



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 11 за 2023 г.



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

**Л.В. Александрова,
М.Ю. Кислицына, Ю.Н. Орлов**

**Методы нахождения
взаимосвязей при обработке
психологических тестов**

Статья доступна по лицензии
[Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)



Рекомендуемая форма библиографической ссылки: Александрова Л.В., Кислицына М.Ю., Орлов Ю.Н. Методы нахождения взаимосвязей при обработке психологических тестов // Препринты ИПМ им. М.В.Келдыша. 2023. № 11. 18 с. <https://doi.org/10.20948/prepr-2023-11>
<https://library.keldysh.ru/preprint.asp?id=2023-11>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

**Л.В. Александрова, М.Ю. Кислицына,
Ю.Н. Орлов**

**Методы нахождения взаимосвязей
при обработке психологических
тестов**

Москва — 2023

Александрова Л.В., Кислицына М.Ю., Орлов Ю.Н.

Методы нахождения взаимосвязей при обработке психологических тестов

В работе описаны подходы и метрики, используемые для интерпретации результатов психологического тестирования. Рассмотрены способы автоматической кластеризации респондентов по группам, отвечающим определенным паттернам специфических способностей. Обсуждается эффект уменьшающейся отдачи Спирмена, который хорошо проявляется при медианном сравнении, но нарушается в других шкалах оценивания.

Ключевые слова: психометрический анализ, корреляция, эффект Спирмена

Alexandrova L.V., Kislitsyna M.Yu., Orlov Yu.N.

Methods of finding relationships in the processing of psychological tests

The paper describes the approaches and metrics used to interpret the results of psychological testing. The methods of automatic clustering of respondents into groups corresponding to certain patterns of specific abilities are considered. The effect of Spearman's diminishing returns is discussed, which is well manifested in the median comparison, but is violated in other evaluation scales.

Keywords: psychometrical analysis, correlation, Spearman effect

Содержание

1. Введение	3
2. Анализируемые данные	6
3. Оценка эффекта Спирмена по точности ответов	7
4. Построение паттернов точности ответов	11
5. Заключение.....	18
Литература	18

1. Введение

Статистический анализ психометрических тестов – стандартная процедура, применяемая для определения психологических коррелятов и кластеризации респондентов по тем или иным признакам. Однако применение стандартных статистических процедур – таких как вычисление средних по группе, коэффициентов корреляции и квантилей распределений – связано с некоторыми неявными допущениями, выполнение которых априори неочевидно. Например, в работе [1] обсуждается возможная нелинейность шкал оценивания интеллекта, а также причины невоспроизводимости результатов психологических оценок с той точностью, с которой формально были сделаны выводы из психометрического анализа. Также важно понимать, что предположения об одинаковом типе интеллекта без его формального математического (то есть допускающего прямое измерение) определения являются неподтвержденными гипотезами. Иными словами, гипотеза о том, что любая группа респондентов представляет собой некоторую выборку из стационарной генеральной совокупности, может оказаться неверной.

В этой связи исследование различных метрик оценивания интеллекта и способов кластеризации респондентов в соответствии с различными генеральными совокупностями представляет практический интерес.

Заметим также, что интерпретация статистики в социальных науках носит эвристический характер. Например, ставится задача определить учебное заведение, подготавливающее самых высококлассных (то есть качественных) специалистов, или, более детально, установить, есть ли связь между некоторым институтом (школой) и качеством ученика или выпускника. Для этого составляются так называемые таблицы сопряженности (см., например, [2], раздел «корреляционный анализ»). Таблица сопряженности – это четырехуровневая таблица A_{ipqs} , которая показывает долю агентов i -го предприятия p -го типа q -го качества, значение которого (качества) относится к s -му классовому интервалу. Вышеуказанная задача состоит в том, чтобы определить, на каком уровне значимости устанавливается связь между предприятием и качеством (для каждого направления обучения и для каждого вида качества). Для решения этой задачи существует довольно большое количество статистических критериев [3], использующих такие понятия, как коэффициенты корреляции, ассоциации, коллигации, контингенции и ряд других. Неоднозначность ранжирования вызвана тем, что использование любых из этих коэффициентов дает формально ответ не на поставленный вопрос, а лишь позволяет на определенном уровне значимости принять или отклонить гипотезу об отсутствии корреляции между учебным заведением и уровнем качества обучения.

Кроме того, само ранжирование допускает различные формулировки. Например, если построить распределения частот, т.е. эмпирических

вероятностей того, что выпускник p -й специальности, подготовленный в i -м учебном заведении, относится к s -му классу по q -му виду качества, то можно ранжировать предприятия по величине этих частот, относящихся к определенному показателю качества. После этого, в частности, можно назвать самым «качественным» то учреждение, у которого максимальная доля выпускников высшего уровня по качеству, и выстроить эти заведения по убыванию доли высококачественных специалистов. Но можно также назвать самым качественным то предприятие, которое выпускает минимальную долю специалистов наихудшего качества. Предлагаемые два варианта оценивания не всегда приведут к одному и тому же результату в виде последовательности рангов. Таким образом, говоря о тенденциях или псевдо-закономерностях, которые можно извлечь из собираемых статистик, следует иметь в виду, что они относятся не столько к изучаемой связи, сколько к способу оценивания.

Трудность с проведением ранжирования существенно возрастает, если типы интеллекта неизвестны. Очевидно, результаты статистического анализа изменятся, если все типы объединить в один и назвать его просто «интеллект». В общем случае получаемая тогда трехиндексная таблица A_{iqs} результатов такого анализа не позволяет построить четырехиндексную таблицу A_{ipqs} . Если же мы объединим также и предприятия – «изготовители» специалистов –, то получаются данные в виде двухиндексной таблицы A_{qs} о результатах тестирования просто каких-то респондентов. Из этих укрупненных данных стараются обычно извлечь информацию в виде условных вероятностей: если объект относится к классу s по качеству q , то какова вероятность того, что он относится к классу s' по качеству q' ?

Таковы, например, данные экзаменационных оценок студентов по Q предметам по S -балльной шкале оценок. Если не изучать причины того, почему была получена та или иная оценка, то есть не вводить «типы студентов», то оцениваемые вероятности получаются недостоверными. Недостоверность проявляется не в статистической неточности оценивания выборочных параметров, а в том, что для другой группы студентов, имеющих иное распределение по «типам», на том же достаточно высоком уровне статистического доверия могут быть получены существенно иные условные вероятности. Следовательно, без знания внутренней структуры множества изучаемых объектов сложно делать количественные обобщения средних результатов измерений на любое его подмножество. Особенно это касается выводов о гипотетической связи измеряемого параметра (например, оценки за экзамен или ответа на тест) и неизмеряемого параметра – способности к определенным действиям, уровня умственного развития и т.п.

В последнее время на интерпретацию корреляций ответов применительно к психологическим тестам обращается пристальное внимание. Так, в работе [4] указывается на недостаточную методологическую обоснованность методов оценки так называемого закона убывающей отдачи Спирмена (Spearman's Law

of Diminishing Returns, SLODR, [5]), согласно которому корреляции между результатами тестирования интеллекта оказываются выше для менее интеллектуальных респондентов. В работах [6, 7] также обсуждаются вопросы, связанные с метриками для оценивания эффекта SLODR. Применительно к психологическим исследованиям эффект состоит в том, что если дать группе респондентов набор задач для выявления различных умственных способностей (например, задачи на выполнение арифметических действий, на пространственное восприятие, на память, на логические рассуждения и ряд других) и затем разделить эту группу по результатам на две подгруппы – лучшую и худшую, то в худшей корреляция между ответами на разные вопросы будет выше, чем в лучшей.

Важно подчеркнуть, что на практике часто возникают ситуации с нарушением отмеченной тенденции, поскольку пока нет общепринятой робастной метрики, в которой эффект имеет математически корректное объяснение. Например, близость к нулю линейного коэффициента корреляции не гарантирует отсутствия взаимосвязи между явлениями. Близость к единице коэффициента ассоциации [2, 3] не гарантирует также и наличия прямой функциональной связи, если среднеквадратичное отклонение этого коэффициента больше его самого.

Обычно отклонение от ожидаемого эффекта SLODR объявляется артефактом, связанным с: различиями в последовательности заданий; самообучением в процессе тестирования; внешними воздействиями; различиями в интеллектуальном составе групп. Тем самым проблема переносится в область интерпретации результатов тестирования в рамках тех или иных моделей интеллекта. Не обсуждая здесь собственно различные теории интеллекта, мы обратим внимание на некоторые аспекты психометрического оценивания респондента по результатам тестирования.

Настоящая работа продолжает исследования, начатые в [8, 9]. В этих работах анализируются результаты одного достаточно репрезентативного психологического теста, описанного в [4, 10]. В работах [8, 9] было показано, что анализируемое множество ответов респондентов неоднородно в статистическом смысле, в связи с чем был предложен способ некоторой их кластеризации с тем, чтобы уже внутри кластера определить более или менее успешные подгруппы. Здесь мы рассмотрим этот подход более детально.

Следует сказать, что существенной проблемой анализа результатов тестирования является отсутствие априорного знания о степени зависимости анализируемых качеств q_1, q_2, \dots, q_Q . Если качества предполагать независимыми, то каждый респондент своими ответами порождает некоторый персональный профиль уровня владения этими качествами. Если затем респондентов объединить в группы по сходным профилям, то появятся типовые «паттерны качеств», которые могли бы служить способом разбиения на группы по типам интеллекта. Однако метрика близости профилей определяется тем, насколько качества зависимы, а выяснить это можно только через ответы респондентов,

что приводит к замкнутому кругу. Было бы интересно построить такие профили хотя бы в «нулевом» приближении, анализируя распределения ответов на тестовые задания в рамках обычных метрик типа L1 или C. Учитывая незавершенность теории интеллекта и связанную с ней неоднозначность интерпретации, рассмотрим, какое содержание можно тем не менее извлечь из результатов тестирования.

2. Анализируемые данные

Кратко опишем состав и формат анализируемых данных. Обрабатывались данные тестирования $N=11335$ респондентов. Каждому респонденту в рамках каждой из 10 различных методик P_i были заданы по 30 вопросов. Оценка каждого ответа имела 4 составляющих: продуктивность, точность, скорость, эффективность (это показатели «качества» q_i).

Область определения оценки различна для каждого из четырех качеств. Продуктивность оценивается целыми числами от 0 до 30. Скорость, точность и эффективность оценивается вещественными числами соответственно от 0 до 6, от 0 до 1 и от 0 до 100. Например, по точности каждый ответ на отдельный вопрос оценивался по системе «верно – 1», «неверно – 0». Итоговая оценка точности по каждой методике получалась как среднее арифметическое отдельных ответов и принимала значение от 0 до 1. Аналогично и остальные шкалы в нашем исследовании перенормируются на $[0; 1]$.

Тестировались следующие специальные качества интеллекта [4, 10]:

- P_1 – способность к пониманию отношений между понятиями, или «аналогии»;
- P_2 – способность к числовому мышлению, или «числовые ряды»;
- P_3 – зрительная память на абстрактные контурные изображения, или «память на фигуры»;
- P_4 – способность конструировать целое по имеющимся частям, или «узоры»;
- P_5 – оценка навыков вычислений, или «арифметический счет»;
- P_6 – оценка памяти на слова, или «вербальная память»;
- P_7 – нахождение лексического эквивалента для знаковой последовательности, или «установление закономерности»;
- P_8 – способность рассуждению по схеме причинно-следственных связей, или «силлогизмы»;
- P_9 – способность к обобщению при анализе вербальной информации, или «исключение слова»;
- P_{10} – способность к пространственному вращению объекта в трех измерениях, или «кубы».

Каждый респондент, таким образом, оценивался 10-мерным вектором точностей, а также 10-мерными векторами скоростей, продуктивностей и эффективностей.

Эмпирические функции распределения точности ответов для каждой методики приведены на Рис. 1.

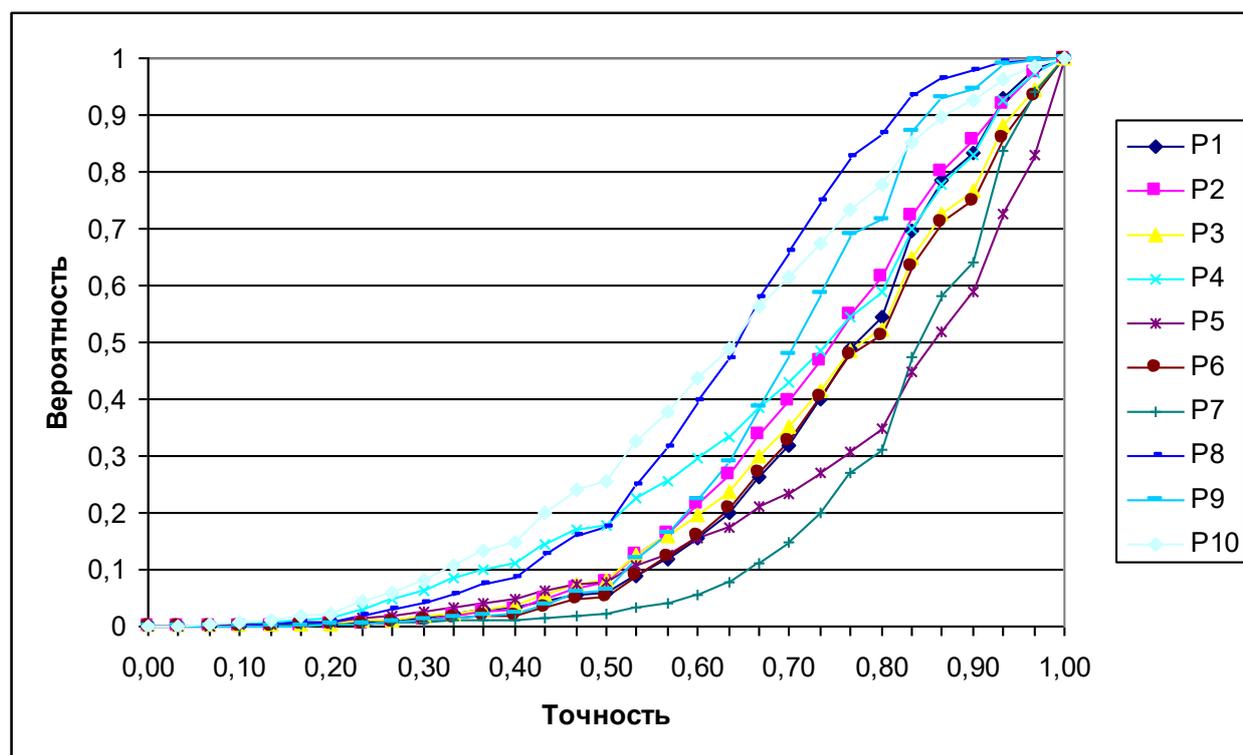


Рис. 1 – Распределения точности ответов по 10 методикам

В работе [8] было показано, что все эти распределения различны, то есть их следует трактовать как выборочные из разных генеральных совокупностей. Следовательно, выборки ответов неоднородны. Но даже если бы они и были однородны, это все равно не имело бы значения, поскольку при построении распределений все ответы были объединены, и потому неизвестно, одни и те же респонденты оказываются в заданных классовых интервалах оценок по рассматриваемым методикам или же это разные люди. Поэтому статистика ответов по множеству в целом не имеет большого значения, а требуется проведение сравнения индивидуальной точности ответов в рамках тех или иных методик по выявлению специфических качеств.

3. Оценка эффекта Спирмена по точности ответов

Проанализируем попарную зависимость точности ответов от применяемых методик для множества респондентов в целом. Рассмотрим сначала попарную ассоциацию между методиками применительно к первой и последней оценкам качества – «двоечникам» и «отличникам». Разбиваем всю шкалу оценок на 4 класса равномерно с шагом 0,25: «плохо», «удовлетворительно», «хорошо», «отлично». Соответственно, первый класс («плохо») идентифицируется числом 1, второй 2, третий 3, четвертый 4.

Положим $a(i; j, k)$ – число респондентов, имеющих оценку i по обеим методикам j и k ; $b(i; j, k)$ – число респондентов, имеющих оценку i по методике j , но не имеющих ее по методике k ; $c(i; j, k)$ – число респондентов, не имеющих оценку i по методике j , но имеющих ее по методике k ; и, наконец, $d(i; j, k)$ – число респондентов, не имеющих оценки i по обеим методикам. Для «двоечников» $i = 1$, а для «отличников» $i = 4$. В этих терминах коэффициент ассоциации определяется формулой

$$L_{jk}(i) = \frac{a(i; j, k)d(i; j, k) - b(i; j, k)c(i; j, k)}{a(i; j, k)d(i; j, k) + b(i; j, k)c(i; j, k)}. \quad (1)$$

Так, например, при сравнении методик 1 и 2 (аналогии и числовые ряды) оказалось, что для «двоечников» $L_{12}(1) = 0,97$, а для «отличников» $L_{12}(4) = 0,55$, что согласуется с представлениями SLODR. Такая же картина наблюдается для всех пар сочетаний методик (см. таблицы 1 и 4): $L_{ij}(1) > L_{ij}(4)$.

Возникает вопрос: имеет ли место указанный эффект для любых интервалов шкал? То есть верно ли, что если $r < s$, то $L_{ij}(r) > L_{ij}(s)$?

Сравнение между собой «хорошистов» и «отличников» приводит к другим результатам: $L_{12}(3) = 0,32 < L_{12}(4)$. Следовательно, эффект SLODR в данной метрике не универсален.

Применительно ко всему массиву данных коэффициенты парной ассоциации результатов тестов по применяемым методикам приведены в таблицах 1-4.

Таблица 1 – Коэффициенты ассоциации для «двоечников» $L_{ij}(1)$

Метод	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1		0,97	0,94	0,78	0,89	0,95	0,95	0,93	0,97	0,80
P2			0,96	0,89	0,94	0,97	0,97	0,94	0,98	0,87
P3				0,78	0,88	0,96	0,95	0,84	0,97	0,86
P4					0,79	0,85	0,89	0,79	0,92	0,74
P5						0,94	0,94	0,87	0,95	0,77
P6							0,97	0,92	0,98	0,91
P7								0,95	0,98	0,90
P8									0,96	0,82
P9										0,94

Как видно, для всех пар методик коэффициенты $L_{ij}(1)$ выше 0,7, то есть весьма высоки и большей частью близки к единице. Лишь некоторые из них меньше 0,8. Наименее связаны между собой ответы по методикам (1, 4), (3, 4), (5, 4), (8, 4), (10, 4), (10, 1), (10, 5), т.е. для «двоечников» методики «узоры» (4)

и «кубы» (10) слабее всего связаны с остальными специальными способностями.

Таблица 2 – Коэффициенты ассоциации для «троечников» L_{ij} (2)

Метод	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1		0,78	0,54	0,59	0,66	0,64	0,79	0,71	0,77	0,45
P2			0,56	0,59	0,75	0,61	0,80	0,65	0,70	0,45
P3				0,46	0,55	0,69	0,62	0,40	0,53	0,39
P4					0,53	0,43	0,63	0,50	0,55	0,54
P5						0,62	0,71	0,58	0,61	0,48
P6							0,80	0,51	0,60	0,40
P7								0,65	0,80	0,51
P8									0,65	0,47
P9										0,44

Во всех без исключения случаях ассоциация «троечников» ниже, чем «двоечников». Для «троечников» наименьшие ассоциации (менее 0,5) проявляются для пар методик (3, 4), (6, 4) и для всех вариантов с методикой 10. Таким образом, и для этой группы методики «узоры» и «кубы» наименее связаны с остальными способностями.

Таблица 3 – Коэффициенты ассоциации для «хорошистов» L_{ij} (3)

Метод	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1		0,32	0,14	0,23	0,26	0,23	0,35	0,04	0,36	0,04
P2			0,21	0,20	0,39	0,16	0,26	0,11	0,25	0,08
P3				0,12	0,15	0,42	0,24	-0,01	0,11	0,03
P4					0,20	0,12	0,13	0,09	0,14	0,17
P5						0,21	0,34	0,00	0,21	-0,01
P6							0,30	0,01	0,17	0,00
P7								-0,15	0,29	-0,09
P8									0,12	0,22
P9										0,08

Ассоциации между различными методиками у «хорошистов» по большей части отсутствуют: в основном они по модулю меньше 0,2. Заметными являются только пары (1, 2), (2, 5), (3, 6), (1, 7), (5, 7) и (1, 9). Пока что наблюдается монотонное убывание ассоциации с ростом качества ответов.

Таблица 4 – Коэффициенты ассоциации для «отличников» $L_{ij}(4)$

Метод	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1		0,55	0,36	0,55	0,51	0,39	0,59	0,61	0,60	0,46
P2			0,43	0,53	0,68	0,38	0,60	0,55	0,45	0,50
P3				0,44	0,41	0,65	0,48	0,30	0,31	0,41
P4					0,55	0,37	0,66	0,50	0,44	0,64
P5						0,44	0,57	0,51	0,45	0,48
P6							0,49	0,37	0,38	0,38
P7								0,61	0,61	0,64
P8									0,57	0,52
P9										0,42

Сравнивая результаты таблиц 3 и 4, видим, что коэффициенты ассоциации «отличников» значительно выше, чем «хорошистов», и близки к «троечникам», хотя в целом несколько меньше этих последних. Тем не менее 8 из 45 пар $L_{ij}(4)$ превосходят $L_{ij}(2)$: это пары (4, 5), (4, 7) и «кубы» (10, 1), (10, 2), (10, 3), (10, 4), (10, 7), (10, 8). Таким образом, наблюдается противоречие эффекту SLODR. Наблюдаемая картина может быть вызвана спецификой групп, из которых состоят респонденты, но для всех пар методик характерна немонотонная зависимость коэффициента ассоциации от качества ответа.

Таким образом, обнаружен эффект немонотонной ассоциативной связи между ответами на тестовые методики без разбиения респондентов на группы.

Полезно рассмотреть еще один способ оценивания эффекта SLODR для сравнения связей внутри первого и последнего интервалов шкалы оценок точности. Пусть $N_1^{(1)}(i)$ есть число респондентов, попавших в первый интервал оценок по методике P_i , а $N_1^{(2)}(i, j)$ есть число респондентов, попавших в первый интервал оценок по обоим методикам P_i и P_j . Обозначим через $v_1^{(2)}(i, j)$ долю множества пересечения в объединении этих множеств:

$$v_1^{(2)}(i, j) = \frac{N_1^{(2)}(i, j)}{N_1^{(1)}(i) + N_1^{(1)}(j) - N_1^{(2)}(i, j)}. \quad (2)$$

Аналогично определим долю пересечения в объединении s множеств для первого и последнего классов интервалов $v_{1,4}^{(s)}(i_1, \dots, i_s)$. Эффект SLODR в этом способе оценивания состоит в том, что ожидаемо должно быть $v_{1,4}^{(s)}(i_1, \dots, i_s) > v_4^{(s)}(i_1, \dots, i_s)$.

Рассмотрим для краткости один из вариантов составления множества пересечения – как последовательного пересечения ответов по методикам P_1 и P_2 , затем P_1 , P_2 и P_3 и, наконец, P_1 , P_2 , P_3 и P_4 . В этом примере важна тенденция, которая оказалась противоположной ожидаемому эффекту. Так,

доля респондентов из низшей группы, которая одинаково плохо ответила на вопросы методик P_1 и P_2 , оказалась равной $\nu_1^{(2)}(1,2)=0,34$, тогда как доля респондентов из высшей группы, одинаково отлично ответивших на вопросы тех же методик P_1 и P_2 , оказалась равной $\nu_4^{(2)}(1,2)=0,52$. Далее, если включить в рассмотрение методику P_3 , то аналогичные доли равны $\nu_1^{(3)}(1,2,3)=0,17$ и $\nu_4^{(3)}(1,2,3)=0,32$. Для доли пересечения по четырем методикам получаем $\nu_1^{(4)}(1,2,3,4)=0,07$ и $\nu_4^{(4)}(1,2,3,4)=0,23$. Во всех примерах стабильно доля одинаковых «отличников» выше доли одинаковых «двоечников».

Таким образом, результаты ранжирования зависят от способа оценки качества.

4. Построение паттернов точности ответов

Для анализа данных были построены распределения респондентов по всем четырем характеристикам их ответов на вопросы, в результате чего в [9] были получены 40 распределений.

Классифицируем оценки ответов следующим образом: выделим 10% от общего числа респондентов с наименьшими и наибольшими значениями оценки (Рис. 2) и назовем их децилями «двоечников» и «отличников» соответственно. Выполним это действие для всех критериев оценок и всех блоков вопросов.

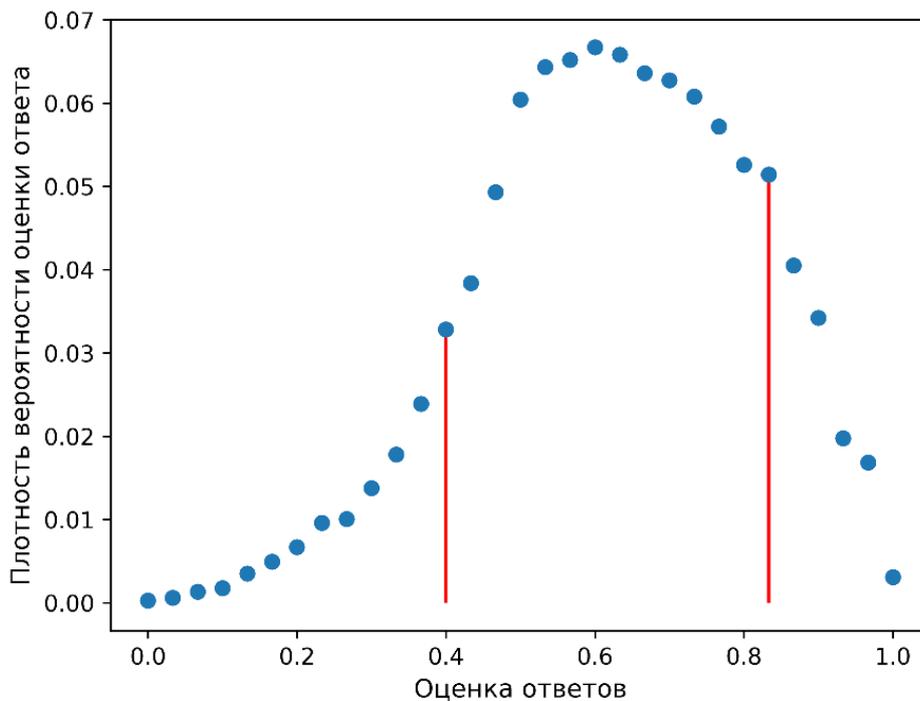


Рис. 2 – Пример распределения плотности вероятности оценки ответа для одного блока вопросов

Для анализа зависимости критериев между собой в блоке найдем пересечение ID респондентов в дециле, то есть определим, какие оценки в одном и том же дециле принадлежат одним и тем же людям. Выполним следующие действия: в выделенном дециле каждой оценки запишем ID респондентов, определим число пересекающихся людей для двух и трех критериев оценивания. В итоге для каждого рассматриваемого пересечения критериев были получены следующие значения: общее число уникальных респондентов в каждом блоке вопросов, число пересекающихся людей по блокам и разница между числом респондентов от пересечения трех и двух оценок. На графиках ниже представлены гистограммы пересечения респондентов по различным критериям оценки для всех блоков вопросов.

Пересечения в дециле респондентов, относящихся к группе «отличников», приведено на Рис. 3 – 6.

Число пересечений респондентов в 10% отличников по совокупности ответов для 1, 3 и 4 оценок

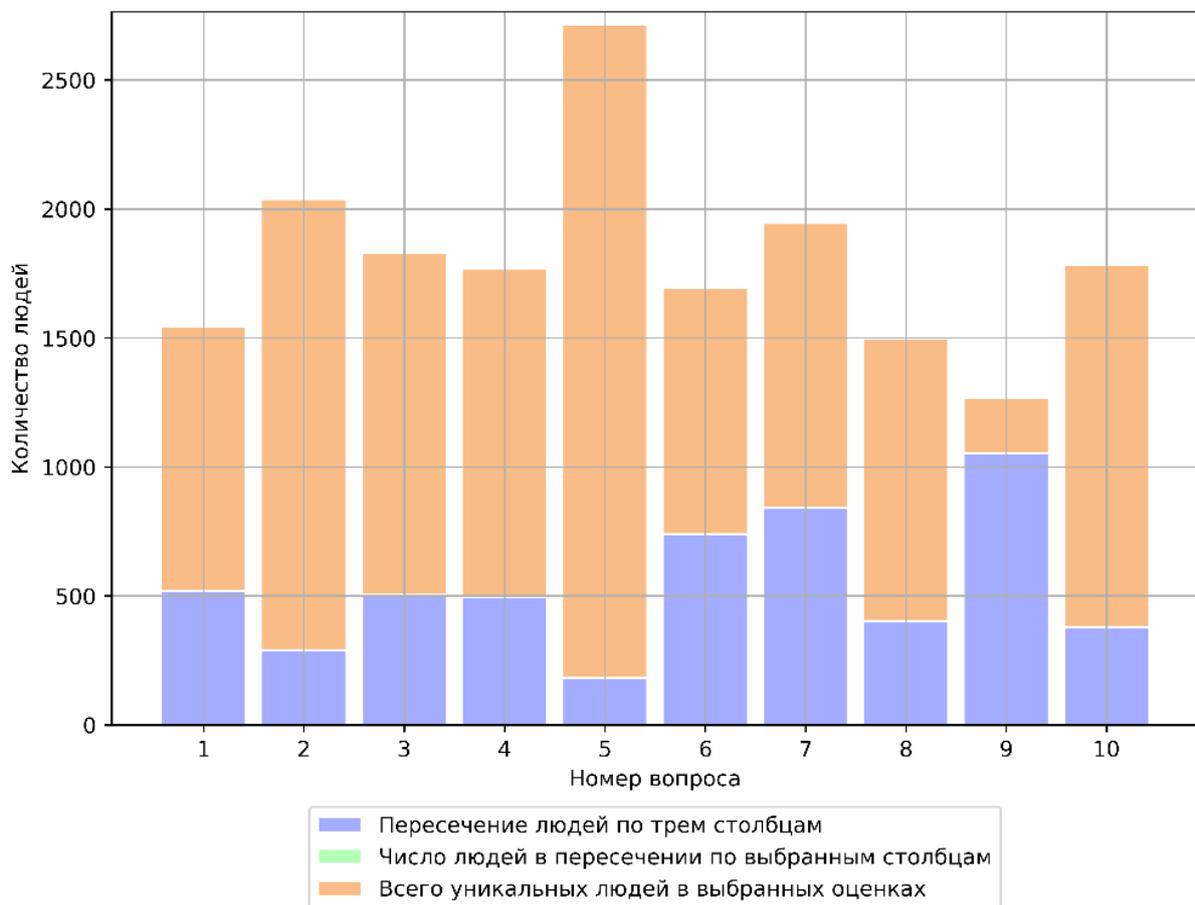


Рис. 3 – Пересечение респондентов по трем критериям (точность, продуктивность, эффективность)

Число пересечений респондентов в 10% отличников по совокупности ответов для 1 и 3 оценки

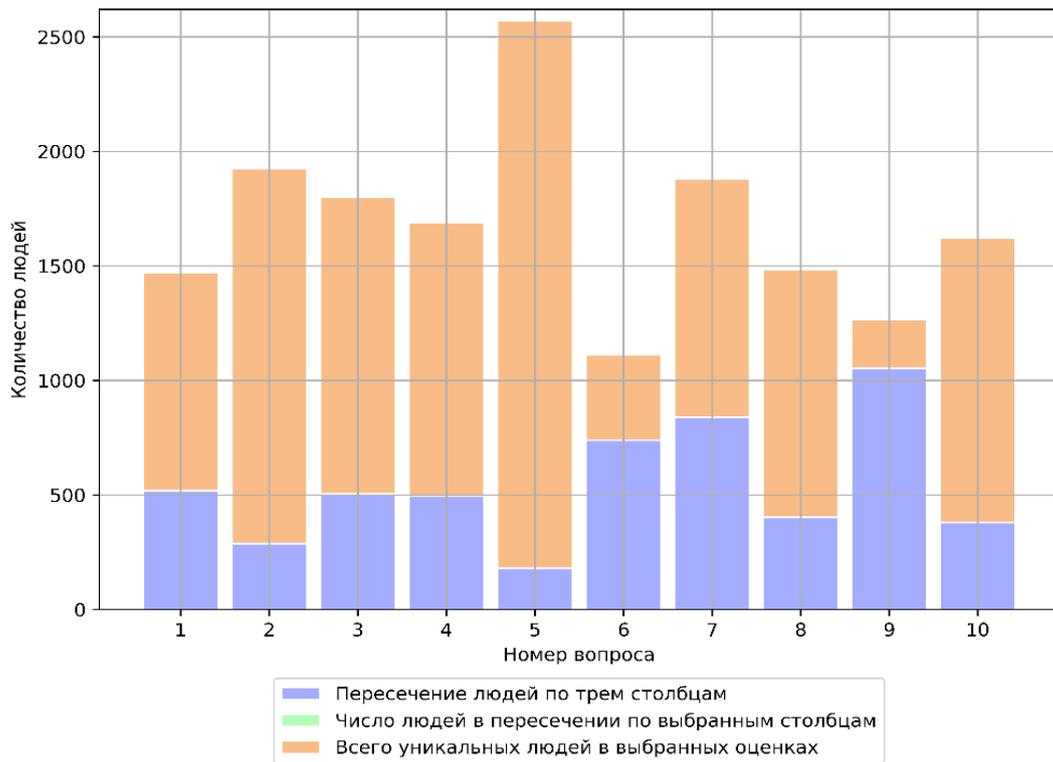


Рис. 4 – Пересечение респондентов по точности и продуктивности

Число пересечений респондентов в 10% отличников по совокупности ответов для 1 и 4 оценки

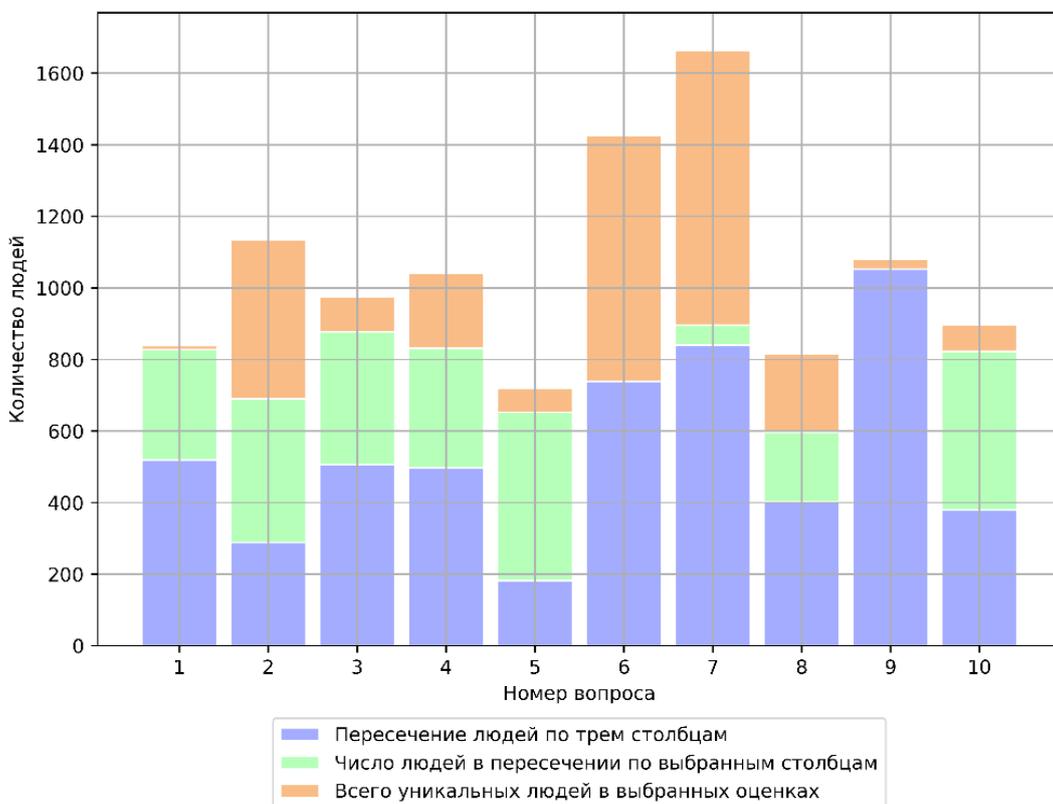


Рис. 5 – Пересечение респондентов по продуктивности и эффективности

Число пересечений респондентов в 10% отличников по совокупности ответов для 3 и 4 оценки

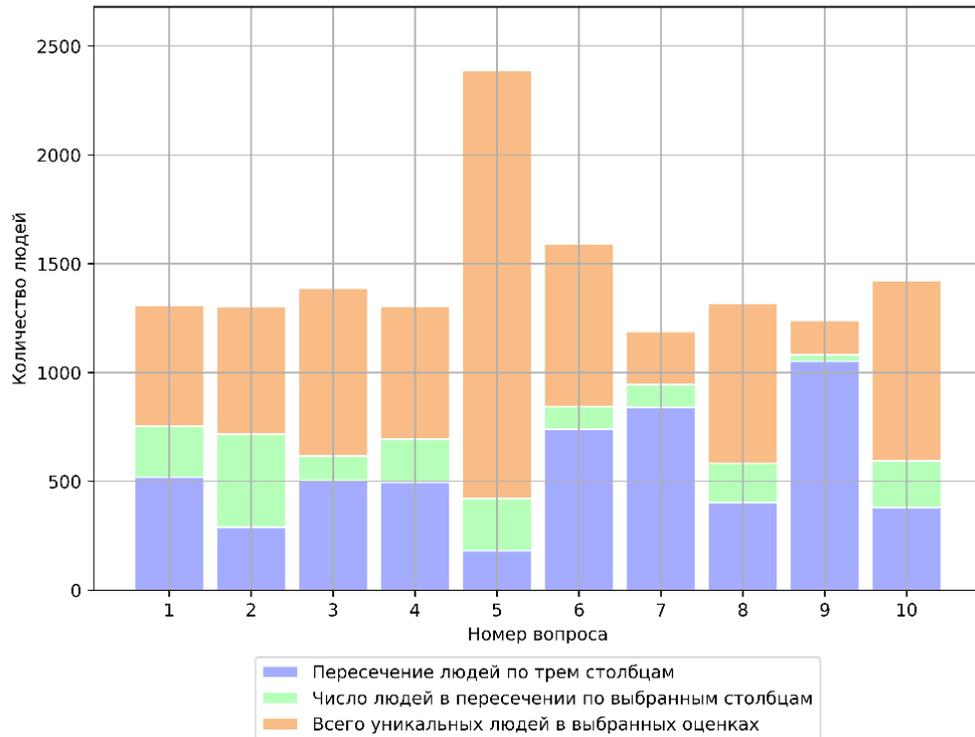


Рис. 6 – Пересечение респондентов по точности и эффективности

Дециль респондентов группы «двоечники» (Рис. 7 – 10).

Число пересечений респондентов в 10% двоечников по совокупности ответов для 1, 3 и 4 оценок

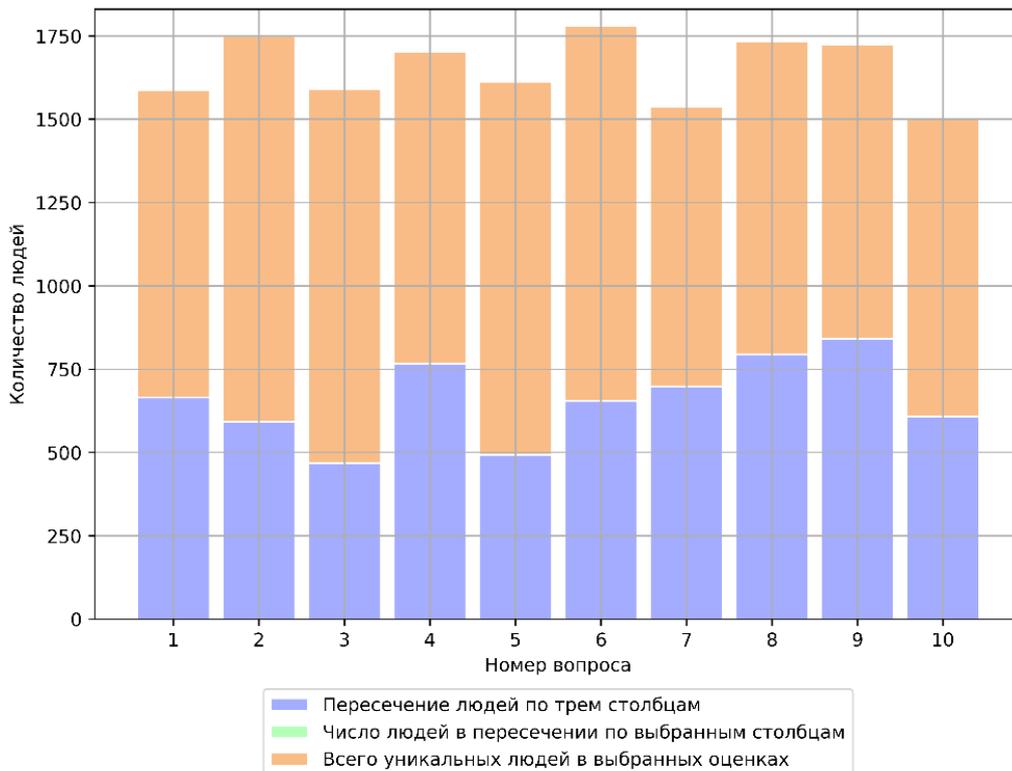


Рис. 7 – Пересечение респондентов по трем критериям

Число пересечений респондентов в 10% двоечников по совокупности ответов для 1 и 3 оценки

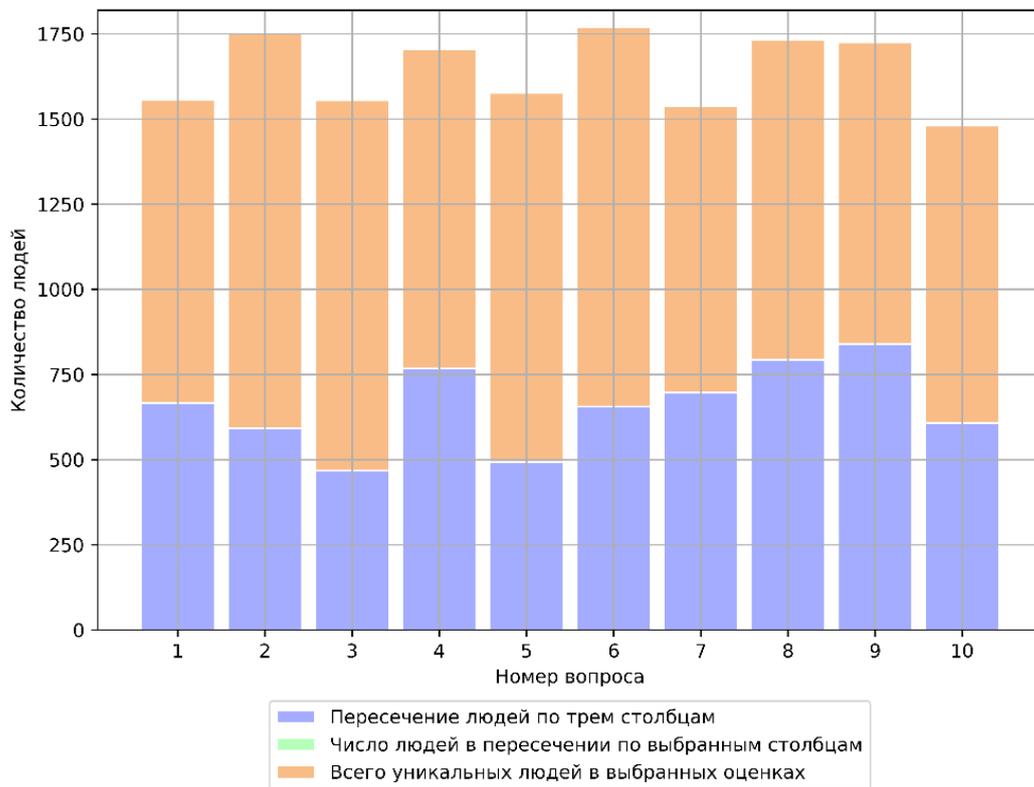


Рис. 8 – Пересечение респондентов по точности и продуктивности

Число пересечений респондентов в 10% двоечников по совокупности ответов для 1 и 4 оценки

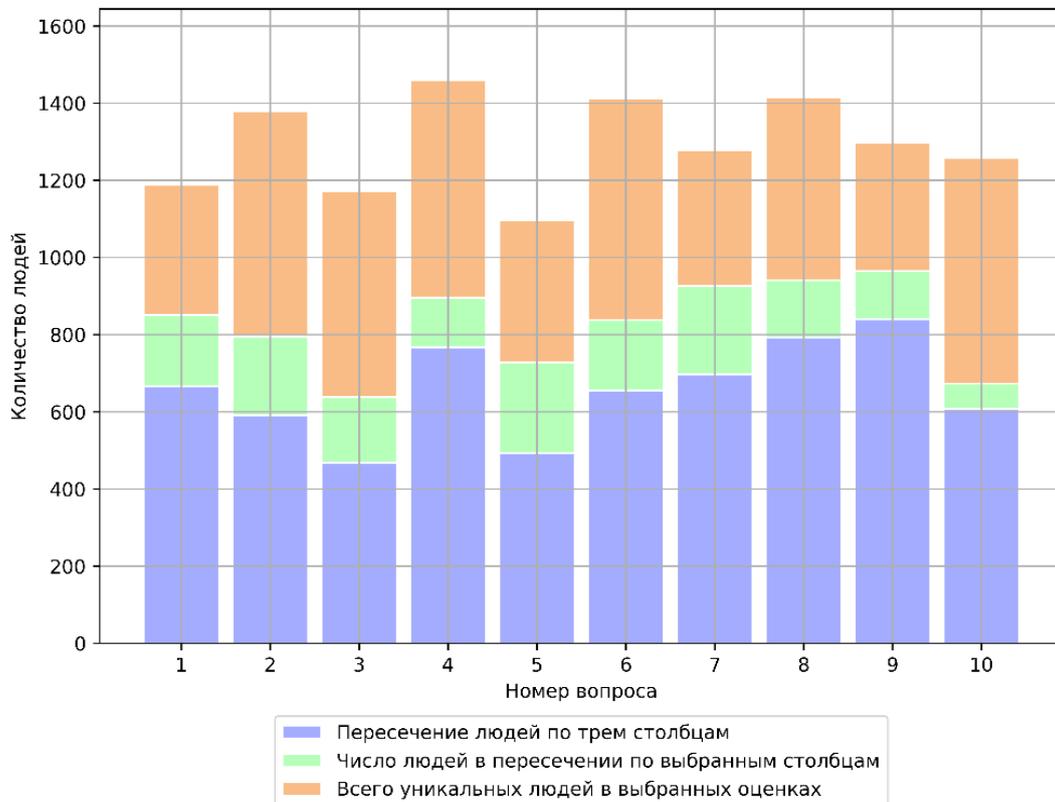


Рис. 9 – Пересечение респондентов по продуктивности и эффективности

Число пересечений респондентов в 10% двоечников по совокупности ответов для 3 и 4 оценки

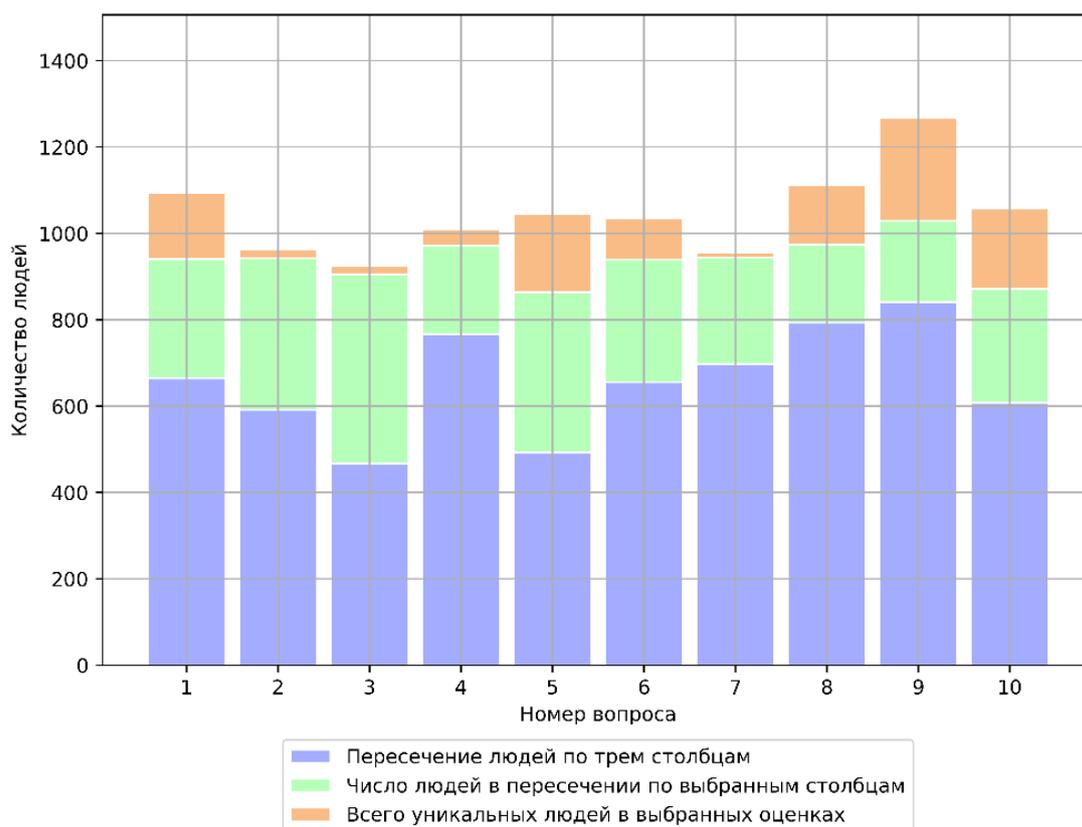


Рис. 10 – Пересечение респондентов по точности и эффективности

Теперь исследуем число полного пересечения респондентов между блоками всех вопросов вместе для отличников и двоечников. Запишем ID людей от пересечения трех критериев оценивания в каждом блоке. Выделим блок вопросов и назовем его главным. ID респондентов главного блока сравниваем с ID остальных блоков (зависимых). Найдем процент пересекающихся респондентов относительно главного блока вопросов.

Таким образом, для «отличников» и «двоечников» были получены следующие значения (Рис. 11 – 12), где числа в клетках приведены в процентах.

В ходе работы было рассчитано число респондентов от полного пересечения всех блоков вопросов. Для отличников оно равно 0, для двоечников 10.

Из полученных данных видно, что число людей среди ответов двоечников с одинаковыми ID больше, чем среди отличников, так что такая метрика соответствует эффекту Спирмена.

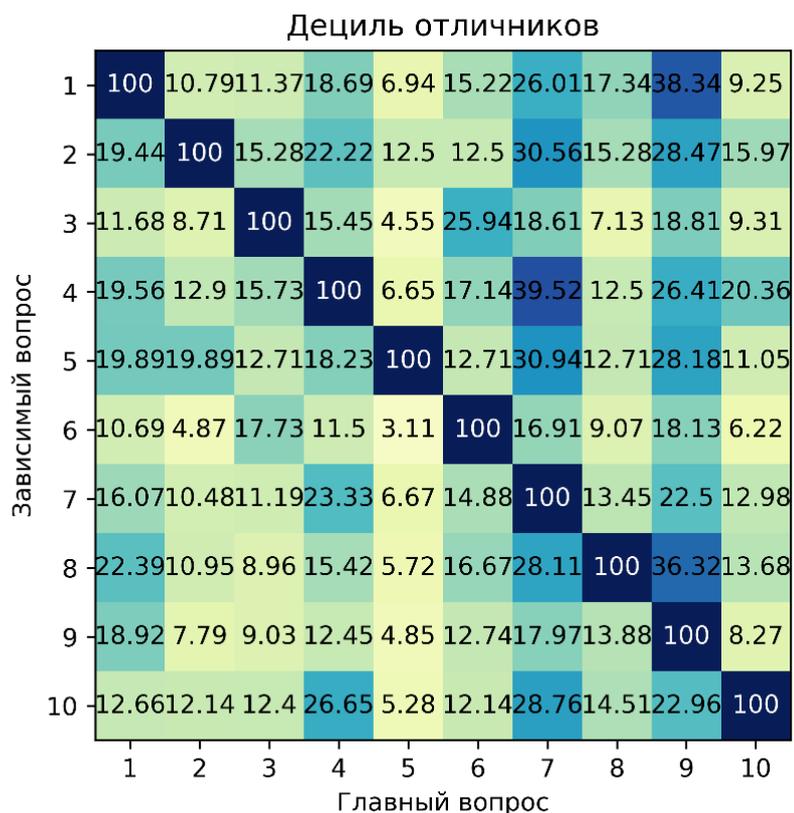


Рис. 11 – Процентные пересечения респондентов между блоками вопросов для «отличников»

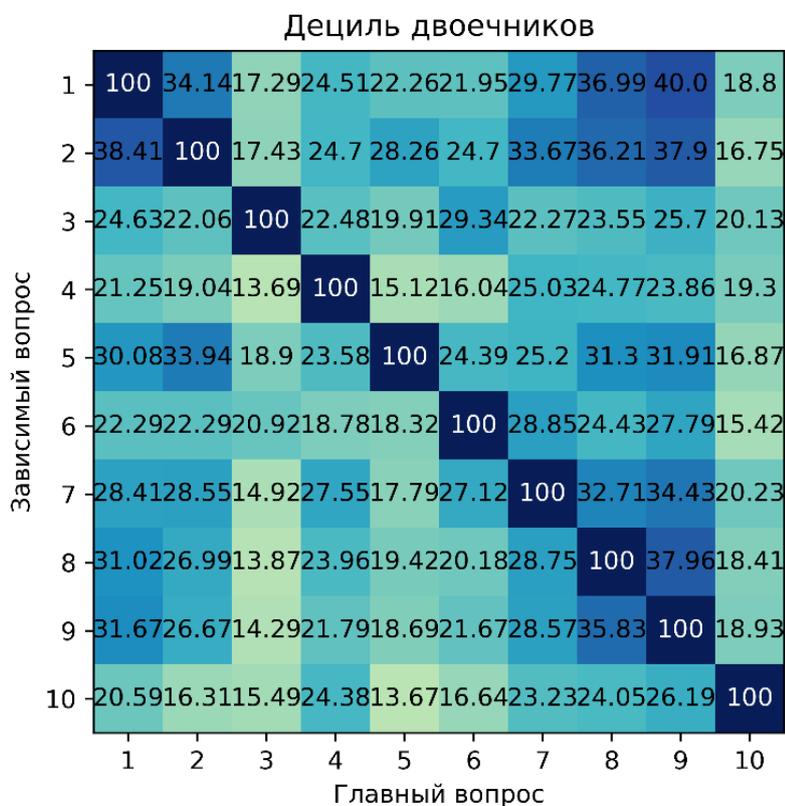


Рис. 12 – Процентные пересечения респондентов между блоками вопросов для «двоечников»

5. Заключение

Проведенный анализ результатов психологического теста показал зависимость выводов или интерпретаций от методов оценивания. Поскольку состав участников статистического эксперимента каждый раз разный, то сравнивать между собой статистики распределений ответов в разных экспериментах некорректно, так как, по сути, это выборки из разных генеральных совокупностей.

По-видимому, эффект SLODR не имеет универсального характера. Он в значительной степени обусловлен составом участников тестирования, зависит от методики оценивания и выбираемых метрик.

Литература

1. Кричевец А.Н., Корнеев А.А., Сугоняев К.В. Проблема однородности шкал интеллектуальных способностей: психометрическая оценка // Вестник Московского Ун-та. Серия 14: Психология, 2021. Т. 14, № 1, с. 144-189.
2. Кобзарь А.И. Прикладная математическая статистика. – М.: Физматлит, 2006. – 816 с.
3. Кендалл М., Стюарт А. Статистические выводы и связи. Том 2. – М.: Наука, 1973. – 899 с.
4. Корнеев А.А., Кричевец А.Н., Ушаков Д.В. Закон убывающей отдачи Спирмена: виды асимметрий распределений и их роль в порождении артефактов // Сибирский психологический журнал, 2019. № 71. С. 24-43.
5. Spearman C. The Abilities of Man. – NY, MacMillan, 1927. – 484 p.
6. Hartmann P., Reuter M. Spearman's Law of Diminishing Returns tested with two methods // Intelligence, 2006. V. 34 (1). P. 47-62.
7. Murray A.L., Dixon H., Johnson W. Spearman's law of diminishing returns: a statistical artifact? // Intelligence, 2013. V. 41 (5). P. 439-451.
8. Воронина М.Ю., Орлов Ю.Н. О статистическом анализе результатов психологического тестирования // Препринты ИПМ им. М.В. Келдыша. 2020. № 29. 28 с.
9. Voronina M.Yu., Orlov Yu.N. Digitization of student's personal characteristics: statistical analysis of psychological tests results and the Spearman effect // Proc. Of the Int. Conference DCCN'2021, Lecture Notes in Computer Sciences, V. 13144. P. 346-357.
10. Сугоняев К.В., Радченко Ю.И. Закон уменьшения отдачи Спирмена: исследование на масштабных российских выборках // Вестник ЮУрГУ, сер. Психология, 2018. Т. 11. № 1. С. 5-21.