Теория

ПОДБОР ПАРАМЕТРОВ МОДЕЛИ ПЕДАГОГИЧЕСКИХ ИЗМЕРЕНИЙ

Олег Деменчёнок
Восточно-Сибирский институт МВД России
AskSystem@yandex.ru

Рассмотрены теоретические основы оценки параметров модели измерения, предложены аналитические зависимости для модели Г. Раша и двухпараметрической модели. Описана технология поиска параметров модели Г. Раша средствами Microsoft Excel. Проведён анализ результатов подбора параметров модели и анализ качества моделей. Показано преимущество оценок параметров модели методом наименьших квадратов для небольших объёмов данных.

Ключевые слова: *тест, тестовый балл, модель Г. Раша, двухпараметрическая модель, метод максимального правдоподобия, метод наименьших квадратов.*

Специфической чертой современного этапа развития теории и практики педагогических измерений следует считать широкое внедрение методов математического моделирования. Математические модели способствуют выполнению двух основных функций теории — объяснению и прогнозированию наблюдаемых результатов тестирования.

Для практического применения моделей в педагогических измерениях необходимо прежде всего определить параметры модели. Затем неизбежно возникает вопрос об адекватности полученной модели — насколько хорошо модель справляется со своими функциями? Подбору параметров и оценке качества модели посвящена данная статья.

Теоретические основы оценки параметров модели измерения

Задача построения математической модели, связывающей уровни подготовленности испытуемых и трудности тестовых заданий с конкретными результатами выполнения теста, может быть сформулирована следующим образом. Из теоретических соображений выбран принципиальный вид зависимости вероятности правильного ответа. Эта зависимость — та или иная модель Item Response Theory (IRT) математической теории измерений (МТИ).

Модель содержит ряд параметров: θ — подготовленность испытуемых, β — трудность заданий, a — значение параметра крутизны графика функции задания или различающая способность задания для двух- и трёхпараметрических моделей, c — мера возможного угадывания правильного ответа на задание, для трёхпараметрической модели. Требуется так выбрать эти параметры, чтобы результаты конкретного тестирования и математическая модель наилучшим образом совпадали.

Решение этой задачи во многом зависит от того, что именно мы условимся считать «наилучшим», т.е. от принятого критерия оптимальности. Можно, например, считать критерием оптимальности минимум максимального отклонения между расчётными и экспериментальными данными; можно потребовать, чтобы в минимум обращалась сумма абсолютных величин отклонений и т.д. При каждом из этих требований мы получим своё решение задачи, свои значения параметров θ и β.

Определённое влияние на значения параметров оказывают также дополнительные условия (ограничения). Например, можно ограничить область поиска, задав максимальное и/или минимальное возможное значение параметра; потребовав, чтобы среднее значение параметра было равно нулю и т.д. Кроме того, на результате подбора параметров может сказаться способ реализации алгоритма поиска. Таким образом, для каждого сочетания критерия оптимальности, принятых ограничений и способа реализации алгоритма будет получен свой «наилучший» набор параметров.

Теоретическую базу всех методов и приёмов, положенных в основу построения эмпирических математических моделей, составляет метод максимального (наибольшего) правдоподобия. В общем виде метод максимума правдоподобия можно сформулировать так: наилучшее описание явления — то, которое даёт наибольшую вероятность получить в результате измерений именно те значения, которые и были фактически получены[1].

В идеальном случае математическая модель полностью соответствует экспериментальным данным. Однако в действительности результаты измерения всегда содержат погрешности. В связи с этим имеет место множество гипотез, не противоречащих опытным данным. Метод максимального правдоподобия состоит в выборе такой гипотезы, при которой вероятность получить в процессе измерения фактически наблюдаемые величины была бы максимальной. В формальной записи этот метод может быть представлен в виде максимума произведения вероятностей всех наблюдаемых независимых событий[2]:

$$p_1 \cdot p_2 \cdot \ldots \cdot p_N = \prod_{i=1}^{N} p_i \to \max. \quad (1)$$

Оценки параметров модели по методу максимального правдоподобия являются асимптотическими (т.е. требуют очень большого объёма данных).

Ещё один известный метод — метод наименьших квадратов — это метод теории ошибок для оценки неизвестных величин по результатам измерений, содержащим случайные ошибки. По этому методу требование наилучшего согласования расчётных x и экспериментальных x_0 данных сводится к тому, чтобы сумма квадратов отклонений между ними обращалась в минимум:

$$\sum_{i=1}^{N} (x_i - x_{0i})^2 \to \min. \quad (2)$$

Метод наименьших квадратов приводит к сравнительно простому математическому способу оценки параметров; кроме того, он допускает довольно веское теоретическое обоснование с вероятностной точки зрения. В классическом учебнике по теории вероятностей[3] математически доказано, что принцип максимума правдоподобия совпадает с методом наименьших квадратов, если различие между моделью и экспериментальными данными вызвано случайными ошибками. Таким образом, метод наименьших квадратов — это частный случай метода максимума правдоподобия.

1
Львовский Б.Н. Статистические методы построения эмпирических формул. М.: Высшая школа, 1988.

2
Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998.

3
Вентцель Е.С. Теория вероятностей. М.: Высшая школа, 2001. 576 с.

Критерии подбора параметров модели Раша по методу максимума правдоподобия

Вернёмся к задаче подбора параметров модели IRT (МТИ). Предположим, из теоретических соображений выбрана модель Г. Раша. По этой модели вероятность результата x при решении i-м тестируемым j-го задания определяется выражением:

$$P_{ij}(x) = \frac{e^{x_{ij}(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}}, \qquad (3)$$

где x_{ij} — результат выполнения задания, $x = 1$ при правильном ответе, $x = 0$ при неправильном ответе; θ_i — уровень подготовленности i-го тестируемого; β_j — уровень трудности j-го задания.

По методу максимального правдоподобия наилучшим будет признан тот набор значений θ и β, при котором произведение расчётных вероятностей фактически полученных результатов максимально:

$$\prod_{i=1}^{n}\prod_{j=1}^{m} P_{ij}(x) = \prod_{i=1}^{n}\prod_{j=1}^{m} \frac{e^{x_{ij}(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} \to \max. \qquad (4)$$

где n — количество испытуемых; m — число тестовых заданий.

Особенность модели Г. Раша заключается в том, что она оперирует только разностью параметров θ и β. Поэтому можно получить бесконечное количество «наилучших» решений, отличающихся на произвольную постоянную величину C:

$$\theta - \beta = (\theta + C) - (\beta + C) = \theta - \beta.$$

Для определённости целесообразно ввести ограничение, например, в виде равенства нулю среднего значения уровня подготовленности всех испытуемых:

$$\bar{\theta} = 0. \qquad (5)$$

В этом случае наилучшее решение будет единственным. Для реализации в вычислительных алгоритмах желательно преобразовать ограничение (5) в неравенство:

$$\left|\bar{\theta}\right| \le \varepsilon, \qquad (6)$$

где ε — некоторая допустимо малая величина, например, равная 0,001.

Выражений (4) и (6) достаточно для непосредственного численного решения задачи подбора параметров модели.

По методу наибольшего правдоподобия можно получить ещё один вариант решения, основанный на том, что в точке экстремума (максимума или минимума) производная функции равна нулю.

Зная, что $e^a e^b = e^{a+b}$, преобразуем числитель условия (4):

$$\frac{e^{\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}(\theta_i-\beta_j)}}{\prod_{i=1}^{n}\prod_{j=1}^{m}\left(1+e^{\theta_i-\beta_j}\right)} \to \max. \quad (7)$$

Введём обозначения: R_i — сумма результатов выполнения тестовых заданий i-м тестируемым (R_i также называют исходным баллом):

$$R_i = \sum_{j=1}^{m} x_{ij}, \quad (8)$$

s_j — сумма результатов выполнения j-го задания всеми тестируемыми;

$$s_j = \sum_{i=1}^{n} x_{ij}. \quad (9)$$

Используя (8) и (9), заменим

$$\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}\theta_i = \sum_{i=1}^{n} R_i\theta_i, \quad (10)$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}\beta_j = \sum_{j=1}^{m} s_j\beta_j. \quad (11)$$

Таким образом, уравнение (7) принимает вид

$$\frac{e^{\sum_{i=1}^{n} R_i\theta_i - \sum_{j=1}^{m} s_j\beta_j}}{\prod_{i=1}^{n}\prod_{j=1}^{m}\left(1+e^{\theta_i-\beta_j}\right)} \to \max. \quad (12)$$

Так как логарифм — монотонно возрастающая функция, то максимум функции правдоподобия совпадает с максимумом её логарифма:

$$\ln\left(\frac{e^{\sum_{i=1}^{n} R_i\theta_i - \sum_{j=1}^{m} s_j\beta_j}}{\prod_{i=1}^{n}\prod_{j=1}^{m}\left(1+e^{\theta_i-\beta_j}\right)}\right) \to \max,$$

$$\sum_{i=1}^{n} R_i\theta_i - \sum_{j=1}^{m} s_j\beta_j - \sum_{i=1}^{n}\sum_{j=1}^{m}\ln\left(1+e^{\theta_i-\beta_j}\right) \to \max. \quad (13)$$

Для нахождения максимума функции правдоподобия найдем и приравняем к нулю частные производные логарифмической функции правдоподобия по каждому из аргументов:

$$\frac{\partial\left(\sum_{i=1}^{n} R_i\theta_i - \sum_{j=1}^{m} s_j\beta_j - \sum_{i=1}^{n}\sum_{j=1}^{m}\ln\left(1+e^{\theta_i-\beta_j}\right)\right)}{\partial\theta_i} =$$

$$= R_i - \sum_{j=1}^{m}\frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}} = 0 \text{ , или}$$

$$R_i - \sum_{j=1}^{m}\frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}} = 0, \quad (14)$$

$$\frac{\partial\left(\sum_{i=1}^{n} R_i\theta_i - \sum_{j=1}^{m} s_j\beta_j - \sum_{i=1}^{n}\sum_{j=1}^{m}\ln\left(1+e^{\theta_i-\beta_j}\right)\right)}{\partial\beta_j} =$$

$$= -s_j + \sum_{i=1}^{n}\frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}} = 0, \text{ или}$$

$$s_j - \sum_{i=1}^{n}\frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}} = 0. \quad (15)$$

Нетрудно заметить, что выражение (14) представляет собой разность между суммой результатов выполнения заданий i-м испытуемым и суммой вероятностей правильных ответов на все задания для того же испытуемого. Аналогично, уравнение (15) представляет собой разность между суммой результатов выполнения задания всеми тестируемыми и суммой вероятностей правильных ответов на то же задание для всех испытуемых.

В итоге мы получаем второй вариант решения по методу наибольшего правдоподобия. Он состоит из n уравнений (14) (по одному уравнению на каждого тестируемого для $i = 1, 2 \ldots n$) и m уравнений (15) (по одному на каждое задание), которые нужно численно решить с учётом дополнительного ограничения, например (6), для нахождения параметров модели Г. Раша. Наличие аналитических зависимостей для частных производных упрощает и существенно ускоряет поиск решения. Поэтому многие алгоритмы подбора параметров модели Г. Раша основаны на уравнениях (14) и (15).

Критерий подбора параметров двухпараметрической модели

В последнее время всё чаще находит практическое применение двухпараметрическая модель измерения, в которой каждое задание характеризуется двумя параметрами — трудностью β и дифференцирующей способностью a. Так, например, в классическом пособии по IRT Ф. Бейкера[4] именно двухпараметрическая модель рассматривается в качестве основной модели измерения.

Вероятность правильного ответа в двухпараметрической модели описывается уравнением:

$$P = \frac{1}{1+e^{-a(\theta-\beta)}}, \qquad (16)$$

где a — дифференцирующая способность задания.

Знак параметра a определяет направление, а численное значение пропорционально тангенсу угла наклона касательной к характеристической кривой тестового задания в точке $\theta = \beta$ (см. рис. 1). Авторы работ по IRT[5,6,7] рекомендуют ограничить область допустимых значений дифференцирующей способности положительными числами $a \in (0, \infty)$. Действительно, задания с дифференцирующей способностью $a \le 0$ непригодны для педагогических измерений:

- при $a < 0$ задание более успешно выполняют менее подготовленные испытуемые. Это противоречит педагогике и теории педагогических измерений, может объясняться некорректной формулировкой или фактическими ошибками в преподаваемом учебном материале;
- при $a = 0$ задание одинаково успешно выполняют все испытуемые. Характеристическая кривая такого задания представляет собой горизонтальную прямую. Дифференцирующая или «различающая» способность задания нулевая.

4 *Baker F.B.* The Basics of Item Response Theory. 2 ed. Hieneman, Portsmouth, New Hempshire, 2001.

5 *Sotaridona L.S., Pornel J.B., Vallejo A.* Some Applications of Item Response Theory to Testing / The Philippine Statistician, 2003 Vol. 52, Nos. 1–4, pp. 81–92.

6 *Fox J.P.* Randomized Item Response Theory Models / Journal of Educational and Behavioral Statistics, Summer 2005, Vol. 30, No. 2, pp. 1–24.

7 *Verhelst N.D.* Item Response Theory / Reference Supplement to the Preliminary Pilot version of the Manual for Relating Language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment. Strasbourg: Council of Europe, 2004. 42 p.

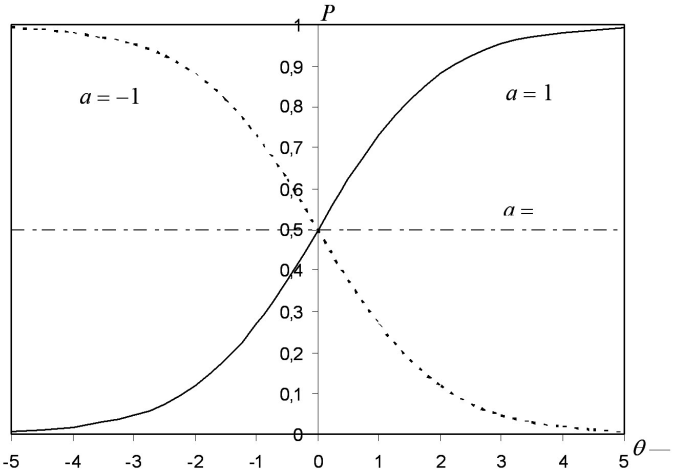

Рис. 1. Характеристическая кривая задания при *a* = −1, 0 и 1

Полезно заметить, что при $a = 1$ двухпараметрическая модель совпадает с моделью Г. Раша.

Чтобы сформулировать критерий оптимальности подбора параметров двухпараметрической модели, нужно знать вероятность результата x. Формула (16) даёт вероятность правильного ответа, т.е. вероятность $P(x = 1)$. Так как x может принимать только одно из двух значений: 1 («правильно») или 0 («неправильно»), то вероятность $P(x = 0)$ неправильного ответа найдём как вероятность противоположного события[8]:

$$P(x=0) = 1 - P(x=1) = 1 - P_{ij}. \qquad (17)$$

Для расчёта вероятности результата x объединим уравнения (16) и (17):

$$P_{ij}(x) = P_{ij}(x=1)^{x_{ij}} \cdot P_{ij}(x=0)^{1-x_{ij}} = P_{ij}^{\,x_{ij}}\left(1 - P_{ij}\right)^{1-x_{ij}},$$

$$P_{ij}(x) = \left(\frac{1}{1+e^{-a_j(\theta_i-\beta_j)}}\right)^{x_{ij}} \left(1 - \frac{1}{1+e^{-a_j(\theta_i-\beta_j)}}\right)^{1-x_{ij}}. \qquad (18)$$

При $x = 1$ (правильный ответ) полученное выражение преобразуется в формулу (16), а при $x = 0$ (неправильный ответ) — в формулу (17):

$$P_{ij}(x=1) = P_{ij}^{\,1}(1-P_{ij})^{1-1} = P_{ij}(1-P_{ij})^{0} = P_{ij}\cdot 1 = P_{ij},$$

$$P_{ij}(x=0) = P_{ij}^{\,0}(1-P_{ij})^{1-0} = 1\cdot(1-P_{ij}) = 1 - P_{ij}.$$

Подставив (18) в выражение (4), запишем критерий максимального правдоподобия для двухпараметрической модели в виде:

$$\prod_{i=1}^{n}\prod_{j=1}^{m} P_{ij}(x) = \prod_{i=1}^{n}\prod_{j=1}^{m}\left(\frac{1}{1+e^{-a_j(\theta_i-\beta_j)}}\right)^{x_{ij}} \cdot \left(1 - \frac{1}{1+e^{-a_j(\theta_i-\beta_j)}}\right)^{1-x_{ij}} \to \max. \qquad (19)$$

8 Два события называются противоположными, если появление одного из них равносильно не появлению другого. Примеры противоположных событий: попадание и промах при стрельбе, «герб» и «цифра» при бросании монеты.

Критерий (18) пригоден также и для модели Г. Раша при $a = 1$.

Критерий подбора параметров модели по методу наименьших квадратов

Сформулируем условие оптимальности подбора параметров модели, используя метод наименьших квадратов. По этому методу приближённые значения параметров находятся из условия минимизации суммы квадратов отклонений между моделью и наблюдаемыми результатами.

Отклонение (различие) между моделью и наблюдаемыми результатами может быть найдено как разность результата выполнения задания x_{ij} и вероятности правильного ответа P_{ij}:

для правильного ответа

$$x_{ij} - P_{ij} = 1 - P_{ij},$$

для неправильного ответа

$$x_{ij} - P_{ij} = 0 - P_{ij} = -P_{ij}.$$

Разность $x_{ij} - P_{ij}$ показывает, насколько расчётная вероятность правильного ответа отличается от идеального значения: для правильного ответа P_{ij} должна стремиться к единице, а для неправильного ответа — к нулю.

Тогда критерий минимума квадратов отклонений для обеих рассматриваемых моделей формулируется в виде условия:

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_{ij} - \frac{1}{1+e^{-a_j(\theta_i-\beta_j)}}\right)^2 \to \min. \qquad (20)$$

В ходе пробных расчётов выявлено, что в ходе поиска решения иногда отмечается существенный рост значений θ. Поэтому условие (6) дополнено вторым ограничением, устанавливающим пределы изменения θ от -5 до $+5$:

$$\left|\theta\right|_{\max} \le 5. \qquad (21)$$

Начальные значения для поиска параметров модели

Критерий оптимальности представляет собой функцию, зависящую от большого числа переменных (в ходе решения все параметры модели являются переменными). Такие функции могут иметь не один максимум или минимум, а несколько. Глобальный максимум (минимум) соответствует максимальному (минимальному) значению критерия. Остальные так называемые локальные, максимумы соответствуют максимальному значению лишь в пределах некоторой окрестности точки экстремума. Если кроме глобального имеются и локальные экстремумы, то нет уверенности, что получен действительно самый «наилучший» результат. Потому важно, с каких начальных значений переменных начинается численное решение. Чем ближе начальные значения θ и β к точке глобального экстремума, тем быстрее и, главное, надёжнее будет получен нужный результат.

В литературе по IRT принято вычисление начальных значений непосредственно по результатам выполнения теста. Так, начальные значения уровней подготовки испытуемых находят по формуле:

$$\theta_i^0 = \ln\frac{p_i}{q_i} = \ln\frac{R_i}{m - R_i}, \quad (22)$$

где p_i и q_i — доля правильных и неправильных ответов i-го тестируемого.

Начальные значения уровней трудности заданий:

$$\beta_j^0 = \ln\frac{q_j}{p_j} = \ln\frac{n - s_j}{s_j}, \quad (23)$$

где p_j и q_j — доля правильных и неправильных ответов на j-е задание теста.

Для двухпараметрической модели начальные значения дифференцирующей способности всех заданий равны единице.

Подбор параметров модели в электронной таблице Microsoft Excel

Все расчёты могут быть выполнены в Microsoft Excel, доступном и широко известном средстве автоматизации вычислений. Excel способен находить экстремум функции одной или нескольких переменных. Кратко опишем технологию поиска параметров модели Г. Раша по методу максимального правдоподобия для представленных в табл. 1 результатов тестирования (принцип решения остальных задач аналогичен). Сначала вводим данные табл. 1 в клетки B2:O35.

Находим сумму баллов каждого тестируемого: в клетку P2 вводим **=СУММ(B2:O2)** и копируем эту формулу в клетки P3:P35. Для суммы баллов по каждому заданию в B36 вводим **=СУММ(B2:B35)** и копируем в C36:O36.

Рассчитываем начальные значения параметров модели:
- для уровней подготовленности испытуемых в Q2 вводим

Таблица 1. **Ввод исходных данных (курсивом выделены номера тестируемых и заданий)**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		*4*	*5*	*7*	*6*	*9*	*8*	*10*	*11*	*13*	*12*	*14*	*15*	*16*	*17*
2	*25*	0	1	0	1	0	0	0	0	0	0	0	0	0	0
3	*4*	1	0	1	0	1	0	0	0	0	0	0	0	0	0
4	*33*	1	0	1	0	0	0	1	0	0	0	0	0	0	0
5	*1*	1	1	1	1	0	0	0	0	0	0	0	0	0	0

=LN(P2/(14–P2)) и копируем в клетки Q3:Q35 (14 — число тестовых заданий);

- для уровней трудности заданий в B37 вводим **=LN((34–B36)/B36)** и копируем в C37:O37 (34 — количество испытуемых).

Определяем вероятности результатов выполнения тестовых заданий, для чего в B42 вводим **=ЕСЛИ(B2>0;1/(1+EXP((–1)·(\$R2-B\$38)));1–1/(1+EXP((–1)·(\$R2–B\$38)))** и копируем в клетки B42:O75. Если значение в клетке B2 больше нуля (задание выполнено правильно), то рассчитывается вероятность правильного ответа по формуле 1/(1+EXP((–1)·(\$R2–B\$38)); в противном случае — вероятность неправильного ответа 1–1/(1+EXP((–1)·(\$R2–B\$38)). Для расчёта параметров двухпараметрической модели вместо сомножителя (–1) следует записать адрес клетки со значением дифференцирующей способности задания.

Зададим параметры модели, равные начальным:

- выделим начальные значения уровней подготовленности испытуемых Q2:Q35 и вставим значения в R2:R35, для чего нажмем кнопку *Копировать*, щелкнем в клетке R2 и выберем в меню *Правка — Специальная вставка — Значения*;
- для уровней трудности выделяем B37:O37 и вставляем значения в B38:O38.

В клетку R36 введем формулу модуля среднего значения уровней подготовленности **=ABS(СРЗНАЧ(R2:R35))** .

Найдем произведение всех вероятностей: в R38 введем **=ПРОИЗВЕД(B42:O75)**. В клетке R38 появляется 1,66431E–58, т.е. начальное значение вероятности всех ответов очень мало — $1{,}66\cdot10^{-58}$. Так как механизм поиска решения Excel требует, чтобы значения критерия и, следовательно, его изменения при подборе параметров не были бы слишком малы, то удобнее искать максимум логарифмической функции правдоподобия. Для этого в клетку R39 введём **= Ln(R38)**.

Проверим, установлен ли механизм поиска решения: выберем в меню *Сервис — Надстройки*, поставим отметку в пункте *Поиск решения* и нажмем *ОК*. Далее нужно выполнить команду *Сервис — Поиск решения* и заполнить диалоговое окно (рис. 2):

1) кликнуть левой клавишей мыши в поле *Установить целевую ячейку* (т.е. клетку, значение которой необходимо максимизировать или минимизировать), переместить указатель мыши и кликнуть на ячейке R39 с формулой критерия оптимальности;

2) указать направление поиска, отметив пункт *Равной максимальному значению*;

3) в поле *Изменяя ячейки* ввести адреса ячеек, значения

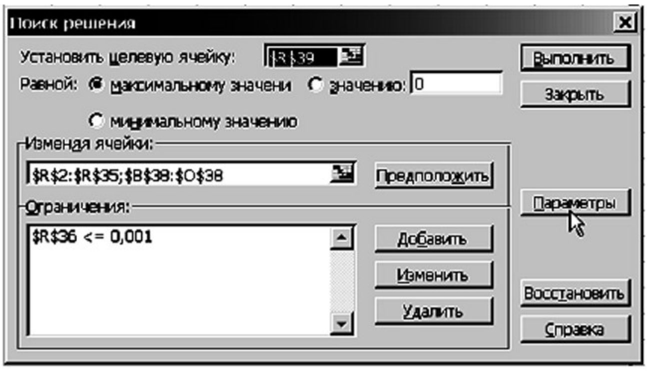

Рис. 2. Окно поиска решения

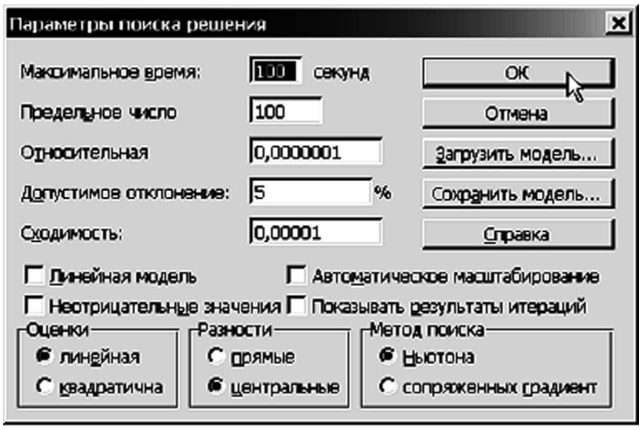

Рис. 3. Рекомендуемые параметры поиска решения

которых будут варьироваться в процессе поиска решения. В нашем случае это R2:R35; B38:O38 (удобно вводить, выделяя клетки с помощью мыши, второй диапазон при нажатой клавише *Ctrl*);

4) с помощью кнопки *Добавить* ввести ограничение: значение B36≤0,001;

5) нажав кнопку *Параметры*, указать параметры поиска в соответствии с рис. 3. Это повысит точность подбора параметров.

После щелчка на кнопке *OK* получим решение поставленной задачи. В клетке R38 находится максимальное значение вероятности всех данных, равное $5{,}09 \cdot 10^{-48}$. Значения уровней подготовленности, найденные по методу наибольшего правдоподобия, записаны в клетках R2:R35, уровней трудности заданий — в ячейках B38:O38.

Анализ результатов подбора параметров модели

Насколько отличаются параметры модели измерения, найденные по разным методикам? Для ответа на этот вопрос сравним результаты решения одной и той же задачи, полученные разными способами. Для примера взята задача подбора параметров модели Г. Раша из известной работы Б.Д. Райта и М.Х. Стоуна[9] (табл. 2). Исходные данные — это результаты выполнения 14 тестовых заданий 34 испытуемыми. Результаты по некоторым задания и испытуемым предварительно были удалены для того, чтобы обеспечить отсутствие строк и столбцов, состоящих только из нулей или только из единиц, поэтому нумерация не сплошная.

[9] *Wright B.D., Stone M.H.* Best Test Design. Chicago: MESA PRESS. 1979.

Таблица 2. **Результаты выполнения теста**

		Тестовые задания													
		4	**5**	**7**	**6**	**9**	**8**	**10**	**11**	**13**	**12**	**14**	**15**	**16**	**17**
Испытуемые	**25**	0	1	0	1	0	0	0	0	0	0	0	0	0	0
	4	1	0	1	0	1	0	0	0	0	0	0	0	0	0
	33	1	0	1	0	0	0	1	0	0	0	0	0	0	0
	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
	27	1	1	1	1	0	0	0	0	0	0	0	0	0	0
	11	0	1	1	1	1	1	0	0	0	0	0	0	0	0
	12	1	1	1	0	1	0	1	0	0	0	0	0	0	0
	17	1	0	1	1	1	1	1	0	0	0	0	0	0	0
	19	1	1	1	1	1	1	0	0	0	0	0	0	0	0
	30	1	1	1	1	1	1	0	0	0	0	0	0	0	0
	2	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	3	1	1	1	1	1	1	1	0	0	1	0	0	0	0
	5	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	6	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	8	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	9	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	13	1	1	0	0	1	1	1	1	0	1	0	0	0	0
	16	1	1	1	1	1	1	0	1	0	0	0	0	0	0
	26	1	1	1	1	1	1	1	0	0	0	0	0	0	0

	28	1	1	1	1	1	1	0	1	0	0	0	0	0	0
	29	1	1	0	1	1	0	1	1	0	0	1	0	0	0
	31	1	1	1	1	1	1	1	0	0	0	0	0	0	0
	10	1	1	1	1	1	1	1	1	0	0	0	0	0	0
	18	1	1	1	1	1	1	1	0	1	0	0	0	0	0
	14	1	1	1	1	1	1	1	1	0	0	0	0	0	0
	32	1	1	1	1	1	1	1	1	0	0	0	0	0	0
	20	1	1	1	1	1	1	1	0	1	0	0	0	0	0
	21	1	1	1	1	1	1	1	1	1	0	0	0	0	0
	22	1	1	1	1	1	1	1	1	0	1	0	0	0	0
	23	1	1	1	1	1	1	1	0	1	0	1	0	0	0
	34	1	1	1	1	1	1	1	0	0	1	1	0	0	0
	15	1	1	1	1	1	1	1	1	1	1	0	0	0	0
	7	1	1	1	1	1	1	1	1	1	1	0	1	0	0
	24	1	1	1	1	1	1	1	1	1	0	0	0	1	1

В той же работе Б.Д. Райта и М.Х. Стоуна приведены результаты подбора параметров модели Г. Раша по алгоритмам PROX и UCON. Алгоритм PROX описан в предыдущей статье[10] автора. Этот упрощённый алгоритм адаптирован для проведения вычислений без использования компьютера.

Алгоритм UCON (от англ. Unconditional maximum likelihood estimation — безусловная оценка по методу максимального правдоподобия) Б.Д. Райт и М.Х. Стоун рекомендуют для компьютерных расчётов. В качестве критерия оптимальности приняты зависимости[11]:

$$\left| \frac{R_i - \sum_{j=1}^{m} p_{ij}}{\sum_{j=1}^{m} p_{ij}(1-p_{ij})} \right| \to \min, \quad (24)$$

$$\left| \frac{s_j - \sum_{i=1}^{n} p_{ij}}{\sum_{i=1}^{n} p_{ij}(1-p_{ij})} \right| \to \min. \quad (25)$$

Очевидно, что числители выражений (24) и (25) соответствуют уравнениям (14) и (15). Следовательно, алгоритм UCON можно рассматривать как способ реализации метода наибольшего правдоподобия

10 *Деменчёнок О.Г.* Определение уровня подготовленности испытуемого с использованием модели Г. Раша // Педагогические измерения, № 1, 2008.

11 *Wright B.D., Stone M.H.* Best Test Design. Chicago: MESA PRESS. 1979.

для модели Г. Раша по варианту решения, связанному с частными производными.

В процессе поиска решения перед очередной итерацией проводится центрирование параметров θ и β, т.е. значения параметров пересчитываются для обеспечения равенства нулю средних значений.

Для оценки параметров модели измерения существуют специализированные компьютерные программы. Наиболее известной является программа Winsteps Д.М. Линека (John M. Linacre). Начальные значения параметров улучшаются с помощью алгоритма PROX, затем для поиска решения используется алгоритм JMLE (второе название алгоритма UCON, от англ. Joint Maximum Likelihood Estimation — объединённая оценка по методу максимального правдоподобия) или предложенный Д.М. Линеком алгоритм XMLE (от англ. Exclusory Maximum Likelihood Estimation — исключительная оценка по методу максимального правдоподобия). XMLE и JMLE используют одинаковые методы оценки параметров модели и приводят к одинаковым результатам, но из-за особенностей способа вычислений XMLE работает быстрее[12]. Хотя программа Winsteps не проводит поиск параметров для двухпараметрической модели, она способна оценивать дифференцирующую способность заданий для случая, когда подготовленность испытуемых и трудность заданий найдены для модели Г. Раша.

Автором проведено определение параметров моделей по данным табл. 2 в программе Winsteps, а также методами максимального правдоподобия (метод МП) и наименьших квадратов (метод НК) в электронной таблице Microsoft Excel. Результаты сведены в табл. 3 и 4.

12 *Linacre J.M.* User's Guide to WINSTEPS – Rasch-Model Computer Programs. Program Manual. Chicago, 2007. 385 p.

Таблица 3. **Уровни подготовленности испытуемых**

Метод расчёта	θ_{25}	θ_4	θ_{33}	θ_1	θ_{27}	θ_{11}	θ_{12}	θ_{17}	θ_{19}	θ_{30}	θ_2	θ_3	θ_5	θ_6	θ_8	θ_9	θ_{13}
Модель Г. Раша																	
Начальные значения	-1,8	-1,3	-1,3	-0,9	-0,9	-0,6	-0,6	-0,3	-0,3	-0,3	0	0	0	0	0	0	0
PROX	-3,8	-2,7	-2,7	-1,9	-1,9	-1,2	-1,2	-0,6	-0,6	-0,6	0	0	0	0	0	0	0
UCON	-3,9	-3,2	-3,2	-2,6	-2,6	-2	-2	-1,2	-1,2	-1,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2	-0,2
XMLE (Winsteps)	-4,2	-3,5	-3,5	-2,8	-2,8	-2,1	-2,1	-1,2	-1,2	-1,2	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1
Метод МП	-4,3	-3,6	-3,6	-2,9	-2,9	-2,2	-2,2	-1,3	-1,3	-1,3	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1	-0,1
Метод НК	-5,0	-5,0	-5,0	-4,1	-4,1	-2,1	-3,8	0,0	-2,0	-2,0	0,0	-1,9	0,0	0,0	0,0	0,0	3,1

Двухпараметрическая модель																	
Метод МП	-3,1	-1,8	-2,4	-2,1	-2,1	-1,1	-1,4	-0,7	-0,7	-0,7	-0,2	-0,4	-0,2	-0,2	-0,2	-0,2	0,4
Метод НК	-5,0	-3,6	-4,1	-2,9	-2,9	-1,2	-3,2	-0,4	-1,2	-1,2	-0,4	-1,1	-0,4	-0,4	-0,4	-0,4	2,8
Метод расчёта	θ_{16}	θ_{26}	θ_{28}	θ_{29}	θ_{31}	θ_{10}	θ_{14}	θ_{18}	θ_{20}	θ_{32}	θ_{21}	θ_{22}	θ_{34}	θ_{15}	θ_{23}	θ_{7}	θ_{24}
Модель Г. Раша																	
Начальные значения	0	0	0	0	0	0,3	0,3	0,3	0,3	0,3	0,6	0,6	0,6	0,9	0,9	1,3	1,3
PROX	0	0	0	0	0	0,6	0,6	0,6	0,6	0,6	1,2	1,2	1,2	1,9	1,9	2,7	2,7
UCON	-0,2	-0,2	-0,2	-0,2	-0,2	0,8	0,8	0,8	0,8	0,8	1,7	1,7	1,7	2,5	2,5	3,3	3,3
XMLE (Winsteps)	-0,1	-0,1	-0,1	-0,1	-0,1	1,1	1,1	1,1	1,1	1,1	2,1	2,1	2,1	3,1	3,1	4,0	4,0
Метод МП	-0,1	-0,1	-0,1	-0,1	-0,1	1,2	1,2	1,2	1,2	1,2	2,3	2,3	2,3	3,3	3,3	4,2	4,2
Метод НК	-0,4	0,0	-0,4	1,8	0,0	1,8	1,8	0,8	0,8	1,8	3,1	3,1	0,3	4,2	3,3	4,9	5,0
Двухпараметрическая модель																	
Метод МП	0,4	-0,2	0,4	-0,6	-0,2	0,6	0,6	0,9	0,9	0,6	1,5	0,8	0,3	2,1	2,0	3,0	3,5
Метод НК	0,6	-0,4	0,6	1,5	-0,4	0,6	0,6	-0,4	-0,4	0,6	3,3	2,8	0,0	3,9	3,3	5,0	4,7

Таблица 4. **Параметры тестовых заданий**

Метод расчёта	**Параметр**	**Номер задания**													
		4	5	7	6	9	8	10	11	13	12	14	15	16	17
Модель Г. Раша															
Начальные значения		-2,8	-2,3	-2,3	-2,0	-2,0	-1,3	-0,9	0,5	1,3	1,5	2,3	3,5	3,5	3,5
PROX		-3,9	-3,3	-3,3	-2,9	-2,9	-2,0	-1,4	0,5	1,5	1,8	2,8	4,3	4,3	4,3
UCON		-4,2	-3,6	-3,6	-3,2	-3,2	-2,2	-1,5	0,8	1,9	2,1	3,2	4,6	4,6	4,6
XMLE (Winsteps)		-4,2	-3,7	-3,7	-3,2	-3,2	-2,2	-1,4	0,8	2,2	2,5	3,7	5,1	5,1	5,1
Метод МП		-4,4	-3,8	-3,8	-3,4	-3,4	-2,3	-1,5	0,8	2,3	2,6	3,8	5,3	5,3	5,3
Метод НК		-5,9	-5,1	-6,0	-4,6	-3,9	-3,1	-1,1	1,2	3,0	3,5	6,6	5,6	5,5	5,5

Двухпараметрическая модель															
Метод МП		-2,8	-2,7	-3,6	-2,8	-1,8	-1,1	-1,1	0,4	1,0	2,5	3,7	3,4	3,4	3,4
	a	1,7	1,2	0,8	0,9	4,0	3,6	1,1	2,6	4,0	0,7	0,7	2,2	2,4	2,4
Метод НК		-4,5	-3,4	-4,5	-3,1	-3,1	-2,1	-0,8	0,2	3,1	3,6	8,8	4,9	5,1	5,1
	a	5,6	8,7	5,6	9,9	1,2	5,4	8,3	9,1	9,3	0,6	0,3	9,2	1,5	1,5
XMLE (Winsteps)	*a*	1,1	1,1	0,6	0,9	1,4	1,4	1,0	1,1	1,3	0,7	0,6	1,2	1,2	1,2

Анализ приведённых в табл. 3 и 4 данных свидетельствует:

1) значения параметров существенно зависят от выбранной модели и метода расчёта;

2) характер изменения параметров θ и β для всех рассмотренных моделей и методов подобен (см. рис. 4). Это подтверждает корреляционный анализ — коэффициент корреляции для уровней подготовленности составляет от 0,92 до 0,99, для уровней трудности — от 0,94 до 0,99, что свидетельствует об очень сильной статистической взаимосвязи;

3) значения дифференцирующей способности заданий, найденные по методу максимального правдоподобия программой Winsteps, хорошо согласованы (коэффициент корреляции 0,90). Однако эти значения слабо согласованы с результатами по методу наименьших квадратов (коэффициент корреляции 0,17);

4) равным начальным уровням подготовленности испытуемых (т.е. одинаковым суммам результатов тестирования R_i) соответствуют равные итоговые значения по всем методикам расчёта для модели Г. Раша, за исключением метода наименьших квадратов;

5) аналогично, равным начальным уровням трудности тестовых заданий (т.е. одинаковым суммам выполнения задания всеми тестируемыми s_j) соответствуют равные итоговые значения по всем методикам расчёта для модели Г. Раша, за исключением метода наименьших квадратов;

6) для двухпараметрической модели равным начальным уровням θ и β могут соответствовать разные итоговые значения, как по методу максимального правдоподобия, так и по методу наименьших квадратов.

Высокая согласованность наборов параметров θ и β для рассмотренных моделей и методов может рассматриваться как доказательство устойчивости и

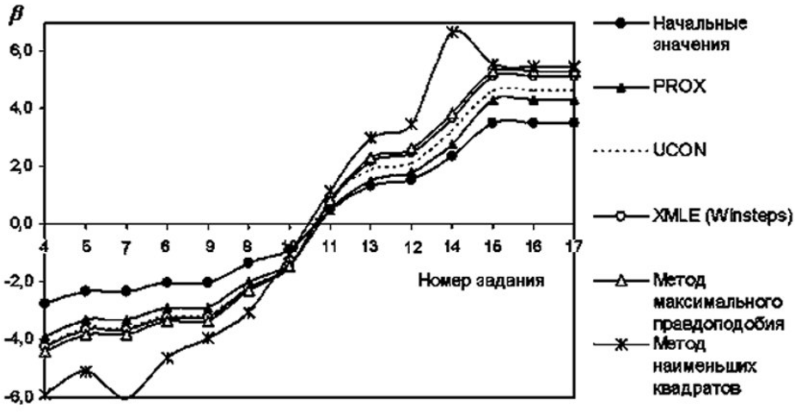

Рис. 4. Уровни трудности заданий (модель Г. Раша)

объективности оценок этих параметров. Так, на рис. 4 приведены графики уровней трудности заданий для модели Г. Раша. Хотя по каждому из методов расчёта получены свои значения *b*, нельзя не заметить сходства графиков: результаты XMLE (Winsteps) практически совпадают с результатами по методу максимума правдоподобия, весьма близки к ним результаты UCON, чуть больше разнятся начальные значения и PROX.

Отмеченные графики очень похожи: на каждом отрезке они практически параллельны. Несколько отличаются результаты по методу наименьших квадратов: при той же общей тенденции отмечается иной характер изменения на отдельных участках. Например, по методу наименьших квадратов уровень трудности 14-го задания больше, чем 15-го; в то время как по всем остальными методам ситуация обратная. Очевидно, эта разница обусловлена отличием критериев подбора параметров.

Несогласованность оценок дифференцирующей способности заданий, на взгляд автора, объясняется тем, что этот параметр играет роль регулятора точной настройки для достижения более полного совпадения расчётных и экспериментальных данных. Поэтому изменение соотношения других параметров модели способно провести к существенной вариации значения дифференцирующей способности задания.

Из полученных результатов следует, что соответствие одинаковых начальных уровней подготовленности (т.е. равных исходных баллов R_i) равным итоговым показателям — это не свойство IRT, и даже не свойство математической модели Г. Раша, а лишь особенность конкретного метода расчёта.

Отметим также, что получены только положительные значения дифференцирующей способности заданий. Это подтверждает соответствие тестовых заданий положениям теории: лучше подготовленные тестируемые успешнее справляются с заданиями. Если бы в результате подбора параметров были бы найдены задания с дифференцирующей способностью $a \le 0$, то эти задания следовало бы удалить ввиду непригодности для педагогических измерений.

Довольно часто встречается утверждение, что исходный балл R_i однозначно определяет уровень подготовленности, причём для нахождения подготовленности отдельного испытуемого предлагается уравнение вида[13]:

$$R_i = \sum_{j=1}^{m} p_{ij}, \qquad (26)$$

т.е уровень подготовленности определяется из условия равенства исходного балла i-го испытуемого и суммы вероятностей правильных ответов на все задания для того же человека.

Для оценки правомерности такого подхода для каждого тестируемого по всем полученным моделям найдена абсолютная величина ошибки ΔR_i:

$$\Delta R_i = \left| R_i - \sum_{j=1}^{m} p_{ij} \right| = \left| R_i - \sum_{j=1}^{m} \left(\frac{1}{1 + e^{-a_j(\theta_i - \beta_j)}} \right) \right|. \qquad (27)$$

Средние ΔR_{cp} и максимальные ΔR_{max} значения ΔR_i приведены в табл. 5.

Таблица 5. **Отклонение между исходным баллом и суммой вероятностей правильных ответов**

Метод расчёта	ΔR_{cp}	$\Delta R_{\max}$
Модель Г. Раша		
Начальные значения	0,8	2,1
PROX	0,4	1,0
UCON	0,1	0,4
XMLE (Winsteps)	0,3	0,8
Метод максимального правдоподобия	0,0	0,0
Метод наименьших квадратов	0,4	2,0
Двухпараметрическая модель		
Метод максимального правдоподобия	0,4	1,5
Метод наименьших квадратов	0,6	1,7

13 *Wright B.D., Stone M.H.* Best Test Design. Chicago: MESA PRESS. 1979.

Очевидно, что равенство исходных баллов R_i и суммы вероятностей правильных ответов обеспечивает только реализация метода максимального правдоподобия для модели Г. Раша средствами Microsoft Excel: только в этом случае $\Delta R_{cp} = \Delta R_{max} = 0$. Во всех остальных случаях использование для определения уровня подготовленности уравнения (26) приводит к ошибке. Поэтому для подбора уровня подготовленности отдельного тестируемого целесообразно выбрать другое условие, например, по методу наименьших квадратов:

$$\sum_{j=1}^{m}\left(x_{ij}-\frac{1}{1+e^{-a_j(\theta_i-\beta_j)}}\right)^2 \to \min. \quad (28)$$

Анализ качества моделей

Какая модель и какой из способов оценки параметров модели измерения лучше? Как уже отмечалось, модель должна как можно лучше объяснять наблюдаемые результаты тестирования. Модель измерения определяет вероятность одного ответа. Следовательно, главный показатель качества модели — степень близости к единице расчётной вероятности результата тестирования. Оценивать это предлагается по средней вероятности фактически полученных результатов выполнения тестовых заданий (для расчётов по модели Г. Раша $a = 1$):

$$\overline{p}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}P_{ij}}{n\cdot m}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{1}{1+e^{-a_j(\theta_i-\beta_j)}}\right)^{x_{ij}}\left(1-\frac{1}{1+e^{-a_j(\theta_i-\beta_j)}}\right)^{1-x_{ij}}}{n\cdot m}. \quad (29)$$

При полном совпадении модели и данных тестирования средняя вероятность равна единице. Тогда средняя ошибка модели Δ может быть определена выражением:

$$\Delta = 1-\overline{p}. \quad (30)$$

Средняя квадратичная ошибка (т.е. стандартное отклонение между расчётом по модели и данными табл. 1):

$$\sigma=\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\delta_{ij}-\overline{\delta}\right)^2}{n\cdot m-1}}, \quad (31)$$

где $\delta_{ij} = x_{ij} - p_{ij}$. — отклонение между моделью и наблюдаемыми результатами (для правильного ответа δ_{ij} обращается в ноль при вероятности правильного ответа, равной единице, а для неправильного ответа δ_{ij} равно нулю при нулевой вероятность правильного ответа); $\overline{\delta}$ — среднее значение δ_{ij}.

Универсальным показателем качества модели является также коэффициент детерминации $K_Д$, измеряющий степень тесноты статистической взаимосвязи расчётных и экспериментальных данных[14]:

14 *Айвазян С.А., Мхитарян В.С.* Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998.

$$K_Д = 1 - \frac{\sigma^2}{\sigma_x^{\ 2}} = 1 - \frac{\sigma^2(n \cdot m - 1)}{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_{ij} - \bar{x}\right)^2}, \qquad (32)$$

где σ_x — стандартное отклонение результатов x_{ij}; $\bar{x}$ — среднее значение x_{ij}.

Показатели качества сведены в табл. 6. Для удобства анализа дополнительно приведены относительные значения показателей по сравнению с результатами при начальных значениях параметров модели.

Очевидно, что начальные значения — это действительно хорошая отправная точка для подбора параметров модели измерения. Средняя вероятность полученных результатов $\bar{p}$ при начальных значениях равна 0,808; т.е. уже при начальных значениях модель на 81% объясняет наблюдаемые результаты тестирования. Подбор параметров модели Г. Раша даёт возможность увеличить $\bar{p}$ до 0,845...0,882 (или на 4,6–9,2%). Подбор параметров двухпараметрической модели повышает среднюю вероятность результатов тестирования до 0,874...0,924, что лучше начального значения на 8,2–14,4%.

Таблица 6. **Показатели качества модели**

Метод расчёта	$\bar{p}$	$\bar{p}/\bar{p}_0$, %	Δ	Δ/Δ_0, %	σ	σ/σ_0, %	$K_д$	$K_д/K_{д_0}$,%
Начальные значения	0,808	100,0	0,192	100,0	0,288	100,0	0,834	100,0
Модель Г. Раша								
PROX	0,845	104,6	0,155	80,7	0,270	93,8	0,854	102,4
UCON	0,854	105,7	0,146	76,1	0,265	92,1	0,859	103,0
XMLE (Winsteps)	0,858	106,3	0,142	73,8	0,266	92,4	0,858	102,9
Метод максимального правдоподобия	0,861	106,7	0,139	72,1	0,264	91,6	0,861	103,2
Метод наименьших квадратов	0,882	109,2	0,118	61,5	0,251	87,0	0,874	104,8
Двухпараметрическая модель								
Метод максимального правдоподобия	0,874	108,2	0,126	65,6	0,248	86,2	0,877	105,1
Метод наименьших квадратов	0,924	114,4	0,076	39,6	0,223	77,3	0,901	108,0

В рамках модели Г. Раша начальную среднюю ошибку D=0,192 можно уменьшить до 0,155...0,118; т.е. примерно на 20–40%. При двухпараметрической модели ошибка ещё меньше – 0,126...0,076, что на 35–60% ниже начальной величины. Таким образом, по сравнению с моделью Г. Раша двухпараметрическая модель даёт выигрыш в снижении начальной величины средней ошибки на 15–20%.

Средняя квадратичная ошибка s является мерой рассеивания ошибок. В данном случае меньшим средним ошибкам соответствуют меньшие значения s. Как и следовало ожидать, средние квадратичные ошибки минимальны для метода наименьших квадратов, поскольку этот метод минимизирует сумму квадратов ошибок. Так как исходные данные для всех методов одни и те же, то знаменатель в формуле (32) неизменен. Поэтому лучшие значения коэффициентов детерминации соответствуют минимальным средним квадратичным ошибкам, т.е. методу наименьших квадратов.

Если сравнивать методы расчёта, то упрощенный алгоритм PROX приводит к относительно низким показателям качества модели. Для модели Г. Раша более высокие показатели качества у алгоритмов UCON, XMLE (Winsteps) и реализации метода максимального правдоподобия в Microsoft Excel. Небольшое различие между результатами по этим методам обусловлено более высокой эффективностью механизма поиска экстремума электронной таблицы Microsoft Excel.

Метод наименьших квадратов позволяет создать наиболее адекватные модели; по сравнению с методом максимального правдоподобия средняя ошибка меньше на 3...6%. Это не означает безусловного преимущества метода наименьших квадратов. Из теории прикладной статистики известно: точные оценки параметров модели по методу максимального правдоподобия требуют очень большого объёма данных. Поэтому при малых выборках с оценками по методу максимального правдоподобия могут конкурировать (и даже превосходить их) оценки по методу наименьших квадратов[15]. Следовательно, полученные результаты надо трактовать так: при небольшом объёме данных (десятки тестируемых) метод наименьших квадратов приводит к более точным моделям, чем метод максимального правдоподобия. При увеличении объёма данных (сотни и тысячи тестируемых) показатели качества моделей, полученных по этим методам, вероятно, сравняются.

15 *Айвазян С.А., Мхитарян В.С.* Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998.

Выводы

1. Значения параметров модели измерения существенно зависят от типа выбранной модели и метода расчёта.
2. Характер изменения параметров θ и β для всех рассмотренных моделей и методов подобен, что может рассматриваться как доказательство устойчивости и объективности оценок этих параметров.
3. Несогласованность оценок дифференцирующей способности заданий a, на взгляд автора, объясняется тем, что этот параметр играет роль регулятора точной настройки для достижения более полного совпадения расчётных и экспериментальных данных. Поэтому изменение соотношения других параметров модели способно провести к существенной вариации значения дифференцирующей способности задания.
4. Соответствие одинаковых равных исходных баллов равным уровням подготовленности — это не свойство IRT, и даже не свойство математической модели Г. Раша, а лишь особенность конкретного метода расчёта.
5. Большинство методов расчёта не обеспечивает равенства исходного балла R_i и суммы вероятностей правильных ответов, поэтому использование этого равенства в качестве условия для определения уровня подготовленности отдельного испытуемого не целесообразно.
6. По сравнению с моделью Г. Раша двухпараметрическая модель даёт выигрыш в существенном снижении средней ошибки, т.е. эта модель лучше объясняет экспериментальные данные. Кроме того, двухпараметрическая модель позволяет выявить непригодные для педагогических измерений задания с дифференцирующей способностью $a \le 0$.
7. При небольшом объёме данных (десятки тестируемых) метод наименьших квадратов приводит к более точным моделям, чем метод максимального правдоподобия. При увеличении объёма данных показатели качества моделей, полученных по этим методам, вероятно, сравняются.