

МЕТОД АНАЛИЗА ЭКСПЕРТНЫХ ЗАКЛЮЧЕНИЙ, ПОЛУЧЕННЫХ В РЕЗУЛЬТАТЕ ЭКСПЕРТИЗЫ МНОЖЕСТВА ОДНОТИПНЫХ ПРОЕКТОВ

Ю.А. Шамсутдинов, вед. науч. сотр. ФГБНУ НИИ РИНКЦЭ, канд. техн. наук, snaps7777@extech.ru

С.Е. Ситников, ст. науч. сотр. ФГБНУ НИИ РИНКЦЭ, sitnikov@extech.ru

С.Е. Захарикова, ст. науч. сотр. ФГБНУ НИИ РИНКЦЭ, itks@extech.ru

В статье предложен и опробован метод анализа экспертиз, выполненных на множестве однотипных проектов. Рассмотрен нетрадиционный подход к теории проверки статистических гипотез, связанный с участием в процессе человека.

Ключевые слова: экспертиза, достоверность, статистический критерий, параметр расхождения.

EVALUATION METHOD OF EXPERT-EXAMINATION OF WIDE VARIETY OF STEREOTYPED PROJECTS

Y.A. Shumsutdinov, Leading Researcher, SRI FRCEC, Doctor of Engineering, snaps7777@extech.ru

S.E. Sitnikov, Senior Researcher, SRI FRCEC, sitnikov@extech.ru

S.E. Zakharikova, Senior Researcher, SRI FRCEC, itks@extech.ru

The authors propose and test a method of analysis of examinations performed on a set of similar projects. They consider an unconventional approach to the theory of testing statistical hypotheses associated with participation of a person in the process.

Keywords: expert-examination, reliability, statistical criterion, difference parameter.

Введение

В статье рассмотрен метод анализа экспертных заключений, выполненных на множестве однотипных проектов. Такого рода задачи возникают, например, при конкурсных отборах НИР (НИОКР) в условиях заданного объема финансирования или при выборе конкретной альтернативы среди множества других. Метод выполняет анализ всего массива экспертиз, проведенных по методу оценочного анкетирования, поэтапно. На каждом этапе экспертизы разбиваются на группы, после анализа которых часть из них – отнесенных к «недостоверно оцененным», подвергается дополнительной экспертизе, далее процесс повторяется до достижения заданной степени достоверности их оценки.

Метод базируется на теории проверки статистических гипотез. В силу специфики природы исходных данных, связанной с участием в процессе человека, ее подходы были использованы не традиционным образом.

При классическом подходе проверка гипотезы строится на вычислении критерия (статистики) и сравнения его с критическим, по результатам чего производится принятие гипотезы, в рассматриваемом здесь приложении – отнесение выборки экспертиз к «достоверно оцененным». Данный подход, в силу сложившейся традиции его применения, предполагает достаточно «жесткое» разделение данных (и только на две группы), поскольку уровень значимости, как правило, выбирается на уровне 1–5%, то есть, 95–99 процентного уровня достоверности принятого решения. Кроме того, сам выбор уровня значимости достаточно произволен. Однако, статистические параметры результатов экспертиз, в силу участия в них человека и, как следствие, большой вариабельности в полученных данных, могут существ-

венно расходиться с требованиями к их однородности (в классическом понимании – соответствии выдвинутой гипотезы). Как частный случай, может оказаться так, что они все, или их большая часть, могут не попасть в группу «достоверно оцененных», что приведет к некорректному или завышенно строгому решению задачи разделения. Целесообразно применение более «мягкого» подхода. С этой целью был применен метод, при котором оценке подвергались не сами значения критериев, а анализировалось его изменение. Поиск характерных участков, где этот параметр существенно менялся, а значит, менялся и характер данных, а именно, их однородность и связанная с этим параметром достоверность экспертизы, позволяет определять границы не на основе проверки их соответствия «абстрактным» (табличным) критическим значениям, а на основе критических значений, которые сформированы самими данными и имеют, таким образом, «естественное происхождение» и непосредственную привязку к конкретному массиву данных.

Вторым отличием от классической теории служило то, что в методе предложено разделение экспертиз на 3 группы – «достоверно оцененные», «условно достоверно оцененные» и «недостоверно оцененные», что хорошо согласовано с современными психологическими представлениями о процессе принятия решения экспертом, в котором также предполагается три различные интерпретации принятого им решения. Соответственно, такое представление данных более объективно отражает специфику работы эксперта и должно быть учтено.

В целом, метод позволяет более осознанно и обоснованно оценивать экспертизы, повышая объективность оценки качества их проведения.

Модель эксперта

Представление эксперта-аналитика [1] на базе модели когнитивного управления, которая близка к традиционному представлению эксперта, принятому в теории экспертных оценок или науке управления, основана на формальных методах. Эта модель рассматривала в качестве схемы представления знаний человека (эксперта), некие понятия и теоретические модели, явно или неявно «навязанные» ему в процессе познаний. Это не всегда согласуется с практической деятельностью экспертов, так как никак не учитывает специфических черт деятельности эксперта, в частности, его компетентности и успешности, о чем было отмечено в работе [2], в которой данная модель была изменена с учетом следующего.

Специалист является экспертом в некоторой области, если он:

– обладает совокупностью эвристик, рутинное применение которых в стереотипных для него ситуациях обеспечивает относительно успешное решение задач;

– способен с относительно высокой степенью успеха отличать нестереотипные (и пограничные) ситуации из своей области компетенции и в случаях, когда его стереотипные эвристики неадекватны ситуации – осуществлять творческое решение задачи.

В предложенной модели, таким образом, предполагается, что знание, которое формирует эксперт, действующий в области стереотипных для него ситуаций, является наиболее достоверным, тогда как в случае нестереотипных ситуаций достоверность может снижаться. Выделяется и третий аспект. За рамками областей более или менее результативной (по достоверности) деятельности, в которой человек выступает как эксперт благодаря репродуктивному или творческому переносу профессиональных знаний и умений, лежат области с меньшей достоверностью результатов, в которых возрастает роль немотивированных эвристик. Грубо, это область «обыденного мышления», когда эксперт уже перестает быть экспертом. Применению неадекватных эвристик способствует [3]:

– недостаток времени или познавательных ресурсов для исчерпывающего анализа ситуации, неопределенность, информационная сложность;

– специфика действий стереотипов, запускаемых автоматически без достаточных оснований (предполагается механизм условного рефлекса);

– «факторы веры» – принцип обязательства или занятой позиции, принцип авторитета, научные парадигмы и нормы научных направлений и школ.

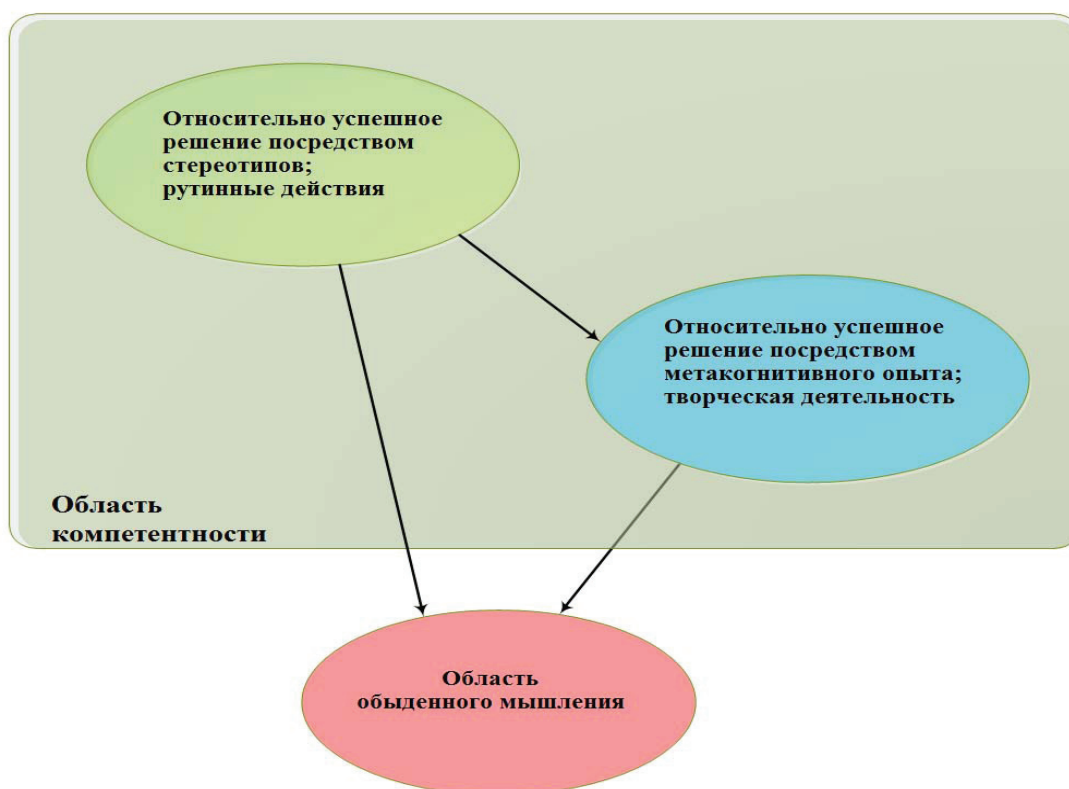


Рис. 1. Обобщенная структура информатики

Таким образом, модельное разбиение областей деятельности эксперта по областям, отличающимся достоверностью результатов, можно представить следующим образом (см. рис. 1).

Такое представление процесса принятия решения экспертом логично транслировать на результат его деятельности и анализировать качество его экспертиз исходя из 3-х возможных вариантов – «достоверно оцененные» (аналог «рутинных действий»), «условно достоверно оцененные» (аналог «творческой деятельности») и недостоверно оцененные (аналог «обыденного мышления»).

Данный подход был учтен при разработке настоящего метода.

Постановка задачи. Рассмотрим наиболее типовую ситуацию. Пусть надо выполнить экспертизу множества проектов – N , выдвинутых на конкурс, по методу оценочного анкетирования. Каждый проект оценивается в баллах из предложенной шкалы, оценку одного проекта выполняют два эксперта, всего их – $2N$. Проект оценивается по итоговому баллу R , рассчитанному как среднее арифметическое значение из двух баллов R_1 и R_2 , соответственно, поставленными первым и вторым экспертами. Требуется выполнить общую оценку достоверности проведенной экспертизы.

Исходные данные для экспертизы. По итогам первичной экспертизы мы имеем следующий набор данных (см. табл. 1).

Выбор параметра для оценки качества экспертизы. Баллы экспертов, включая средние баллы, не могут выступать в качестве параметра оценки достоверности экспертизы в силу того, что они, по определению, являются критериями оценки качества самой работы и не имеют отношения к качеству ее экспертизы.

Оценочным параметром, таким образом, может выступать только расхождение в оценках экспертов по каждой конкретной работе – Δ_i , являющейся случайной величиной.

Таблица 1

Данные первичной экспертизы

Проекты	Балл эксперта 1 (R_1)	Балл эксперта 2 (R_2)	Средний Балл (R)	Расхождение (Δ_i)
1	R_{11}	R_{21}	R_1	Δ_1
2	R_{12}	R_{22}	R_2	Δ_2
...
N	R_{1N}	R_{2N}	R_N	Δ_N

Где $\Delta_i = |R_{1i} - R_{2i}|$.

Допустимость статистической обработки этого параметра требует пояснения и обоснования, поскольку правомерность использования статистических методов должно предполагать использование в качестве исходной информации данных одной природы, в то время как исходные значения расхождений получены на экспертизах, выполненных как на множестве «разнородных» по теме проектов, так и с привлечением множества «разнородных» экспертов и, на первый взгляд, это выглядит разнородной информацией.

Обоснованием могут служить следующие рассуждения. Несмотря на то, что каждый проект является тематически узкоспециализированным, в совокупности с проводящими экспертизу учеными «по этой же специальности», этот тандем – «проект–пара экспертов», позволяет абстрагироваться от уникальности (узкой специализации) темы проекта, поскольку этому соответствует такая же «уникальность» специалистов. Это позволяет свести ситуацию с проведением экспертизы «множества разных проектов с привлечением множества разных экспертов» (случай разнородных данных) к ситуации проведения экспертизы «однотипных проектов однотипными экспертами» (случай однородных данных). Этот ход рассуждений можно довести до логически завершенного – выполнение «экспертизы множества проектов (все по одной тематике) двумя экспертами в области этой тематики» или выполнение «экспертизы одного проекта множеством пар экспертов (все специалисты в области тематики проекта)» (оба случая эквивалентны), когда природа однородности расхождения представляется еще более очевидной.

Таким образом, выбранный параметр расхождения допустимо использовать для оценки качества экспертиз.

Интерпретация параметра расхождения. Параметр расхождения образуется как разница в баллах, выставленных двумя экспертами по одному и тому же проекту, соответственно, величина этой разницы является мерой рассогласованности мнений экспертов или, интерпретируя ее иначе, величиной «неопределенности» при принятии решения одним условным экспертом, когда он выставляет балл, равный среднеарифметическому значению баллов 2-х экспертов, дополняя это решение величиной расхождения, выражая этим свою меру неопределенности или сомнения относительно принятого решения. В этом случае мы видим, что величина расхождения может выступать параметром, характеризующим результат принятого экспертом решения в соответствии с принятой его моделью:

– если Δ мала, условный эксперт принимал решение по «совокупности эвристик, рутинное применение которых обеспечило ему относительно успешное решение задачи» и, таким образом, эту экспертизу можно рассматривать как «достоверно оцененную»;

– если Δ среднее, условный эксперт принимал решение по «стереотипным эвристикам для неадекватной ситуации – осуществляя творческое решение задачи», эту экспертизу можно рассматривать как «условно достоверно оцененную»;

– если Δ большое, решение было принято на уровне «обыденного мышления», эту экспертизу можно рассматривать как «недостоверно оцененную».

Таким образом, параметр расхождения отражает качество экспертизы и может быть использован в методе их разграничения по данному правилу.

Определение взаимосвязи однородности данных с качеством экспертизы и формулирование метода разграничения групп экспертиз. Анализ материалов статистической теории в приложении к оценке экспертизы, выполняемой на условиях, которые были сформированы в постановке задачи, не дает однозначных рекомендаций по выбору конкретных критериев и методов их применения. В этом смысле имеется определенная свобода, как в выборе критериев, так и методов. Выборку расхождений, как и выборку значений баллов экспертов, можно рассматривать в самом общем случае как произвольно распределенные случайные величины, то есть не подчиняющихся какому-либо закону распределения (например, нормальному), поэтому целесообразным видится поиск непараметрических критериев, свободных от привязки к конкретным законам распределения. Среди таких критериев наиболее мощными считаются критерии Смирнова и Лемана–Розенבלата, менее мощным, но часто используемым, Манна–Уитни (Вилкоксона) и др. [4].

Данные критерии могут быть использованы при проверке на однородность двух независимых выборок, в качестве которых могут быть рассмотрены данные оценок 2-х экспертов, на основе которых вычислены значения расхождения. Логика применения критериев здесь может быть следующая. Принимаем, что данные оценок, полученные на экспертизах, имеющих меньшее значение расхождений, являются более однородными. Это может быть объяснено следующим образом. В идеальном случае мнения экспертов совпали и на всей выборке $i = 1, N$, величина $\Delta_i = 0$. Тогда функции распределения, полученные по рассмотренным экспертизам, однозначно считаются однородными. По мере роста расхождений (при вовлечении в рассмотрение данных, полученных на проектах с большим расхождением) логично предположить, что однородность данных будет ухудшаться, соответственно, величина статистики оценки однородности должна расти. В свою очередь, величина расхождений является характеристикой достоверности экспертизы и, таким образом, мы имеем прямую зависимость однородности с достоверностью.

Тогда алгоритм разбиения экспертиз на группы может быть выполнен следующим образом.

Вся совокупность данных (две выборки оценок экспертов, см. табл. 2) ранжируются по величине расхождения и поэтапно вычисляется статистика на выборочных значениях интервала расхождений, которые выбираются начиная от окрестностей в пределах нулевого значения до максимального.

Здесь данные ранжированы согласно значениям Δ_j , где $\Delta_j < \Delta_{j+1}$, соответственно, к выборке « m » относятся экспертизы, имеющие расхождение в оценках менее или равные величине Δ_m , к выборке « k » – экспертизы, попадающие в интервал $\Delta_m < \Delta_j < \Delta_k$, и т. д. В общем случае, величина интервала расхождений может быть выбрана одинаковой для всех выборок. На каждой выборке вычисляется статистика по одному из критериев (например, Смирнова) – Q_m, Q_k, \dots, Q_z .

Предполагаемый характер изменения статистики, таким образом, согласно рассмотренной модели принятия решения экспертом и прямой зависимости оценки однородности со степенью достоверности, может тяготеть либо к варианту монотонного роста статистики (вариант 1), либо к варианту скачкообразных изменений значений статистики (вариант 2), происходящих на границах между разными областями достоверности принятых решений (см. рис. 2). Возможны промежуточные варианты.

То есть, определение границ возможно либо в окрестностях резкого изменения величины статистики, отслеживаемой по ее приращению при переходе от одной выборки к другой (вариант 2), либо, если явно выраженных участков не наблюдается, границы определяются классическим образом, то есть путем сравнения статистик с критическим значением с заданным уровнем значимости, определенным заказчиком (вариант 1). Например, для разделения экспертиз (при втором варианте) на «достоверно» и «условно-достоверно» оцененные может быть выбран уровень значимости на уровне 80 % – доверительного интервала, а для разделения «условно-достоверных» и «недостоверных» 95 % – интервал.

Таблица 2

Выборки оценок экспертов

Пр-ты	R_1	R_2	Δ_i	Выборка «m»	Выборка «k»	...	Выборка «N»	Значение критерия			
								«m»	«k»	...	«N»
~	$R_{1\sim}$	$R_{2\sim}$	Δ_1	1				Q_m			
~	$R_{1\sim}$	$R_{2\sim}$	Δ_2	2							
...							
~	$R_{1\sim}$	$R_{2\sim}$	Δ_m	m				Q_k			
~	$R_{1\sim}$	$R_{2\sim}$	Δ_{m+1}		1						
...						
~	$R_{1\sim}$	$R_{2\sim}$	Δ_k		k			Q_{\sim}			
~	$R_{1\sim}$	$R_{2\sim}$	Δ_{k+1}			1					
...					
~	$R_{1\sim}$	$R_{2\sim}$	Δ			~		Q_z			
~	$R_{1\sim}$	$R_{2\sim}$	$\Delta_{\dots+l}$				1				
...				
~	$R_{1\sim}$	$R_{2\sim}$	Δ_N				z				

Статистика

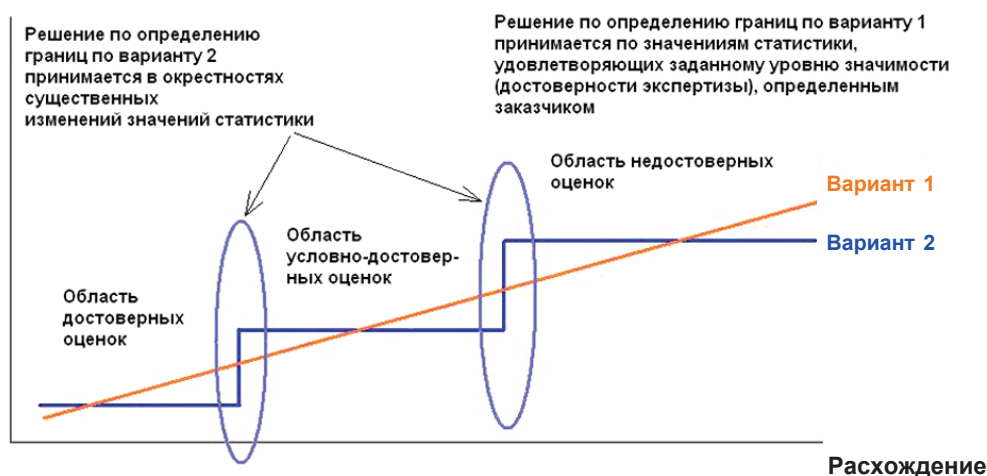


Рис. 2. Предполагаемые (гипотетические) варианты изменения статистики

Можно допустить, что принятые предположения, в первую очередь, об прямой зависимости роста неоднородности данных с ростом параметра расхождения, могут в практических приложениях не проявиться явно или полученные результаты окажутся противоречивыми. Тогда, дополнительным анализом может быть оценка данных на базе параметрических критериев, например, коэффициента корреляции, вычисленном на соответствующих выборках, что разрешало бы возможные противоречия и повышало достоверность полученных результатов.

Проверка метода. Предложенный метод анализа экспертиз был апробирован на реальных результатах экспертизы научных проектов, выдвинутых на конкурс по заданию Минобрнауки. Метод проведения экспертизы была аналогичным рассмотренному в постановке задачи – метод оценочного анкетирования с привлечением на проект 2-х экспертов и выставлением итоговой оценки по среднеарифметическому из 2-х баллов. Заданный диапазон оценок 0...100 баллов.

Был проведен анализ 2-х экспертиз:

– экспертизы 168 проектов в области естественных и технических наук (далее – выборка 1);

– экспертизы 500 проектов в междисциплинарных областях (выборка 2).

Обе экспертизы проводились по одной и той же методике.

Оценки экспертов распределились в пределах 20...100 баллов, диапазон значений расхождения в пределах 0...60 баллов.

В качестве непараметрических критериев использовались следующие: Смирнова, омега-квадрат (Лемана–Розенблатта) и Вилкоксона (Манна–Уитни). Выбор этих методов преследовал цель получить более обоснованную и достоверную оценку выдвинутого в теоретической части метода анализа экспертиз, поскольку, два первых из перечисленных статистических метода проверки гипотез относятся к «мощным», третий же к часто используемому, что определило их выбор. Привлечение всех трех, с другой стороны, учитывало тот факт, что каждый метод имеет свои особенности применения и, чтобы их нивелировать, проверка данных выполнялась по всем трем методам.

Для возможности использования дополнительного параметра анализа – коэффициента корреляции (КК), распределение исходных оценок было проверено на соответствие нормальному закону по критерию Колмогорова [1].

Оценка на нормальность распределения данных выполнялась по формуле

$$\sqrt{n}D_n(\theta^*) = \sqrt{n} \sup_x |F_n(x) - F(x; \theta^*)|,$$

где $F_n(x)$ и $F(x; \theta^*)$ – соответственно, коммулятивные эмпирические и теоретические частоты.

По полученному значения максимума, непосредственно значение критерия вычислялось по формуле

$$\lambda = \frac{D}{\sqrt{n}},$$

который и сравнивался с критическим на предмет проверки гипотезы H_0 , то есть, соответствия распределения оценок экспертов нормальному закону распределения. В отличие от «классического» критерия Колмогорова, где в параметре $\theta = (m, \sigma^2)$ используются истинные значения математического ожидания и дисперсии, был использован параметр $\theta^* = (\bar{x}, s^2)$ где эти величины были заменены на выборочные. То есть, примененный критерий был «типа Колмогорова» и поэтому квантили статистик были выбраны с учетом этого и имели меньшие значения, по сравнению с «классическими». Величины квантилей для разных уровней значимости приведены в табл. 3.

Полученные значения статистики на группах данных сведены в табл. 4.

Как видно, практически на всех интервалах данных, величина $Q_{i...j}$ для обоих выборок имеет значения, меньше квантиля на уровне значимости 0,05 (за исключением одного значения для 2-й выборки), поэтому, с небольшим допущением, с данным уровнем значимости мы можем принять гипотезу, согласно которой наши данные согласованы с нормальным законом распределения и применение параметра – коэффициента корреляции для дополнительного анализа допустимо.

Таблица 3

Величины квантилей для разных уровней значимости

Уровень значимости λ	0,15	0,1	0,05	0,025	0,01
Квантили для $\sqrt{n}D_n(\theta^*)$	0,775	0,819	0,895	0,955	1,035

Таблица 4

Статистика на группах данных

Значение критерия, $Q_{i...j}$	Интервалы, $\Delta_{i...j}$				
	$\Delta_{0...10}$	$\Delta_{10...20}$	$\Delta_{20...30}$	$\Delta_{30...45}$	$\Delta_{45...60}$
Выборка 1	0,71	0,58	0,29	0,81	0,67
Выборка 2	0,68	0,87	0,33	0,74	1,20

Обработка данных Значения двух выборочной статистики Смирнова [6] вычислялись по выражению

$$\lambda' = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \times \max |F_{n_1}(x) - F_{n_2}(x)|,$$

где $F_{n_1}(x)$ и $F_{n_2}(x)$ – эмпирические функции распределения, построенные по двум выборкам с объемами n_1 и n_2 .

Значения двух выборочной статистики по критерию типа омега-квадрат (Лемана–Розенблатта) [5] вычислялись по формуле

$$A = \frac{n_1 n_2}{n_1 + n_2} \int_{-\infty}^{\infty} (F_{n_1}(x) - F_{n_2}(x))^2 dH_{n_1+n_2}(x),$$

где $H_{n_1+n_2}(x)$ – эмпирическая функция распределения, построенная по объединенной выборке:

$$H_{n_1+n_2}(x) = \frac{n_1}{n_1 + n_2} F_{n_1}(x) + \frac{n_2}{n_1 + n_2} F_{n_2}(x).$$

Статистика Z двух выборочного критерия Манна–Уитни [7] определялась следующим образом

$$Z = \frac{U - EU}{\sqrt{DU}} = \frac{R_2 - \frac{n_2(n_2+1)}{2} - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2}{12}(n_1 + n_2 + 1)}}$$

где U – число инверсий, EU – среднее число инверсий, DU – дисперсия инверсий.

Полученные значения статистик на указанных интервалах сведены в табл. 5.

В таблице λ'_1 , A_1 и Z_1 – соответственно, статистики Смирнова, Лемана–Розенблатта и Манна–Уитни для 1-й выборки, с индексом два – для 2-й.

Таблица 5

Статистика на указанных интервалах

Критерий	$\Delta i \dots j$				
	$\Delta_{0 \dots 10}$	$\Delta_{10 \dots 20}$	$\Delta_{20 \dots 30}$	$\Delta_{30 \dots 45}$	$\Delta_{45 \dots 60}$
λ'_1	0,101	0,106	0,353	0,308	0,600
λ'_2	0,039	0,1	0,065	0,141	0,260
A_1	0,063	0,049	0,706	0,090	1,130
A_2	0,052	0,151	0,069	0,231	0,215
Z_1	0,130	0,087	3,010	0,564	0,000
Z_2	0,252	0,562	0,011	0,860	0,197

Объединенный график для выборки 1 представлен на рис. 3, для выборки 2 – на рис. 4.

Предварительный анализ данных по графикам показывает несостоятельность критерия Манна–Уитни – сильный разброс значений и нулевое значение на выборке 1 на интервале 45–60 (резкое падение на выборке 2), что существенно расходится с величинами статистик, полученных по 2-м другим критериям. Следует отметить, что и в работе [6] подчеркивалось, что именно 2 критерия – Смирнова и Лемана–Розенблатта, являются состоятельными в отличия от множества других, включая критерий Манна–Уитни. Исходя из этого, будем опираться в дальнейшем анализе на данные, полученные по данным 2-х критериев, а данные по критерию Манна–Уитни – исключим.

Обобщенный график тогда для обоих выборок, представленный на рис. 5 отражает данные по двум критериям – Смирнова и Лемана–Розенблатта.

Анализ результатов. Из анализа графиков на рис. 3, 4 и 5 мы видим, что на обоих выборках проявляются 3 характерных участка, на границе которых происходят практически синхронные и существенные изменения статистик по обоим критериям.

Для выборки 1 это:

- 0 ... 20;
- 20 ... 30–45;
- более 30–45.

Для выборки 2:

- 0 ... 10;
- 10 ... 30–45;
- Более 30–45.

Не будем пока акцентировать внимание на факте смещенности данных у обоих выборок (график на рис. 5 это хорошо выявляет) и на том, что значения для второй выборки оказались меньшего значения. Это будет проанализировано ниже. Здесь только заметим, что эти расхождения не противоречат предложенному методу и имеют логичное объяснение, которое будет представлено ниже.

Визуальный анализ графиков показывает три явно выраженных участка, неоднородность данных в которых, оцененных по величине рассогласования, существенно отличаются. Можно сказать, что через эти данные рассмотренная модель эксперта проявляется прямым образом.

Так, при анализе данных 1-й выборки мы наблюдаем, что в проектах, где расхождение во мнении экспертов не превысило 20 баллов, значение статистики (по обоим критериям) минимально, что свидетельствует об однородности данных, или, как следует из предыдущих рассуждений, о достоверности экспертиз, то есть, эти экспертизы могут быть отнесены к «достоверно оцененным». Соотнеся эту оценку с моделью эксперта, здесь условный эксперт принимал решение в «области стереотипных для него ситуаций, в которых оценка считается наиболее достоверной». Для второй выборки это интервал расхождений составил 0 ... 10 баллов.

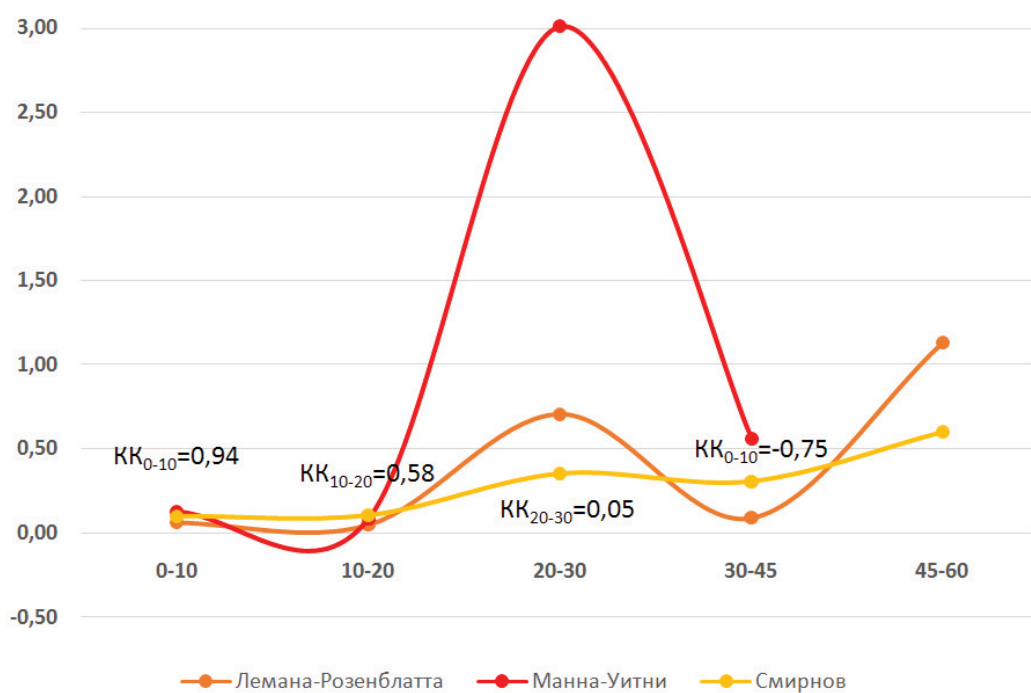


Рис. 3. Общий график статистик Лемана–Розенблатта, Манна–Уитни и Смирнова на выборке 1

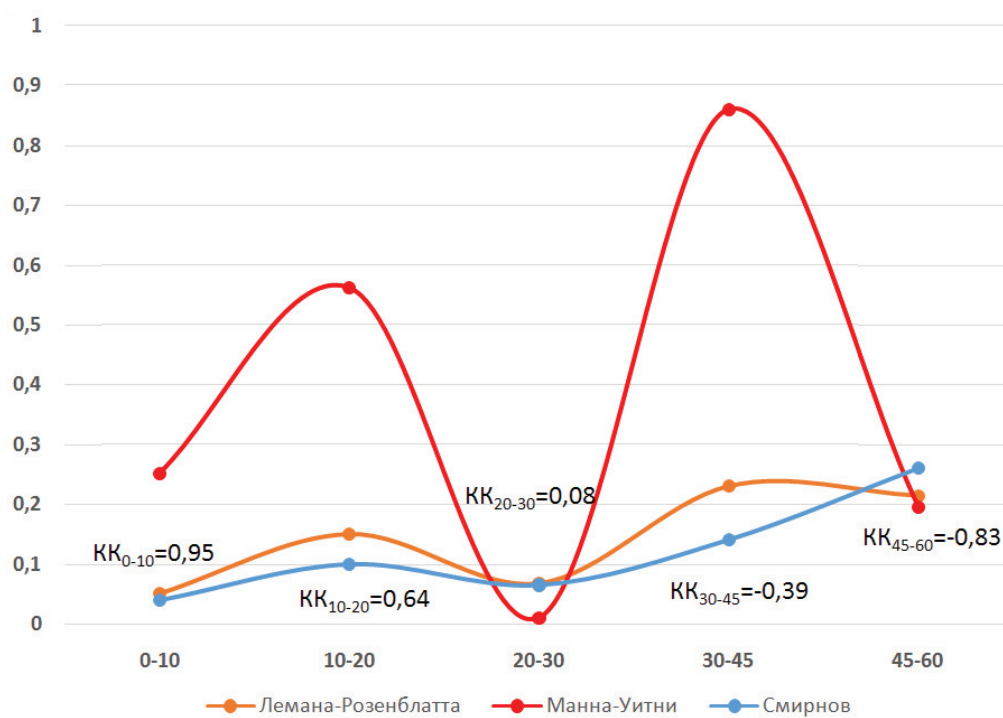


Рис. 4. Общий график статистик Лемана–Розенблатта, Манна–Уитни и Смирнова на выборке 2

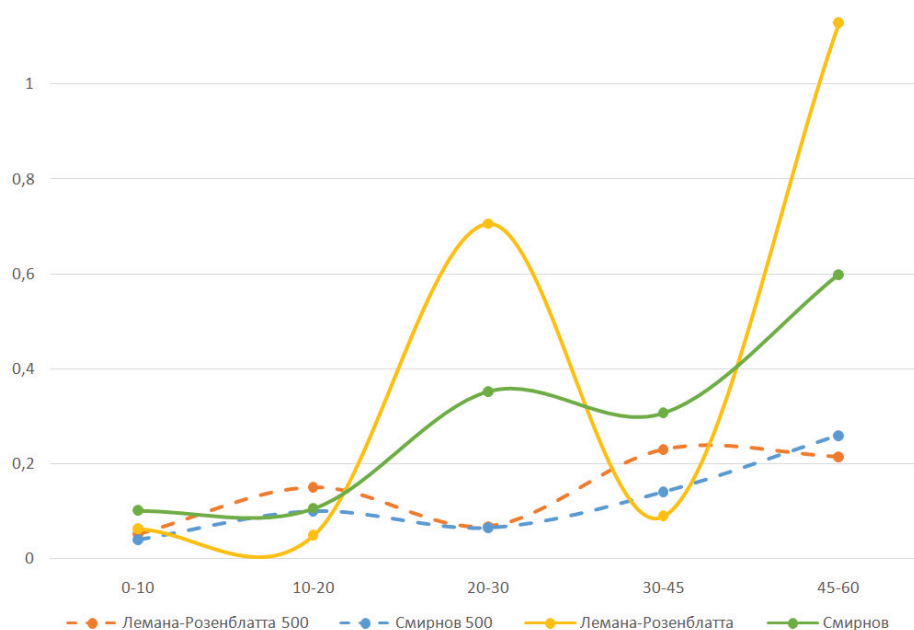


Рис. 5. Общий график статистик Смирнова и Лемана–Розенблатта для двух выборок

В экспертизах с расхождением в пределах 20 ... 40 (второе значение взято примерно средним по рассмотренному интервалу 30 ... 45) видно нарушение однородности данных, т. е. значение статистики возрастает, при этом с явной границей справа, где ее значения несколько снижаются. Соотнеся эту оценку с моделью эксперта, здесь у условного эксперта «стереотипные эвристики уже неадекватны ситуации – он осуществлял творческое решение задачи». Для второй выборки это интервал – 10–30.

Наконец в экспертизах, имеющих величину расхождения в 40 и более баллов, видна явная неоднородность в распределении оценок. Соотнеся эту оценку с моделью эксперта, это, очевидно, область «обыденного мышления», когда эксперт уже перестает быть экспертом. Для второй выборки это соответствует величине расхождения в 30 и более баллов.

Дополнительным фактором, свидетельствующим о непротиворечивости предложенному подходу полученных результатов, служат величины коэффициентов корреляции, также оцененных на данных интервалах. Как видно (рис. 3, 4), на проектах, отнесенных к достоверно оцененным, $КК \approx 0,95$, что свидетельствует о том, что информация о проекте (не важно какого качества была его содержательная часть) была исчерпывающей, знания экспертов соответствовали темам проектов и субъективные факторы сказались минимально, так что мнения практически совпали. На проектах, отнесенных к оцененным «недостоверно», $КК \approx -0,80$, что так же свидетельствует о зависимости мнений экспертов, только отрицательной. Эта зависимость имеет противоположную (в отличии от рассмотренной на достоверно оцененных проектах) природу. Если попытаться интерпретировать этот результат с позиции модели эксперта, оценившего проект на основе «обыденного мышления», то оно сформировано на основе того, что на мнение одного из экспертов (или, возможно, обоих) повлияли факторы, способствующие проявлению неадекватных эвристик (недостаток времени или познавательных ресурсов для исчерпывающего анализа ситуации, неопределенность, информационная сложность, специфика действий стереотипов, запускаемых автоматически без достаточных оснований, принцип обязательства или занятой позиции, принцип авторитета, научные парадигмы и нормы научных направлений и школ). То есть, эта зави-

симось имеет в своей основе «неадекватность» или противоречивость оценки, что и проявилось в отрицательном значении КК. В области «условно достоверных экспертиз» КК имеет промежуточное значение.

Поясним логичность обнаруженных отличий в выявленных интервалах и поведении статистик на различных выборках. Как указывалось, ранее, 1-я выборка была сформирована на 168 оценках экспертиз проектов по естественно научным и техническим темам, 2-я – на 500 оценках экспертиз проектов по междисциплинарным темам.

1. По поводу меньших значений величин статистик на выборке из 500 проектов (выборка 2). Каждый балл, выставленный экспертом может рассматриваться как некая сумма составляющих. Например, итоговый балл суммировался из баллов, выставленных по 7-10 критериям (статус научного руководителя, научный задел, уровень коллектива, пр.), в каждом критерии учитывались дополнительные параметры оценки (количество патентов, статей, индекс цитирования, пр.). Если допустить вклад составляющих баллов в итоговую оценку как равноценное, то согласно центральной предельной теореме, при росте числа наблюдений, распределение частот конечной наблюдаемой величины стремится к нормальному распределению. То есть, распределения оценок в обеих выборках (первого и второго эксперта) имея, отличные друг от друга законы вероятности распределений, при увеличении выборки стремятся к нормальному закону распределения. Тем самым, появляется дополнительный фактор, который повышает их однородность в общем смысле – обе выборки «нормализуются». В этом смысле уменьшение величин статистик, отражающих величину однородности данных, на всем интервале расхождений выглядит закономерным.

2. Смещенность данных. Проекты второй выборки (по междисциплинарным темам) имели гораздо больший разброс в мнениях экспертов, поскольку для экспертиз привлекались эксперты из смежных областей, и у них естественным образом было больше экспертиз, попавших в интервалы с большим расхождением. Это прямым образом отразилось в смещении графиков, так, что области «условно достоверных» и «недостоверно оцененных» имели оценку неоднородности уже на меньших интервалах расхождений.

Таким образом объясняются как меньшие значения статистик на большей выборке, так и смещенность оценки однородности (достоверности), связанная с субъективными факторами (в данном случае, с другим характером проектов, представленных на экспертизу). В целом же, метод показал согласованные результаты в обоих случаях.

В итоге, мы видим, что проверка предложенного метода согласуется с представлением результата экспертиз в виде явно выраженных 3-х групп, каждая из которых характеризуется (в смысле «достоверности») определенным значением статистики, которое в определенных окрестностях анализируемого параметра (расхождения) относительно существенно меняется, определяя границы разделения групп. Практическим следствием такого подхода является анализ 3 групп, к каждой из которых могут быть применены различные действия:

- проекты со статусом «достоверно оцененные» могут быть приняты как прошедшие экспертизу безусловно и не требующие дополнительных проверок;
- проекты со статусом «недостоверно оцененные» безусловно требуют дополнительной экспертизы;
- проекты со статусом «условно достоверно оцененные» требуют уточнения или дополнительного анализа, что может быть определено исходя из временных, организационных или финансовых ресурсов, выделенных на экспертизу, но в любом случае, они больше относятся к «достоверным», нежели к «недостоверным» экспертизам.

Выводы. В работе предложен и опробован метод анализа экспертиз, выполненных на множестве однотипных проектов. Метод использует изменение величины статистического критерия, применяемого в теории проверки статистических гипотез, в качестве определяющего параметра для оценки и разделения экспертиз по качеству их проведения. В методе учтена модель эксперта, предполагающая разделение качества принятого им решения на

три категории, что дает более объективную оценку проведенной экспертизы и позволяет производить ее оценку с большей долей достоверности. Практическая проверка метода на реальных данных показала его состоятельность, требующую, безусловно, дополнительной проверки на подобных задачах.

В статье приведены результаты, полученные при выполнении работ в рамках Государственного задания 2015/Н7 Минобрнауки России по теме № 3246.

Список литературы

1. Абрамова Н.А., Вассунов И.В. О влиянии формализации на адекватность когнитивной модели управленческой ситуации // Труды 5-й международной конференции «Когнитивный анализ и управление развитием ситуаций» CASC' 2005/ИПУ РАН. Москва. 2005, стр. 47–51.
2. Абрамова Н.А. О проблеме рисков из-за человеческого фактора в экспертных методах и информационных технологиях // Проблемы управления. 2007. № 2. Available at: <http://cyberleninka.ru/article/n/o-probleme-riskov-iz-za-chelovecheskogo-faktora-v-ekspertnyh-metodah-i-informatsionnyh-tehnologiyah> (дата обращения: 12.08.2014).
3. Челдани Р. Психология влияния. Санкт-Петербург. Питер, 2001, 270 стр.
4. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении. Научный журнал КубГАУ, № 97(03), 2014.
5. Орлов А.И. Состоятельные критерии проверки абсолютной однородности независимых выборок. Журнал «Заводская лаборатория. Диагностика материалов». 2012. Т. 78. № 11. С. 66–70.
6. Орлов А.И. Математика случая: Вероятность и статистика – основные факты. Учебное пособие. М.: МЗ-Пресс, 2004.
7. Алексеева Н.П. Введение в статистические методы. Учебное пособие по прикладной статистике. СПбГУ, 2010.

References

1. Abramova N.A., Vassunov I.V. (2005) *O vliyaniy formalizatsii na adekvatnost' kognitivnoy modeli upravlencheskoy situatsii. Trudy 5-y mezhdunarodnoy konferentsii «Kognitivnyy analiz i upravlenie razvitiem situatsiy» CASC' 2005* [On the influence of formalization on the adequacy of the cognitive model of managerial situation. Proceedings of the 5th International Conference «Cognitive analysis and development management situations» CASC '2005]. IPU RAN [ICS RAS]. Moscow, pp. 47–51.
2. Abramova N.A. (2007) *O probleme riskov iz-za chelovecheskogo faktora v ekspertnykh metodakh i informatsionnykh tekhnologiyakh* [On the problem of risks due to human factors in expert methods and information technologies]. *Problemy upravleniya* [Control], no 2. Available at: <http://cyberleninka.ru/article/n/o-probleme-riskov-iz-za-chelovecheskogo-faktora-v-ekspertnyh-metodah-i-informatsionnyh-tehnologiyah> (date of treatment: 08/12/2014).
3. Cheldani R. (2001) *Psikhologiya vliyaniya* [Psychology of influence]. St. Petersburg, 270 p.
4. Orlov A.I. (2014) *Neparametricheskie kriterii soglasiya Kolmogorova, Smirnova, omega-kvadrat i oshibki pri ikh primeneni* [Nonparametric Kolmogorov-Smirnov test, and omega-square errors in their application]. *Nauchnyy zhurnal KubGAU* [Science magazine KubGAU], no. 97(03).
5. Orlov A.I. (2012) *Sostoyatel'nye kriterii proverki absolyutnoy odnorodnosti nezavisimyykh vyborok* [Wealthy testing criteria of absolute uniformity independent samples]. *Zhurnal «Zavodskaya laboratoriya. Diagnostika materialov»* [Journal «Plant laboratory. Diagnosis materials»], vol. 78, no. 11, pp. 66–70.
6. Orlov A.I. (2004) *Matematika sluchaya: Veroyatnost' i statistika – osnovnye fakty. Uchebnoe posobie* [Mathematics cases: Probability and Statistics – Key Facts: Textbook]. MZ-Press [MH-Press]. Moscow.
7. Alekseev N.P. (2010) *Vvedenie v statisticheskie metody. Uchebnoe posobie po prikladnoy statistike* [Introduction to statistical methods. Tutorial on Applied Statistics]. SpbGU [St. Petersburg State University].