCALE

The article was received by editorial board on 19.01.2018, in the final version – 19.02.2018.

Golik Felix V., Novgorod Branch of the Russian Presidential Academy of National Economy and Public Administration; 31 Germana Street, Veliky Novgorod, Novgorodskaya Oblast, 173003, Russian Federation, Doct. Sci. (Engineering), Professor, ORCID <https://orcid.org/0000-0002-4208-456X>, e-mail: felix.golik@mail.ru

The article aims at elaborating procedures for identifying coherence of experts' judgments (CEJ) in the form of expert estimates obtained from group assessment with numerical scales, as well as defining probabilistic characteristics of the proposed procedures and results of their assessment. The object of the study is experts' estimates obtained from group assessment with the use of a numerical scale. The subject is CEJ assessment. An average value, median and variation coefficient of experts' estimates were used as measures of CEJ estimates with the use of a numerical scale. However, it is not possible to verify the adequacy of these indicators correctly since the law of probability distribution of estimates is unknown as there is a link between the degree of CEJ and the distribution law type. With the complete CEJ the distribution is degenerate, and in its absence the distribution is flat. In this regard, it is proposed to verify the fact of coherence presence instead of its quantitative assessment. The procedure is reduced to testing the hypothesis about the law of distribution of experts' estimates. If the distribution differs insignificantly from the flat one, then a decision of CEJ lack is made. Otherwise, it is considered that the experts' opinions are coordinated. In order to solve the problem, the author used the system of combinatorial placement theory, probability theory and mathematical statistics. Estimates of the probability characteristics of CEJ identification procedures are obtained through computer simulation. The research results include the elaboration of constructive methods of CEJ identification with the use of a numerical scale for assessment; procedures of identifying CEJ in relation to specific objects and group assessment in general; estimates of the probability characteristics of assessment procedures through computer simulation. The results can be applied in management decision-making, forecasting and long-term planning in different fields.

Keywords: expert estimate, assessment scale, coherence of expert judgments (CEJ), "empty boxes" criterion, statistical hypothesis, significance level, confidence probability, Pearson curves, total variation distance, computer modeling, decision-making, information technologies

Введение. Экспертное оценивание объектов применяется в тех случаях, когда получить их количественные характеристики затруднительно или невозможно. При этом типично использование экспертных оценок в числовой форме (обычно – в баллах). Однако для того, чтобы понять, насколько этим оценкам

экспертов можно «доверять», важно определить степень согласованности суждений экспертов (ССЭ). При низком уровне ССЭ результаты экспертного оценивания могут быть поставлены под сомнение. Несмотря на то, что вопросы оценки ССЭ обсуждаются в ряде работ [1, 3, 8, 9, 13, 15, 16, 21 и др.], некоторые направления этой темы остаются разработанными недостаточно полно. Поэтому общей целью настоящей статьи является попытка обосновать целесообразность использования нового метода оценки согласованности, в котором вместо количественной оценки ССЭ проверяется сам факт ее наличия.

Общая характеристика предметной области. Возрастающая сложность управления современными экономическими системами требует проведения анализа целей и задач деятельности, средств их достижения, оценки влияния факторов на эффективность и качество работы. Перечисленные задачи относятся к классу неструктурированных и частично структурированных. Они не поддаются формализации, что приводит к необходимости применения специальных методов для их решения, к числу которых относятся экспертные методы.

Экспертиза как способ получения информации традиционно используется при принятии решений. Однако научные исследования по ее теоретическому обоснованию были начаты сравнительно недавно, всего несколько десятилетий назад [3, 23]. Результаты этих и последующих исследований позволяют сделать вывод о том, что современная теория экспертного оценивания является научным методом анализа сложных неформализуемых проблем [14, 15, 18, 20, 36].

Экспертные методы непрерывно развиваются и совершенствуются. Однако, несмотря на успехи, достигнутые в последние годы в разработке и практическом использовании метода экспертных оценок, существует ряд проблем и задач, требующих дальнейших исследований и практической проверки. Одна из таких проблем связана с оценкой ССЭ.

Проверка ССЭ является стандартной процедурой обработки данных, полученных методом группового экспертного оценивания [1, 19]. Для оценки ССЭ вычисляется количественная мера, характеризующая степень близости индивидуальных оценок экспертов. Вид меры зависит от принятой шкалы оценивания [3, с. 39; 20, с. 85]. При измерении в порядковой шкале в качестве показателя ССЭ используют коэффициенты корреляции и конкордации Кендалла [20, с. 184], медиану Кемени [20, с. 152], энтропийный коэффициент конкордации [3, с. 430]. Если оценка производится по числовой шкале, то применяют методы точечного и интервального оценивания среднего значения, медианы, коэффициента вариации [3, с. 122; 30, 31].

Найденные в результате анализа экспертных данных показатели ССЭ являются случайными величинами, поскольку рассчитываются на основании экспертных оценок. Они зависят от многих факторов и содержат случайную составляющую. В связи с этим возникает проблема оценки значимости этих показателей, корректное решение которой, по-видимому, известно только для коэффициентов корреляции и конкордации Кендалла [11], используемых при экспертизах в порядковой шкале. При оценке значимости показателей ССЭ, применяемых в числовых шкалах, делается допущение о нормальности распределения оценок экспертов [3, с. 123]. Это допущение может быть справедливым при большом числе экспертов. Но в общем случае, когда нормальность оценок не доказана, его применение оправдывается лишь возможностью использования классических методов расчета границ доверительных интервалов. При этом неизбежно возникает сомнение в обоснованности выводов, которые делаются на основании результатов, полученных при некорректных допущениях. На наш взгляд, общего конструктивного решения проблемы оценки значимости показателей ССЭ в числовой шкале¹, за исключением частного случая доказанной нормальности распределения оценок экспертов, не существует, поскольку закон распределения оценок и уровень ССЭ связаны между собой. Действительно, если оценки экспертов равномерно распределены в пределах числовой шкалы оценивания $[A, B]$, то однозначно можно утверждать, что суждения экспертов не согласованы. Полная ССЭ наступает, если оценки всех экспертов совпали. При этом распределение вырожденное. Следовательно, распределение зависит от степени ССЭ и изменяется в широких пределах. Тогда границы доверительного интервала для показателя согласованности придется находить с учетом конкретного вида распределения оценок экспертов (отличного от нормального), что связано с решением задач, до сих пор не решенных в рамках классической теории математической статистики.

Если проблему оценки значимости показателей ССЭ не удастся решить, то ее можно обойти, изменив постановку задачи. Как было сказано выше, вероятностное распределение оценок экспертов и уровень ССЭ связаны между собой. Следовательно, судить о ССЭ можно как по показателям согласованности, так и по виду закона распределения оценок. Учитывая, что при полной несогласованности оценки распределены по равномерному закону, то по степени «близости» распределения к равномерному можно судить о степени ССЭ. Именно этот подход явно или косвенно используется при построении непараметрических критериев согласованности, в частности коэффициента конкордации Кендалла.

Если не учитывать степень согласованности, а говорить лишь о её наличии или отсутствии, то задача сводится к проверке статистической гипотезы H_0 о законе распределения вероятности оценок экспертов. Если принимается решение, что оценки распределены по равномерному закону, то суждения

¹ Имеются в виду такие показатели, как среднее значение, медиана, коэффициент вариации.

экспертов не согласованы. При отклонении гипотезы считаем, что суждения согласованы. Результат проверки гипотезы H_0 можно трактовать как обнаружение или не обнаружение факта наличия согласованности экспертных оценок или коротко – обнаружение согласованности экспертизы.

Поэтому конкретной целью настоящей работы является разработка процедур обнаружения ССЭ (экспертных оценок), полученных при проведении групповых экспертиз в числовых шкалах; оценка вероятностных характеристик предложенных процедур и получаемых при их применении результатов. Достижение поставленной цели позволит повысить достоверность оценки ССЭ и сократить затраты времени на исследования, поскольку если ССЭ не обнаружена на заданном уровне значимости, то и последующий анализ степени согласованности не имеет смысла.

Как следует из цели исследования, основной задачей данной работы является разработка процедуры проверки статистической гипотезы о равномерности распределения оценок экспертов. Арсенал критериев, применяемых при решении этой задачи, достаточно велик. Различают критерии «специальные», непараметрические и энтропийные [13]. Специальные критерии, «приспособленные» для проверки именно гипотезы о равномерном распределении. К ним относятся критерии Шермана, Кимбелла, Морана, Ченга – Спиринга и др. В классе непараметрических критериев чаще всего используют критерии Колмогорова и Крамера – Мизеса – Смирнова (омега-квадрат). [4, 13, 17, 21, 29, 35]. Статистики указанных критериев распределены по законам, существенно зависящим от объема выборки и аппроксимируются «стандартными» распределениями при достаточно жестких ограничениях¹. В настоящей работе предложено проверять гипотезу о равномерном распределении по комбинаторному критерию «пустых ящиков», который отличается простотой и наглядностью [12]. Кроме того, статистика критерия достаточно точно аппроксимируется нормальным распределением уже при малых объемах выборки, что позволяет применять классические методы проверки статистических гипотез. В соответствии с этим критерием, для проверки гипотезы H_0 о том, что выборка, состоящая из n независимых величин $x_1, x_2, \dots, x_n$, извлечена из генеральной совокупности с равномерным распределением, разбивают интервал наблюдения $[A, B]$ на N полуинтервалов $(z_{k-1}, z_{k-1} + \Delta]$, $k = 1, 2, \dots, N$, $z_0 = A$, $\Delta = (B-A) / N$ и подсчитывают число ξ полуинтервалов, не содержащих ни одного элемента выборки $x_1, x_2, \dots, x_N$. Гипотеза H_0 принимается, если статистика $\xi \leq c_\alpha$ и отклоняется, если $\xi > c_\alpha$. Постоянная c_α выбирается из условия, что вероятность ошибки первого рода, т. е. вероятность отклонить верную гипотезу H_0 , равна заданному значению α ².

Постановка задачи. Пусть X_{ji} – численная оценка свойств j -го объекта, данная i -м экспертом, принадлежит интервалу $[A, B]_j$, $j = 1, 2, \dots, K$, $i = 1, 2, \dots, N$. Будем считать, что оценки экспертов j -го объекта взаимно независимы, и каждый эксперт предьявляет оценку каждого объекта.

Проверку гипотезы о ССЭ можно выполнить по оценкам конкретного объекта или по оценкам всех объектов в совокупности. В первом случае решение сводится к проверке статистических гипотез

$$H_{0j} : F_j(x) = F_0(x), \alpha,$$

$$H_{1j} : F_j(x) \neq F_0(x), \alpha, j = 1, 2, \dots, K,$$

где $F_j(x)$ – функция распределения вероятности экспертных оценок j -го объекта; $F_0(x)$ – функция равномерного распределения вероятности; α – уровень значимости. В результате получим K решений, индикаторы которых обозначим

$$h_j = \begin{cases} 0, & \text{если } H_{0j} \text{ отклонена} \\ 1, & \text{если } H_{0j} \text{ принята} \end{cases} \quad j = 1, 2, \dots, K.$$

Каждое решение подтверждает ($h_j = 0$) или отклоняет ($h_j = 1$) допущение о ССЭ при оценке свойств j -го объекта.

Вторую задачу можно решить путем объединения индивидуальных решений h_j или посредством объединения экспертных оценок X_{ji} , принадлежащих индивидуальным интервалам $[A, B]_j$, $j = 1, 2, \dots, K$, в общем интервале $[A, B]$ и проверки гипотезы $H_0: F(x) = F_0(x), \alpha$, где $F(x)$ – функция распределения объединенных оценок экспертов.

Обнаружение согласованности результатов экспертизы j -го объекта. Рассмотрим процедуру проверки согласованности экспертизы свойств j -го объекта. Результаты таких проверок будем называть индивидуальными решениями.

¹ Статистики критерия Колмогорова при объеме выборки $n > 20$ с достаточной для практики точностью можно аппроксимировать распределением χ^2 с двумя степенями свободы или распределением Колмогорова [4, с. 80]. Для статистики $n\omega^2$ существует предельное распределение при $n \rightarrow \infty$ [4, с. 83].

² Здесь, в отличие от [12, с. 9], приведена упрощенная формулировка процедуры проверки гипотезы по критерию «пустых ящиков», адаптированная к проверке гипотезы о равномерном распределении.

Для обеспечения общности дальнейших рассуждений преобразуем экспертные оценки $X_{j,i}$ произвольной метрической числовой шкалы оценивания в интервале $[A, B]$ к интервалу $[0, 1]$, применив к оценкам $X_{j,i}$, линейное преобразование: $x_{j,i} = (X_{j,i} - A) / (B - A)$.

Разобьем интервал $[0, 1]$ на N полуинтервалов равной длины точками $z_k = k \Delta$, $k = 1, 2, \dots, N$, $z_0 = 0$, $\Delta = 1 / N$. Тогда вероятность $P[z_{k-1} < x \leq z_k] = F_0(z_k) - F_0(z_{k-1}) = 1 / N$, где

$$F_0(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

– функция равномерного распределения вероятности случайной величины.

Для проверки гипотезы H_{0j} необходимо рассчитать вероятность $p_{j,k}$ распределения статистики ξ_j , равной числу полуинтервалов $(z_{k-1}, z_k]$, в которые не попала ни одна из экспертных оценок j -го объекта. В [12, с. 10] доказано, что эта вероятность равна:

$$p_{j,k} = P[\xi_j = k] = \frac{N!}{k!(N-k)!} \left(1 - \frac{k}{N}\right)^N \sum_{l=0}^{N-k} \frac{(-1)^l (N-k)!}{l!(N-k-l)!} \left(1 - \frac{l}{N-k}\right)^N. \quad (1)$$

Учитывая, что все N экспертов оценивают каждый объект, распределение (1) не зависит от номера объекта и $p_{j,k} = p_k$.

Проверка гипотезы H_0 сводится к сравнению статистики ξ_j^* с ее критическим значением c_α . Здесь ξ_j^* – число пустых полуинтервалов $[z_{k-1}, z_k]$, $k = 1, 2, \dots, N$, не содержащих оценок $x_{j,i}$ j -го объекта, а c_α – критическое значение статистики ξ на уровне значимости α :

$$c_\alpha = \sum_{k=0}^{N-1} \text{if}(F_k \leq 1 - \alpha < F_{k+1}, k, 0), \quad (2)$$

где $F_k = \sum_{l=0}^k p_l$ – функция распределения вероятности случайной величины ξ , $\text{if}(\text{cond}, x, y)$ – логическая функция, которая равна x , если истинно логическое условие cond (не равно нулю), иначе равна y .

В таблице 1 приведены критические значения c_α , рассчитанные по формуле (2).

Таблица 1 – Критические значения статистики ξ

Уровень значимости α	Количество экспертов N																			
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
0,1	1	1	1	2	2	3	3	4	4	5	5	5	6	6	7	7	8	8		
0,05	1	1	2	2	3	3	4	4	5	5	5	6	6	7	7	8	8	8		
0,01	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	8	9	9		

Если $\xi_j^* \leq c_\alpha$, то принимается гипотеза H_{0j} . При $\xi_j^* > c_\alpha$ гипотеза H_{0j} отклоняется. Это означает, что генеральная совокупность случайных величин, из которой извлечена выборка $V_j = \{x_{j,i}, i = 1, 2, \dots, N\}$, распределена по закону, значимо отличающемуся от равномерного и суждения экспертов относительно свойств j -го объекта согласованы.

Индивидуальные решения равны индикаторам решений, принятых при проверке гипотез H_{0j} относительно j -го объекта:

$$h_j = \begin{cases} 0, & \xi_j^* > c_\alpha, \\ 1, & \xi_j^* \leq c_\alpha, \end{cases} j = 1, 2, \dots, K, \quad (3)$$

Вследствие дискретности статистики ξ , вероятность ошибки первого рода $P_1(\alpha) = P[\xi > c_\alpha | H_0 \text{ верна}]$ может существенно превышать уровень значимости α , в соответствии с которым рассчитывается постоянная c_α . Это означает, что велика вероятность ошибочно принять решение о согласованности экспертизы, когда на самом деле суждения экспертов несогласованны. Приемлемые по надежности индивидуальные решения (3) можно получить лишь при достаточно большом числе N экспертов и (или) малой вероятности α (рис. 1). При этом следует иметь в виду, что при уменьшении α растет вероятность отклонения альтернативной гипотезы H_1 – правильного решения о наличии согласованности.

Большое число экспертов обычно встречается при оценке согласованности суждений респондентов в социологических исследованиях. Именно для подобных исследований и можно рекомендовать использование индивидуальных решений.

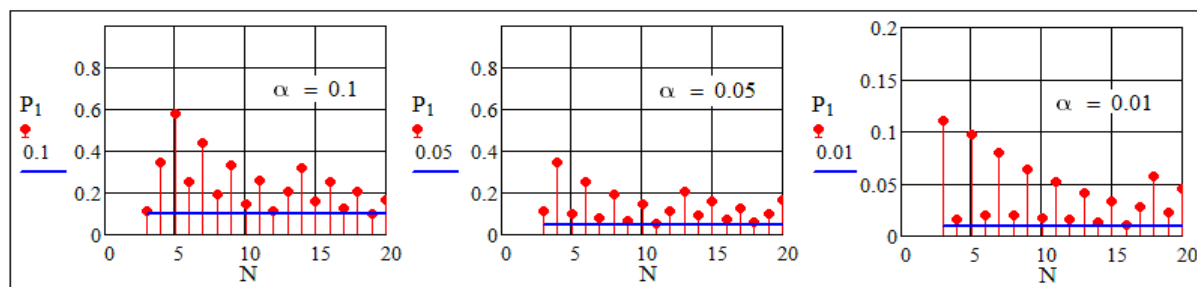


Рисунок 1 – Зависимость вероятности ошибки первого рода от числа N экспертов при различных значениях уровня значимости α

Аппроксимация вероятностного распределения p_k . Точная формула (1) распределения статистики ξ неудобна для расчетов вероятности p_k при больших N . Возникают трудности и при нахождении квантиля c_α путем решения уравнения $F_k = \alpha$. Задача упростится, если удастся аппроксимировать распределение p_k каким-либо стандартным непрерывным распределением. В настоящей работе исследуется аппроксимация кривыми Пирсона, универсальные свойства которых хорошо известны [16, 22]. Для подбора аппроксимирующей кривой необходимо найти моменты статистики ξ и по ним рассчитать коэффициенты Пирсона β_1, β_2 [7].

Расчетные формулы для математического ожидания m и дисперсии μ_2 известны [12, с 12]:

$$m = N \left(1 - \frac{1}{N} \right)^N, \quad (4)$$

$$\mu_2 = N(N-1) \left(1 - \frac{2}{N} \right)^N + N \left(1 - \frac{1}{N} \right)^N - N^2 \left(1 - \frac{1}{N} \right)^{2N}. \quad (5)$$

Центральные моменты третьего и четвертого порядка найдем численно:

$$\mu_3 = \sum_{k=0}^N (k-m)^3 p_k, \quad (6)$$

$$\mu_4 = \sum_{k=0}^N (k-m)^4 p_k \quad (7).$$

Коэффициенты Пирсона рассчитываются по формулам:

$$\beta_1 = \frac{\mu_3}{\mu_2}, \quad (8)$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}. \quad (9)$$

«Каппа» Пирсона равна

$$\kappa = \frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)}. \quad (10)$$

Для качественной оценки близости распределения p_k к нормальному, найдем в системе координат коэффициентов Пирсона расстояние от точки с координатами (β_1, β_2) до точки $(0, 3)$, соответствующей нормальному распределению [7]

$$R = \sqrt{\beta_1^2 + (\beta_2 - 3)^2}. \quad (11)$$

Результаты численных расчетов по формулам (8...11) приведены на рисунке 2.

На основании анализа результатов расчета можно сделать следующие выводы:

1) при всех значениях N , при которых проводились расчеты, «каппа» Пирсона отрицательна $\kappa(N) < 0$, следовательно, распределение p_k можно аппроксимировать кривой Пирсона типа I;

2) коэффициент $\beta_1(N)$ близок к нулю, а коэффициент $\beta_2(N)$ не слишком сильно отличается от 3, что говорит о возможности аппроксимации нормальным распределением;

3) с ростом N вывод 2) находит все большее подтверждение, расстояние $R(N)$ убывает и стремится к 0, что совпадает с утверждением об асимптотической нормальности распределения p_k [12, с. 32].

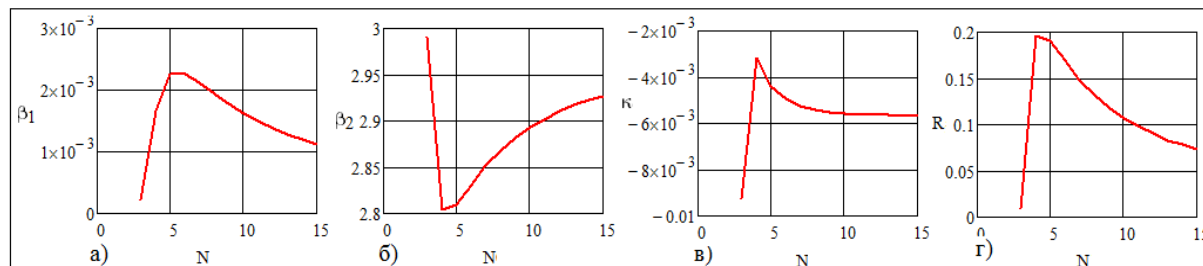


Рисунок 2 – Зависимость параметров кривых Пирсона от количества полуинтервалов N : а) $\beta_1(N)$, б) $\beta_2(N)$, в) $\kappa(N)$, г) $R(N)$

Аппроксимация нормальным распределением имеет существенные преимущества по сравнению с распределением Пирсона – в частности, благодаря свойству инвариантности относительно линейных преобразований, которое будет нами использовано в дальнейшем.

Для оценки точности аппроксимации распределения p_k нормальным законом рассчитаем расстояние полной вариации σ в вероятностном пространстве (σ -метрику) [10, с. 33]. Значения σ ограничены интервалом $[0; 1]$. При $\sigma = 0$ ошибки аппроксимации нет, а при $\sigma = 1$ между случайными величинами существует предельно возможное различие. По величине σ можно судить насколько сильно распределения случайных величин X и Y отличаются между собой по сравнению с предельно возможным случаем.

Расстояние полной вариации задается соотношением:

$$\sigma(X, Y) = \frac{1}{2} \int |d(F_X(x) - F_Y(y))|,$$

где F_X, F_Y – функции распределения вероятностей случайных величин X и Y соответственно.

В нашем случае

$$\sigma = \frac{1}{2} \sum_{k=0}^N |p_k - fN(k)|, \quad (12)$$

где

$$fN(x) = \frac{1}{\sqrt{2\pi\mu_2}} \exp\left(-\frac{(x-m)^2}{2\mu_2}\right), \quad (13)$$

$fN(x)$ – плотность распределения вероятности нормальной случайной величины; m – математическое ожидание и дисперсия μ_2 рассчитываются по формулам (4) и (5) соответственно.

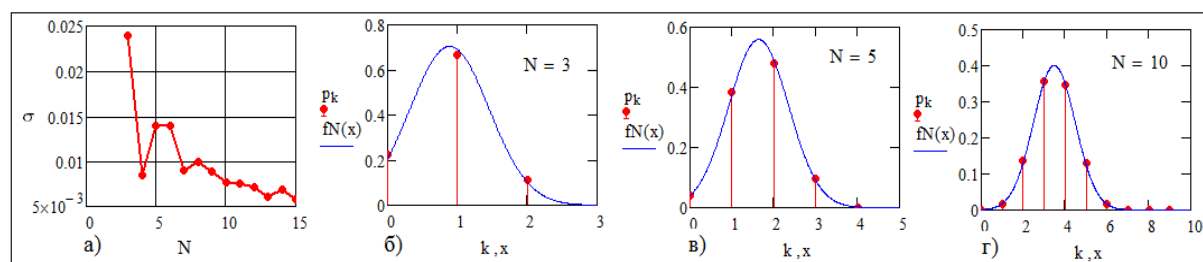


Рисунок 3. Зависимость погрешности аппроксимации σ от N и графики распределений вероятностей $p_k, fN(x)$: а) $\sigma(N)$; б) $N=3, \sigma=0,024$; в) $N=5, \sigma=0,014$; г) $N=10, \sigma=7,611 \cdot 10^{-3}$

Расчеты показывают, что даже при $N=3$ погрешность аппроксимации невелика и равна $\sigma = 0,024$ (или 2,4 %). С ростом N ошибка аппроксимации убывает. Графики на рисунке 3 наглядно иллюстрируют хорошее совпадение вероятностей p_k и значений плотности $fN(x)$ в точках $x = k$. Таким образом, распределение p_k с достаточной для практики точностью можно аппроксимировать нормальным законом $fN(x)$.

Обнаружение согласованности результатов экспертизы для совокупности объектов. Комплексную проверку ССЭ можно выполнить одним из трех вариантов. 1) По индивидуальным решениям h_j . 2) По индивидуальным статистикам ξ_j^* . 3) По экспертным оценкам X_{ji} , объединенным в одном интервале. Рассмотрим эти методы подробнее.

Обнаружение ССЭ по индивидуальным решениям h_j . Сумма индивидуальных решений $h = \sum_{(j)} h_j$ равна числу объектов, суждения экспертов о свойствах которых оказались несогласованными

(приняты нулевые гипотезы). Тогда $g = K - h$ – количество согласованных решений и $p^* = g / K$ – оценка

вероятности принятия согласованной оценки. В силу сделанного ранее допущения о взаимной независимости решений h_j , можно считать, что число g распределено по биномиальному закону с параметрами p , K . Тогда границы доверительного интервала генеральной вероятности p равны [6, с. 332]:

$$p_{1,2} = \frac{p^* + \frac{1}{2} \cdot \frac{u_{1-\alpha/2}^2}{K} \mp u_{1-\alpha/2} \sqrt{\frac{p^*(1-p^*)}{K} + \frac{1}{4} \cdot \frac{u_{1-\alpha/2}^2}{K^2}}}{1 + \frac{u_{1-\alpha/2}^2}{K}}, \quad (14)$$

где $u_{1-\alpha/2}$ – квантиль нормального распределения, α уровень значимости.

В соответствии с принципом большинства, решение о ССЭ по всем объектам принимается, если генеральная вероятность $p > 0,5$, то есть для нижней границы доверительной области выполняется неравенство $p_1 > 0,5$. Вероятность p_1 зависит от оценки p^* и числа объектов K . Численно решив уравнение $p_1(p^*, K) = 0,5$, найдем корни p_0 и K_0 . Решение о согласованности принимается, если $p^* > p_0(K)$ (рис. 4 г).

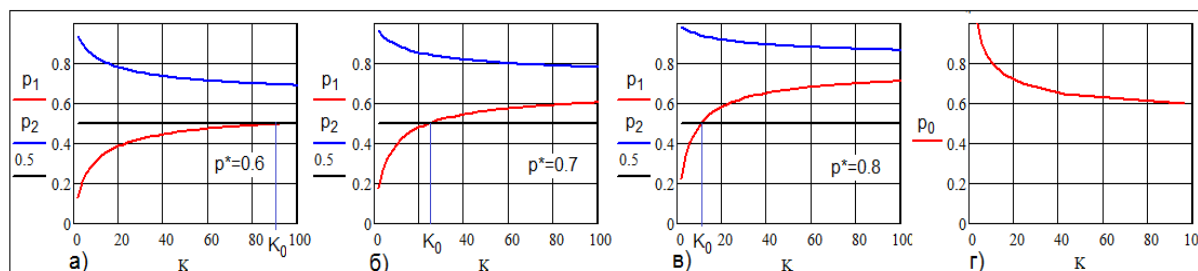


Рисунок 4 – Границы доверительной области для генеральной вероятности p (а, б, в) при $\alpha=0,05$ и зависимость критической вероятности p_0 от числа K объектов (г)

Минимальное число K_0 объектов, при котором принимается решение о согласованности, зависит от вероятности p^* (рис. 4). Так, при $p^* = 0,6$ $K_0 = 96$, а при $p^* = 0,7$ $K_0 = 24$. Даже при $p^* = 1$ K должно быть не меньше $K_0 = 4$. Количество объектов, подвергающихся экспертизе, обычно сравнительно невелико, что ограничивает возможность применения данного метода.

Обнаружение согласованности по индивидуальным статистикам ξ_j . Введем новую статистику $\zeta = \frac{1}{K} \sum_{(j)} \xi_j$, где ξ_j как и прежде равно числу полуинтервалов $(z_{k-1}, z_k]$, $k = 1, 2, \dots, N$, не содержащих экспертных оценок $x_{j,i}$ j -го объекта. Выше было показано, что вероятностное распределение статистик ξ_j можно считать нормальным. Полагая, что распределение индивидуальных решений не зависит от номера j объекта, получим, что статистика ζ распределена по нормальному закону

$$fN(x) = \frac{\sqrt{K}}{\sqrt{2\pi\mu_2}} \exp\left(-\frac{(x-m)^2}{2\mu_2} K\right), \quad (15)$$

где m и μ_2 по-прежнему рассчитываются по формулам (4) и (5) соответственно.

Решение о согласованности принимается по результатам проверки статистической гипотезы H_0 о равенстве генерального среднего m_0 гипотетическому значению m при известной генеральной дисперсии μ_2/K :

$$H_0: m_0 = m, \alpha, \quad (16)$$

здесь m – математическое ожидание (4) статистики ξ , α – уровень значимости.

Если гипотеза H_0 (16) не отклоняется, то принимается решение о несогласованности суждений экспертов. Согласованность считают обнаруженной, если гипотеза H_0 отклонена. Гипотеза проверяется по одностороннему критерию и не отклоняется, если $u_{\text{расч}} \leq u_{1-\alpha}$. Здесь $u_{1-\alpha}$ – квантиль нормального распределения; $u_{\text{расч}}$ – расчетное отклонение, распределенное по нормальному закону и равное

$$u_{\text{расч}} = \frac{\zeta^* - m}{\sqrt{\mu_2}} \sqrt{K}; \quad (17)$$

$\zeta^* = \frac{1}{K} \sum_{(j)} \xi_j^*$, ξ_j^* – фактическое (наблюдаемое) значение статистики ξ относительно j -го объекта.

Вероятность ошибки первого рода, как и в случае индивидуальных оценок, вследствие дискретности статистики ζ может отличаться от α . Но это отличие существенно меньше (ср. рис. 1 и рис. 5). Вероятность слабо зависит от N и достаточно хорошо совпадает с теоретическими значениями уже при $K = 5$ (рис. 6).

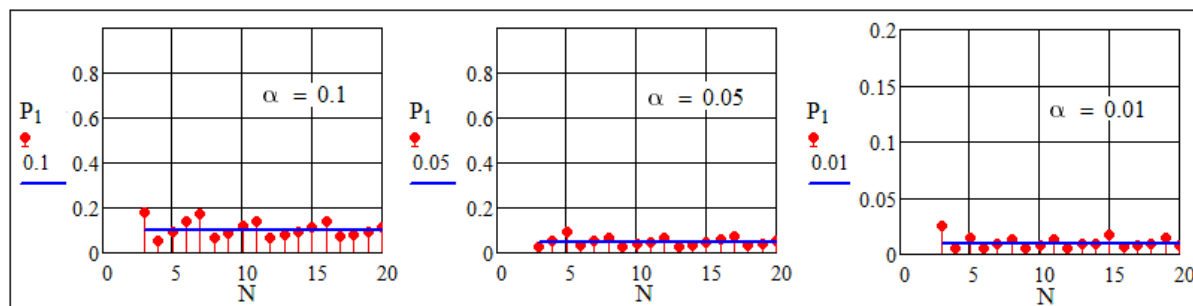


Рисунок 5 – Зависимость вероятности ошибки первого рода от числа N экспертов при различных значениях уровня значимости α и количестве объектов $K = 3$

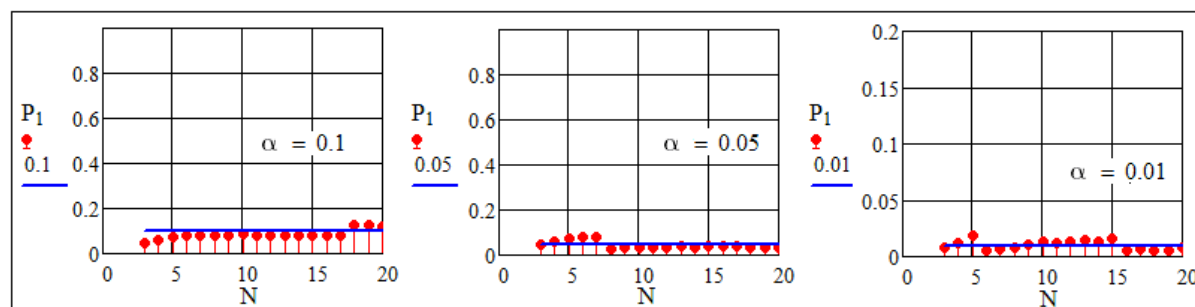


Рисунок 6 – Зависимость вероятности ошибки первого рода от числа N экспертов при различных значениях уровня значимости α и количестве объектов $K = 5$

Обнаружение согласованности по объединенным экспертным оценкам. Объединение экспертных оценок состоит в преобразовании матрицы оценок $\mathbf{X} = \{x_{ji}\}$, $j = 1, 2, \dots, K$, $i = 1, 2, \dots, N$, в вектор $\mathbf{Z} = [z_1, z_2, \dots, z_n, \dots, z_Q]$, $z_{n(j,i)} = x_{ji}$. Индексы n элементов z_n вектора $\mathbf{Z}$ связаны с индексами элементов x_{ji} исходной матрицы $\mathbf{X}$ соотношением $n(j, i) = j + (i - 1) \cdot K$, $n = 1, 2, \dots, Q$, $Q = K \cdot N$.

В результате получаем, что совокупность K объектов рассматривается как один объект, а эквивалентное число экспертов равно Q . В этом случае применима процедура (3) проверки гипотезы о согласованности относительно эквивалентного объекта.

Имитационная модель. В настоящей работе все численные расчеты выполнялись в программе MathCad. В этой же среде была реализована имитационная модель, посредством которой найдены вероятностные характеристики процедур обнаружения согласованности экспертиз. Модель имитирует формирование оценок экспертов, распределенных по равномерному и нормальному законам. По выборке равномерно распределенных чисел проверялась работоспособность предложенных процедур и оценивалась вероятность P ошибки первого рода. Случайные величины с нормальным распределением имитировали ситуацию согласованной экспертизы. Оценивалась вероятность D обнаружения согласованности.

Рассмотрим подробнее основные процедуры, реализованные в модели.

Условные обозначения параметров модели:

I – объем выборки;

N – число экспертов;

K – число объектов;

μ – математическое ожидание нормальной случайной величины;

s – среднеквадратическое отклонение нормальной случайной величины;

α – уровень значимости.

Формирование массивов экспертных оценок:

$x^{\triangleleft} = \text{rnorm}(N, \mu, s)$ – матрица случайных чисел размера $N \times K \cdot I$ с нормальным распределением; функция $\text{rnorm}()$ возвращает вектор из N случайных чисел, имеющих нормальное распределение со средним μ и среднеквадратическим отклонением s ; $i = 0, 1, \dots, K \cdot I - 1$.

$y^{\triangleleft} = \text{runif}(N, 0, 1)$ – матрица случайных чисел размера $N \times K \cdot I$ с равномерным распределением; функция $\text{runif}()$ возвращает вектор из N случайных чисел, распределенных в интервале $[0, 1]$.

Элементом выборки является матрица случайных чисел размером $N \times K$.

Границы интервалов и подсчет пустых полуинтервалов:

$\Delta = 1 / N$, $z_n = \Delta(n + 1)$ – границы интервалов, $n = 0, 1, \dots, N - 1$.

$$d0_i = \sum_{n=0}^{N-1} if(0 < v_{n,i} \leq z_0, 0, 1), \quad d1_i = \sum_{n=0}^{N-1} if(z_{n-1} < v_{n,i} \leq z_n, 0, 1),$$

$v_{n,i} = y_{n,i}$ при имитации несогласованных и $v_{n,i} = x_{n,i}$ согласованных оценок.

$$\xi_i = if(d0_i = N, 1, 0) + \sum_{n=1}^{N-1} if(d1_{n,i} = N, 1, 0) - \text{количество пустых полуинтервалов. Вектор статисти-}$$

стик $\xi = (\xi_i), i = 0, 1, \dots, K \cdot I - 1$.

Критические значения и индикаторы решений:

Критические значения c_α рассчитываются по формуле (2). А индикаторы решений находим из выражения (3). Вектор решений $h = (h_i), i = 0, 1, \dots, K \cdot I - 1$.

Результаты компьютерного моделирования. Моделирование проводилось с целью оценки вероятностных характеристик процедур обнаружения согласованности экспертиз. Во всех случаях объем выборки I принимался равным 10^4 .

Характеристики обнаружения согласованности по индивидуальным решениям h_j . Решение о согласованности принимается по величине относительной частоты обнаружения согласованности при экспертизе K объектов

$$p_i^* = 1 - \frac{1}{K} \sum_{n=i \cdot K}^{(i+1)K-1} h_n.$$

Затем по формуле (14) рассчитывается нижняя граница p_1 доверительного интервала генеральной вероятности, и полученное значение сравнивается с критической вероятностью 0,5. Рассчитывается вероятность

$$P = \frac{1}{I} \sum_{(i)} if(p_i^* > 0,5; 1; 0).$$

Вероятность P интерпретируется следующим образом:

- если имитируются экспертные оценки с равномерным распределением ($v=y$), то это вероятность ошибки первого рода: $P = P_1$;
- если экспертные оценки распределены по нормальному закону ($v=x$), то это вероятность правильного обнаружения согласованности: $P = D$.

Результаты моделирования подтвердили предварительные выводы: приемлемые для практики значения вероятности D можно получить лишь при большом числе объектов и (или) большом уровне значимости α .

Характеристики обнаружения согласованности по индивидуальным статистикам ξ . Решение об обнаружении согласованности принимается по результатам проверки гипотезы (17) относительно среднего значения статистики ξ , оценка которого равна:

$$m_i^* = \frac{1}{K} \sum_{n=i \cdot K}^{(i+1)K-1} \xi_n.$$

Расчетное отклонение согласно (16) равно

$$u_i^* = \frac{m_i^* - m}{\sqrt{\mu_2}} \sqrt{K},$$

где m и μ_2 – математическое ожидание и дисперсия статистики ξ , рассчитываются по формулам (4) и (5) соответственно.

По индикаторам $h1_i = if(u_i^* > u_{1-\alpha/2}, 0, 1)$ результатов проверки гипотезы (15) вычисляется вероятность

$$P = 1 - \frac{1}{I} \sum_{(i)} h1_i.$$

При $v=x$ (оценки экспертов распределены по нормальному закону) $P = D$ – вероятность правильного обнаружения согласованности. Если $v=y$, то $P = P_1$ – вероятность ошибки первого рода.

Результаты оценки вероятности P_1 приведены на рисунках 5 и 6. Эти вероятности примерно равны теоретическим значениям, равным уровням значимости α , поэтому детально не анализировались.

Вероятность D рассчитывалась при двух значениях среднеквадратического отклонения s оценок экспертов: $s = 1/6$ и $s = 1/7$. Это позволило косвенно оценить чувствительность метода. Очевидно, чем

меньше s , тем выше степень согласованности оценок и тем больше должна быть вероятность D , в чем легко убедиться, попарно сравнив рисунки 7 а), б), в) с рисунками 7 г), д), е). Отметим, что нерегулярный характер кривых на рисунке 7 обусловлен дискретностью статистик ξ и ζ .

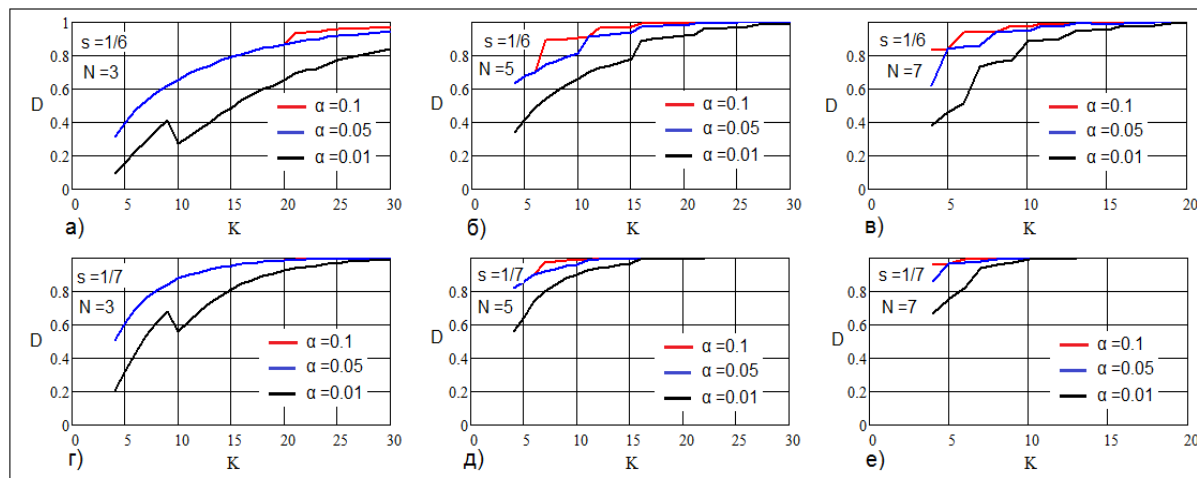


Рисунок 7 – Зависимость вероятности D от числа K объектов при различном числе N экспертов, s и α

Характеристики обнаружения согласованности по объединенным экспертным оценкам. В этом случае моделируется ситуация, когда Q экспертов, число которых равно $N \cdot K$, оценивают свойства одного объекта, то есть реализуется индивидуальное оценивание. Вероятности D и P_1 , как и прежде, являются частными случаями вероятности P

$$P = 1 - \frac{1}{I} \sum_{(i)} h_i,$$

где $h_i = if(\xi_i > c_{\alpha}, 0, 1)$, а статистики ξ рассчитывают отдельно для нормально ($v = x$) и равномерно ($v = y$) распределенных экспертных оценок.

При индивидуальном оценивании велика вероятность ошибки первого рода (см. рис 1). Вероятность пропуска факта согласованности тоже может оказаться значительной. Поэтому следует учитывать суммарную ошибку, вероятность которой в соответствии с критерием идеального наблюдателя равна

$$E = P_1 + (1 - D).$$

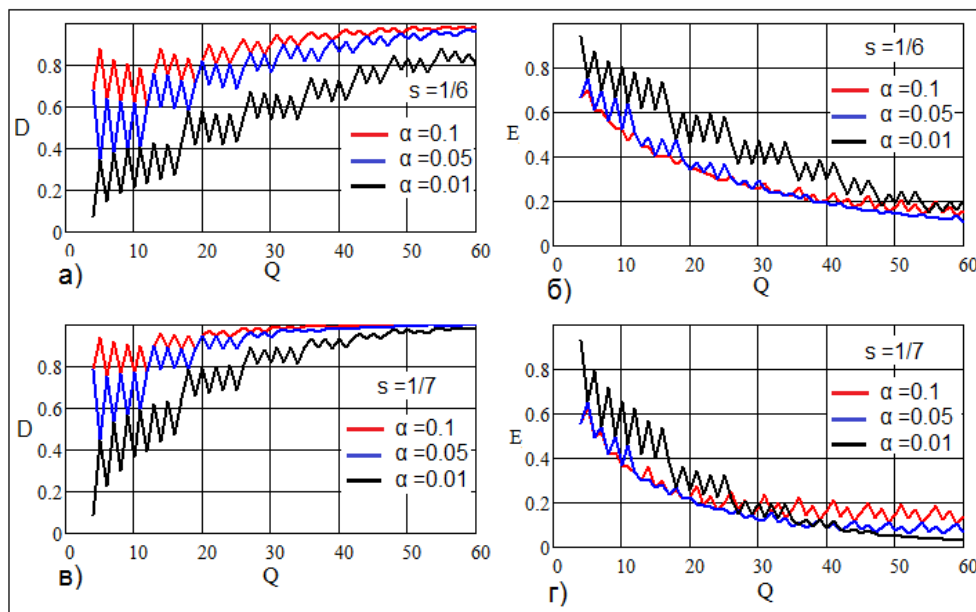


Рисунок 8 – Зависимость вероятностей D и E от числа Q экспертов при различных s и α

Вероятностные характеристики, приемлемые для практики, получаются уже при $Q > 30 \dots 40$. Эти условия не слишком обременительны. Действительно, $Q = 40$, когда, например, четыре эксперта оценивают свойства десяти объектов.

Пример расчета. Рассмотрим порядок расчета на конкретном примере. В качестве экспертных оценок используем данные о результатах аттестации студентов одной из учебных групп. Аттестация проводилась по балльной шкале. В соответствии с принятой в учебном заведении системой для каждой n -й дисциплины установлен свой диапазон шкалы оценивания $[0; L_n]$. Поэтому для обеспечения равной весовой значимости дисциплин, оценки $b_{i,n}$ i -го студента по n -й дисциплине необходимо нормировать: $x_{i,n} = b_{i,n} / L_n$, $i = 1, 2, \dots, K$, $n = 1, 2, \dots, N$. Нормированные оценки приведены в таблице, размещенной в Интернете как дополнительный материал (http://hi-tech.asu.edu.ru/files/other/pdf_001.pdf). Отдельные фрагменты этой таблицы представлены в таблице 2. Данные содержат оценки $K = 34$ студентов по $N = 10$ дисциплинам. Диапазон шкалы оценивания $[A = 50; B = 100]$ ¹. Будем рассматривать i -го студента как i -объект C_i , а n -ую дисциплину D_n как n -го эксперта $Э_n$. Строго говоря, представление «дисциплины» в качестве «эксперта» не совсем корректно, поскольку результаты оценивания зависят не только от мнения преподавателя (эксперта), но и от многих других факторов – сложности дисциплины, качества преподавания, личных взаимоотношений студент – преподаватель и пр. Но для иллюстрации порядка расчета это не существенно.

Предварительные вычисления. Прежде всего, определим границы полуинтервалов. Для этого найдем шаг $\Delta = (B - A) / N = 5$ и рассчитаем границы $z_0 = A$, $z_n = z_0 + n\Delta$, $n = 1, 2, \dots, N$. Численные значения приведены в заголовке таблицы 3. Из таблицы 1 находим критическое значение c_α статистики ξ_i . При $N = 10$ и $\alpha = 0,05$ $c_\alpha = 4$. Это означает, что решение об отсутствии ССЭ ($h_i = 1$) принимается, если $\xi_i \leq 4$.

Таблица 2 – Нормированные оценки знаний студентов

Объекты	Эксперты, $Э_n$									
	Э1	Э2	Э3	Э4	Э5	Э6	Э7	Э8	Э9	Э10
C1	52,00	69,50	87,50	50,00	85,00	84,67	75,00	69,33	100,00	70,00
C2	88,00	76,00	100,00	89,33	85,00	71,33	92,00	70,00	93,33	96,67
C3	80,00	70,00	85,00	70,00	95,00	73,33	92,00	76,67	93,33	96,67
...	...	...	...	...	...	...	...	...	...	...
C9	80,00	50,00	69,00	50,00	90,00	56,00	51,00	70,00	50,00	69,33
C10	96,00	97,50	97,50	89,33	90,00	91,33	95,00	90,00	93,33	100,00
...	...	...	...	...	...	...	...	...	...	...

Обнаружение согласованности результатов экспертизы i -го объекта. Процедура сводится к вычислению статистик ξ_i и индикаторов h_i . Рассмотрим три варианта расчета количества пустых интервалов: «ручным» методом, без применения вспомогательных вычислительных средств; с использованием возможностей программы MS Excel; в программе MathCad.

«Ручной» метод. Просматривая последовательно оценки каждого объекта, размещаем их в соответствующих интервалах. При этом можно записывать не численные значения, а «имена» экспертов, оценки которых попали в интервал. Так в интервал $[50 \dots 55]$ относительно объекта C1 попали оценки экспертов Э1 и Э4 (см. табл. 3). Выполнив эту достаточно утомительную процедуру для всех объектов и всех интервалов, находим количество ξ_i пустых полуинтервалов и определяем значения индикаторов h_i .

Таблица 3 – Результаты «ручного» заполнения полуинтервалов

Объекты	Границы полуинтервалов										ξ_i	h_i
	[50...55]	(55...60]	(60...65]	(65...70]	(70...75]	(75...80]	(80...85]	(85...90]	(90...95]	(95...100]		
C1	Э1, Э4			Э2, Э8, Э10	Э7		Э5, Э6	Э3		Э9	4	1
C2				Э8	Э6	Э2	Э5	Э1, Э4	Э7, Э9	Э3, Э10	3	1
C3				Э2	Э4, Э6	Э1, Э8	Э3		Э5, Э7, Э9	Э10	4	1
...	...	...	...	...	...	...	...	...	...	...	...	...
C9	Э2, Э4, Э7, Э9	Э6		Э3, Э8, Э10		Э1			Э5		5	0
C10								Э4, Э5, Э8	Э6, Э7, Э9	Э1, Э2, Э3, Э10	7	0
...	...	...	...	...	...	...	...	...	...	...	...	...

¹ В таблице (http://hi-tech.asu.edu.ru/files/other/pdf_001.pdf) приведены результаты с учетом пересдач, т. е. нет неудовлетворительных оценок.

Заполнение полуинтервалов в таблице Excel. Процедура существенно упрощается благодаря применению функции СЧЕТЕСЛИ (диапазон; критерий). Эта функция подсчитывает количество ячеек, содержащих числа, удовлетворяющих условию. Пример ее использования приведен на рисунке 9.

	A	B	C	D	E	F	G	H	I	J	K	L
1		Э1	Э2	Э3	Э4	Э5	Э6	Э7	Э8	Э9	Э10	
2	C1	52.00	69.50	87.50	50.00	85.00	84.67	75.00	69.33	100.00	70.00	
3												
4		Границы полуинтервалов										
5	Объекты	[50...55]	[55...60]	[60...65]	[65...70]	[70...75]	[75...80]	[80...85]	[85...90]	[90...95]	[95...100]	ξ_i
6	C1	2	0	0	3	1	0	2	1	0	1	4
7												

$=\text{СЧЕТЕСЛИ}(B2:K2,"<=55")$
 $=\text{СЧЕТЕСЛИ}(B2:K2,"<=65")-\text{СЧЕТЕСЛИ}(B2:K2,"<=60")$
 $=\text{СЧЕТЕСЛИ}(B6:K6,"=0")$

Рисунок 9 – Подсчет числа пустых полуинтервалов в программе Excel

Расчеты в MathCad. Возможности программы значительно шире, чем MS Excel. Все расчеты легко автоматизировать. Листинг расчета количества пустых интервалов ξ и индикаторов h приведен на рисунке 10.

y := Импорт данных из Excel D:\...\Пример.xls		$N := 10$ количество экспертов $K := 34$ количество объектов $n := 0..N-1$ $j := 0..K-1$
$A_s := 1..N-1$ $Z_s := \Delta \cdot (s+1) + A$ границы интервалов $d01_i := \sum_n \text{if}(A \leq y_{i,n} \leq Z_0, 1, 0)$ вектор количество оценок в первом интервале i-го объекта $d02_i := \text{if}(d01_i = 0, 1, 0)$ вектор пустых первых интервалов вспомогательные преобразования $d03 := d02^T$ $d13 := d12^T$ $d3 := \text{stack}(d03, d13)$		$\Delta := \frac{B-A}{N} = 5$ ширина интервала $d11_{i,s} := \sum_n \text{if}(Z_{s-1} < y_{i,n} \leq Z_s, 1, 0)$ матрица количества оценок в s-м интервале i-го объекта $d12_{i,s} := \text{if}(d11_{i,s} = 0, 1, 0)$ матрица пустых s-х интервалов
$\xi_i := \sum d3^{(i)}$ вектор индивидуальных статистик ξ		$\xi := \sum_i \xi_i = 164$ $h_i := \text{if}(\xi_i \leq 4, 1, 0)$ вектор индикаторов h $h := \sum_i h_i = 13$

Рисунок 10 – Программа расчета показателей ССЭ при индивидуальной экспертизе

Результаты обнаружения индивидуальной ССЭ для всех объектов приведены в таблице, размещенной в Интернете в качестве дополнительного материала (http://hi-tech.asu.edu.ru/files/other/pdf_001.pdf).

Обнаружение ССЭ групповой экспертизы. Обнаружение по индивидуальным индикаторам h_j . Частота согласованных индивидуальных решений равна (см. рис. 10) $p^* = 1 - h / K = 1 - 13 / 34 = 0,618$. Нижняя граница доверительного интервала, рассчитанная по формуле (14) при уровне значимости $\alpha = 0,05$ и объеме выборки $K = 34$ равна 0,275. Эта вероятность меньше 0,5, следовательно, для принятия решения о ССЭ недостаточно информации.

Обнаружение по индивидуальным статистикам ξ_i . Статистика $\zeta = \frac{1}{K} \sum_{(i)} \xi_i = 164 / 34 = 4,824$.

Математическое ожидание числа пустых интервалов (4) $m_1 = 3,487$, а дисперсия (5) $\mu_2 = 0,993$. Расчетное отклонение (17) равно $u_{расч} = \frac{4,824 - 3,487}{\sqrt{0,993}} \sqrt{34} = 7,823$. Критическое значение $u_{1-\alpha/2} = 1,96$ при $\alpha = 0,05$.

Расчетное отклонение больше критического. Гипотеза о равномерном распределении экспертных оценок отклоняется. Следовательно, суждения экспертов согласованы. При этом достигнутый уровень значимости (p-value) равен $2,665 \cdot 10^{-15}$ [28].

Обнаружение по объединенным экспертным оценкам. В результате объединения матрица оценок преобразуется в вектор, размерность которого равна $Q = KN = 340$. При этом число полуинтервалов также равно Q . Ясно, что при таком большом числе интервалов подсчет пустых полуинтервалов «ручным» методом и средствами Excel становится малопродуктивным. Однако в программе MathCad решение находится достаточно просто (см. рис. 11). Расчетное отклонение больше критического, гипотеза о равномерном распределении оценок отклоняется. Принимается решение «суждения экспертов согласованы».

$$\begin{aligned}
 Q &:= N \cdot K = 340 & q &:= 0 \dots Q-1 & s &:= 1 \dots Q-1 & \Delta &:= \frac{B-A}{Q} = 0.147 & \text{ширина интервала} & z_q &:= \Delta \cdot (q+1) + A & \text{границы интервалов} \\
 d0 &:= \sum_i \left(\sum_n \text{if}(A \leq y_{1,n} \leq z_0, 1, 0) \right) = 17 & \text{количество оценок в первом интервале} & d1_s &:= \sum_i \left(\sum_n \text{if}(z_{s-1} < y_{1,n} \leq z_s, 1, 0) \right) & \text{матрица количества оценок в s-м интервале} \\
 \xi &:= \text{if}(d0 = 0, 1, 0) + \sum_s \text{if}(d1_s = 0, 1, 0) = 266 & \text{количество пустых интервалов} & m_1 &:= Q \cdot \left(1 - \frac{1}{Q}\right)^Q = 124.895 & \text{математическое ожидание числа пустых интервалов} \\
 \mu_2 &:= Q \cdot (Q-1) \cdot \left(1 - \frac{2}{Q}\right)^Q + Q \cdot \left(1 - \frac{1}{Q}\right)^Q - Q^2 \cdot \left(1 - \frac{1}{Q}\right)^{2Q} = 33.07 & \text{дисперсия числа пустых интервалов} \\
 u_p &:= \frac{\xi - m_1}{\sqrt{\mu_2}} = 24.537 & \text{расчетное отклонение} & p1_v &:= 1 - \text{pnorm}(u1_p, 0, 1) = 0 & \text{достигнутый уровень значимости}
 \end{aligned}$$

Рисунок 11 – Программа расчета показателей ССЭ при групповой экспертизе

Обсуждение результатов. 1. Примером расчета подтверждена реализуемость предложенных процедур обнаружения ССЭ.

2. Методика расчета позволяет оценить адекватность полученных результатов на заданном или достигнутом уровне значимости.

Оценивая итоги анализа конкретных данных, приведенных в качестве дополнительного материала (http://hi-tech.asu.edu.ru/files/other/pdf_001.pdf), следует отметить, что согласованность групповой экспертизы подтверждена на очень высоком уровне значимости. В достоверности этого вывода легко убедиться, сравнив гистограмму распределения числа оценок по интервалам с плотностью равномерного распределения (рис. 12а). Отличие распределения оценок от равномерного еще наглядней подтверждается сравнением кумулятивной кривой и функции распределения равномерного закона (рис. 12б).

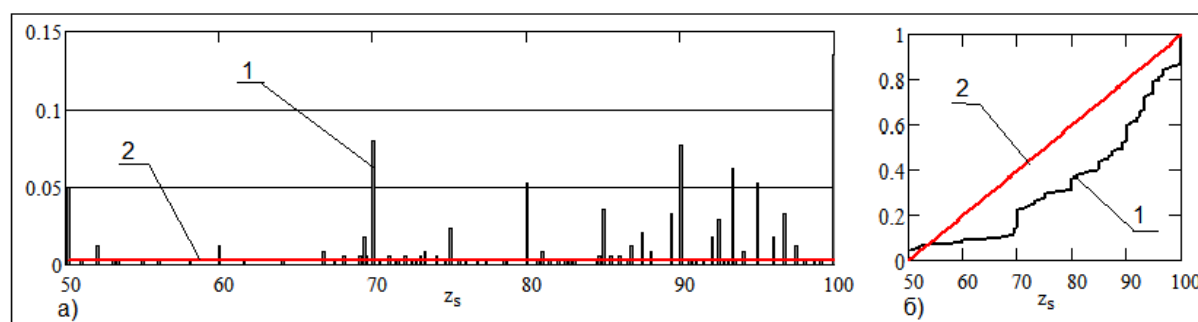


Рисунок 12 – Распределение числа оценок экспертов по интервалам; а) 1– гистограмма, 2– плотность распределения; б) 1– кумулятивная кривая, 2 – функция распределения

Для сравнения полученных результатов с традиционными по данным из приложенной таблицы (http://hi-tech.asu.edu.ru/files/other/pdf_001.pdf) был рассчитан коэффициент конкордации Кендалла. Он оказался равным $W = 0,476$ при расчетном отклонении $\chi^2_{расч} = 156,974$ и критическом значении $\chi^2_{кр} = 47,4$ на уровне значимости $\alpha=0,05$. Принято считать, что суждения экспертов согласованы, если $W > 0,5$. Следовательно, экспертные оценки по данным из приложенной таблицы (http://hi-tech.asu.edu.ru/files/other/pdf_001.pdf) не согласованы. Это противоречит сделанным ранее выводам. Расхождение объясняется, на наш взгляд, тем, что при расчете коэффициента конкордации числовые оценки преобразуются в ранги. Это приводит к потере информации, и, следовательно, к снижению достоверности и точности результата. Кроме того, критерий $W > 0,5$ не имеет какого-либо серьезного статистического обоснования. Смысловая интерпретация коэффициента W достаточно произвольна, что затрудняет принятие решения. Предложенный автором подход исключает неопределенность, поскольку по результатам исследования принимается одно из конкретных решений: суждения экспертов согласованы или не согласованы.

Заключение. Научная новизна исследования состоит в предложении вместо количественной оценки степени ССЭ использовать альтернативную оценку, состоящую в обнаружении факта согласованности путем проверки статистической гипотезы о вероятностном распределении экспертных оценок. Эту гипотезу предложено проверять по комбинаторному критерию «пустых ящиков». В свою очередь, это дает возможность найти корректные, строго обоснованные вероятностные характеристики процедур обнаружения согласованности.

Основными результатами проведенного исследования являются разработанные конструктивные методы обнаружения ССЭ при использовании числовой шкалы оценивания. Предложены процедуры обнаружения ССЭ относительно конкретных объектов (индивидуальная согласованность) и относительно групповой экспертизы в целом. Путем компьютерного имитационного моделирования найдены оценки вероятностных характеристик процедур.

Результаты работы ориентированы на использование при решении широкого круга задач, когда применяется неформализуемая информация, полученная в результате экспертного оценивания. Такие ситуации возникают при прогнозировании и долгосрочном планировании в различных сферах деятельности: экономики и финансов [1, 3, 8, 9, 14, 15, 25], социологии [24, 27], экологии [32, 33], архитектуре [34], медицине [5], международных отношений и многих других.

Применение предложенных методов позволит сократить затраты времени на исследования, поскольку если ССЭ не обнаружена на заданном уровне значимости, то и последующий анализ степени согласованности не имеет смысла. В случае обнаружения согласованности на заданном уровне значимости анализ может быть продолжен с применением других известных методов [18].

Список литературы

1. Афоничкин А. И. Управленческие решения в экономических системах : учебник / А. И. Афоничкин, Д. Г. Михаленко. – Санкт-Петербург : Питер, 2009. – 480 с. – (Серия «Учебник для вузов»). – ISBN 978-5-388-00405-5.
2. Балдин К. В. Концептуальные основы применения метода экспертного оценивания в предпринимательстве : монография / К. В. Балдин. – Ухта : УГТУ, 2016. – 110 с.
3. Бешелев С. Д. Математико-статистические методы экспертных оценок / С. Д. Бешелев, Ф. Г. Гурвич. – Москва : Статистика, 1980. – 263 с.
4. Большев Л. Н. Таблицы математической статистики / Л. Н. Большев, Н. В. Смирнов. – Москва : Наука, 1983. – 416 с.
5. Брумштейн Ю. М. Парк медоборудования в медицинских учреждениях: анализ целей и подходов к управлению; состава информации, учитываемой при принятии решений / Ю. М. Брумштейн, Е. В. Пфандер // Прикаспийский журнал: управление и высокие технологии. – 2017. – № 3. – С. 19–35.
6. Вентцель Е. С. Теория вероятностей / Е. С. Вентцель. – Москва : Наука, 1964. – 576 с.
7. Голик Ф. В. Аппроксимация кривыми Пирсона плотности распределения суммы независимых одинаково распределенных случайных величин / Ф. В. Голик // Кибернетика и программирование. – 2017. – № 2. – С. 17–41. – DOI: 10.7256/2306-4196.0.022583. – Режим доступа: http://e-notabene.ru/kp/article_22583.html (дата обращения 29.09.2017).
8. Данелян Т. Я. Формальные методы экспертных оценок / Т. Я. Данелян // Экономика, статистика и информатика. – 2015. – № 1. – С. 184–187.
9. Евланов Л. Г. Экспертные оценки в управлении / Л. Г. Евланов, В. А. Кутузов. – Москва : Экономика, 2013. – 466 с.
10. Золотарев В. М. Современная теория суммирования независимых случайных величин / В. М. Золотарев. – Москва : Наука, 1986. – 416 с.
11. Кендэл М. Ранговые корреляции / М. Кендэл. – Москва : Статистика, 1975. – 45 с.
12. Колчин В. Ф. Случайные размещения / В. Ф. Колчин, Б. А. Севастьянов, В. П. Чистаков. – Москва : Наука, 1976. – 224 с.
13. Лемешко Б. Ю. Непараметрические критерии согласия : руководство по применению / Б. Ю. Лемешко. – Москва : НИЦ ИНФРА-М, 2014. – 163 с.
14. Литвак Б. Г. Экспертные технологии управления / Б. Г. Литвак. – 2-е изд. – Москва : Дело, 2004. – 398 с.
15. Лукашин Ю. П., Рахлина Л. И. Современные направления статистического анализа взаимосвязей и зависимостей / Ю. П. Лукашин, Л. И. Рахлина. – Москва : ИМЭМО РАН, 2012. – 54 с. – ISBN 978-5-9535-0332-7.
16. Митропольский А. К. Техника статистических вычислений / А. К. Митропольский. – Москва : Наука, 1971. – 576 с.
17. Орлов А. И. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении / А. И. Орлов // Научный журнал КубГАУ. – 2014. – № 97 (03). – С. 1–28.
18. Орлов А. И. Прикладная статистика / А. И. Орлов. – Москва : Экзамен, 2006. – 671 с.
19. Орлов А. И. Структура непараметрической статистики (обобщающая статья) / А. И. Орлов // Заводская лаборатория. Диагностика материалов. – 2015. – № 7, т. 81. – С. 62–72.
20. Орлов А. И. Организационно-экономическое моделирование: теория принятия решений : учебник / А. И. Орлов. – Москва : КНОРУС, 2011. – 568 с.
21. Переверзев Е. С. Случайные процессы в параметрических моделях надежности / Е. С. Переверзев. – Киев : Наук. думка, 1987. – 240 с.
22. Рыжиков Ю. И. Управление запасами / Ю. И. Рыжиков. – Москва : Наука, 1969. – 344 с.
23. Статистические методы анализа экспертных данных. Ученые записки по статистике / отв. ред. Т. В. Рябушкин. – Москва : Наука, 1977.
24. Толстова Ю. Н. Анализ социологических данных / Ю. Н. Толстова. – Москва : Научный мир, 2000.
25. Усманова З. А. Системный анализ факторов и процессов принятия решений при управлении портфелем проектов коммерческого банка с учетом развития информационно-телекоммуникационных технологий / З. А. Усманова, А. А. Ханова // Прикаспийский журнал: управление и высокие технологии. – 2017. – № 2. – С. 58–71.
26. Четыркин Е. М. Статистические методы прогнозирования / Е. М. Четыркин. – 2-е изд., перераб. и доп. – Москва : Статистика, 2007. – 650 с.

27. Шуткин А. Н. Теоретические и экспериментальные модели прогнозирования профессиональной пригодности работников экстремальных профессий / А. Н. Шуткин // Прикаспийский журнал: управление и высокие технологии. – 2017. – № 3. – С. 36–45.
28. Arsham H. Kuiper's P-value as a measuring tool and decision procedure for the goodness-of-fit test / H. Arsham Kuiper's // J. Appl. Statist. – 1988. – Vol. 15, № 2. – P. 131–135.
29. Bagdonavicius V. Non-parametric tests for complete sample / V. Bagdonavicius, J. Kruopis, M. Nikulin. – ISTE-Wiley : Hoboken, 2011. – 308 p.
30. Durre, Alexander. Spatial sign correlation / Alexander Durre, Daniel Vogel, Roland Fried // Journal of Multivariate Analysis. – 2015. – № 135. – P. 89–105
31. Genton Marc G. Robustness properties of dispersion estimators / Marc G. Genton, Yanyuan Ma // Statistics & Probability, Letters. – 1999. – № 44. – P. 343–350.
32. Gong, Zaiwu. On Consistency Test Method of Expert Opinion in Ecological Security Assessment / Zaiwu Gong, Lihong Wang // International journal of environmental research and public health. – 2017. – Vol. 14, № 9.
33. Huber, Julia. Two experts, three opinions: volatile organic compounds' testing methods and regulative systems / Julia Huber, Daniel-Johann, Franziska Hesser // European Journal of wood and wood Products. – 2018. – Vol. 76, issue 1. – P. 5–12.
34. Tsuboi Tsuboi, Shirou. Method for the proposal and evaluation of urban structures for compact cities using an expert system / Tsuboi Tsuboi, Shirou; Ikaruga, Shinji; Kobayashi, Takeshi // Frontiers of Architectural research. – 2016. – Vol. 5, № 4. – P. 403–411.
35. Zhang J. Powerful goodness-of-fit tests based on the likelihood ratio / J. Zhang // Journal of the Royal Statistical Society: Series B. – 2002. – V. 64, № 2. – P. 281–294.
36. Veen, Duco. Proposal for a Five-Step Method to Elicit Expert Judgment / Veen, Duco; Stoel, Diederick; Zondervan-Zwijenburg, Marielle // Frontiers in Psychology. – 2017. – Vol. 8, № 2110.

References

1. Afonichkin A. I., Mihalenko D. G. *Upravlencheskie resheniya v ekonomicheskikh sistemakh* [Management solutions in economic systems], Saint Petersburg, Piter Publ., 2009. 480 p.
2. Baldin K. V. *Kontseptualnye osnovy primeneniya metoda ekspertnogo otsenivaniya v predprinimatelstve* [Conceptual bases of application of a method of expert estimation in business], Uhta, UGTU Publ. House, 2016. 110 p.
3. Beshelev S. D., Gurvich F. G. *Matematiko-statisticheskie metody ekspertnykh otsenok* [Mathematical and statistical methods of expert evaluations], Moscow, Statistika Publ., 1980. 263 p.
4. Bolshev L. N., Smirnov N. V. *Tablitsy matematicheskoy statistiki* [Tables of Mathematical Statistics], Moscow, Nauka Publ., 1983. 416 p.
5. Brumshteyn Yu. M., Pfander E. V. Park medoborudovaniya v meditsinskikh uchrezhdeniyakh: analiz tseley i podkhodov k upravleniyu; sostava informatsii, uchityvaemoy pri prinyatii resheniy [The equipment stock in Russian medical institutions: the analysis of goals and approaches to administration; as well as information to be used in decision-making]. *Pri-kaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Control and High Technologies], 2017, no. 3, pp. 19–35
6. Ventcel E. S. *Teoriya veroyatnostey* [Probability Theory], Moscow, Nauka Publ., 1964. 576 p.
7. Golik F. V. Approksimatsiya krivymi Pirsona plotnosti raspredeleniya summy nezavisimyykh odinakovo raspredelennykh sluchaynykh velichin [Approximation by Pearson curves of the distribution density of a sum of independent identically distributed random variables]. *Kibernetika i programmirovaniye* [Cybernetics and Programming], 2017, no. 2, pp. 17–41. DOI: 10.7256/2306-4196.0.0.22583. Available at: http://e-notabene.ru/kp/article_22583.html (accessed 29.09.2017).
8. Danelyan T. Ya. Formalnye metody ekspertnykh otsenok [Formal methods of expert evaluation]. *Ekonomika, statistika i informatika* [Economics, Statistics and Informatics], 2015, no. 1, pp. 184–187.
9. Evlanov L. G., Kutuzov V. A. *Ekspertnye otsenki v upravlenii* [Expert assessments in management], Moscow, Ekonomika Publ., 2013, 466 p.
10. Zolotarev V. M. Sovremennaya teoriya summirovaniya nezavisimyykh sluchaynykh velichin [The modern theory of summation of independent random variables], Moscow, Nauka Publ., 1986. 416 p.
11. Kendehl M. *Rangovye korrelyatsii* [Rank correlations], Moscow, Statistika Publ., 1975. 45 p.
12. Kolchin V. F., Sevastyanov B. A., Chistakov V. P. *Sluchaynye razmeshcheniya* [Random placements], Moscow, Nauka Publ., 1976. 224 p.
13. Lemesko B. Yu. *Neparametricheskie kriterii soglasiya: Rukovodstvo po primeneniyu* [Nonparametric acceptance criteria], Moscow, NIC INFRA Publ., 2014. 163 p.
14. Litvak B. G. *Ekspertnye tekhnologii upravleniya* [Expert technologies of management], Moscow, Delo Publ., 2004. 398 p.
15. Lukashin Yu. P., Rahlina L. I. Sovremennyye napravleniya statisticheskogo analiza vzaimosvyazey i zavisimostey [Modern directions of statistical analysis of relationships and dependencies], Moscow, IMEHO RAN Publ. House, 2012. 54 p. ISBN 978-5-9535-0332-7.
16. Mitropolskiy A. K. *Tekhnika statisticheskikh vychisleniy* [Technique statistical calculations], Moscow, Nauka Publ., 1971. 576 p.
17. Orlov A. I. Neparametricheskie kriterii soglasiya Kolmogorova, Smirnova, omega-kvadrat i oshibki pri ikh primeneni [Non-parametric criteria of Kolmogorov's agreement Smirnov's agreement omega-square and errors in their application] *Nauchnyy zhurnal KubGAU* [Scientific Journal of KubSAU], 2014, no. 97 (03), pp. 1–28.
18. Orlov A. I. *Prikladnaya statistika* [Applied Statistics], Moscow, Ekzamen Publ., 2006. 671 p.

19. Orlov A. I. Struktura neparametricheskoy statistiki (obobshchayushchaya statya). [Structure of nonparametric statistics (generalizing article)]. *Zavodskaya laboratoriya. Diagnostika materialov* [Industrial laboratory. Diagnostics of materials], 2015, no. 7, vol. 81, pp. 62–72.
20. Orlov A. I. *Organizatsionno-ekonomicheskoe modelirovanie* [Organizational and economic modeling], Moscow, KNORUS Publ., 2009.
21. Pereverzev E. S. *Sluchaynye protsessy v parametricheskikh modelyakh nadezhnosti* [Random processes in parametric models of reliability], Kiev, Nauk. dumka Publ., 1987. 240 p.
22. Ryzhikov Yu. I. *Upravlenie zapasami* [Inventory Management], Moscow: Nauka Publ., 1969. 344 p.
23. Ryabushkin T. V. (ex. ed.) *Statisticheskie metody analiza ekspertnykh dannykh. Uchenye zapiski po statistike* [Statistical methods of analysis of expert data. Scientific notes on statistics], Moscow, Nauka Publ., 1977.
24. Tolstova Yu. N. *Analiz sotsiologicheskikh dannykh* [Analysis of sociological data], Moscow, Nauchnyy mir Publ., 2000.
25. Usmanova Z. A., Khanova A. A. Sistemnyy analiz faktorov i protsessov prinyatiya resheniy pri upravlenii portfelem proektov kommercheskogo banka s uchetom razvitiya informatsionno-telekommunikatsionnykh tekhnologiy [The system analysis of factors and processes of decision-making at control over commercial bank projects portfolio in the environment of information and telecommunication technologies development]. *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Control and High Technologies], 2017, no. 2, pp. 58–71.
26. Chetyrkin E.M. *Statisticheskie metody prognozirovaniya* [Statistical methods of forecasting], Moscow, Statistika Publ., 2007. 650 p.
27. Shutkin A. N. Teoreticheskie i eksperimentalnye modeli prognozirovaniya professionalnoy prigodnosti rabotnikov ekstremalnykh professiy [Theoretical and experimental forecasting model of professional suitability of employees of extreme professions]. *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Control and High Technologies], 2017, no. 3, pp. 36–45.
28. Arsham H. Kuiper's P-value as a measuring tool and decision procedure for the goodness-of-fit test. *J. Appl. Statist.*, 1988, vol. 15, no.2, pp. 131–135.
29. Bagdonavicius V., Kruopis J., Nikulin M. *Non-parametric tests for complete sample*, ISTE-Wiley, Hoboken Publ., 2011. 308 p.
30. Durre, Alexander; Vogel, Daniel; Fried, Roland. Spatial sign correlation. *Journal of Multivariate Analysis*, 2015, no. 135, pp. 89–105.
31. Genton Marc G., Ma Yanyuan. Robustness properties of dispersion estimators. *Statistics & Probability, Letters*, 1999, no. 44, pp. 343–350.
32. Gong, Zaiwu; Wang, Lihong. On Consistency Test Method of Expert Opinion in Ecological Security Assessment. *International journal of environmental research and public health*, 2017, vol. 14, no. 9.
33. Huber, Julia; Daniel-Johann; Hesser, Franziska. Two experts, three opinions: volatile organic compounds' testing methods and regulative systems. *European Journal of wood and wood Products*, 2018, vol. 76, issue 1, pp. 5–12.
34. Tsuboi Tsuboi, Shirou; Ikaruga, Shinji; Kobayashi, Takeshi. Method for the proposal and evaluation of urban structures for compact cities using an expert system. *Frontiers of Architectural research*, 2016, vol. 5, no. 4, pp. 403–411.
35. Zhang J. Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B*, 2002, vol. 64, no. 2, pp. 281–294.
36. Veen, Duco; Stoel, Diederick; Zondervan-Zwijnenburg, Marielle. Proposal for a Five-Step Method to Elicit Expert Judgment. *Frontiers in Psychology*, 2017, vol. 8, no. 2110.

РЕДАКЦИОННЫЙ КОММЕНТАРИЙ К СТАТЬЕ

Задачи экспертного оценивания различных объектов (в т.ч. и возможных вариантов управленческих, технических и иных видов решений) используется очень широко: прежде всего, в тех случаях, когда эти объекты нельзя оценить по совокупности измеримых параметров на основе некоторых интегральных критериев.

При этом важно определить, насколько согласованы мнения экспертов. Если эти мнения сильно различаются, то усредненные (в т.ч. и с применением экспертных коэффициентов) результаты оценивания использовать не целесообразно, т.к. они вряд ли являются показательными. При плохой согласованности мнений экспертов типичными решениями специалистов, организующих экспертное оценивание, обычно являются следующие. (1) Исключение из числа экспертов тех лиц, мнения которых резко отличаются от остальных – если этих лиц немного. (2) Формирование экспертной группы заново и повторное проведение экспертного оценивания.

Основная идея данной статьи является оригинальной, методы достижения поставленной цели используются классические. Изложение носит достаточно подробный характер. Использование математического аппарата и программных средств представлено исчерпывающе.

Теоретическая часть работы дополнена конкретным примером. Расчеты по нему доведены до получения итоговых результатов.

По статье в целом можно сделать следующие замечания. 1) Большое количество источников в списке литературы типа книг, причем некоторые из них имеют достаточно давние года издания. 2) В работе вообще не рассматриваются функциональные возможности программных пакетов, предназначенных для поддержки принятия решений (в т.ч. на основе данных экспертного оценивания), а также пакетов статистического анализа данных. Между тем, на практике большинство пользователей обычно ориентируется на возможности уже существующих общераспространенных программных средств.