

УДК 621.791(047.1)(476)

П. С. Серенков,  
В. М. Романчак,  
Н. Н. Иванова

## ИССЛЕДОВАНИЕ ДОСТОВЕРНОСТИ МЕТОДОВ ЭКСПЕРТНОГО ОЦЕНИВАНИЯ

Это вторая статья цикла по исследованию, анализу и прогнозированию достоверности экспертных оценок показателей качества. Предметом исследования выступают факторы «управления» процессом оценивания, в частности, методы его организации и обработки результатов опроса экспертов.

Установлено, что из всех факторов «управления» наибольшее влияние на достоверность экспертных оценок оказывают тип оценочной шкалы, способ представления объектов эксперту, способ оценивания. Исследование степени влияния факторов проводилось методом морфологического ящика. Реализована модель отсеивающего эксперимента, направленная на выявление методов экспертного оценивания, имеющих наибольший относительный уровень доверия. Представлены характеристики воспроизводимости и смещения по каждому из методов, определены их особенности и закономерности. Приведены рекомендации по выбору и применению альтернативных методов экспертного оценивания.

This is the second article in the cycle on examinations, analysis and forecasting of the authenticity of quality index expert assessments. The subject of examination is presented by factors of «management» of assessment process, in particular, methods of its organization and results of experts survey.

It is determined, that from all «management» factors the most impact on expert assessments authenticity comes from assessment scale type, method of object providing to an expert and assessment method. Examination of the factors impact degree was carried out using morphologic box method. We realized an eliminating experiment model destinated to reveal expert assessment methods that have the maximum relative confidence level. This article provides characteristics of reproducibility and shift for each method and determines their particularities and regularities. It also provides recommendations on selection and application of alternate expert assessment methods.

### Проблемная ситуация

В предыдущей статье цикла [1] с помощью применения системного подхода к процессу экспертного оценивания было установлено пять основных источников потерь достоверности экспертных оценок (рис. 1): контекст модели, входы модели, структура процесса, ресурсы (подходы, методы и средства, шкалы, персонал), выход.

Проведен разведочный анализ процесса экс-

пертного оценивания с целью исследования влияния на достоверность результата «входов» модели, а именно структуры и объема оцениваемой информации одного объекта (однопараметрический, многопараметрический) и количества оцениваемых объектов. Также обоснован метод определения достоверности экспертных оценок («выхода» процесса), основанный на применении методов обработки измерительной информации, приведенных в стандартах серии СТБ ИСО 5725 (рис. 1).



Рис. 1. Экспертное оценивание как процесс преобразования «входа» в «выход»

В настоящей статье предметом нашего внимания выступают факторы «управления» процесса оценивания, в частности, методы его организации и обработки результатов опроса.

В настоящее время существует огромное многообразие комбинаций факторов «управления», что, в свою очередь, создает комплексные источники потерь достоверности, количественному анализу которых посвящена статья.

**Исследование влияния факторов «управления» процессом экспертного оценивания на достоверность результатов**

На начальном этапе путем опроса экспертов было выявлено пять факторов «управления», влияющих на достоверность результатов процесса экспертного оценивания. Факторы были ранжированы по степени значимости и построена диаграмма Парето [2], позволившая выделить среди них наиболее значимые (рис. 2).



Рис. 2. Диаграмма Парето по типам источников потерь достоверности результатов экспертного оценивания (в % от общего числа)

- 1 — способы предъявления объектов;
- 2 — способы оценивания; 3 — типы оцениваемой шкалы; 4 — сложность объекта оценивания; 5 — методы обработки оцениваемой информации

Метод Парето позволяет утверждать, что три фактора из всех представленных оказывают наибольшее влияние на достоверность экспертных оценок: 1) тип оценочной шкалы, 2) способ предъявления объектов эксперту, 3) способ оценивания.

В качестве метода исследования степени влияния отобранных экспертами факторов был выбран метод морфологического ящика [3]. Структура морфологического ящика «Методы экспертного оценивания» представлена на рисунке 3. Каждая ячейка морфологического ящика характеризуется тремя свойствами (координатами):

- тип оценочной шкалы,

- способ предъявления объектов оценивания,
- способ оценивания.

**Тип оценочной шкалы**

Анализировались два типа шкал: априорные и апостериорные шкалы.

Апостериорные шкалы широко используются в психофизических исследованиях [4]. Отличительная особенность шкал этого типа: эксперты формируют ее по своим собственным правилам непосредственно в процессе последовательного предъявления эксперту ряда альтернативных объектов оценивания. Единственное ограничение – шкала должна быть цифровой. Установлено, что возможность экстраполировать шкалу в обе стороны приводит к тому, что шкала в представлении эксперта становится «неравномерной». На концах шкалы, где свойства объектов выражены наиболее сильно и наиболее слабо, интервалы соседних оценок «расширяются» [4]. Очевидно, это формирует методическую составляющую неопределенности оценок в этой шкале.

Априорные шкалы общеприняты в квалиметрии [5]. Их отличительная особенность в том, что они формируются заранее, по определенным правилам: обычно задают минимальное и максимальное значения шкалы (например, 1 и 10 баллов соответственно), цену деления шкалы (например, 1 балл). Установлено, что априорная шкала также является «неравномерной», только на этот раз на границах шкалы интервалы соседних оценок «сжимаются», что также формирует методическую составляющую неопределенности оценок [4].

**Способ предъявления объектов на оценивание**

Объекты могут быть предъявлены эксперту двумя способами: упорядоченно или в случайном порядке. В первом случае объекты выстраиваются строго по возрастанию или убыванию оцениваемого свойства. Во втором случае объекты предъявляются рандомизированно.

**Способ оценивания**

По способу оценивания объектов различают методы непосредственного оценивания и парных сравнений.

Метод непосредственного оценивания предполагает, что каждому объекту эксперт присваивает определенную оценку. Метод парных сравнений предполагает, что эксперту предъявляются объекты парами, и он оценивает степень различия между ними. При этом возможны два способа реализации данного метода: 1) сравнение «каждого с предыдущим», 2) сравнение «каждого с одним», причем в качестве последнего может быть выбран наихудший или наилучший по характеризваемым свойствам объект. Эксперт при оценивании может отвечать на вопросы: насколько свойство одного объекта превосходит свойство другого? или во сколько раз свойство

одного объекта превосходит свойство другого? То есть могут оцениваться как разность свойств, так и их отношение.

Метод морфологического ящика выбран потому, что позволяет скомпоновать все возможные комбинации методов организации процесса и оценивания результатов. После комбинирования всех возможных вариантов было получено 12 возможных разновидностей методов экспертного оценивания (рис. 3).

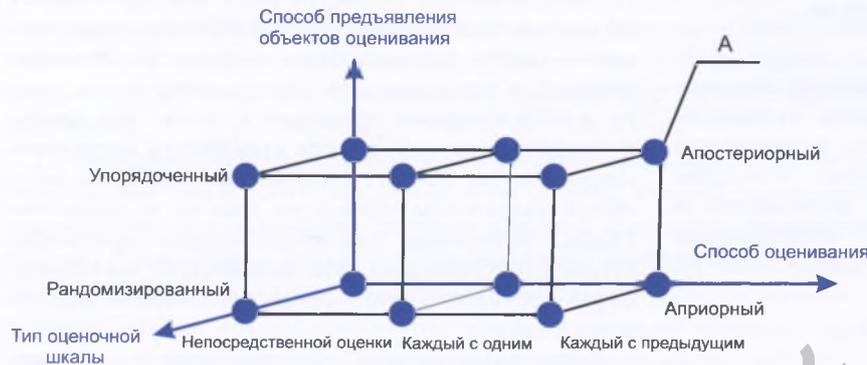


Рис. 3. Морфологический ящик «Методы экспертного оценивания»

A – пример точки морфологического ящика: метод парных сравнений, способ оценивания – «каждый с предыдущим» с использованием апостериорной шкалы, порядок предъявления объектов оценивания – упорядоченный

Для ранжирования полученных 12 комбинаций был спланирован и реализован отсеивающий эксперимент.

**Планирование отсеивающего эксперимента**

В предыдущей статье [1] был обоснован и реализован оригинальный модельный эксперимент по исследованию процесса экспертного оценивания объектов – геометрических фигур. На основании тех же доводов в данном случае объектами оценивания также выбраны геометрические фигуры – круги различного диаметра. Оцениваемым параметром выступает величина круга, которую косвенно моделирует значение его площади. Массив геометрических фигур, участвующих в эксперименте, приведен на рисунке 4.



Рис. 4. Массив фигур типа «круг», участвующих в эксперименте по экспертному оцениванию

**Примечание**

Различные значения оцениваемого параметра фигуры можно интерпретировать как различную степень выраженности свойств оцениваемого объекта, т. е. различный уровень качества оцениваемого объекта.

Следует подчеркнуть, что выбор объекта только типа «круг» связан с чистотой эксперимента. Круг – однопараметрический объект оценивания, позволяющий минимизировать воздействия сторонних влияющих факторов.

Респондентами при выполнении экспериментов, как и в предыдущей серии исследований [1], выступали представители тех же фокус-групп.

**Реализация эксперимента**

В качестве примера рассмотрим исследование одного из 12 методов морфологического ящика: метода парных сравнений с использованием апостериорной шкалы, способ оценивания – «каждый с одним», порядок предъявления объектов оценивания – рандомизированный.

Так как шкала апостериорная, то каждый респондент выбирает самостоятельно удобную для себя шкалу оценивания. Для каждого эксперта эксперимент проводится в два тура. Для этого заранее заготовлены две колоды карточек с изображенными кругами различного размера (оцениваемыми объектами), перемешанными в случайном порядке. Задача каждого эксперта в каждом туре: оценить различие величин сравниваемых кругов, характеризуя его соответствующим числом априорной шкалы, и занести в таблицу для записи результатов.

Результаты опроса двух экспертов по двум турам представлены в таблице 1.

Обработка результатов эксперимента проводилась в 4 этапа.

**Этап 1**

Так как исходные данные (таблица 1) характеризуют различие величин геометрических фигур, абсолютные бальные оценки, характеризующие абсолютную величину каждой фигуры, находились путем решения системы уравнений:

$$(x_i - x_j)/h = r^{ij}, j \neq i, \tag{1}$$

где  $h$  – константа, которая зависит только от данного альтернативного плана.

Для нашего случая система уравнений имеет вид:

$$\begin{cases} x_2 - x_1 = r^{21} \\ x_3 - x_1 = r^{31} \\ \dots \\ x_{15} - x_1 = r^{151} \end{cases}, \tag{2}$$

Таблица 1

Лист записи результатов экспертного оценивания оценок различия величин геометрических фигур

Номера фигур для парного сравнения	1 тур		Номера фигур для парного сравнения	2 тур	
	Оценка степени различия величин фигур			Оценка степени различия величин фигур	
	1-й эксперт	2-й эксперт		1-й эксперт	2-й эксперт
14-1	300	80	14-15	10	5
12-1	280	70	8-15	60	70
6-1	180	45	9-15	40	60
13-1	300	90	11-15	10	85
2-1	50	10	2-15	220	50
10-1	290	80	3-15	210	30
8-1	280	65	5-15	80	10
7-1	180	40	10-15	40	30
9-1	270	75	7-15	60	80
11-1	320	90	1-15	250	70
3-1	60	30	4-15	120	30
5-1	100	55	12-15	20	20
4-1	80	60	13-15	50	40
15-1	320	95	6-15	60	95

где  $x_1, x_2, \dots, x_{15}$  – неизвестные, обозначающие оценки величин фигур или их площади;  $r^{21}, r^{31}, \dots, r^{151}$  – балльные оценки экспертов ( $r^{ij}$ ), характеризующие различие величин геометрических фигур.

Параметры шкалы: масштабный коэффициент  $h = 1$ , условное начало шкалы  $a = 0$ .

Для данных первого тура первого эксперта (таблица 1) решение системы уравнений (2) выглядит следующим образом:

$$\begin{cases} x_1 = 0 \\ x_2 = 50 + x_1 \\ x_3 = 60 + x_1 \\ \dots \\ x_{15} = 320 + x_1 \end{cases} \quad (3)$$

Пример результатов решения системы уравнений (3) представлен в таблице 2.

Таблица 2

Результаты решения системы уравнений для первого тура первого респондента

Обозначение фигуры	Оценка, балл
1	50
2	60
...	...
15	320

Аналогичные системы уравнений были составлены и решены для данных, полученных каждым экспертом в каждом туре для каждого из 12 методов экспертного оценивания (соответствуют точкам морфологического ящика, рис. 3).

Этап 2 – Нормализация данных

Оценки площадей кругов и реальные значения площадей кругов выражаются в различных единицах и в различных шкалах, поэтому рационально для целей сопоставления и анализа нормализовать полученные оценки. Нормализация экспертных оценок (таблица 2) осуществлялась по формуле

$$n_i = \frac{x_i}{\sum_{i=1}^m x_i}, \quad (4)$$

где  $n_i$  – значение нормализованной оценки площади  $i$ -й фигуры;  $x_i$  – оценка площади  $i$ -й фигуры,  $m$  – количество оцениваемых фигур.

В таблице 3 в качестве примера представлены результаты нормализации для первого тура опроса первого эксперта.

Таблица 3

Результаты нормализации экспертных оценок

Обозначение фигуры	Нормализованная оценка, $n_i$
1	0
2	0,011
...	...
15	0,179

Этап 3 – Определение степени доверия к результатам экспертного оценивания

Степень доверия к экспертным оценкам, полученным тем или иным методом, определялась через значение среднего квадратического отклонения (СКО) результатов экспертного оценивания, полученных в процессе реализации модельного эксперимента. Эксперимент для каждого метода ( $i, i = 1, \dots, 12$ ) и каждого уровня оцениваемой величины ( $j, j = 1, \dots, 15$ ) проводился в соответствии с иерархическим планом (рис. 5), рекомендуемым СТБ ИСО 5725-3.

В общем случае схему расчета СКО метода можно представить следующим образом (рис. 5).

СКО метода как аналог СКО воспроизводимости  $S_R$  определялось по формуле (1), приведенной в СТБ ИСО 5725-2,

$$S_R = \sqrt{S_L^2 + S_r^2}, \quad (5)$$

где  $S_L$  – СКО, вызванное сменой экспертов;  $S_r$  – СКО в условиях повторяемости, где каждый эксперт оценивал величину площади круга в двух турах.

СТБ ИСО 5725-3 допускает использование методики дисперсионного анализа ANOVA для определения оценок составляющих суммарной дисперсии воспроизводимости метода. В нашем случае дисперсионный анализ был проведен при помощи программного обеспечения Excel.

Рассмотрим пример проведения дисперсионного анализа для метода «рандомизированного, парных сравнений, каждый с предыдущим для априорной шкалы».

Для каждого круга (уровня оцениваемой величины  $j, j = 1, \dots, 15$ ) была составлена матрица, куда вписывались нормализованные оценки площади фигуры каждым экспертом в каждом туре оценивания. В таблице 4 представлены оценки площади первого круга  $j = 1$  (рис. 4).

Далее с помощью надстройки Excel «Анализ данных» нами был проведен дисперсионный анализ значений матрицы, результаты которого представлены в таблице 5.

Таблица 4

Матрица оценок площади первого круга методом «рандомизированным, парных сравнений, каждый с предыдущим для априорной шкалы»

Эксперты № тура	Эксперт 1	Эксперт 2
Тур 1	0,0101	0,0421
Тур 2	0,0136	0

Таблица 5

Результаты дисперсионного анализа экспертных оценок первого круга методом «рандомизированным, парных сравнений, каждый с предыдущим для априорной шкалы»

Источник вариации (Компонент дисперсии)	Значение среднего квадрата MS	Аналог компонента дисперсии
Межгрупповая	$MS_{\text{эффект}} = 0,000085$	$S_L$
Внутригрупповая	$MS_{\text{ошибка}} = 0,000446$	$S_r$

В данной таблице представлены основные компоненты дисперсии, вызванной межгрупповым и внутригрупповым компонентами дисперсии. В нашем случае значения  $MS_{\text{эффект}}$  и  $MS_{\text{ошибка}}$  соответственно являются аналогами  $S_L$  (СКО, вызванное сменой экспертов) и  $S_r$  (СКО, вызванное участием каждого эксперта в двух турах оценивания).

Подобным образом дисперсионный анализ был проведен для каждого оцениваемого круга по каждому методу. Некоторые результаты для априорной шкалы представлены на рисунке 6.

По таблице 5 и рисунку 6 можно проследить ряд интересных закономерностей:

- вклад внутригрупповой составляющей дисперсии (дисперсии каждого эксперта ( $S_r$ : тур 1 – тур 2)) в СКО метода ( $S_R$ ) в целом больше чем вклад межгрупповой составляющей дисперсии (дисперсии от группы экспертов ( $S_L$ : эксперт 1 – эксперт 2));
- наибольшее различие между внутригрупповой и межгрупповой дисперсиями наблюдается

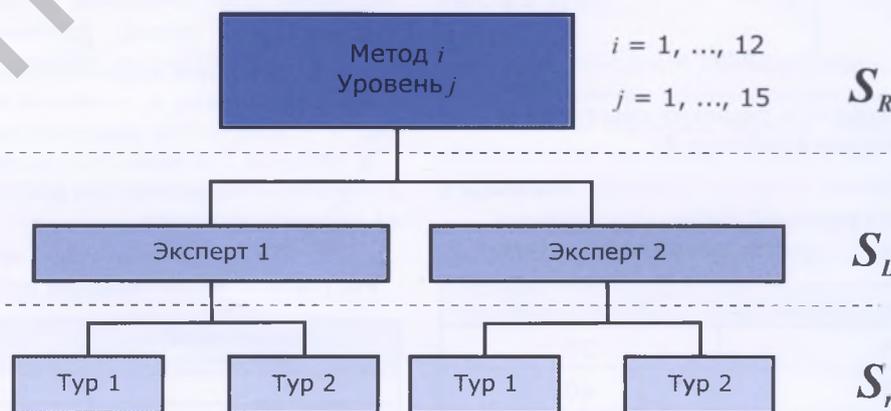


Рис. 5. Схема расчета СКО методов экспертного оценивания

в рандомизированных методах парных сравнений ( $S_r = (2...3) \cdot S_L$ ), наименьшее – в методах непосредственной оценки.

- упорядоченность предъявления объектов на оценивание в целом уравнивает значения межгрупповой и внутригрупповой дисперсий ( $S_r \approx S_L$ );

- значения  $S_L$  и  $S_r$  в целом не зависят от величины площади оцениваемого круга.

**Результаты исследований аспектов факторов управления**

По приведенному выше алгоритму были проведены исследования всех 12 разновидностей методов морфологического ящика (рис. 3). Некоторые результаты исследования для первого эксперта, представленные на рисунках 7 – 9, позволяют проследить закономерности появления потерь достоверности различных методов экспертного оценивания.

Из рисунка 7 следует, что вид оцениваемой шкалы не имеет существенного влияния на СКО результатов экспертного оценивания (рассеяние относительно аппроксимирующей прямой незначительно). Следовательно, эксперт может сам выбирать для себя удобную шкалу, и это не скажется на результатах оценки.

Из рисунка 8 следует, что в случае представления объектов на оценивание случайным образом наблюдается существенно большее рассеяние оценок (большее СКО результатов экспертного оценивания), чем в случае, когда объекты представляются упорядоченно.

Из рисунка 9 следует, что методы парных сравнений при упорядоченном способе представления объектов на оценку показывают меньшее значение СКО, чем в случае представления объектов случайным образом. В то же время способ «непосредственного оценивания» не показывает значимых различий значений СКО при упорядоченном и рандомизированном способах представления объектов на оценку.

На рисунке 10 представлены итоговые результаты комплексного исследования достоверности экспертных оценок методом морфологического ящика. Для ранжирования рассматриваемых методов по признаку «степень достоверности результатов оценивания», на наш взгляд, достаточно сравнить значения СКО для каждого из исследуемых методов- $S_R$ .

Отметим выявленные закономерности:  
– упорядочение объектов оценивания всегда повышает достоверность результатов (уменьшает величину СКО);

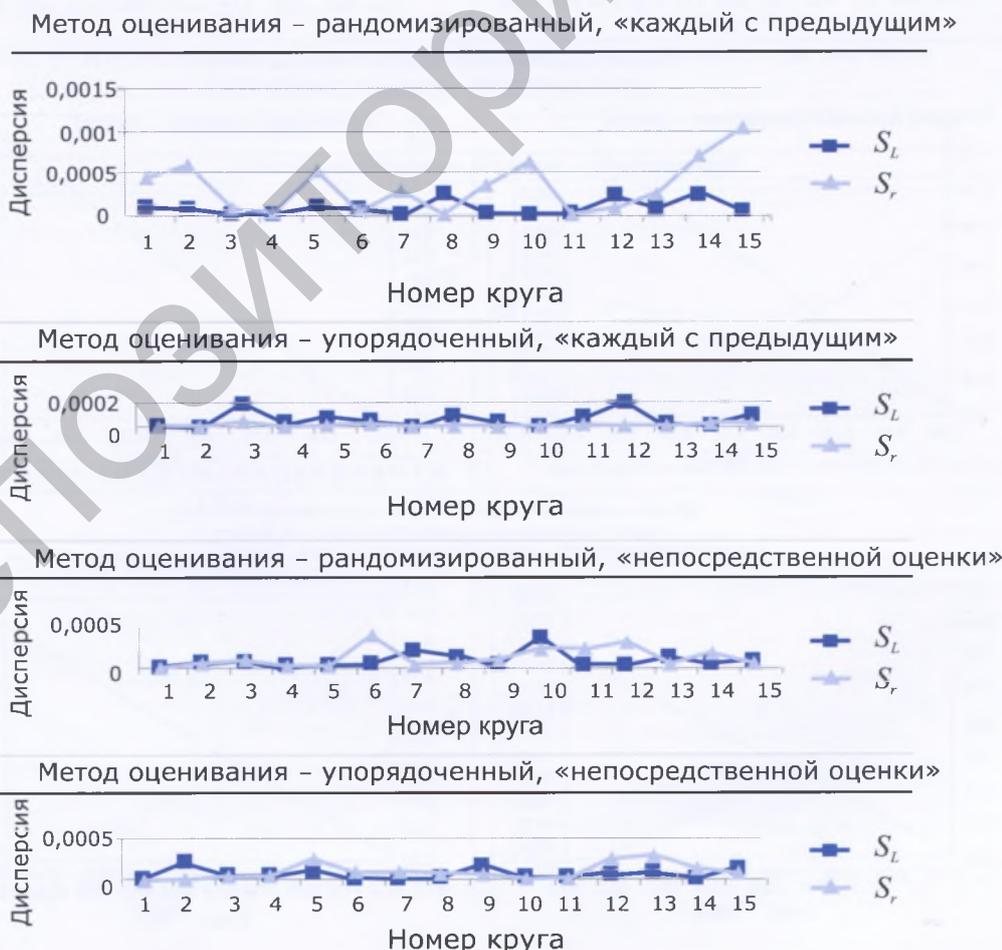


Рис. 6. Результаты дисперсионного анализа

– наиболее предпочтительными (с наименьшим СКО) являются методы парных сравнений с использованием апостериорной и априорной шкал, способ оценивания – «каждый с предыдущим» или «каждый с одним»; способ предъявления объектов на оценивание – упорядоченный; – при прочих равных условиях способ оценивания «каждый с предыдущим» всегда хуже (большее СКО), чем способы «каждый с одним» или «непосредственной оценки». Это объясняется тем, что эксперт, реализуя способ «каждый с предыдущим», теряет «ноль», чего не наблю-

дается в двух других случаях.

В рамках данного эксперимента параллельно с СКО методов проводилось определение смещения экспертных оценок от «истинного» значения площади круга. В целом закономерности поведения смещения методов аналогичны закономерностям поведения их СКО. Смещения методов при упорядоченном способе представления объектов на оценивание примерно в 2 раза меньше, чем при рандомизированном. Отношение смещения метода оценивания к СКО  $\Delta/\sigma$  варьируется от 0,5 до 5.

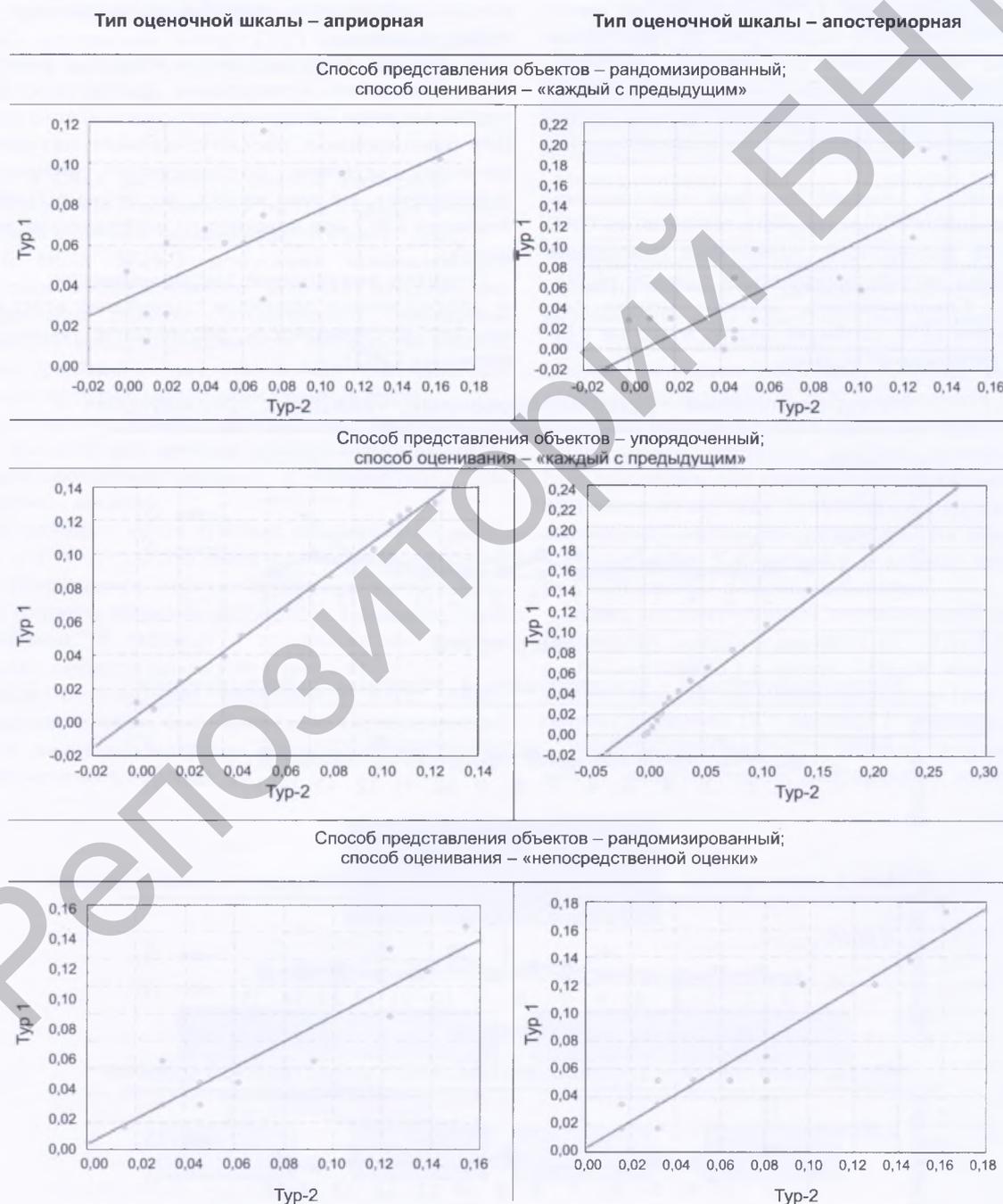


Рис. 7. Исследование влияния типа оценочной шкалы (априорная – апостериорная)

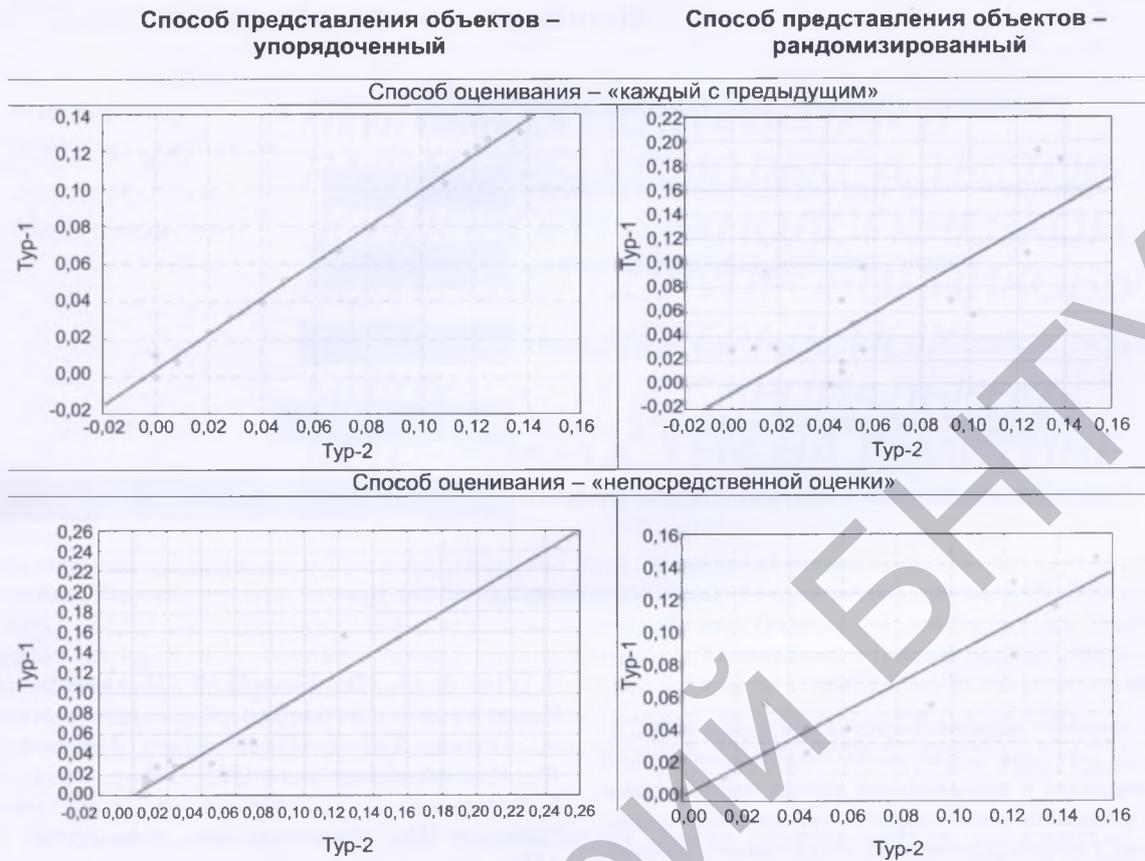


Рис. 8. Исследование влияния способа представления объектов эксперту

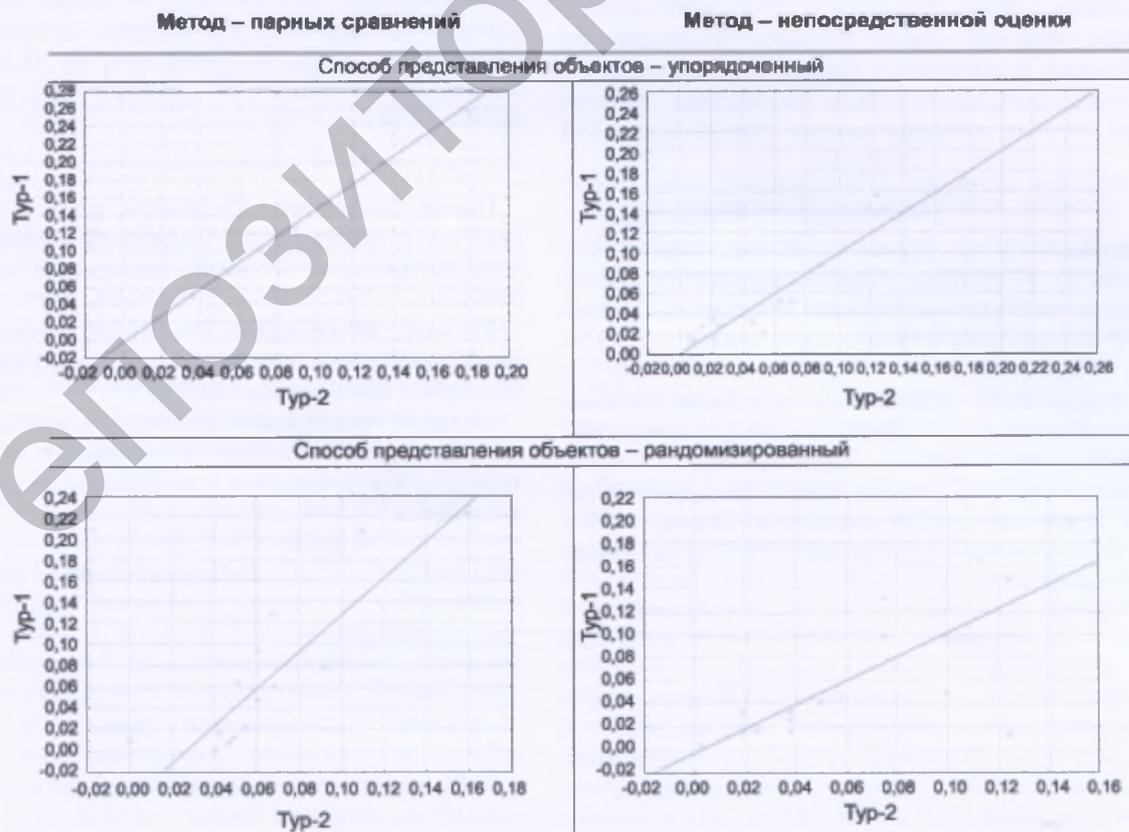


Рис. 9. Исследование влияния способа оценивания объекта

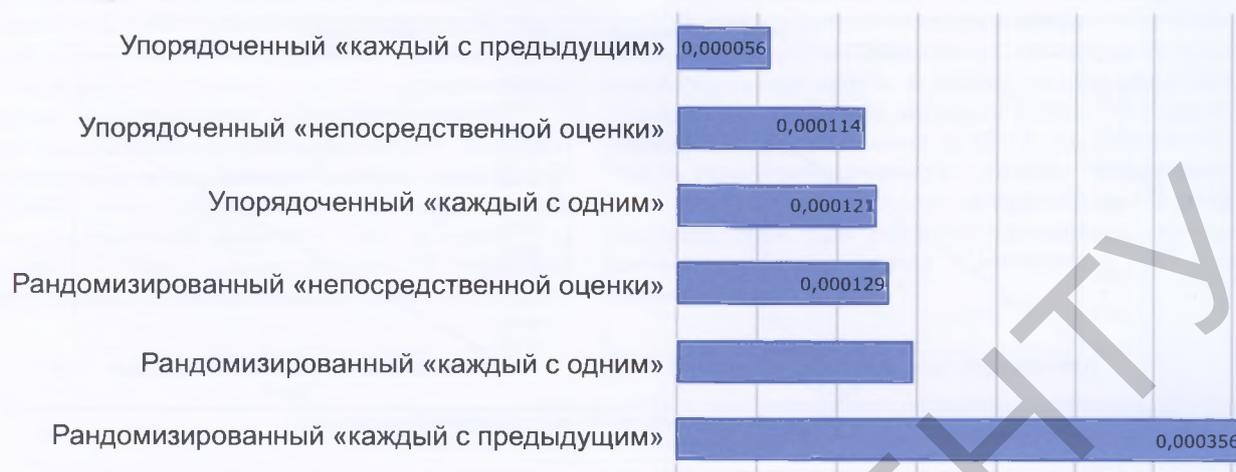
Значение  $S_R$  методов

Рис. 10. Результаты расчета дисперсии воспроизводимости метода для априорной шкалы

**Выводы**

На основании комплексного исследования существующих методов экспертного оценивания можно рекомендовать к применению метод парных сравнений с использованием априорной или апостериорной шкал, способ оценивания – «каждый с одним», порядок представления объектов на оценивание – упорядоченный. При невозможности упорядочения объектов рекомендуется использовать методы непосредственной оценки, порядок представления объектов на оценивание – рандомизированный.

Общая неопределенность экспертных оценок для различных методов может отличаться в 2–3 раза.

*Список использованной литературы*

1. Серенков П. С., Романчук В. М., Гиль Н. Н., Телебук О. И. Анализ и прогнозирование достоверности экспертных оценок показателей качества // Метрология и приборостроение. – 2015. – № 3. – С. 30–37.
2. Огвоздин В. Ю. Управление качеством. Основы теории и практики: Учебное пособие, 6-е издание, М., Изд. «Дело и Сервис», 2009, 304 с.
3. Альтшуллер Г. С. Найти идею. Введение в теорию решения изобретательских задач. – 3-е изд., дополненное. – Петрозаводск: Скандинавия, 2003. – 240 с.

4. Гусев А. Н., Измайлов Ч. А., Михалевская М. Б. Измерение в психологии: общий психологический практикум. Серия «Практикум»: 2-й выпуск / – М.: Изд-во «Смысл». – 1987.

5. Азгальдов Г. Г., Райхман Э. П. О квалиметрии. – М.: Издательство стандартов, 1972, 172 с.

6. Серенков П. С., Гуревич В. Л., Романчук В. М., Янушкевич А. В. Методы менеджмента качества. Методология управления риском стандартизации. – П. С. Серенков, В. Л. Гуревич, В. М. Романчук, А. В. Янушкевич. – Минск, 2012. – 243 с.

**Павел Степанович Серенков**, доктор технических наук, заведующий кафедрой «Стандартизация, метрология и информационные системы» БНТУ;

**Василий Михайлович Романчук**, кандидат физико-математических наук, доцент кафедры «Инженерная математика» БНТУ;

**Наталья Николаевна Иванова**, магистр технических наук, преподаватель кафедры «Стандартизация, метрология и информационные системы» БНТУ

Дата поступления 11.01.2016