ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ ТЕСТОВЫХ ВОПРОСОВ НА ОСНОВЕ ТОЧЕЧНОГО КОЭФФИЦИЕНТА

Калмыков Н. С.[1]**,**
(*ФГБУН Институт проблем управления им. В.А. Трапезникова РАН, Москва*)
Сидельников Ю. В.[2]
(*ФГБУН Институт проблем управления им. В.А. Трапезникова РАН; МАИ, Москва*)

Экспериментально проверяется наличие объективной составляющей сложности задачи. С этой целью введен и экспериментально исследован так называемый «точечный коэффициент», позволяющий оценить уровень сложности задачи заданной в виде тестового вопроса. Именно на основе ответов на эти вопросы оценивается уровень сложности задачи. Предложены и экспериментально апробированы гипотезы, характеризующие динамику изменений значений точечного коэффициента. Также проверена устойчивость полученных выводов, базирующихся на результатах расчетов по двум качественно различным коэффициентам относительно числа опрашиваемых участников. Экспериментально найден и проверен критерий для определения тестового вопроса как задачи третьего уровня сложности. Основанием этого является характер изменения диапазона разброса значений гистограммного коэффициента. Проведено экспериментальное исследование наличия когнитивных искажений при оценивании экспертами некоторых задач тестового характера. Предложены возможные направления дальнейших разработок.

Ключевые слова: экспертные оценки, тестовый вопрос, точечный коэффициент.

1. Введение

Актуальность данной темы исследования обоснована следующим фактом. В литературе по экспертным оценкам часто описывают креативные методы. Так, к примеру, в статье [9] разработан сценарный метод построения дорожных карт для стратегического планирования и прогнозирования, в котором само построение сценариев опирается на такой креативный метод как «Шесть шляп». Также в научной литературе даны описания экс-

---

[1] *Никита Сергеевич Калмыков, м.н.с.* (kalmikov-nik@bk.ru).
[2] *Юрий Валентинович Сидельников, д.т.н., профессор* (sidelnikovy@mail.ru).

периментов, в которых анализируют ответы экспертов на тестовые вопросы на основании того или иного показателя ошибки или точности [5]. Мы же путем экспериментов должны выявить наличие объективной составляющей сложности тестового вопроса как задачи на основе введения и использования различных коэффициентов. Кроме того, в статье авторов [3] по рассматриваемой теме, представленной в публикации как первой из двух частей данного исследования, было указано, что для подтверждения уровня правдоподобности полученных выводов на основе экспериментальных исследований необходимо, в частности, варьировать вид коэффициента. Поясним последнее утверждение. В первой части того же исследования [3] было рассмотрено шесть основных факторов, от которых зависит корректность выводов экспериментальных исследований. Один из факторов – вид коэффициента, на основе которого вычисляется усредненное значение всех экспертных оценок сложности конкретной задачи, заданной в виде тестового вопроса.

Цель экспериментального исследования: выяснить на основе тестовых экспериментов наличие объективно существующих уровней и подуровней сложности задач, сформулированных в виде тестовых вопросов.

Для достижения поставленной цели предлагается решить следующую совокупность задач**:**

1. Ввести и экспериментально апробировать на основе имеющейся статистики качественно новый, так называемый «точечный коэффициент», позволяющий оценить уровень сложности задачи заданной в виде тестового вопроса.
2. Сопоставить результаты расчетов как по введенному в статье точечному коэффициенту, так и рассмотренному ранее гистограммному коэффициенту [3] для подтверждения уровня правдоподобности выводов относительно наличия объективной составляющей сложности задач.

Для пояснения сути второй задачи введем **Постулат №1.**

Полагаем, что найдутся такие задачи, у которых можно выделить как субъективную, так и объективную составляющую их сложности.

То, что существует субъективная составляющая сложности задачи, очевидно. Для призера математических олимпиад и обычного школьника сложность одной и той же стандартной тестовой задачи, которые обычно задают в школе на дом, различна. И в этом случае, по сути, говорят о субъективной составляющей этой стандартной тестовой задачи. Но есть ли и объективная составляющую сложности задачи. По-видимому, да. Но как ее измерять? В случае когда мы имеем одного оценщика и тестовую задачу, мы можем сопоставить его оценку X и истинное значение тестового вопроса, заданного в виде задачи Y, используя тот или иной показатель ошибки. Но это ошибка конкретного эксперта-оценщика, а не оценка сложности задачи. Для решения последней задачи надо снять субъективную составляющую. Но для этого необходимо учесть:

– усредненное значение всех экспертных оценок и их точности для каждой из задач, заданной в виде тестового вопроса;

– зависимость получаемого результата (вывода) от вида коэффициента используемого при расчете.

А также проверить статистическую устойчивость полученного расчета (вывода) от количества экспертов-оценщиков.

Поясняя, можно указать, что показатели ошибки используются для оценки X ответа одного конкретного эксперта при наличии истинного значения ответа Y на тестовый вопрос, заданный в виде задачи. Таким образом, необходимо ввести качественно новые показатели – коэффициенты. Но в этом случае уровень правдоподобности вывода относительно наличия объективной составляющей уровней сложности зависит от:

I. коэффициента, на основе которого проводились расчеты уровней сложности задач;

II. количества экспертов, которые давали ответ на тестовый вопрос.

При сопоставлении мы должны использовать два различных, но сопоставимых коэффициента – точечный и ранее введенный и описанный в первой части исследования гистограммный коэффициент [3]. Здесь мы вводим **Постулат №2**. Сопоставление результатов, полученных при расчетах на основе двух качественно различных коэффициентов, позволит выявить устойчивость выводов на основе расчетов по этим коэффициентам. Полагая, что если результаты, полученные при расчетах на основе введенных коэффициентов, будут совпадать или близки, то это будет означать, что полученный вывод относительно наличия объективной составляющей уровней сложности задач либо не зависит, либо слабо зависит от вида коэффициента, на основе которого и был получен этот вывод. Точнее, вывод относительно уровня правдоподобности (корректности) выводов относительно наличия (существования) уровней и подуровней объективной составляющей уровней сложности задач.

Таким образом, в результате экспериментальных исследований мы можем получить некое утверждение относительно наличия (существования) уровней и подуровней объективной составляющей уровней сложности задач.

3. Проанализировать и предложить гипотезы, характеризующие динамику изменений значений точечного коэффициента на основе тестовых вопросов и массива экспертных оценок, введенных и рассмотренных в монографии одного из авторов [5] и в статье [3]. Вывести и обосновать по ним заключения.

Поясним эту задачу. Это необходимо для того, чтобы можно было показать статистическую устойчивость выводов, рассчитываемых на основе этого коэффициента, от количества участвующих экспертов. При этом необходимо отметить, что исходным эмпирическим материалом для последующего анализа служит массив оценок, полученных в качестве ответов на 9 различных, независимых по своему содержанию тестовых вопросов. Сами вопросы описаны во втором разделе данной статьи. В рамках данного экспериментального исследования были использованы оценки ста сорока шести участников, а общее число оценок как ответов от опрашиваемых участников – 2146 (1073 гистограммных и столько же точечных оценок).

4. Исследовать влияние когнитивных искажений на результаты ответов экспертов.

5. Сопоставить ранжировки вопросов на основе оценочных ответов экспертов по двум качественно различным коэффициентам для обоснования уменьшения уровня значимости зависимости рассматриваемого вывода от вида введенных коэффициентов.

6. Обозначить дальнейшие исследования по рассматриваемой теме.

2. Введение точечного коэффициента

Введем точечный коэффициент, на основе которого можно определять уровень сложности задачи в виде тестового вопроса. Точнее, коэффициент, на основании которого вычисляется усредненное значение экспертных числовых (точечных) оценок уровня сложности задачи, заданной в виде тестового вопроса. Для краткости в дальнейшем будем его называть «точечный коэффициент», по аналогии с гистограммным коэффициентом, введенным в первой части исследования авторов [3].

Данный коэффициент рассчитывается следующим образом:

$PA^j(x^j_1, x^j_2, \ldots, x^j_{N^j})$ – точечный коэффициент, характеризующий сложность j-го вопроса как тестовой задачи. При расчете данного коэффициента мы используем информацию, полученную от N^j участников, ответивших на j-й вопрос и которые дали точечные (числовые) оценки $(x^j_1, x^j_2, \ldots, x^j_{N^j})$.

Коэффициент, как функция, представлен следующим образом:

$PA^j(x^j_1, x^j_2, \ldots, x^j_{N^j}) = |(ME^j(x^j_1, x^j_2, \ldots, x^j_{N^j}) - y^j) / y^j|$, где:

– $ME^j(x^j_1, x^j_2, \ldots, x^j_{N^j})$ – медиана числового ряда $(x^j_1, x^j_2, \ldots, x^j_{N^j})$ всех оценок (ответов) N^j участников по j-му вопросу;

– x^j_{i} – точечная (числовая) оценка i-го участника (ответ) по j-му тестовому вопросу;

– y^j – истинное значение ответа на j-й тестовый вопрос;

– $i = 1, 2, \ldots, N^j$; N^j – общее число участников, ответивших на j-й вопрос;

– $j = 1, 2,\ldots K$; K – общее число вопросов.

Рассмотрим обобщение коэффициента $PA^j(x^j_1, x^j_2, \ldots, x^j_{N^j}) = |(ME^j(x^j_1, x^j_2, \ldots, x^j_{N^j}) - y^j) / y^j|$ для произвольного вопроса. Тогда значение параметра $|ME(x_1, x_2, \ldots, x_N) - y|$, может изменяться от 0 до произвольного числа R. Таким образом, область значения обобщенного коэффициента $|ME(x_1, x_2, \ldots, x_N) - y) / y|$ как функции, принимает значения в отрезке от 0 до величины $|R / y|$.

Постулат №3. Полагаем, что чем больше отклоняется значение $ME^j(x^j_1, x^j_2, \ldots, x^j_{N^j})$ от y^j, тем вопрос сложнее, и, соответственно, чем ближе соответствующая разность к нулю, тем вопрос проще.

Таким образом, чем ближе значение коэффициента $PA^j(x^j_1, x^j_2, \ldots, x^j_{N^j})$ к нулю, тем уровень сложности j-го вопроса меньше. Это аналогично динамике гистограммного коэффициента [3].

3. Описание тестовых вопросов

Исходным эмпирическим материалом для последующего анализа служит массив из 1073 гистограммных и стольки же точечных оценок, полученных от 147 участников при ответе на 9 различных, независимых по своему содержанию вопросов, как тестовых задач для участников.

Перечислим эти тестовые вопросы как задачи для участников:

1. Какое количество времени электропоезд метрополитена в Москве затрачивал на проезд по кольцу (в минутах) в 1979 году?
2. Сколько вагонов в составе электропоезда метрополитена в Москве на линии Беляево – Медведково было в 1979 году?
3. Каким было население Африки в 1975 году (в млн чел.)?
4. Какова была в Москве доля мужчин (в процентном соотношении) в 1970 г.?
5. Каким было население Филиппин в 1970 году?
6. Какова продолжительность жизни Н.Г. Чернышевского (в годах)?
7. Какое количество времени электропоезд метрополитена в Москве затрачивал на проезд от станции Киевская до станции Щелковская (в минутах) в 1979 году?
8. Какова длина корпуса люминесцентных ламп в вашей рабочей комнате (в см)?
9. Сколько граммов весит предложенная вам для оценки деревянная вазочка?

4. Гипотезы, характеризующие динамику изменений значений точечного коэффициента

Зафиксируем предварительное условие. Рассмотренные выше вопросы примем в качестве тестовых задач для участников; количественный состав участников и весь массив ранее полученных оценок.

На данном массиве ответов, полученном при оценивании экспертами вопросов, рассмотрим пять гипотез, которые характеризуют динамику изменения значений уже не гистограммного [3], а точечного коэффициента $PA^j_i(N^j_i)$ в зависимости от изменения объемов выборок без возвращений из N^j_i выбранных случайным образом ответов экспертов. Здесь i – номер выборки, $i = 1, 2, ..., m_j$; m_j – количество выборок для j вопроса, а $N^j_i < N^j$.

При этом мы используем тот же механизм получения случайной выборки объемом N^j_i из генеральной совокупности, состоящей из N^j ответов экспертов на j-й вопрос, который описан в статье авторов [3]. Повторим его описание.

1. По каждому j-му вопросу случайным образом составляется i-я выборка без возвращения объемом N^j_i из натуральных чисел от 1 до N, где N соответствует количеству участников, ответивших на данный вопрос; $i = 1, 2, ..., m_j$, а m_j – количество выборок для j-Го вопроса). Данные выборки были созданы при помощи функции lotto на языке программирования Visual Basic, возвращающей псевдослучайные числа без повторов по заданным верхней границе интервала, нижней границе и количеству псевдослучайных чисел, которые отбираются из заданного интервала, соответствующего количеству экспертов, ответивших на конкретный вопрос [2].
2. Каждое из полученных псевдослучайных натуральных чисел соответствует порядковому номеру эксперта и сопоставляется со значением ответа данного эксперта.
3. Данная процедура повторяется 999 раз для создания новых выборок для каждого j-го вопроса и каждой i-й выборки объемом N^j_i.

Нумерацию гипотез будем продолжать, исходя из их нумерации в статье авторов [3].

При этом мы, исходя из того, что гипотезы №1÷4 аналогичны не только по структуре описания, но полученным результатам. Но характеризуют динамику изменения значений не рассматриваемого в данной статье точечного коэффициента, а лишь только гистограммного.

Гипотеза №5. Полагаем, что содержание гипотез, характеризующих динамику изменения значений точечного коэффициента $PA^j_i(N^j_i)$ в зависимости от изменения объемов выборок из N^j_i, выбранных случайным образом ответов экспертов будет подобна аналогичным показателям гистограммного коэффициента.

Гипотеза №6. Полагаем, что для любого j-го тестового вопроса найдутся такое количество участников тестового опроса $K6$ и значение (N^j_{i6}) объема выборок, что при расчете точечного коэффициента по выборкам (N^j_i) значение $\min PA^j_i(N^j_i)$ как функции будет монотонно возрастать с увеличением объема этих выборок, начиная со значения N^j_{i6}, и стремиться к значению PA^j для всей совокупности для j-го вопроса.

Гипотеза №7. Полагаем, что для любого j-го тестового вопроса найдутся такое количество участников тестового опроса $K7$ и значение (N^j_{i7}) объема выборок, что, при расчете точечного коэффициента по выборкам (N^j_i) значение $\max PA^j_i(N^j_i)$ как функции будет монотонно убывать с увеличением объема этих выборок, начиная со значения N^j_{i7}, и стремиться к значению PA^j для всей совокупности для j-го вопроса.

Гипотеза №8. Полагаем, что для любого j-го тестового вопроса, найдутся такое количество участников тестового опроса $K8$ и такое значение (N^j_{i8}) объема выборок, что диапазон разброса значений точечного коэффициента $[\min_i PA^j_i(N^j_i), \max_i PA^j_i(N^j_i)]$ как функции будет монотонно убывать с увеличением объема i-й выборки, начиная со значения N^j_{i8}, а в предельном случае – стремиться к нулю для всей совокупности для j-го вопроса.

Гипотеза №9. Полагаем, что для любого j-го тестового вопроса найдется такое количество участников тестового опроса $K9$, что медиана ряда значений коэффициента PA^j с увеличением объема этой выборки будет хаотично, но устойчиво стремиться к значению PA^j для всей совокупности для j-го вопроса.

В случае если гипотезы №6÷9 будут подтверждены экспериментально, то, базируясь на них, можно получить информацию, характеризующую динамику изменений значений точечного коэффициента и правдоподобное обоснование утверждения относительно устойчивости полученных выводов, базирующиеся на результатах расчетов точечного коэффициента $PA^j(x^j_1, x^j_2, \ldots, x^j_{Nj})$, относительно числа опрашиваемых участников.

5. Экспериментальное обоснование гипотез на основе апробации точечного коэффициента

Проанализируем динамику изменений значений точечного коэффициента и обоснуем гипотезы №6÷9, рассчитав значения точечного коэффициента на большом массиве данных, полученных при ответах экспертов на 9 тестовых вопросов. Для этого рассчитаем эти значения коэффициентов $PA^j(x^j_1, x^j_2, \ldots, x^j_{Nj})$ для этих гипотез и представим полученную информацию в таблицах 1–9.

Таблица 1. Результаты проверки гипотез на основе расчета коэффициентов PA^j_i по вопросу №1

Номер вопроса 1	**Дисконтированные значения объемов выборок на первый вопрос N^1_i. (Общее число ответов: 146)**													
	20	30	40	50	60	70	80	90	100	110	120	130	140	Значение PA^1_{146}
min PA^1_i	0,05	0,14	0,15	0,21	0,21	0,21	0,21	0,22	0,22	0,24	0,26	0,28	0,31	**0,38**
ME значений ряда оценок PA^1_i	0,34	0,34	0,38	0,38	0,38	0,38	0,38	0,38	0,38	0,38	0,38	0,38	0,38	
max PA^1_i	0,57	0,55	0,55	0,48	0,47	0,41	0,40	0,38	0,38	0,38	0,38	0,38	0,38	
Диапазон: max PA^1_i – min PA^1_i	0,52	0,41	0,41	0,28	0,26	0,21	0,19	0,16	0,16	0,14	0,12	0,10	0,07	

Таблица 2. Результаты проверки гипотез на основе расчета коэффициентов $PA^j{}_i$ по вопросу №2

Номер вопроса 2	**Дисконтированные значения объемов выборок на второй вопрос $N^2{}_i$ (Общее число ответов: 123)**											
	20	30	40	50	60	70	80	90	100	110	120	Значение $PA^2{}_{123}$
min $PA^2{}_i$	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,07	0,14	0,14	0,14	**0,14**
ME значений ряда оценок $PA^2{}_i$	0,14	0,14	0,14	0,14	0,14	0,14	0,14	0,14	0,14	0,14	0,14	
max $PA^2{}_i$	0,43	0,43	0,36	0,14	0,14	0,14	0,14	0,14	0,14	0,14	0,14	
Диапазон max $PA^2{}_i$ – min $PA^2{}_i$	0,43	0,43	0,36	0,14	0,14	0,14	0,14	0,07	0,00	0,00	0,00	

Таблица 3. Результаты проверки гипотез на основе расчета коэффициентов $PA^j{}_i$ по вопросу №3

Номер вопроса 3	**Дисконтированные значения объемов выборок на третий вопрос $N^3{}_i$ (Общее число ответов: 140)**												
	20	30	40	50	60	70	80	90	100	110	120	130	**Значение $PA^3{}_{140}$**
min $PA^3{}_i$	0,01	0,01	0,02	0,12	0,20	0,25	0,25	0,29	0,29	0,32	0,37	0,42	0,44
ME значений ряда оценок $PA^3{}_i$	0,46	0,46	0,46	0,46	0,46	0,44	0,44	0,44	0,44	0,44	0,44	0,44	
max $PA^3{}_i$	1,21	1,03	0,87	0,75	0,78	0,69	0,62	0,56	0,55	0,55	0,50	0,50	
Диапазон max $PA^3{}_i$ – min $PA^3{}_i$	1,20	1,02	0,85	0,62	0,57	0,44	0,37	0,27	0,26	0,23	0,12	0,07	

Таблица 4. Результаты проверки гипотез на основе расчета коэффициентов $PA^j{}_i$ по вопросу №4

Номер вопроса 4	**Дисконтированные значения объемов выборок на четвертый вопрос $N^4{}_i$ (Общее число ответов: 141).**												
	20	30	40	50	60	70	80	90	100	110	120	130	**Значение $PA^4{}_{141}$**
min $PA^4{}_i$	0,01	0,02	0,03	0,05	0,05	0,05	0,06	0,07	0,07	0,07	0,07	0,07	**0,07**
ME значений ряда оценок $PA^4{}_i$	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,07	
max $PA^4{}_i$	0,11	0,10	0,09	0,09	0,09	0,09	0,09	0,09	0,09	0,09	0,07	0,07	
Диапазон max $PA^4{}_i$ – min $PA^4{}_i$	0,10	0,08	0,06	0,05	0,05	0,05	0,03	0,02	0,02	0,02	0,00	0,00	

Таблица 5. Результаты проверки гипотез на основе расчета коэффициентов $PA^j{}_i$ по вопросу №5

Номер вопроса 5	**Дисконтированные значения объемов выборок на пятый вопрос $N^5{}_i$ (Общее число ответов: 146)**													
	20	30	40	50	60	70	80	90	100	110	120	130	140	**Значение $PA^5{}_{146}$**
min $PA^5{}_i$	0,00	0,02	0,06	0,04	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,15	0,26	**0,30**
ME значений ряда оценок $PA^5{}_i$	0,28	0,28	0,28	0,28	0,30	0,28	0,30	0,30	0,30	0,30	0,30	0,30	0,30	
max $PA^5{}_i$	0,48	0,44	0,39	0,35	0,35	0,30	0,30	0,30	0,30	0,30	0,30	0,30	0,30	
Диапазон max $PA^5{}_i$ – min $PA^5{}_i$	0,48	0,43	0,33	0,31	0,24	0,19	0,19	0,19	0,19	0,19	0,19	0,15	0,04	

Таблица 6. Результаты проверки гипотез на основе расчета коэффициентов $PA^j{}_i$ по вопросу №6

Номер вопроса 6	**Дисконтированные значения объемов выборок на шестой вопрос $N^6{}_i$ (Общее число ответов: 141)**												
	20	30	40	50	60	70	80	90	100	110	120	130	**Значение $PA^6{}_{141}$**
min $PA^6{}_i$	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,02	0,07	0,07	0,08	0,08	
ME значений ряда оценок $PA^6{}_i$	0,08	0,08	0,08	0,08	0,08	0,08	0,08	0,08	0,08	0,08	0,08	0,08	
max $PA^6{}_i$	0,26	0,25	0,18	0,18	0,17	0,16	0,16	0,15	0,13	0,11	0,11	0,10	**0,08**
Диапазон max $PA^6{}_i$ – min $PA^6{}_i$	0,26	0,25	0,18	0,18	0,16	0,16	0,15	0,13	0,07	0,05	0,02	0,02	

Таблица 7. Результаты проверки гипотез на основе расчета коэффициентов $PA^j{}_i$ по вопросу №7

Номер вопроса 7	**Дисконтированные значения объемов выборок на седьмой вопрос $N^7{}_i$(Общее число ответов: 144)**												
	20	30	40	50	60	70	80	90	100	110	120	130	**Значение $PA^7{}_{144}$**
min $PA^7{}_i$	0,00	0,00	0,06	0,06	0,06	0,09	0,14	0,15	0,18	0,18	0,18	0,18	
ME значений ряда оценок $PA^7{}_i$	0,26	0,25	0,26	0,26	0,27	0,27	0,27	0,28	0,28	0,28	0,28	0,28	
max $PA^7{}_i$	0,62	0,59	0,58	0,53	0,53	0,53	0,53	0,49	0,41	0,40	0,36	0,29	**0,28**
Диапазон max $PA^7{}_i$ – min $PA^7{}_i$	0,62	0,59	0,52	0,47	0,47	0,44	0,39	0,34	0,24	0,22	0,19	0,12	

Таблица 8. Результаты проверки гипотез на основе расчета коэффициентов $PA^j{}_i$ по вопросу №8

Номер вопроса 8	**Дисконтированные значения объемов выборок на восьмой вопрос $N^8{}_i$(Общее число ответов: 44)**							
	10	15	20	25	30	35	40	**Значение $PA^8{}_{44}$**
min $PA^9{}_i$	0,00	0,01	0,00	0,01	0,00	0,01	0,01	**0,01**
ME значений ряда оценок $PA^8{}_i$	0,05	0,03	0,03	0,03	0,01	0,01	0,01	
max $PA^8{}_i$	0,21	0,13	0,11	0,09	0,09	0,05	0,03	
Диапазон max $PA^8{}_i$ – min $PA^8{}_i$	0,21	0,12	0,10	0,08	0,09	0,04	0,02	

Таблица 9. Результаты проверки гипотез на основе расчета коэффициентов $PA^j{}_i$ по вопросу №9

Номер вопроса 9	**Дисконтированные значения объемов выборок на девятый вопрос $N^9{}_i$ (Общее число ответов: 42).**							
	10	**15**	**20**	**25**	**30**	**35**	**40**	**Значение $PA^9{}_{42}$**
min $PA^9{}_i$	0,00	0,00	0,00	0,00	0,00	0,00	0,09	**0,09**
ME значений ряда оценок $PA^9{}_i$	0,21	0,13	0,09	0,09	0,09	0,09	0,09	
max $PA^9{}_i$	0,71	0,64	0,43	0,41	0,40	0,23	0,09	
Диапазон max $PA^9{}_i$ – min $PA^9{}_i$	0,71	0,64	0,43	0,41	0,40	0,23	0,00	

6. Обоснование предложенных гипотез и соответствующие заключения

Подробно опишем результаты, подтверждающие гипотезы №6÷9, исходя из данных таблиц 1–9.

Гипотеза № 6 подтверждается, так как для любого j-го тестового вопроса нашлось такое количество участников тестового опроса $K6$ и значение $N^j{}_{i6}$ объема выборок, что при расчете точечного коэффициента по выборкам $(N^j{}_i)$ значение $\min PA^j{}_i(N^j{}_i)$ как функции с увеличением объема выборок монотонно возрастает для всех вопросов, начиная со значения $N^j{}_{i6}$. Кроме того, значение $\min PA^j{}_i(N^j{}_i)$ как функции стремится к значению $PA^j{}_i$ для всей совокупности для j-го вопроса.

Для дальнейших исследований и выдвижения новых гипотез рассмотрим этот результат более детально.

Так, для вопросов №1÷4, а также №6, №7 и №9 значение $\min PA^j{}_i(N^j{}_i)$ как функции с увеличением объема выборок монотонно возрастает, начиная с минимального объема выборки. Лишь для вопросов №5 и №8 монотонный рост этой функции с увеличением объема выборки начинается с $i = 60$ (вопрос №5) и $i = 30$ (вопрос №8).

Выявлена следующая особенность: отсутствие совпадения значения $\min PA^j{}_i(N^j{}_i)$ со значением $PA^j{}_i$ для всей совокупности вопросов с самого начала объема i-й выборки.

Но при этом выявлены совпадения значений $\min PA^j{}_i(N^j{}_i)$ со значением $PA^j{}_i$ для вопросов: №2 – начиная с $i = 100$; для №4 – с $i = 90$; для №6 – с $i = 120$; для №8 – с $i = 35$; для №9 – с $i = 40$.

Гипотеза №7 подтверждается, так как для любого j-го тестового вопроса нашлось такое количество участников тестового опроса $K7$ и значение $N^j{}_{i7}$ объема выборок, что при расчете точечного коэффициента по выборкам $(N^j{}_i)$ значение $\max PA^j{}_i(N^j{}_i)$ как функции с увеличением объема этих выборок монотонно убывает для всех вопросов, начиная со значения $N^j{}_{i7}$. Кроме того, значение $\max PA^j{}_i$ стремится к значению PA^j для всей совокупности для j-го вопроса.

Для дальнейших исследований и выдвижения новых гипотез рассмотрим этот результат более детально.

Так, для вопроса №9 значение $\max PA^j{}_i(N^j{}_i)$ как функции строго монотонно убывает, начиная с минимального объема выборки.

Отдельно отметим, что значение $\max PA^j{}_i(N^j{}_i)$ как функции с увеличением объема выборки монотонно убывает, начиная с минимального объема выборки для вопросов №1 и №2; а также №4÷9. А для вопроса №3 начало монотонного убывания с увеличением объема выборки находится при i =60.

Выявлена особенность – отсутствие тенденции совпадении значения $\max PA^j{}_i(N^j{}_i)$ со значением $PA^j{}_i$ для всей совокупности вопросов с самого начала объема i-й выборки. Но есть совпадения значений аналогичных характеристик для вопросов: №1 – начиная с i = 90; для №2 –с i = 50; для №4 – с i = 20; для №5 – с i = 70; для №9 – с i = 40.

Гипотеза №8 подтверждается, так как для любого j-го тестового вопроса нашлось такое количество участников тестового опроса $K8$ и такое значения $N^j{}_{i8}$ объемов выборок, что диапазон разброса значений точечного коэффициента $[\min_i PA^j{}_i(N^j{}_i), \max_i PA^j{}_i(N^j{}_i)]$ как функции монотонно убывает с увеличением объема i-й выборки для всех вопросов, начиная со значения $N^j{}_{i8}$, а в предельном случае – стремится к нулю по всей совокупности для j-го вопроса.

Для дальнейших исследований и выдвижения новых гипотез рассмотрим этот результат более детально.

Так, для вопросов №3, №8, №9 диапазон разброса значений строго монотонно убывает. При этом диапазон разброса значений точечного коэффициента как функции монотонно убывает, начиная с минимального объема выборки практически для всех вопросов. Точнее, для вопросов №1÷7 и №9. И лишь для вопроса №8 имеется незначительное отклонение в монотонности убывания, когда при дисконтированном значении объема выборки, равном 25, диапазон $[\max PA^8{}_i - \min PA^8{}_i]$ равен 0,08, а при значении объема выборки 30 диапазон $[\max PA^8{}_i - \min PA^8{}_i]$ равен 0,09. Что практически не нарушает монотонного характера.

Гипотеза №9 подтверждается, так как для любого j-го тестового вопроса нашлось такое количество участников тестового опроса $K9$, что медиана ряда значений точечного коэффициента PA^j с увеличением объемов выборок хаотично, но устойчиво стремится к значению PA^j для всей совокупности для j-го вопроса.

Для дальнейших исследований и выдвижения новых гипотез рассмотрим этот результат более детально.

В экспериментах наблюдается, причем с самого минимального объема i-й выборки, полное совпадение значений медиан ряда точечных коэффициентов $PA^j{}_i$ со значением PA^j для всей совокупности. Например, для вопросов №2, №4 и №6.

Такого рода тенденция проявляется и для других вопросов, но уже не с минимального объема. Так, совпадение значений медиан ряда точечных коэффициентов со значением коэффициента PA^j для всей совокупности по следующим номерам вопросов: №1 – начиная с i = 40; для №3 – с i = 70; для №5 – с i = 80; для №7 – с i = 90; для №8 – с i = 30; для №9 – с i = 20.

Таким образом, гипотезы №6÷9, характеризующие динамику изменений значений точечного коэффициента, подтвердились.

Кроме того, сопоставляя рассматриваемые подтвержденные гипотезы и аналогичные подтвержденные гипотезы, характеризующие динамику изменений значений гистограммного коэффициента [3], получаем подтверждение **гипотезы №5.**

7. Исследование влияния когнитивных искажений на ответы экспертов на тестовые вопросы

Понятие «когнитивности» в настоящее время широко используется в научной сфере. Так, в работе [13] предлагается подход к социальной системе как «живой» когнитивной системе и применение данного подхода к современным тенденциям развития глобального информационного общества. Само понятие когнитивного искажения было введено Амосом Тверски и Даниэлем Канеманом в 1972 году [11] на основе исследования числовых оценок экспертов на поставленные авторами вопросы. На основе данного исследования авторы указали и продемонстрировали отличие полученных ответов экспертов от ожидаемых ответов на основе рационального выбора. Тверски и Канеман объяснили эти расхождения в суждениях и принятии решений в терминах эвристики. По мнению авторов, эвристика проста в процедурах вычисления, но иногда приводит к «серьёзным и систематическим ошибкам» [12, стр. 1125], вызывающим склонность экспертов к стереотипному мышлению.

Экспертные предубеждения в технологическом прогнозировании достаточно широко обсуждаются в научном сообществе. Один из самых свежих материалов поднимает проблему влияния предубеждений экспертов на процесс создания технологического прогноза [8].

Основываясь на понятии когнитивных искажений, выдвинутых Канеманом и Тверски, мы провели исследование рассматриваемого нами списка вопросов и сделали следующее заключение:

Вопросы №1 «Какое количество времени электропоезд метрополитена в Москве затрачивал на проезд по кольцу (в минутах) в 1979 году?» и №7 «Какое количество времени электропоезд метрополитена в Москве затрачивал на проезд от станции Киевская до станции Щелковская (в минутах) в 1979 году?» сами по себе являются типичными вопросами второго уровня сложности. Одна из причин, на основании которой можно обосновать этот уровень, заключается в том, что количество потенциально истинных вариантов ответов на них мало. Однако данные вопросы имеют признаки когнитивных искажений, существенно повышающих их субъективную степень сложности для экспертов. Постановка данных вопросов неявно отсылала экспертов к схеме московского метрополитена, на которой искажены пропорции длин маршрутов проезда в метро относительно фактических расстояний, что повлияло на результаты ответов экспертов. В данном случае было выявлено наличие у экспертов так называемых «когнитивных карт», отличающихся от истинного значения, причем отличие когнитивных карт было сконцентрировано на одном уровне у большей массы ответов экспертов. Выполняя адаптивную функцию, когнитивные искажения в основном приводят к более быстрому решению задач, однако из-за неуместного применения навыков эксперты в свои ответы вносят ошибку, которую нам удалось отловить и наглядно проиллюстрировать.

Для вопроса №1 эксперты существенно увеличили время прохождения поезда по кольцевой линии, так как на схеме московского метрополитена она отображена с существенными искажениями пропорций в сторону увеличения, а для вопроса №7 – занизили, так как схема искажала пропорции расстояний в сторону уменьшения между станциями метро (рис. 3). При этом эксперты игнорировали другие возможные решения задачи, выбирая «легкий» путь и наглядно демонстрируя прочную взаимосвязь с уже устоявшимися когнитивными картами, так как схема московского метрополитена – это наиболее известный образ, который они стали использовать при решении данной задачи.

Рис. 1. Схема московского метрополитена

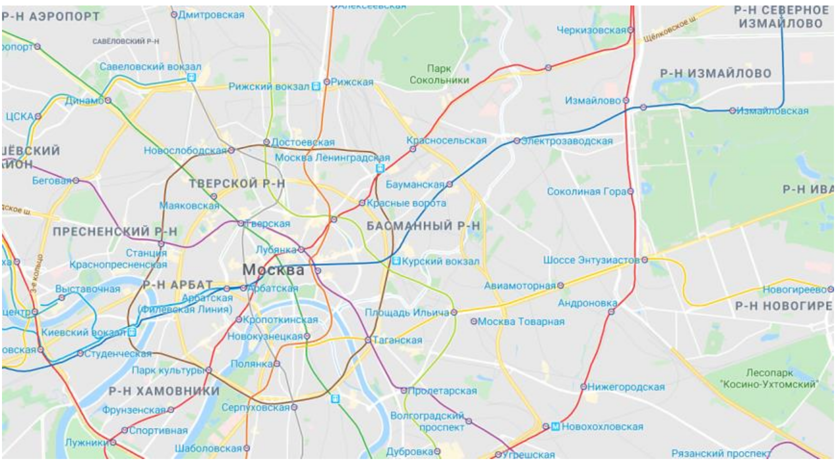

Рис. 2. Карта центра Москвы

Сопоставляя два рисунка – схему московского метрополитена (рис. 1) и карту центра Москвы, отражающую реальные пропорции фактических расстояний между станциями (рис. 2) – можно увидеть существенные искажения.

Таким образом, необходимо провести разграничение типов вопросов на объективный с искажениями и объективный без искажений уровня сложности.

8. Ранжировка вопросов

Сформируем отдельную таблицу, характеризующую соотношения упорядоченных значений коэффициентов $KH^j_i(N'^j, N^j)$ и $PA^j_i(x^j_1, x^j_2, \ldots, x^j_{N^j})$ по всей совокупности ответов экспертов на j-й вопрос с номером вопроса. При этом упорядочение задается от меньшего значения соответствующего коэффициента к большему.

Таблица 10. Ранжировка вопросов по возрастанию степени сложности от минимальной (1) к максимальной (9) исходя из значений коэффициентов KH^{ji} и PA^j_i

№ позиции (ранжировка по возрастанию коэффициентов)	Значение KH^j_i	№ вопроса	Значение PA^j_i	№ вопроса
1	36,36	8	0,01	8
2	43,97	4	0,07	4
3	44,72	2	0,08	6
4	47,95	5	0,09	9
5	49,65	6	0,14	2
6	57,14	9	0,28	7
7	57,53	1	0,3	5
8	58,33	7	0,38	1
9	67,14	3	0,44	3

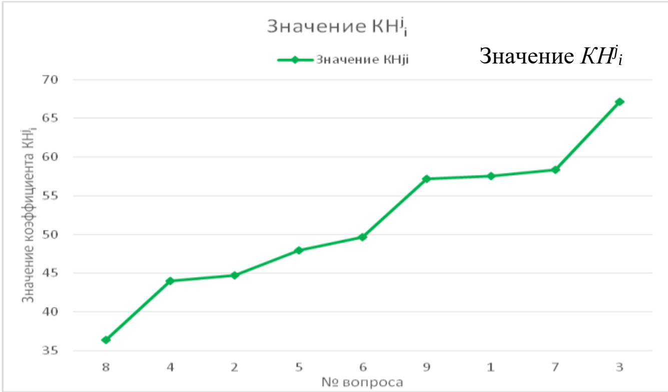

Рис. 3. Графическое представление уровня значений коэффициента $KH^j{}_i$ (от меньшего значения коэффициента к большему). На оси абсцисс указан номер вопроса, на оси ординат – значение соответствующего коэффициента

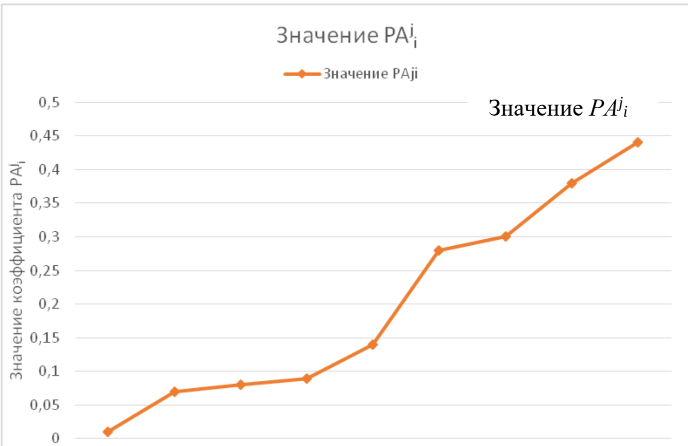

Рис. 4. Графическое представление уровня значений коэффициента $PA^j{}_i$ (от меньшего значения коэффициента к большему). На оси абсцисс указан номер вопроса, на оси ординат – значение соответствующего коэффициента

На основе ранжировки вопросов сразу по обоим коэффициентам можно сделать следующие выводы:

I. Самый низкий уровень сложности вопросов или, другими словами, объективно самыми легкими для экспертов оказались следующие вопросы:

№8. Какова длина корпуса люминесцентных ламп в вашей рабочей комнате (в см)?

№4. Какова была в Москве доля мужчин в 1970 г. (в процентном соотношении)?

В этом случае значения обоих коэффициентов $KH^j_i(N'^j, N^j)$ и PA^j_i оказались наименьшими, и в этом смысле в большинстве своем эксперты показали значения наиболее близкие к истинным.

Кроме того, полагаем, что данный результат однозначно характеризует данные вопросы как вопросы второго уровня сложности [6].

II. Более сложными оказались вопросы:

№2. Сколько вагонов в составе электропоезда метрополитена в Москве на линии Беляево – Медведково было в 1979 году?

№6. Продолжительность жизни Н. Г. Чернышевского (в годах)?

№9. Сколько граммов весит предложенная вам для оценки деревянная вазочка?

Хотя и на эти вопросы эксперты дали оценку достаточно близкую к истинной.

III. Самыми сложными для экспертов оказались следующие вопросы:

№1. Какое количество времени электропоезд метрополитена в Москве затрачивал на проезд по кольцу (в минутах) в 1979 году?

№3. Каким было население Африки в 1975 году (в млн чел.)?

№7. Какое количество времени электропоезд метрополитена в Москве затрачивал на проезд от станции Киевская до станции Щелковская (в минутах) в 1979 году?

В этом случае значения коэффициентов KH^j_i и PA^j_i оказались наибольшими, и в этом смысле эксперты дали свои оценки, наиболее далекие от истинных значений.

Более трудна трактовка уровня сложности пятого вопроса: «Каким было население Филиппин в 1970 году?». Это связано с тем, что исходя из значения коэффициента PA^j_i он достаточно сложный, но исходя из значения коэффициента KH^j_i не является таковым.

Таким образом, мы предлагаем следующую **гипотезу №10**:

Разночтение трактовки уровня сложности пятого вопроса по различным критериям связано с тем, что коэффициент $KH^j_i(N'^j, N^j)$ характеризует не только уровень его сложности, как коэффициент PA^j_i, но и интегральную меру уверенности совокупности экспертов.

Отличие между пятым и третьим вопросами состоит в том, что эксперты при оценивании вопроса №3 не всегда достаточно адекватно воспринимали степень трудности этого вопроса по сравнению с вопросом №5. Другими словами, усредненная степень понимания экспертами сложности вопроса №5 была больше, чем аналогичный показатель для вопроса №3. Именно это и привело к более высокой оценке сложности третьего вопроса и различиям в позициях рассмотренной ранжировки этих вопросов по разным критериям.

Эти различия в иерархии между интегральной оценкой степени уверенности совокупности экспертов в степени трудности задачи этого же вопроса, выраженной значением коэффициента $KH^j_i(N'^j, N^j)$, и интегральной оценкой уровня сложности этого же вопроса, выраженной значением коэффициента PA^j_i,не является простой проблемой и требует дополнительного исследования.

Для выяснения уровня взаимосвязи между двумя ранжировками, отмеченными в таблице 10, рассмотрим два вектора значений коэффициентов по возрастанию номеров вопросов. В первом векторе ($X1$) укажем значения коэффициентов KH^j_i, во втором векторе ($X2$) укажем значения коэффициентов PA^j_i:

$X1$ = [36,36; 43,97; 44,72; 47,95; 49,65; 57,14; 57,53; 58,33; 67,14],

$X2$ = [0,01; 0,07; 0,14; 0,3; 0,08; 0,09; 0,38; 0,28; 0,44].

Попробуем испытать (опровергнуть или нет) предположение, что нет никакой связи между признаками (значениями двух коэффициентов). То есть вторая последовательность $X2$ значения коэффициентов $PA^j{}_i$ случайна по отношению к первой $X1$.

Как обычно, если вторая последовательность выбирается чисто случайно, то близкие к единице (по абсолютной величине) значения коэффициента ранговой корреляции имеют малую вероятность, а типичные значения этого коэффициента лежат около нуля.

Для этого нами была проверена связь между двумя векторами $X1$ и $X2$ путем расчета ранговой корреляции Спирмена и ранговой корреляции Кендалла. Для расчетов использовалась библиотека SciPy версии 1.3.2 в программном продукте Jupyter Notebook [10].

Сначала рассчитаем корреляцию Спирмена.

Распределение коэффициента ранговой корреляция Спирмена ρ при разных значениях K (число элементов ранжировки) можно найти в специальных сборниках статистических таблиц, например в таблицах Л.Н. Большева и Н.В. Смирнова [1, с. 153] либо использовать уже существующие программные продукты. На этом основании рассчитаем значение коэффициента ранговой корреляция Спирмена:

$\rho = 0{,}7833333333333333$.

Напомним, что коэффициент может изменяться в диапазоне от –1 до +1. При полной согласованности между элементами двух последовательностей коэффициент равен единице. Равенство коэффициента минус единице наблюдается при полной рассогласованности между элементами двух последовательностей.

Результат: уровень взаимосвязи между двумя отмеченными ранжировками $X1$ и $X2$ близок к единице на основе расчета коэффициента корреляция Спирмена.

Согласно трактовке Л.Е. Полякова [7] при использовании коэффициента ранговой корреляции можно условно трактовать тесноту связи между признаками следующим образом. Так, значения коэффициента корреляции 0,3 и менее – показатели слабой тесноты связи; значения более 0,4, но менее 0,7 – показатели умеренной тесноты связи, а значения 0,7 и более – показатели высокой тесноты связи.

Таким образом, в рамках данной трактовки между двумя векторами имеется высокая теснота связи.

Следующая задача, которую мы должны решить: «с какой степенью надежности мы можем полагаться на заключение о том, что в совокупности, из которой произведен выбор, существует корреляция, если получен данное значение выборочного коэффициента корреляции ρ. То есть необходимо проверить существенность наблюдавшейся корреляции рангов…» [4].

Рассчитаем P-значение для тестовой гипотезы, нулевая гипотеза которой состоит в том, что два набора данных некоррелированы, и имеет ту же размерность.

На основе стандартных расчетов SciPy версии 1.3.2 в программном продукте Jupyter Notebook получаем значение статистической значимости [10]

$P = 0{,}012519873019449882$, что меньше табличного.

Следовательно, такое значение $\rho \geq 0{,}78333$ могло появиться случайно лишь с вероятностью 0,012, что практически невозможно.

Это заставляет нас считать, что гипотеза независимости признаков опровергается имеющимися наблюдениями. И тем самым доказана положительная связь между значениями точечного и гистограммного коэффициентов.

К вышеизложенным выводам необходимо сделать замечание. Так как соответствующее число элементов ранжировки менее 10, а оно определяется через K – общее число вопросов, то нет необходимости вводить поправку на непрерывность.

Аналогичным образом для анализа уровня взаимосвязи можно использовать коэффициент ранговой корреляции Кэндэла.

Коэффициент ранговой корреляции Кендалла равен

$\tau = 0{,}6666666666666666$.

Далее, рассчитаем P-значение для тестовой гипотезы, нулевая гипотеза которой состоит в том, что два набора данных не коррелированы, и имеет ту же размерность.

$P = 0{,}012665343915343916$.

Таким образом, получаем аналогичный вывод.

9. Возможные направления дальнейших разработок

1. Для подтверждения правдоподобности выводов необходимо провести дополнительные экспериментальные исследования на другом массиве испытуемых и массиве данных, полученных по результатам опроса. Тем самым, варьировать эти массивы как факторы, от которых зависит уровень правдоподобности гипотез и соответствующих выводов.

2. Необходима экспериментальная проверка введенной нами гипотезы №10. Мы полагаем, что сам вид экспертной оценки в виде гистограммы дает информацию о распределенной степени уверенности эксперта в своей оценке. В связи с этим будут проведены дополнительные исследования относительно соотношения между уровнем сложности тестовых задач и интегральной меры уверенности совокупности экспертов в понимании этих задач.

3. Необходимо дополнительное экспериментальное исследование уровня влияния когнитивных искажений на значения гистограммных $KH^j_i(N'^j, N^j)$ и точечных $PA^j_i(N^j_i)$ коэффициентов для любого j-го тестового вопроса. Данное исследование, как мы полагаем, может выявить еще и взаимосвязь степени уверенности экспертов с ограничениями эвристических принципов.

4. Необходимо выдвижение новых гипотез на основе более детального исследования результатов полученных при анализе таблиц 1–9.

10. Заключение

1. На основе тестовых экспериментов было обнаружено наличие объективной составляющей уровней сложности задач.

2. Для подтверждения уровня правдоподобности выводов относительно наличия объективной составляющей сложности задач был введен и экспериментально исследован точечный коэффициент. Он позволяет оценить уровень сложности задачи заданной в виде тестового вопроса. В результате исследования было обнаружена устойчивость полученных выводов относительно наличия объективной составляющей уровней сложности задач:

– от числа опрашиваемых участников;

– от вида коэффициента, на основе которого и был получен этот вывод.

3. Экспериментально выявлен критерий для определения тестового вопроса как задачи третьего уровня сложности, когда спектр возможных вариантов ответов очень широк. Эксперимент показал подтверждение условий указанных в найденном критерии, что только у одного из девяти вопросов, а именно третьего, как впоследствии подтвердилось самого трудного, на основании расчета гистограммного и точечного коэффициентов. Кроме того, как было показано в первой части данного исследования [3] есть явно выраженные отличия в характере изменения диапазона разброса значений именно гистограммного коэффициента. Таким образом, полагаем, что критерием для определения тестового вопроса как задачи третьего уровня сложности является специального вида изменения диапазона разброса значений именно гистограммного коэффициента $KH^j_i(N'^j, N^j)$. Этот диапазон разброса значений строго монотонно убывает с увеличением объема i-й выборки, но именно для третьего вопроса начиная с объема выборки равной 90, а у всех остальных вопросов данный диапазон строго монотонно убывал, начиная с минимальных значений объемов выборок равным 10 либо 20. (См. обоснование гипотезы №3 для гистограммного коэффициента [3]).

4. Предложены тестовые задачи, для которых, на основе экспериментальных исследований, показано наличие существенного влияния когнитивных искажений на процесс принятия решений экспертами. Наглядно проиллюстрировано, что выполняя адаптивную функцию, когнитивные искажения хотя и приводят к более быстрому решению задач, но вносят ошибки.

5. Предложен план дальнейших исследований по теме: экспериментальные исследования тестовых вопросов.

Литература

1. БОЛЬШЕВ Л.Н., СМИРНОВ Н.В. *Таблицы математической статистики.* – М.: «Наука», 1965. – С. 153.

2. *Генерация псевдослучайных чисел в среде Visual* Basic [Электронный ресурс] // Случайная выборка. – URL: https://www.planetaexcel.ru/techniques/2/94/ (дата обращения: 15.12.2019).
3. КАЛМЫКОВ Н.С., СИДЕЛЬНИКОВ Ю.В. *Введение и экспериментальное исследование гистограммного коэффициента как инструмента для изучения явления сверхдоверия в среде экспертов* // Управление большими системами. – 2020. – Вып. 87. – С. 101–130.
4. *Ранговые корреляции. – Зарубежные статистические исследования.* – М.: «Статистика», 1975. – 216 с.
5. СИДЕЛЬНИКОВ Ю.В. *Системный анализ технологии экспертного прогнозирования*. – М.: Изд-во МАИ-ПРИНТ «МАИ», 2007. – 348 с.
6. СИДЕЛЬНИКОВ Ю.В., САЛТЫКОВ С.А. *Процедура установления соответствия между задачей и методом* // Экономические стратегии. – 2008. – №7(65). – С. 102–109.
7. *Статистические методы исследования в медицине и здравоохранении* / Под ред. проф. Л.Е. Полякова. – Л.: Медицина, 1971. – 200 с.
8. BONNARCOSI A., APREDA R., FANTONI G. *Expert biases in technology foresight. Why they are a problem and how to mitigate them* // Technological Forecasting and Social Change. – February, 2020. – Vol. 151.
9. CHENG M.N., WONG JANE W.K., CHEUNG C.F., LEUNG K.H. *A scenario-based roadmapping method for strategic planning and forecasting: A case study in a testing, inspection and certification company* // Technological Forecasting and Social Change. – October, 2016. – Vol. 111. – P. 44–62.
10. JUPYTER NOTEBOOK *Программный компонент*. – URL: https://jupyter.org/ (дата обращения 19.04.2020).
11. KAHNEMAN D., TVERSKY A. *Subjective Probability: A Judgment of Representativeness* // Cognitive psychology. – 1972. – No. 3. – The Hebrew University, Jerusalem, 1972. – P. 430-454.
12. TVERSKY A., KAHNEMAN D. *Judgement under uncertainty: Heuristics and biases* // Sciences journal. – 1974. – Vol. 185. – P. 1124–1131.

13. VEITAS V., WEINBAUM D. *Living Cognitive Society: A 'digital' World of Views* // Technological Forecasting and Social Change. – 2016. – Vol. 114. – P. 16–26.

EXPERIMENTAL STUDIES ON THE TEST QUESTIONS BASED ON THE POINT RATIO

Nikita Kalmykov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, junior researcher (kalmikov-nik@bk.ru).
Yuri Sidelnikov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, leading researcher; Moscow Aviation Institute, Moscow, professor
(sidelnikovy@mail.ru).

Abstract: The article proposes the existence of an objective component of the complexity of the problem which is experimentally verified and proposes the point ratio which is introduced and experimentally studied. For this purpose, the so-called "point coefficient" is introduced and experimentally investigated, which allows us to estimate the level of complexity of the problem given in the form of a test questions. This coefficient is based on the answers to these questions that the level of complexity of the problem is estimated. Hypotheses describing the dynamics of changes in the values of the point coefficient are proposed and experimentally tested. Also, the stability of the obtained conclusions based on the results of calculations for two qualitatively different coefficients relative to the number of respondents was checked. The criterion for determining the test question as a problem of the third level of complexity is experimentally found and tested. The reason for this is the nature of the change in the range of the spread of the values of the histogram coefficient. An experimental study of the presence of cognitive distortions in the assessment of some test tasks by experts was conducted. Possible directions for further development are proposed.

Keywords: expert appraisals, test question, point ratio.

УДК 303.833.5+303.832.32
ББК 65.054. 3+ 16.23+ 87.256.643
DOI: 10.25728/ubs.2021.89.2

Статья представлена к публикации членом редакционной коллегии В.Н. Бурковым.

*Поступила в редакцию 31.08.2020.
Опубликована 31.01.2021.*