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

Препринты ИПМ • Препринт № 29 за 2020 г.

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

Воронина М.Ю., Орлов Ю.Н.

О статистическом анализе результатов психологического тестирования

Рекомендуемая форма библиографической ссылки: Воронина М.Ю., Орлов Ю.Н. О статистическом анализе результатов психологического тестирования // Препринты ИПМ им. М.В.Келдыша. 2020. № 29. 28 с. http://doi.org/10.20948/prepr-2020-29
URL: http://library.keldysh.ru/preprint.asp?id=2020-29

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

М.Ю. Воронина, Ю.Н. Орлов

О статистическом анализе результатов психологического тестирования

Москва — 2020

Воронина М.Ю., Орлов Ю.Н.

О статистическом анализе результатов психологического тестирования

В работе представлен анализ результатов психологического теста, позволяющий провести кластеризацию испытуемых по группам, отвечающим определенным паттернам специфических способностей. Показано, что так называемый эффект уменьшающейся отдачи Спирмена не является универсальным и зависит от состава выборки и методики оценивания. Построены паттерны классов точности ответов на вопросы 10 различных типов, выделяющих специфические характеристики интеллекта.

Ключевые слова: психометрический анализ, паттерны точности ответов, корреляция, эффект Спирмена

Voronina M.Yu., Orlov Yu.N.

On the statistical analysis of psychology testing results

In this paper the results of analysis of a psychological test that allows clustering of subjects into groups that meet certain patterns of specific abilities are presented. It is shown that the so-called Spearman's diminishing return effect is not universal and depends on the composition of the sample and the evaluation method. Patterns of accuracy classes of answers to 10 different types of questions that highlight specific characteristics of intelligence are constructed.

Keywords: psychometrical analysis, accuracy patterns, correlation, Spearman effect

Работа выполнена при поддержке гранта РФФИ № 18-01-00619.

Содержание

1. Введение 3
2. Состав и основные свойства анализируемых данных 7
3. Оценка наличия общего фактора 11
4. Оценка эффекта SLODR 13
5. Построение паттернов точности ответов 18
6. Краткий анализ других качеств 26
7. Заключение 27
Литература 28

1. Введение

Статистическая обработка результатов тестирования группы каких-либо объектов на предмет обладания теми или иными свойствами – довольно стандартная процедура, проводимая во многих областях деятельности. Часто она применяется в чисто техническом смысле: определить, имеется ли связь между предприятием-изготовителем (которых, допустим, I) и качеством выпускаемых им приборов (типов которых, допустим, P). Для этого делается определенная градуировка качества прибора. Качество может быть объективно измеряемой многомерной характеристикой прибора, а может быть ранговой величиной. Пусть количество определяемых «качеств» равно Q, по каждому из них шкала разбита на S_q классовых интервалов, $q=1,2,...,Q$. После этого берется достаточно большое количество N исследуемых приборов каждого типа с каждого предприятия, т.е. всего $N \cdot I \cdot P$ объектов, и определяется категория качества каждого из них.

Дальнейшие действия по интерпретации собранной статистики носят эвристический характер. Например, ставится задача определить предприятие, изготавливающее самые качественные приборы, или, более детально, установить, есть ли связь между изготовителем и качеством прибора. Для этого составляются так называемые таблицы сопряженности (см., например, [1], раздел «корреляционный анализ»). Таблица сопряженности – это четырехуровневая таблица A_{ipqs}, которая показывает долю приборов 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 изготовителем и качеством прибора (для каждого типа прибора и для каждого вида качества). Для решения этой задачи существует довольно большое количество статистических критериев [2], 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 изготовителем и уровнем качества. Тем не менее, принято считать, если нет иных соображений, что из близости к нулю коэффициента корреляции (или иного другого из вышеперечисленных) следует отсутствие связи между двумя типами данных что, вообще говоря, неверно. Отсюда следует только, что мала линейная связь между ними. При этом может быть высока связь нелинейного характера. Сравнение же между собой собственно величин, допустим, выборочных коэффициентов корреляции также не дает оснований утверждать, что связи между значениями величин действительно больше или меньше, можно говорить лишь о возможности такой связи.

Кроме того, перечисленные индикаторы не имеют свойства метрик, то есть если при одном способе ранжирования будет установлен некоторый порядок по силе связи, например, фактор I_1 влияет на качество прибора типа P_1 сильнее, чем на P_2, то при другом способе этот порядок может и измениться.

В-третьих, само ранжирование допускает различные формулировки. Например, если построить распределения частот, т.е. эмпирических вероятностей того, что прибор p-го типа, изготовленный на i-ом предприятии, 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 предприятие, у которого максимальная доля приборов высшего уровня по качеству, и выстроить предприятия по убыванию доли высококачественных приборов. Но можно также назвать самым качественным то предприятие, которое выпускает минимальную долю приборов 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 псевдозакономерностях, которые можно извлечь из собираемых статистик, следует иметь в виду, что они относятся не столько к изучаемой связи, сколько к способу оценивания. Это всегда необходимо оговаривать, чтобы не формулировать результат анализа в виде «закона природы». Последнее возможно только в случае использования корректных метрик для анализа свойств объектов.

Трудность с проведением ранжирования существенно возрастает, если типы приборов неизвестны. Очевидно, результаты статистического анализа изменятся, если все типы приборов объединить в один и назвать его просто «прибор». 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 заводы-изготовители (фактически разные заводы могут трактоваться как разные типы приборов, которые мы уже объединили), то получаются данные в виде двухиндексной таблицы A_{qs} о результатах тестирования просто каких-то приборов. 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'? Также интерес представляет вопрос о том, можно ли по результатам тестирования приборов определить количество предприятий, на которых они были изготовлены? Последняя задача представляет собой стандартную формулировку факторного анализа.

Описанная ситуация встречается довольно часто не только в технических, но и гуманитарных областях. 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 (лень, нелюбовь к данному предмету, непонимание, слабая работоспособность и т.д.), то есть не вводить «типы приборов», то оцениваемые вероятности получаются недостоверными. Недостоверность проявляется не в статистической неточности оценивания выборочных параметров, с этим формально все в порядке при достаточно большой выборке, 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 идеале задача тестирования состоит в разработке таких заданий, решение которых однозначно можно было бы трактовать как обладание определенными способностями. Проблема, однако, в том, что способности известны не априори, а выявляются именно по результатам тестирования, то есть косвенно. В противном случае их можно было бы измерить непосредственно и тесты не проводить. Поэтому результаты тестирования как в технических, так и условно гуманитарных областях неоднозначны. Далее мы будем обсуждать результаты психологических тестов, что сильно отличается от тестирования прибора. Например, тест радиоприемника проводится прямым включением оного в сеть, и если на каких-то диапазонах частот он не работает, то фиксируется неполадка. Эту неполадку нельзя обнаружить, допустим, взвешиванием прибора или измерением других его габаритов, пересчитыванием деталей, из которых он состоит, и т.п. Тестирование же людей имеет целью ранжировать их по некоторой не измеряемой непосредственно величине, называемой, допустим, интеллектом, относительно свойств которого нет общепринятого мнения. И если выявлены корреляции в ответах на вопросы из формально разных областей, то свидетельствует ли это о том, что вопросы частично схожи, или о том, что схожи способности отвечающих людей, или о том и о другом вместе?

В последнее время на интерпретацию корреляций ответов применительно к психологическим тестам обращается пристальное внимание. Так, в работе [3] указывается на недостаточную методологическую обоснованность методов оценки так называемого закона убывающей отдачи Спирмена (Spearmen's Law of Diminishing Returns, SLODR, [4]), согласно которому корреляции между результатами тестирования интеллекта оказываются выше для менее интеллектуальных респондентов. В работах [5, 6] также обсуждаются вопросы, связанные с метриками для оценивания эффекта SLODR.

Интуитивно эффект SLODR обусловлен тем, что незнание достигается легче знания, поэтому, например, коэффициент корреляции «двоечников» по двум или более предметам ожидаемо более высокий, чем коэффициент корреляции «отличников». 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.

Впрочем, этот эффект в литературе был известен и до Спирмена и сформулирован не столь научно, но зато доходчиво. Эффект: «Все счастливые семьи похожи друг на друга, каждая несчастливая семья несчастлива по-своему» – Л.Н. Толстой. И объяснение эффекта: «Горе от ума» – А.С. Грибоедов. Поэтому правильнее было бы называть SLODR «эффектом Толстого – Грибоедова (ЭТГ)», но в математической статистике мы сохраним использование терминологии, принятой в этой области.

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 [1, 2] 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.

Например, в двухфакторной модели Спирмена предполагается, что существует некоторый общий фактор (general factor g), то есть общий интеллект или уровень развития неких общих способностей, позволяющий испытуемому вообще отвечать на какие-либо вопросы, а также и специфический фактор s, определяющий уровень развития специальной способности. В так называемых когнитивных теориях интеллекта предполагается, что уровень интеллекта определяется скоростью и эффективностью обработки информации. Также существуют этологические теории, согласно которым интеллект определяется способностью адаптации к внешним условиям. Кроме того, есть и другие теории.

Не обсуждая здесь собственно различные теории интеллекта, мы обратим внимание на некоторые аспекты психометрического оценивания респондента по результатам тестирования.

Существенной проблемой анализа результатов тестирования является отсутствие априорного знания о степени зависимости анализируемых качеств $q_1, q_2, ... q_Q$. 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. Приближение это состоит главным образом в том, что не учитываются какие-либо сторонние по отношению к тесту факторы, сопутствующие тому или иному профилю.

Учитывая незавершенность теории интеллекта и связанную с ней неоднозначность интерпретации, рассмотрим, какое содержание можно, тем не менее, извлечь из результатов тестирования. Мы будем анализировать данные достаточно репрезентативного психологического теста, описанного в [3, 7].

Цель работы состоит в установлении корреляционных зависимостей между результатами опросов респондентов, не сгруппированных заранее по тому или иному признаку. Также целью является формализация построения паттернов типовых распределений показателей качества, что важно для интерпретации результатов тестирования посредством выделения подгрупп и проверке для них наличия эффекта SLODR.

2. Состав и основные свойства анализируемых данных

Анализировались данные тестирования N=11335 респондентов. Каждому респонденту в рамках каждой из 10 различных методик (это «типы P_i» в терминах п. 1) были заданы по 30 вопросов. Оценка каждого ответа имела 4 составляющих: продуктивность, точность, скорость, эффективность (это «качества» q_i). По каждому из четырех качеств ответ на отдельный вопрос оценивался по системе «верно – 1», «неверно – 0». Итоговая оценка по каждой методике получалась как среднее арифметическое отдельных ответов и принимала значение от 0 до 1 с естественным шагом, равным 1/30. Далее мы будем рассматривать главным образом оценку качества «точность».

Тестировались следующие специальные качества интеллекта [3, 7]: 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.

Если строить выборочную функцию распределения (ВФР) точности ответов для каждой методики, то шаг при численном сравнении равен 1/30. На рис. 1 приведены ВФР ответов по каждой методике.

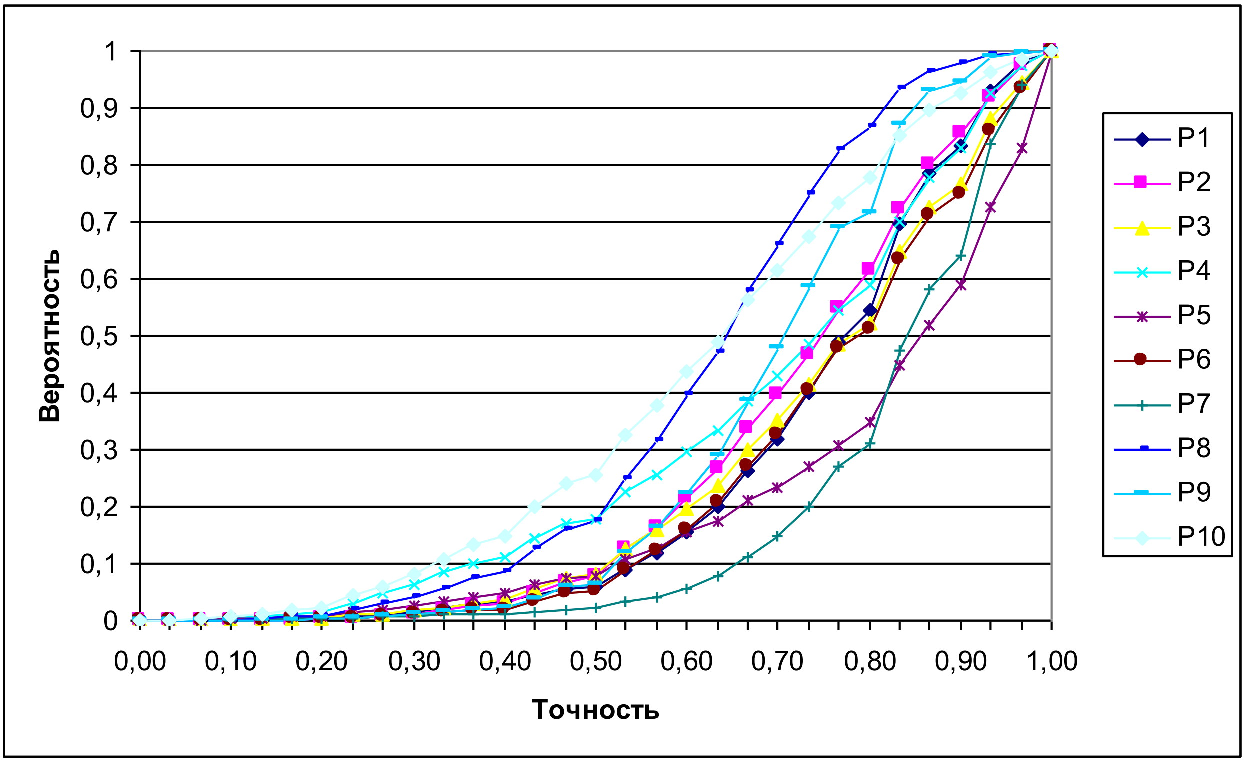

Рис. 1 – ВФР точности ответов по 10 методикам

Насколько различны распределения ответов по разным методикам? Согласно [8], ВФР, построенные по N данным (в нашем случае N=11335), следует считать различными, если расстояние между ними в норме C превосходит так называемый согласованный уровень значимости $\varepsilon^*(N)$, являющийся решением уравнения

$$1 - K\left(\sqrt{\frac{N}{2}}\varepsilon\right) = \varepsilon\,, \qquad (1)$$

где $K(z)$ есть табулированная функция Колмогорова. Для выборки в 11 тыс. данных величина $\varepsilon^*(N)$ равна приблизительно $\varepsilon^*(N) = 0{,}02$. Матрица расстояний между ВФР, которые изображены на рис. 1, приведена в таблице 1. Из этих данных следует, что все распределения различны. Наиболее близки между собой распределения ответов по методикам 3 и 6 (память на фигуры и вербальная память), но и они отличаются больше, чем это требуется для однородных выборок. Следовательно, выборки ответов неоднородны. Но даже если бы они и были однородны, это все равно не имело бы значения, поскольку при построении ВФР все ответы объединены, и потому не известно, 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.

Таблица 1 – Расстояния в норме *C* между ВФР точностей ответов

Метод	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	**0**	0,08	0,07	0,14	0,27	0,09	0,23	0,35	0,20	0,30
P2		**0**	0,09	0,10	0,28	0,11	0,30	0,28	0,15	0,23
P3			**0**	0,10	0,21	**0,04**	0,22	0,34	0,22	0,26
P4				**0**	0,26	0,14	0,28	0,28	0,17	0,19
P5					**0**	0,19	0,11	0,52	0,42	0,43
P6						**0**	0,21	0,35	0,24	0,29
P7							**0**	0,56	0,42	0,48
P8								**0**	0,19	0,09
P9									**0**	0,22

Чтобы соотнести ответы респондентов с классовыми интервалами, определяющими уровень качества ответов, можно поступить двояко. Во-первых, шкалу ответов можно равномерно разбить по числу респондентов, т.е. сделать поквантильное разбиение. Такое разбиение делается тогда, когда «пробники» (то есть люди) считаются одинаковыми, а с их помощью делается попытка что-то узнать о внешних условиях. В случае психологического тестирования интерес представляют собственно респонденты, поэтому более адекватным является другой способ классификации – через равномерное разбиение шкалы ответов по значению качества ответа.

Если изучать выборочную плотность функции распределения (ВПФР), то сравнение происходит в норме L1, для чего шкалу точности надо разбить на оптимальное число классовых интервалов согласно [9]. Если это число не зависит от методики P_i, то плотности распределений сравнимы между собой в единой шкале точности. Если же для каждой методики требуется свое разбиение, то для сравнения придется выбрать разбиение с наихудшей точностью, то есть с самыми крупными интервалами.

Для данного конкретного теста выяснилось, что оптимальное число классовых интервалов для всех методик совпадает и равно 16. Следовательно, качество ответов можно сравнить по выборочной плотности в одинаковой оптимальной шкале точности. Эта шкала примерно в 2 раза крупнее, чем для ВФР (16 интервалов против 30), что соответствует априорным значениям расстояний в нормах C и L1.

На рис. 2 приведены выборочные плотности распределений ответов по разным методикам, призванным выявить специфические особенности интеллекта.

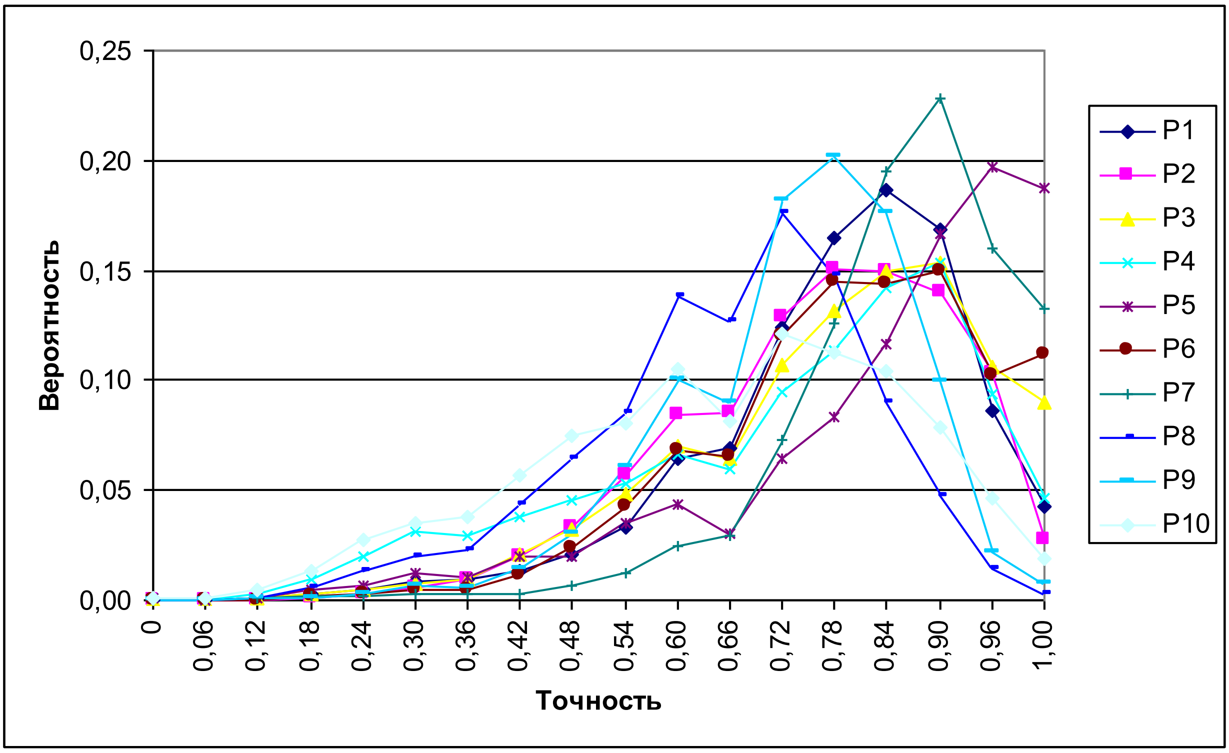

Рис. 2 – ВПФР точности ответов по 10 методикам

Удобство оценивания вероятности качества ответа в терминах классовых интервалов состоит в том, что построенные ВПФР можно трансформировать в формат, позволяющий группировать респондентов. Пусть шкала ответов равномерно разбита на n классовых интервалов. Для $P=10$ типов методик имеем массив $n^P=16^{10}$ интервалов, в которые будут попадать ответы x_{ij} i-го респондента на совокупность вопросов j-го типа. Всего таких ответов будет $N \cdot P$ штук. Каждому респонденту будет отвечать его индивидуальный профиль, представляемый в виде P-мерного вектора, компоненты которого суть натуральные числа, равные номерам классовых интервалов, в которые попали ответы x_{ij}, $j=1,2,...,P$ по каждой из методик. Поскольку надо сгруппировать эти профили по степени их близости в норме L1, то для репрезентативности такой группировки желательно, чтобы размерность массива интервалов была не больше количества ответов. Следовательно, для построения типовых паттернов число интервалов n определяется из условия $n^P = NP$ или

$$n=(NP)^{1/P}. \qquad (2)$$

Применительно к изучаемому массиву данных получается, что $n=3{,}2$. Следовательно, шкалу ответов надо разбить на 3 или, возможно, 4 класса. Если тогда окажется, что большинство индивидуальных ответов принадлежит не миллиону классов, а, скажем, десяти, то тем самым будет осуществлена группировка респондентов по близким паттернам ответов.

3. Оценка наличия общего фактора

Проанализируем возможность объяснения корреляций между ответами на вопросы методик P_i наличием некоторого общего фактора. Например, пусть изменение случайной величины, представляемой рядом значений y по методике P_2, обусловлено двумя случайными факторами – величинами z, отвечающими методике P_1 и некоторыми величинами x, так что следовало бы рассмотреть тройственную регрессию $\tilde{y}_n - \bar{y} = \alpha(z_n - \bar{z}) + \beta(x_n - \bar{x})$. Здесь $\tilde{y}_n$ обозначает регрессионную аппроксимацию значений y_n, а коэффициенты регрессии определяются формулами

$$\alpha = \frac{C_{zy}\sigma_x^2 - C_{zx}C_{xy}}{\Delta},\ \beta = \frac{C_{xy}\sigma_z^2 - C_{xz}C_{zy}}{\Delta},\ \Delta = \sigma_x^2\sigma_z^2 - C_{xz}^2 . \tag{3}$$

В формуле (3) C_{ij} обозначают коэффициенты ковариации, а σ_i^2 – дисперсии. Коэффициенты корреляции R_{ij} связаны с коэффициентами ковариации формулой $R_{ij} = \dfrac{C_{ij}}{\sigma_i \sigma_j}$.

Если величины z и x не являются независимыми, то в регрессионной модели $\tilde{z}_n - \bar{z} = \gamma(x_n - \bar{x})$ коэффициент γ будет значимо отличаться от нуля. Предположим теперь, что фактор x недоступен для наблюдения, и вместо (3) используется упрощенная модель $\tilde{y}_n - \bar{y} = u(z_n - \bar{z})$, где $u = \dfrac{C_{zy}}{\sigma_z^2}$. В этом случае ошибки упрощенной модели оказываются коррелированными с регрессорами:

$$\langle y - \tilde{y}, z\rangle = \langle y - \bar{y} - u(z - \bar{z}), z\rangle = \langle y - \bar{y} - u(z - \bar{z}), z - \tilde{z}\rangle =$$
$$= \langle y - \bar{y} - u(z - \bar{z}), z - \bar{z} - \gamma(x - \bar{x})\rangle = \sigma_y \sigma_z R_{xz}\left(R_{xz}R_{zy} - R_{xy}\right) \neq 0.$$

Отметим, однако, что если корреляция ошибок регрессии с самими регрессорами близка к нулю, это еще не свидетельствует об отсутствии скрытого фактора x, поскольку может реализоваться ситуация, когда близко к нулю выражение $R_{xz}R_{zy} - R_{xy}$.

Если остатки регрессии образуют гауссов белый шум (нормально распределенный процесс с независимыми приращениями), то можно теоретически определить значимость коэффициента корреляции R. Исходным положением является то, что при отсутствии корреляции между объясняемой и объясняющей переменными статистика $R\sqrt{N-2}/\sqrt{1-R^2}$ по выборке длины N имеет t-распределение Стьюдента с $N-2$ степенями свободы. Следовательно, коэффициент корреляции значим на уровне ε (т.е. гипотеза о равенстве нулю генерального коэффициента корреляции отвергается с вероятностью $1-\varepsilon$), если

$$\frac{|R|\sqrt{N-2}}{\sqrt{1-R^2}} > t_{1-\varepsilon}(N-2)\,. \qquad (4)$$

Как правило, к выводу о коррелированности ошибок и регрессоров приводят наблюдения за регрессорами, выполненные с определенной неточностью, т.е. к этому эффекту приводят ошибки измерения наблюдаемых величин, вызванные третьим фактором. Само по себе влияние третьего фактора неустранимо, но можно попытаться снизить корреляцию объясняющей переменной и ошибки модели введением новых переменных – это так называемый метод инструментальных переменных.

Пусть рассматривается модель $y(n) = \bar{y} + \alpha\left(z(n) - \bar{z}\right) + \varepsilon(n)$, связывающая две стационарных случайных величины. Если выяснилось, что между z и ε существует значимая корреляция, надо подобрать некоторые другие регрессоры, от которых требуется, чтобы они хорошо коррелировали бы с z, но давали бы близкую к нулю корреляцию с ε. Выберем некоторую переменную w (если, конечно, она найдется среди наблюдаемых величин), обладающую свойством нулевой корреляции с ε, и построим регрессию старых регрессоров на эту переменную: $\tilde{z}(n) = \bar{z} + \beta\left(w(n) - \bar{w}\right)$. Затем эту аппроксимационную величину $\tilde{z}$ принимаем за инструментальную переменную и строим регрессию y на $\tilde{z}$. По построению, остатки регрессии не коррелируют с регрессорами.

В таблице 2 приведены значения коэффициентов корреляции между точностями ответов респондентов по разным методикам.

Таблица 2 – Коэффициенты корреляции точностей ответов

Метод	Р2	Р3	Р4	Р5	Р6	Р7	Р8	Р9	Р10
Р1	0,49	0,30	0,42	0,38	0,31	0,44	0,51	0,51	0,36
Р2		0,35	0,43	0,51	0,33	0,45	0,46	0,40	0,39
Р3			0,34	0,31	0,50	0,34	0,28	0,29	0,32
Р4				0,36	0,29	0,47	0,42	0,39	0,49
Р5					0,34	0,39	0,38	0,35	0,34
Р6						0,36	0,31	0,32	0,31
Р7							0,45	0,44	0,43
Р8								0,52	0,42
Р9									0,37

Из формулы (4) следует, что все эти корреляции значимы на уровне значимости менее 10^{-4}, т.е. гипотеза о равенстве истинных корреляций нулю отвергается с вероятностью, практически равной единице. Однако для модели парной регрессионной аппроксимации результатов одной методики через другую обнаруживается, во-первых, весьма слабая детерминация возможной зависимости на уровне 0,1-0,2 и, во-вторых, практически отсутствующая корреляция на уровне 10^{-3} остатка аппроксимации с регрессором. Это показывает, что не любой результат тестирования интеллектуальных способностей можно интерпретировать с использованием скрытого g-фактора. Результат корреляционного анализа может сильно зависеть от состава участников тестирования. Далее мы рассмотрим другие оценки зависимости ответов, как без разделения на подгруппы, так и с разделением.

4. Оценка эффекта SLODR

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

$$L_{jk}(i) = \frac{a(i; j,k)d(i; j,k) - b(i; j,k)c(i; j,k)}{a(i; j,k)d(i; j,k) + b(i; j,k)c(i; j,k)}. \qquad (5)$$

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 3 и 6): $L_{ij}(1) > L_{ij}(4)$.

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: $L_{12}(3)=0{,}32 < L_{12}(4)=0{,}55$. Следовательно, эффект SLODR в данной метрике не универсален.

Применительно ко всему массиву данных коэффициенты парной ассоциации результатов тестов по применяемым методикам приведены в таблицах 3-6.

Таблица 3 – 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 4 – 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 5 – 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 6 – 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 5 и 6, 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 (рис. 3).

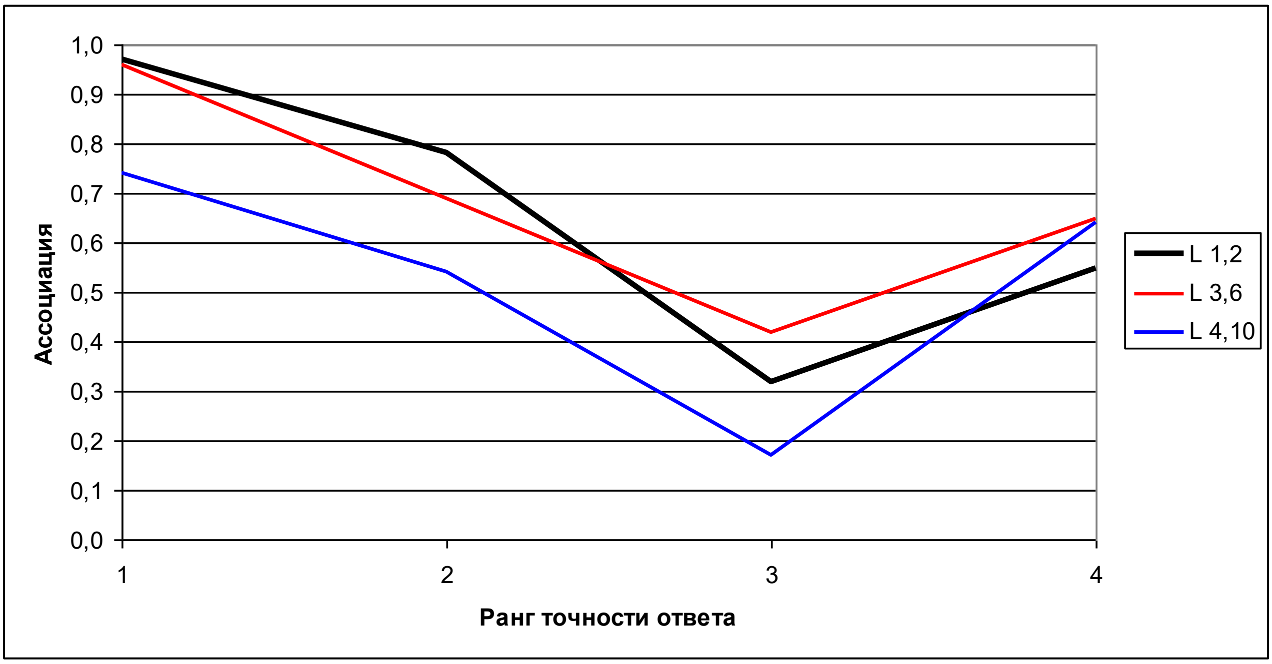

Рис. 3 – Примеры зависимости коэффициента ассоциации от качества ответа

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

Сравним теперь полученные результаты с оценками связи посредством корреляционной функции. Обозначим ВПФР из Рис. 2 через $f_i(j)$, где j нумерует классовый интервал, а i отвечает методике, и введем также парную ВПФР $F_{ik}(j,l)$. Определим корреляционную функцию формулой

$$R_{ik}(j,l) = F_{ik}(j,l) - f_i(j)\, f_k(l). \tag{6}$$

Парная ВПФР $F_{ik}(j,l)$ строится как эмпирическая доля пар ответов одного и того же респондента, одновременно относящихся к интервалу j для методики i и к интервалу l для методики k. Близость к нулю значений корреляционной функции означает слабую связь между данными классами, поскольку тогда парная ВПФР приблизительно равна произведению унарных ВПФР, что характерно для независимых случайных процессов. Если же в классы j и l часто попадают (или не попадают) одни и те же респонденты, то $F_{ik}(j,l)$ начинает значительно отличаться от произведения унарных ВПФР.

Например, применительно к методикам 1 и 2 в шкале из четырех уровней качества ответов получаем следующий результат (таблица 7). Сумма чисел по строкам и по столбцам этой таблицы равна нулю. На диагонали стоят значения корреляционных функций для групп, находящихся в одинаковых классовых интервалах. Из этой таблицы следует, что наибольшая по абсолютной величине корреляционная связь имеется между «отличниками», а на диагонали она монотонно убывает. Таким образом, получен совершенно иной результат по сравнению с анализом коэффициентов ассоциаций (5).

Таблица 7 – Значения корреляционной функции для пары методик (1, 2)

	$f_1(j)$	0,009	0,066	0,364	0,561
$f_2(j)$		**1**	**2**	**3**	**4**
0,006	**1**	0,002	0,003	-0,001	-0,004
0,092	**2**	0,002	0,019	0,001	-0,022
0,415	**3**	-0,001	0,012	0,037	-0,048
0,487	**4**	-0,003	-0,034	-0,037	0,074

Рассмотренный пример построения ранжирования силы связей на основе формул (5) и (6) показывает, что наличие или отсутствие эффекта является не объективным фактом, а трактовкой результатов и зависит как от способа оценивания, так и от разбиения всего множества респондентов на подгруппы. Это свидетельствует о чувствительности эффекта к методу оценивания и к структуре выборки.

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 обеим методикам P_i и P_j. Обозначим через $\nu_1^{(2)}(i,j)$ долю множества пересечения в объединении этих множеств:

$$\nu_1^{(2)}(i,j) = \frac{N_1^{(2)}(i,j)}{N_1^{(1)}(i) + N_1^{(1)}(j) - N_1^{(2)}(i,j)}. \tag{7}$$

Аналогично определим долю пересечения в объединении s множеств для первого и последнего классовых интервалов $\nu_{1,4}^{(s)}(i_1,...,i_s)$. Эффект SLODR в этом способе оценивания состоит в том, что ожидаемо должно быть $\nu_1^{(s)}(i_1,...,i_s) > \nu_4^{(s)}(i_1,...,i_s)$.

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».

В этом разделе мы рассмотрели три различных способа ранжирования связи между ответами в пределах определенных интервалов качества, каждый из которых отличался от традиционного анализа коэффициента корреляции, использование которого будет исследовано ниже на примерах разделения всего множества респондентов на подгруппы. Выяснилось, что результаты ранжирования зависят от способа оценки качества.

5. Построение паттернов точности ответов

Рассмотрим теперь все множество респондентов как объединение непересекающихся подгрупп в соответствии с результативностью их средних ответов на 10 тематических заданий. Выделение таких подгрупп основано на следующей процедуре кластеризации векторов.

Каждый респондент характеризуется 10-мерным вектором точности ответов $\mathbf{a}_n=\left(a_{1,n},...,a_{10,n}\right)$. При выбранной шкале оценки ответов величины a_{in} принимают значения из множества натуральных чисел $\{1,2,3,4\}$. Длиной вектора ответов будем называть сумму его компонент:

$$A_n=\sum_{i=1}^{10}a_{in}\,. \tag{8}$$

Таким образом, возможные длины векторов принимают целочисленные значения от 10 до 40. Распределение респондентов по длинам этих векторов показано на Рис. 4. Правый квантиль порядка 0,9 равен 28, так что почти 90 % ответов относятся к категориям «хорошо» и «отлично», т.е. имеют сумму от 30 и выше.

Результаты ответов разных респондентов также сравниваются в норме L1:

$$d_{kn}=\sum_{i=1}^{10}\left|a_{in}-a_{ik}\right|. \tag{9}$$

В идеале респонденты относятся к одной группе, если векторы точностей их ответов покомпонентно совпадают. Однако такой подход неэффективен, поскольку тогда число классов по порядку величины часто оказывается равным числу испытуемых. Поэтому для выделения класса мы используем, во-первых, имеющуюся шкалу качества оценок и, во-вторых, введем уровень близости компонент векторов точности разных респондентов при ответе на вопросы одной и той же тематики. Положим, что компоненты векторов из одного класса не могут различаться более чем на 1, причем суммарное различие векторов (норма (9)) не превосходит заданного уровня. Кластеризация будет считаться проведенной корректно, если распределение расстояний между любыми двумя половинами выделенной группы обладает свойством стационарности (1).

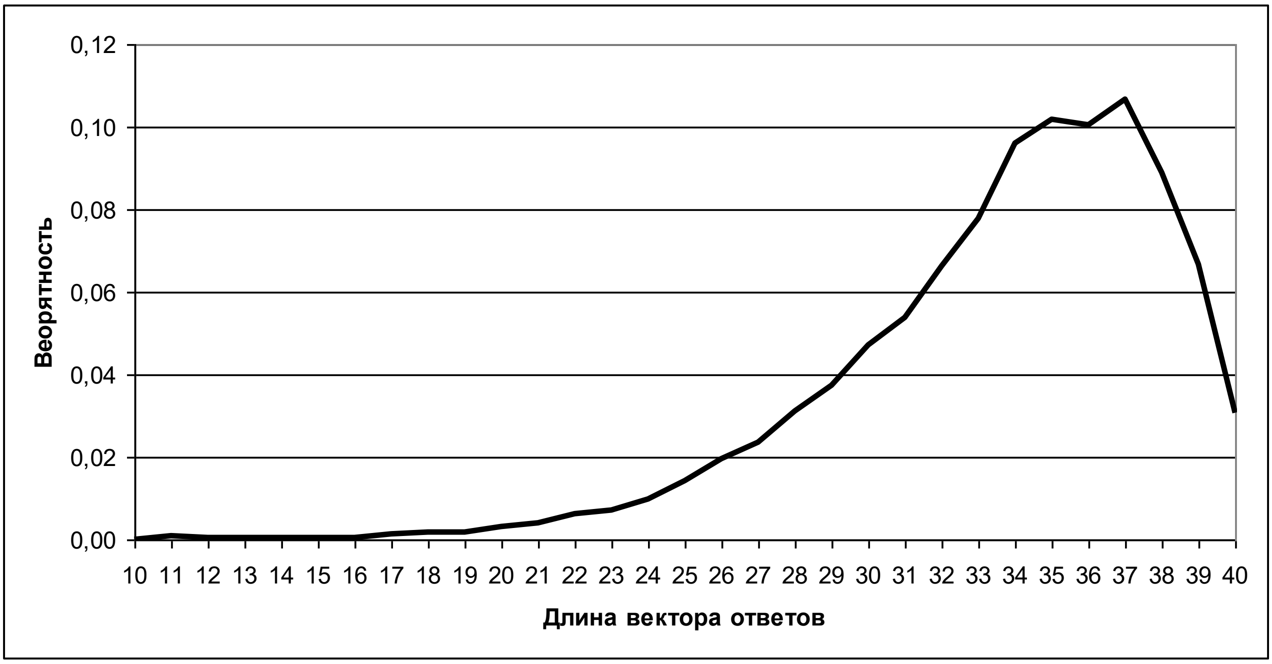

Рис. 4 – Распределение респондентов по суммарной точности ответов

Среднее значение вектора по группе будем называть паттерном группы.

Выяснилось, что примерно половина респондентов образуют три однородных паттерна – «успешных», «середняков» и «плохишей». Первые имеют оценки по всем методикам от 3 до 4, вторые – от 2 до 3 и третьи – от 1 до 2. При этом в первую группу вошло 5319 человек, во вторую 215 и в третью всего 40. Каждая из этих групп может быть в свою очередь распределена по двум подгруппам – верхней и нижней – относительно середины длины вектора оценки A_n. Паттерн «отличников» образован векторами «успешных» с длиной $A_n \geq 36$. Таких респондентов оказалось 3955 или 35 %. Паттерн «хорошистов» составлен из 1364 векторов с условиями $36 > A_n \geq 30, a_{in} \geq 3$. «Сильные середняки» образованы 129 векторами с условиями $30 > A_n \geq 26, 3 \geq a_{in} \geq 2$, а «слабые середняки» образованы 86 векторами с $26 > A_n > 20, 3 \geq a_{in} \geq 2$. Имеется также 15 «неуспевающих», паттерн которых отвечает условиям $A_n \geq 16,\ a_{in} \leq 2$, и 25 «аутсайдеров» с паттерном $A_n < 16,\ a_{in} \leq 2$. Эти паттерны приведены на Рис. 5.

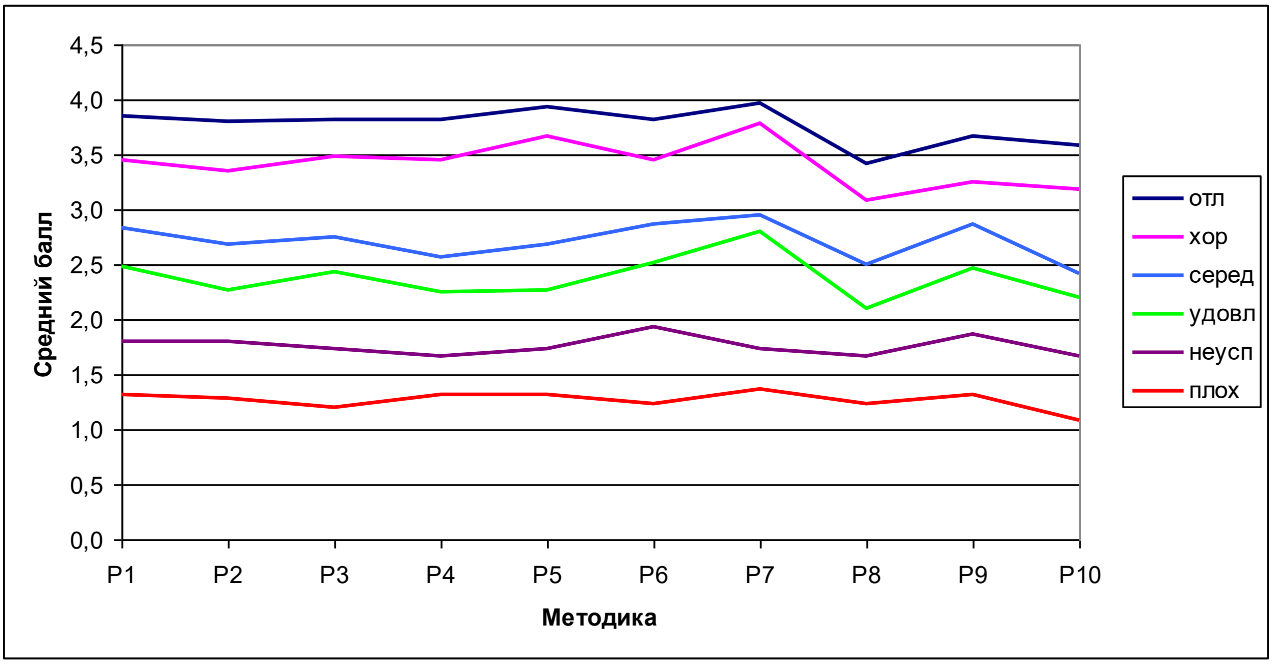

Рис. 5 – Однородные паттерны групп респондентов

Из Рис. 5 видно, что представленные паттерны имеют определенное сходство. Например, во всех группах задача Р2 решается хуже, чем Р1, задача Р8 – хуже, чем Р7, задача Р9 – лучше, чем Р8, а задача Р10 – хуже, чем Р9. Отметим также, что в данном примере «отличники» решают задачи Р6 хуже, чем задачи Р5, а «двоечники» и «середняки» напротив, лучше.

Сравним между собой корреляционные матрицы для респондентов, отвечающих группе «отличников» и совокупной группе «неуспевающих». Коэффициенты корреляции между средними ответами по разным методикам приведены для некоторых групп в таблицах 8-10. Из сравнения этих таблиц с результатами корреляционного анализа для всего множества респондентов (таблица 2) следует, что в группах корреляции существенно меньше, а для некоторых пар методик и вовсе пренебрежимо малы.

Таблица 8 – Коэффициенты корреляции для «неуспевающих»

Метод	Р2	Р3	Р4	Р5	Р6	Р7	Р8	Р9	Р10
Р1	0,35	0,20	0,00	0,05	0,30	0,20	0,31	0,15	0,22
Р2		0,14	0,15	0,20	0,45	0,35	0,25	0,20	0,25
Р3			0,08	0,14	0,20	0,20	0,17	0,47	0,36
Р4				0,25	0,10	0,20	0,29	0,16	0,07
Р5					0,45	0,05	0,25	0,40	0,25
Р6						0,20	0,20	0,35	0,65
Р7							0,20	0,15	0,22
Р8								0,27	0,13
Р9									0,40

Таблица 9 – Коэффициенты корреляции для «середняков»

Метод	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	0,28	-0,12	0,13	0,09	0,15	0,26	0,08	0,11	0,00
P2		0,00	0,13	0,30	0,03	0,19	0,26	0,05	-0,03
P3			-0,03	-0,04	0,18	-0,03	-0,04	0,04	0,06
P4				0,24	-0,08	-0,01	-0,01	-0,03	-0,03
P5					0,12	0,13	0,21	-0,03	-0,06
P6						0,19	0,14	0,20	-0,06
P7							0,14	0,15	0,14
P8								0,14	0,13
P9									0,07

Таблица 10 – Коэффициенты корреляции для «отличников»

Метод	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	0,16	0,06	0,13	0,09	0,07	0,12	0,16	0,20	0,10
P2		0,13	0,12	0,19	0,08	0,10	0,15	0,13	0,14
P3			0,08	0,07	0,26	0,05	0,04	0,05	0,09
P4				0,12	0,05	0,10	0,11	0,08	0,18
P5					0,08	0,11	0,08	0,10	0,08
P6						0,06	0,08	0,09	0,08
P7							0,09	0,12	0,12
P8								0,18	0,16
P9									0,12

В целом корреляции «двоечников» выше, чем корреляции «отличников», что соответствует эффекту SLODR, однако 7 пар (примерно 15 %) этот эффект нарушают. Что касается сравнения «середняков» и «отличников», то здесь нарушений эффекта SLODR гораздо больше: имеем 27 пар нарушителей из 45 (60 %). Следовательно, группировка респондентов по точности ответов на основе близости их к паттернам равномерной успешности (Рис. 5) не дает оснований однозначно идентифицировать эффект уменьшающейся отдачи.

Рассмотрим теперь другую половину респондентов, которые имеют специфичные предпочтения по типам вопросов. Различных способов выбрать k из 10 тем существует C_{10}^{k}, а общее их количество равно $2^{10}-1$. Естественно, рассматривать более 1000 паттернов не эффективно. Поэтому мы отберем те, которые содержат достаточно большое количество респондентов. Для этого мы объединим некоторые темы вопросов. Так, вопросы тем 1, 7 и 8 относятся к общему логическому мышлению и могут быть объединены в класс I «логика».

Вопросы тем 2 и 5 относятся к «математике» (класс II), вопросы тем 3, 4 и 10 – к мышлению пространственными образами (класс III), а вопросы тем 6 и 9 – к вербальным способностям (класс IV). В результате анализа вместе с отобранными выше однородными паттернами образовалось 20 основных паттернов, к которым относится примерно 90 % респондентов. Типы этих паттернов и их доли в общем множестве приведены в таблице 11.

Таблица 11 – Список основных паттернов групп респондентов

№ паттерна	Доля паттерна	Класс темы	I	II	III	IV
1	0,004	Класс точности	1-2	1-2	1-2	1-2
2	0,019		2-3	2-3	2-3	2-3
3	0,469		3-4	3-4	3-4	3-4
4	0,002		3-4	1-2	3-4	3-4
5	0,005		3-4	3-4	1-2	3-4
6	0,020		2-3	3-4	3-4	3-4
7	0,050		3-4	2-3	3-4	3-4
8	0,082		3-4	3-4	2-3	3-4
9	0,076		3-4	3-4	3-4	2-3
10	0,018		3-4	2-3	2-3	2-3
11	0,017		2-3	3-4	2-3	2-3
12	0,007		2-3	2-3	3-4	2-3
13	0,015		2-3	2-3	2-3	3-4
14	0,038		3-4	3-4	2-3	2-3
15	0,016		3-4	2-3	3-4	2-3
16	0,028		3-4	2-3	2-3	3-4
17	0,010		2-3	3-4	3-4	2-3
18	0,019		2-3	3-4	2-3	3-4
19	0,009		2-3	2-3	3-4	3-4
20	0,004		3-4	3-4	1-2	2-3

К первому типу относятся рассмотренные выше однородные паттерны. Второй тип образуют паттерны для респондентов, не знающих какую-то одну тему из четырех. Отметим, что знающих только какую-то одну тему из четырех нет. Третий тип включает респондентов, которые только одну тему знают хуже, чем остальные три. К четвертому типу относятся паттерны тех, кто только одну тему знает лучше, чем остальные три. Пятый тип образуют паттерны респондентов, знающих лучше половину тематик. К шестому типу относится один паттерн сильно неоднородных способностей.

Из неоднородных паттернов следует выделить те, которые встречаются чаще других. Первую тройку образуют паттерны тех, кто знает темы трех классов из четырех на хорошо и отлично, а тему оставшегося класса знает удовлетворительно. Эти оставшиеся классы в порядке убывания: III, IV, II. Следом за ними идут два паттерна тех, кто темы двух классов знает на хорошо и отлично, а темы других двух классов – удовлетворительно. Пары тем, которые знают отлично, следующие: I, II и I, IV.

Структура множества респондентов определяется долями основных паттернов. Результаты тестирования разных групп сравнимы между собой в статистическом смысле, если они имеют близкую структуру по основным паттернам. Возможно, что расхождения в оценках условных вероятностей связи между качеством ответов на вопросы установленных тематик в разных экспериментах обусловлены различным составом групп испытуемых.

Паттерны групп, выделенных в таблице 11, приведены на Рис. 6 – 9.

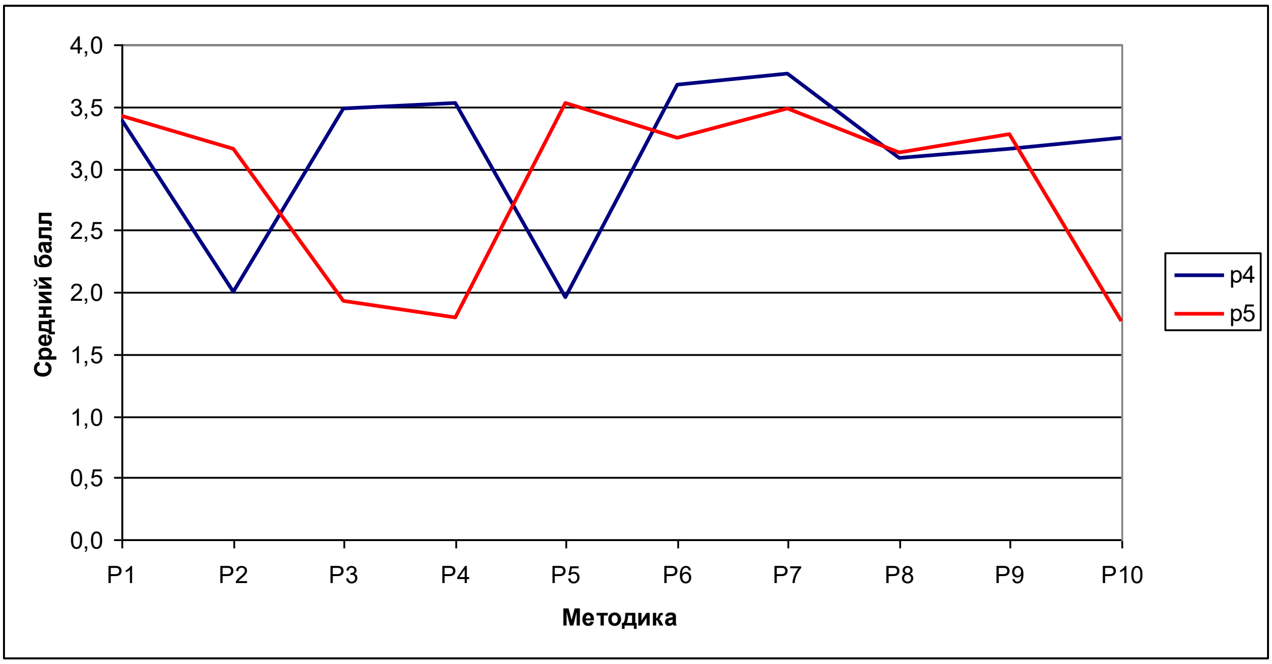

Рис. 6 – Паттерны групп, не знающих темы одного класса

На долю паттернов, приведенных на рис. 6, приходится 0,7 % общего числа респондентов. Это связано с тем, что «двоечников» вообще мало в исходном множестве данных. Наиболее высокий балл в этих паттернах приходится на методику P_7. Далее идут методики P_6, P_5, P_4. Из тех вопросов, по которым испытуемые показали высокие результаты (выше 3 по 4-балльной шкале), хуже всего решаются задачи типа P_8 («силлогизмы»).

Паттерны «более умных» и «менее умных» на рис. 7-8 вместе составляют примерно 30 % респондентов. Паттерны типовых «середняков» на Рис. 9 составляют около 10 % испытуемых.

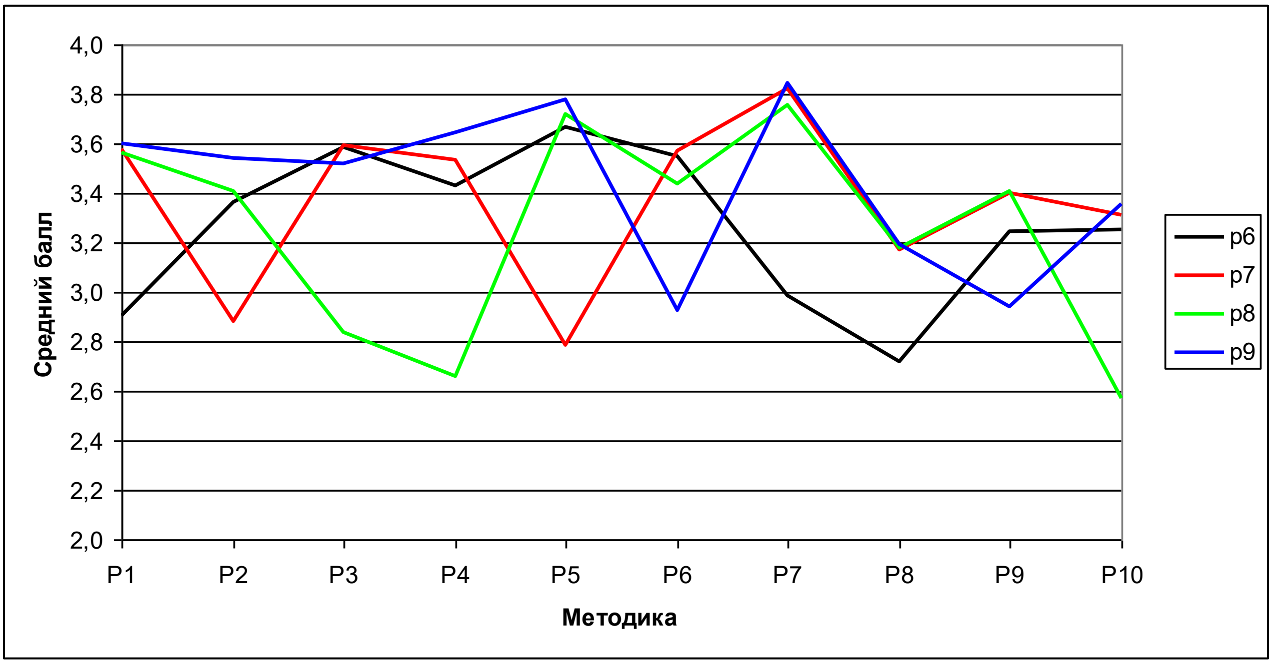

Рис. 7 – Паттерны групп, знающих одну тему хуже остальных

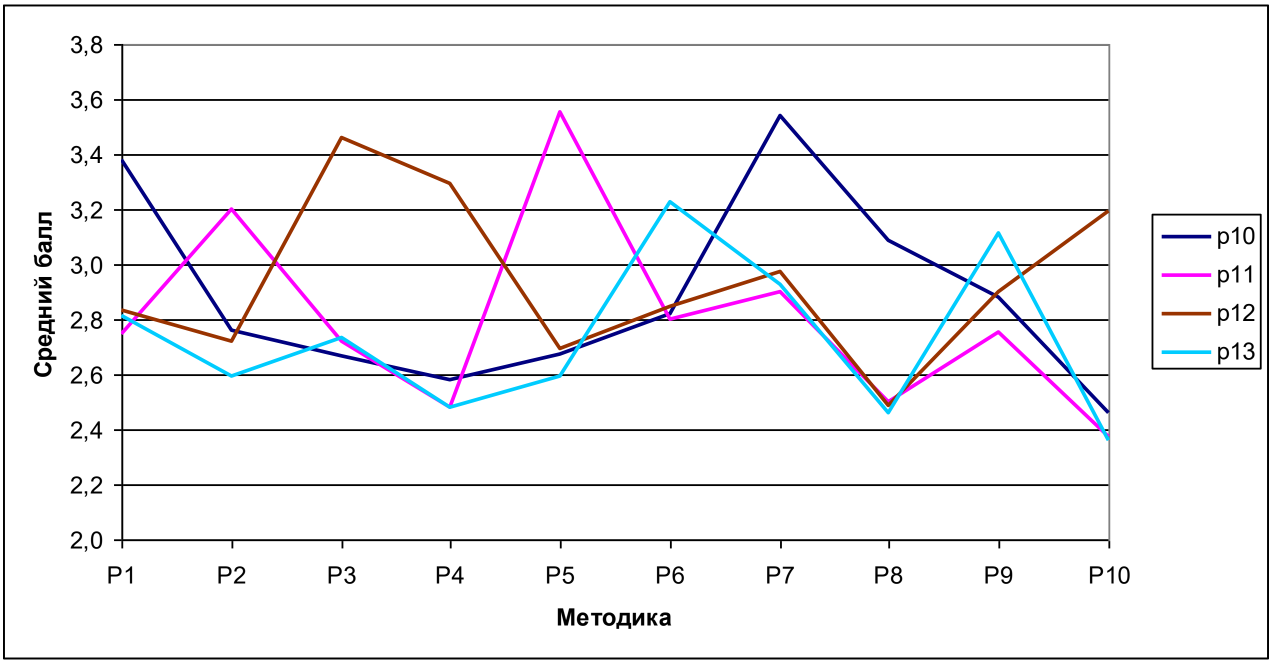

Рис. 8 – Паттерны групп, знающих одну тему лучше остальных

Сравнивая средние баллы на рис. 7 и 8, видим, что если тематический класс, в котором респонденты хорошо разбираются, один, как на рис. 8, то лучшие ответы имеют меньшие баллы, чем в случае рис. 7, когда респонденты хорошо ориентируются в большинстве тематических классов. Косвенно это может свидетельствовать в пользу интегрального фактора интеллекта, который тем выше, чем в больших областях человек разбирается на высоком уровне. Это подтверждается также и на однородных паттернах (рис. 5), где отличники имеют гораздо более высокий уровень оценок, чем знатоки отдельных тем.

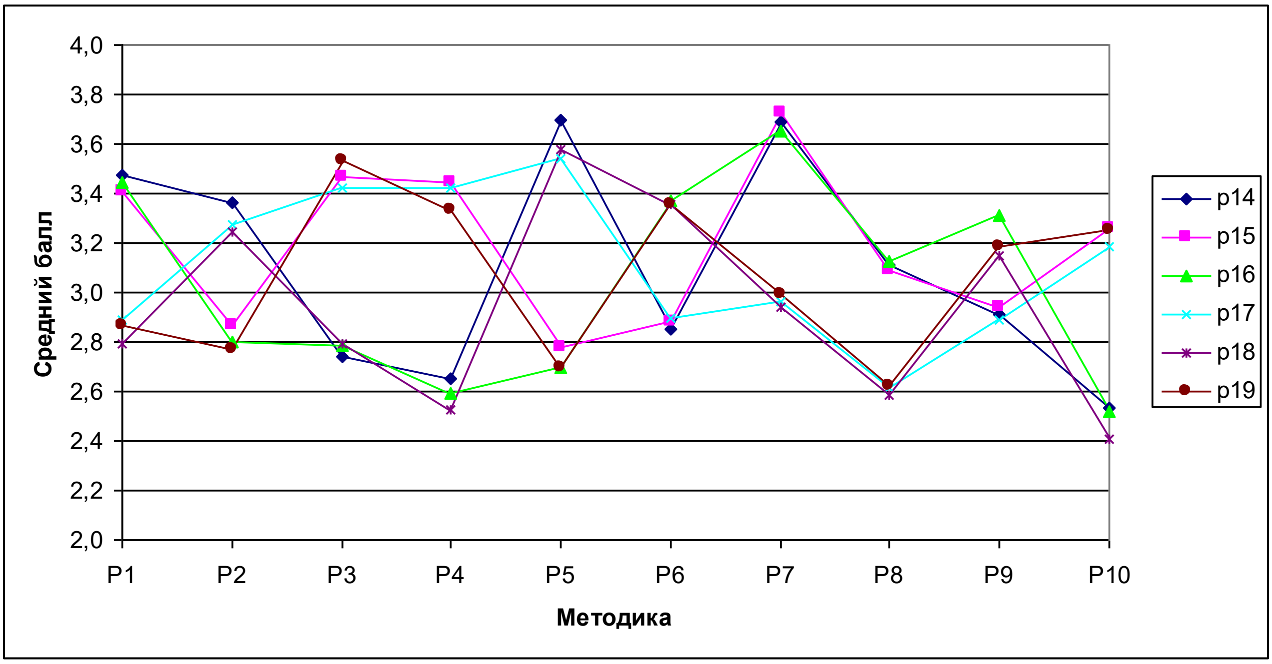

Рис. 9 – Паттерны групп, знающих лучше половину тем

Сравним теперь коэффициенты корреляции между методиками в двух разных группах – менее успешной (паттерны рис. 8) и более успешной (паттерны рис. 7). Эти данные приведены в таблицах 12-13. Коэффициенты корреляции в выделенных группах также оказались довольно малыми. Примерно 20 из 45 пар имеют корреляции, близкие к нулю, так что эти величины сравнивать между собой не очень осмысленно. Для остальных пар получились следующие результаты: в 16 случаях корреляции в менее успешной группе больше, чем в более успешной, в соответствии с эффектом SLODR. Однако в 9 случаях наблюдается противоположная картина.

Таблица 12 – Коэффициенты корреляции для «менее успешной» группы

Метод	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	0,20	0,10	0,03	0,15	-0,12	0,01	0,15	0,10	0,06
P2		0,01	0,08	0,05	-0,02	0,08	0,05	0,08	0,01
P3			0,16	0,04	0,14	0,21	0,00	0,10	0,15
P4				0,14	0,00	0,14	0,12	0,06	0,12
P5					-0,02	0,00	0,14	0,04	0,11
P6						0,01	0,10	-0,01	0,01
P7							0,01	0,09	0,09
P8								0,06	-0,08
P9									0,00

Таблица 13 – Коэффициенты корреляции для «более успешной» группы

Метод	P2	P3	P4	P5	P6	P7	P8	P9	P10
P1	0,14	0,02	0,13	0,13	0,03	0,12	0,14	0,08	0,06
P2		0,06	0,08	0,18	0,06	0,08	0,10	0,02	0,09
P3			0,11	0,01	0,06	0,10	-0,02	0,06	0,05
P4				0,20	0,06	0,08	0,11	0,01	0,15
P5					0,03	0,10	0,08	-0,05	0,06
P6						-0,02	0,01	0,03	0,03
P7							0,08	0,12	0,15
P8								0,01	0,11
P9									0,03

Таким образом, группировка респондентов по типовым паттернам ответов привела к существенному снижению корреляций по сравнению с общей массой испытуемых. Это связано с тем, что на уровне стационарной близости выборочных распределений отклонения ответов отдельных испытуемых от паттерна в целом случайны и не носят выраженного характера.

6. Краткий анализ других качеств

Рассмотрим усредненные показатели других качеств для блоков методик, описанных выше. Сравним между собой оценки продуктивности, скорости, точности и эффективности ответов полной совокупности респондентов без распределения их по группам. Соответствующие плотности распределения вероятностей в равномерной шкале оценивания, состоящей из 10 классовых интервалов, приведены на Рис. 10.

Из Рис. 10 следует, что для всех блоков вопросов прослеживается определенная связь между распределениями различных качеств. Так, распределение эффективности (некоторого кумулятивного качества) имеет моду, расположенную левее максимумов остальных распределений. Она находится примерно в области 0,4. Это – наиболее вероятное значение данного качества. Правее нее расположена мода продуктивности, соответствующее значение равно 0,6. Мода точности, как уже было указано раньше, находится в области значений 0,8. Скорость же имеет наибольшую вероятность в последнем классовом интервале. Связано ли такое эквидистантное расположение мод с методиками оценивания или оно обусловлено особенностями данной конкретной выборки, сказать сложно. Но заметим, что без учета параметра «скорость» остальные три распределения довольно близки по форме и могут быть объединены посредством сдвига по оси оценивания.

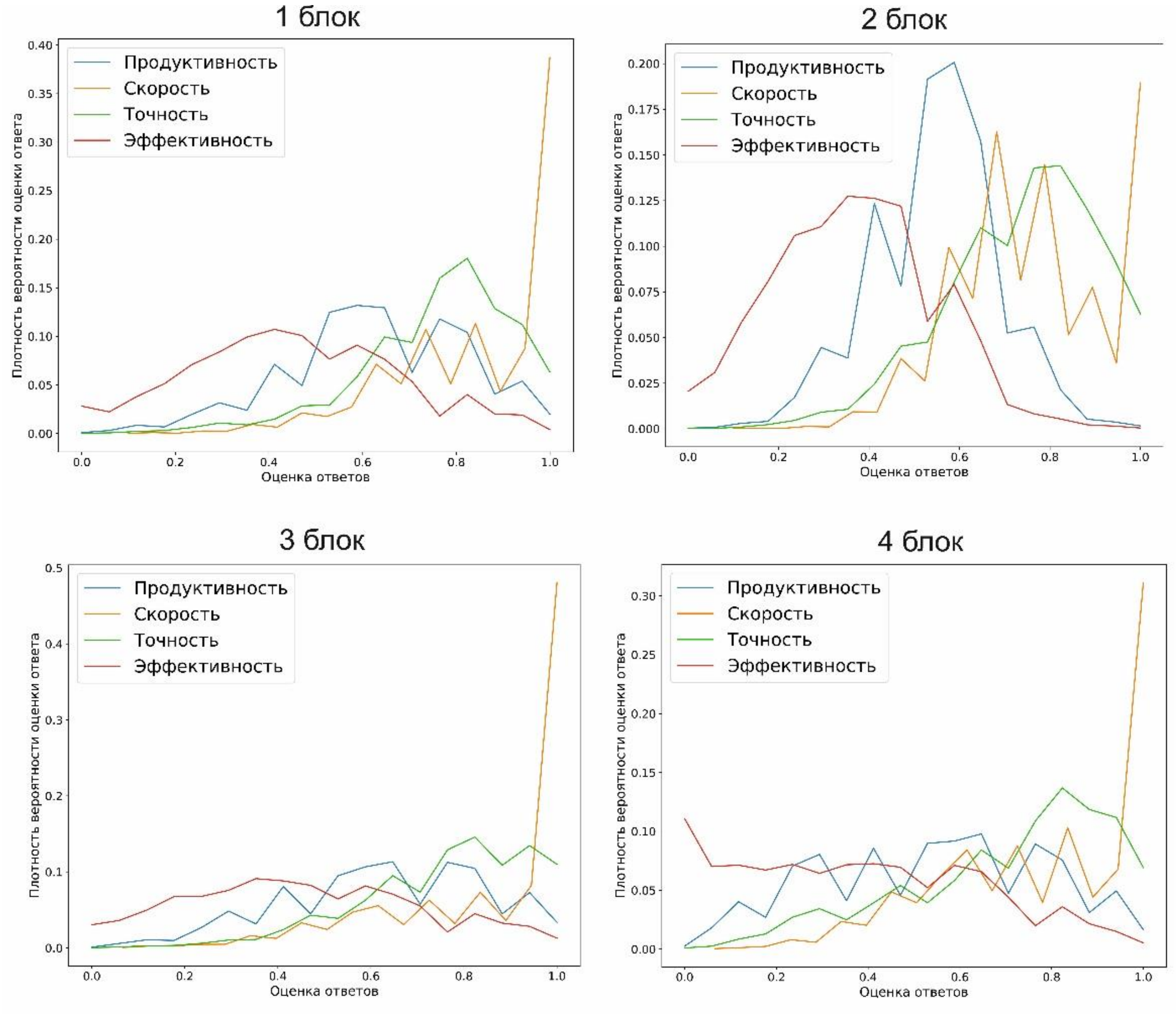

Рис. 10 – Плотности распределения оценок по различным качествам

7. Заключение

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

По-видимому, интегральная оценка результата теста в виде эффективности или иного качества, интерпретируемого как фактор интеллекта, полезна для тех случаев, когда ответы респондента не укладываются в стандартные паттерны, которые были нами отобраны как наиболее представительные. Для тех же, кто соответствует паттерну, индикатором является сам паттерн. Естественно, что метрика в виде длины паттерна может быть применена, но она позволяет ранжировать между собой однородные паттерны. Если же длина одинакова, но разными являются ответы по отдельным методикам, то вводить одно число для оценки разных качеств испытуемых, видимо, не следует.

В заключение скажем несколько слов об эффекте SLODR. По-видимому, он в значительной степени обусловлен составом участников тестирования и не имеет универсального характера: во-первых, его зависимость от уровня качества ответов немонотонна и, во-вторых, имеется заметное количество не отвечающих ему примеров.

Литература

1. Кобзарь А.И. Прикладная математическая статистика. – М.: Физматлит, 2006. – 816 с.
2. Кендалл М., Стюарт А. Статистические выводы и связи. Том 2. – М.: Наука, 1973. – 899 с.
3. 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.
4. Spearmen C. The Abilities of Man. – NY, MacMillan, 1927. – 484 p.
5. Hartmann P., Reuter M. Spearmen’s Law of Diminishing Returns tested with two methods // Intelligence, 2006. V. 34 (1). P. 47-62.
6. Murray A.L., Dixon H., Johnson W. Spearmen’s law of diminishing returns: a statistical artifact? // Intelligence, 2013. V. 41 (5). P. 439-451.
7. 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.
8. Орлов Ю.Н. Оптимальное разбиение гистограммы для оценивания выборочной плотности распределения нестационарного временного ряда // Препринты ИПМ им. М.В. Келдыша. 2013. № 14. 26 с.