

Анализ алгоритма PROX для модели Г. Раша

Олег Деменчёнок

Восточно-Сибирский институт МВД России

AskSystem@yandex.ru

Опубликовано ж. «Педагогические Измерения» № 3, 2009.

Рассмотрены основные подходы к нахождению уровней подготовленности тестируемых и трудности заданий в соответствии с моделью Раша. Описан алгоритм упрощенного решения PROX, а также технология его реализации средствами Microsoft Excel. Показано, что принятое в PROX упрощение решения ведёт к снижению точности обработки результатов тестирования.

Ключевые слова: тест, Item Response Theory (IRT)¹, модель Раша, вероятность правильного ответа, уровень подготовленности, уровень трудности задания, оценивание параметров модели, проверка нормальности распределения.

Читатель, интересующийся педагогическими измерениями, безусловно, знаком с публикациями, посвящёнными тем или иным аспектам Item Response Theory. Таких публикаций немало в последние годы. Как известно, нет ничего практичнее хорошей теории. Однако Item Response Theory ещё очень далеко до широкого применения в отечественной педагогической практике. На взгляд автора, причина этого заключается не столько в сложности теории, сколько в новизне и непривычности идей IRT для педагогического сообщества. Любое нововведение встречается скептически. И чем меньше новая теория согласуется со сложившимся комплексом взглядов, тем сильнее противодействие. Предлагаемый IRT подход к измерению подготовленности обучаемых принципиально отличается от других теорий педагогических измерений, что существенно тормозит её распространение. Для раскрытия потенциала IRT нужно помочь педагогу понять и принять эту теорию.

Модель Раша

Основой современного подхода к педагогическому измерению является математическая модель. Существует множество определений математической модели, которые, как правило, сводятся к описанию объекта исследования средствами математического аппара-

¹ На русский язык IRT В.С.Аванесов переводит как математическую теорию измерений (МТИ). См.: Педагогические измерения, № 3, 2007. с.3.

та. Электронный словарь CliffsNotes² даёт, пожалуй, наиболее лаконичное определение математической модели: это уравнение, выражающее идею.

Применительно к педагогическому тестированию главная идея сформулирована датским математиком Георгом Рашем: "Вероятность правильного ответа определяется уровнем подготовленности испытуемого и мерой трудности тестового задания". Именно эта идея дала жизнь всем современным математическим моделям педагогического измерения. Одна из наиболее известных моделей предложена самим Георгом Рашем. Он счёл возможным ограничить модель одним единственным параметром: вероятность правильного ответа P зависит только от разности между уровнем подготовленности тестируемого θ и уровнем трудности задания β ³. Поэтому его модель называют однопараметрической, хотя она оперирует двумя переменными величинами.

Г.Раш выбрал для математической модели логистическую функцию⁴ вида:

$$f(t) = \frac{e^t}{1 + e^t}, \quad (1)$$

где $e \approx 2,72$ – основание натурального логарифма.

Логистическая функция является монотонно возрастающей (рис.1). Начальная стадия роста логистической кривой приблизительно соответствует экспоненте (показательная функция). Затем рост замедляется, проходит почти линейную фазу и, наконец, и практически останавливается.

Учитывая, что $t = \theta - \beta$, модель Раша обычно записывается в виде:

$$P = \frac{e^{\theta - \beta}}{1 + e^{\theta - \beta}}. \quad (2)$$

Модель Раша также можно записать в эквивалентной форме:

$$P = \frac{1}{1 + e^{-(\theta - \beta)}}. \quad (3)$$

Если значение t положительно (уровень подготовленности тестируемого больше уровня трудности задания), то вероятность правильного ответа P больше вероятности ошибки, т.е. $P > 0,5$. Чем больше разность между уровнем подготовленности и уровнем трудности задания, тем легче испытуемому решить задание. И действительно, с увеличением t вероятность правильного ответа возрастает. Так как вероятность не может

² на языке оригинала "mathematical model is an equation representing an idea";

<http://www.cliffsnotes.com/WileyCDA/Section/Earth-Science-Glossary.id-305499,articleId-57021.html>

³ Rasch G. Probabilistic models for some intelligence and attainment tests. - Copenhagen, Denmark: Danish Institute for Educational Research, 1960.

⁴ Логистической называется функция с S-образной (сигмоидной) формой графика, принимающая значения в интервале от нуля до единицы (Электронный учебник по статистике. Москва, StatSoft; http://www.statsoft.ru/home/textbook/glossary/GLOSS_L.HTML)

быть больше единицы, вероятность правильного ответа будет неограниченно приближаться к единице (асимптотически стремиться к единице).

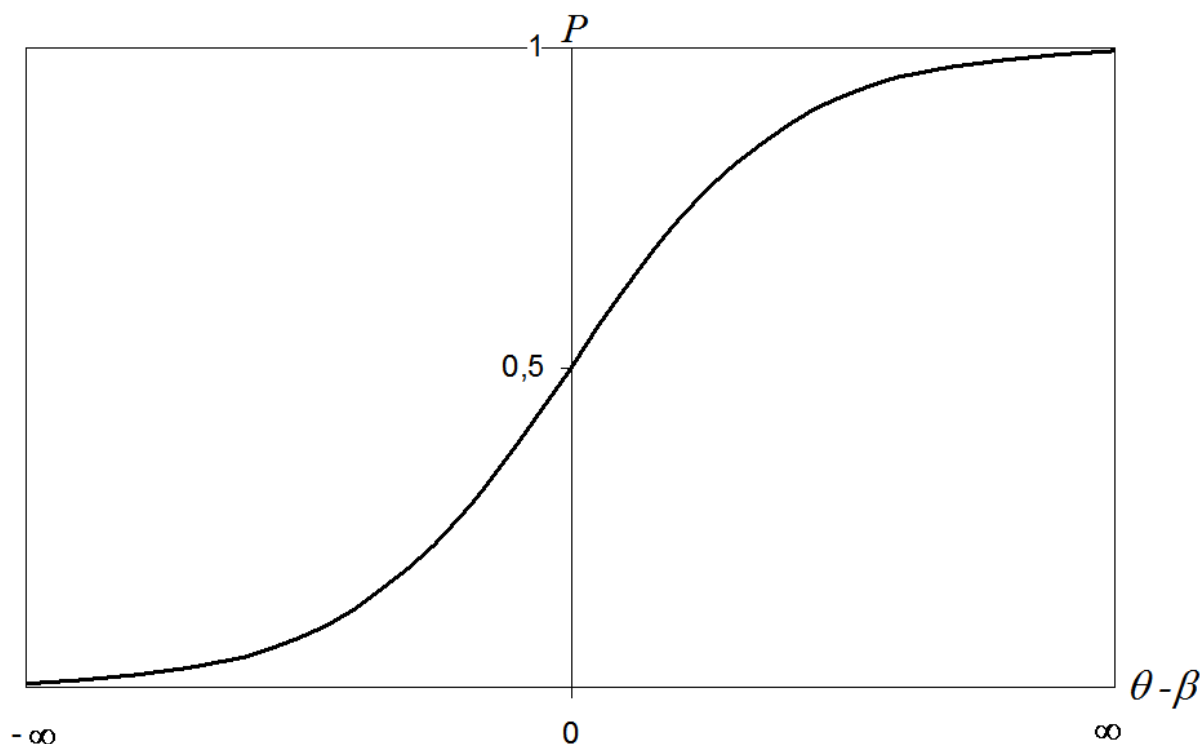


Рис.1. Характер взаимосвязи между параметром модели Раша и вероятностью правильного ответа

Если уровень подготовленности меньше уровня трудности задания, то вероятность правильного ответа $P < 0,5$; и с увеличением модуля разности P уменьшается, стремясь к нулю. Характер взаимосвязи между параметром модели и вероятностью правильного ответа приведён на рис.1.

Отметим несколько существенных свойств модели Раша.

1. Значения вероятности правильного ответа изменяются в интервале от 0 до 1.
2. В случае, когда уровень подготовленности испытуемого намного превосходит уровень трудности задания, вероятность правильного ответа близка к единице.
3. В случае, когда уровень подготовленности испытуемого намного меньше уровня трудности задания, вероятность правильного ответа близка к нулю.
4. В случае равенства уровня подготовленности и уровня трудности задания вероятность правильного ответа $P = 0,5$.
5. Модель Раша с точностью до сотых совпадает со значениями интегральной функции нормированного нормального распределения при изменении масштаба аргумента с коэффициентом 1,7. В этом случае ошибка аппроксимации (расхождение между графиком модели Г.Раша и интегральной кривой нормального распределения) не превышает 0,01:

$$\left| \frac{e^t}{1+e^t} - \frac{1}{2\pi} \int_{-\infty}^{\frac{t}{1,7}} e^{-\frac{\left(\frac{t}{1,7}\right)^2}{2}} dt \right| \leq 0,01 \quad \text{или} \quad \left| \frac{e^t}{1+e^t} - F\left(\frac{t}{1,7}\right) \right| \leq 0,01, \quad (4)$$

где $t = \theta - \beta$; F – функция распределения стандартной нормальной случайной величины, среднее значение \bar{t} которой равно нулю, а стандартное отклонение σ равно единице⁵.

Для большей наглядности на рис.2 приведены графики модели Г.Раша и функция распределения стандартной нормальной случайной величины.

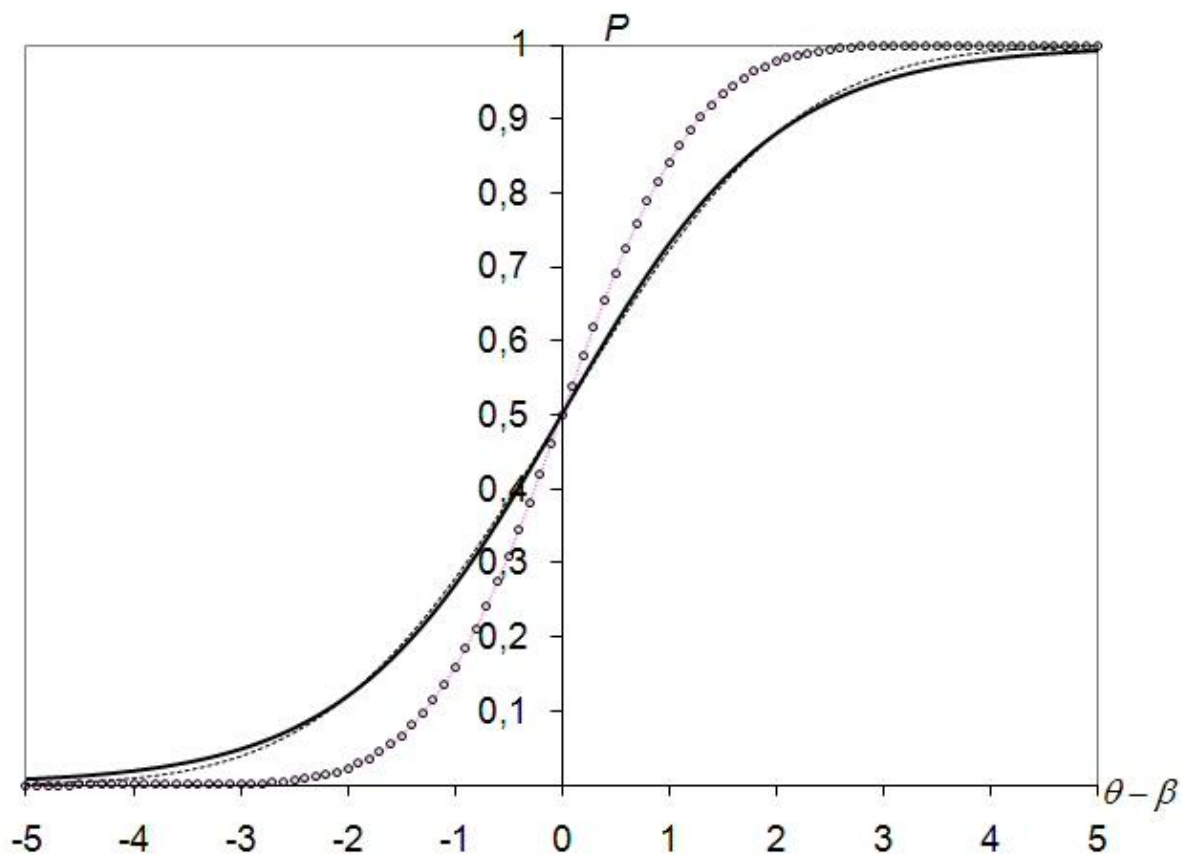


Рис.2. Сравнение графиков модели Г.Раша и функции распределения стандартной нормальной случайной величины: ———— - модель Г.Раша, оооооооо - интегральная кривая нормального распределения; - - - - - интегральная кривая нормального распределения при изменении масштаба аргумента с коэффициентом 1,7

Теоретические основы измерений с использованием модели Раша

Парадоксально, но факт: расчет по модели Раша даже теоретически не может совпасть с результатами тестирования. Объясняется это тем, что в соответствии с пред-

⁵ Кибзун А.И. и др. Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами. – М.: ФИЗМАТЛИТ, 2002. – 224 с.

ставленной моделью Раша результат выполнения тестового задания x может принимать только одно из двух значений – 1 («правильно») или 0 («неправильно»). А расчётная вероятность правильного ответа никогда не равна нулю или единице, она может принимать только дробные значения. Подобно этому, результат подбрасывания монеты – «орёл» (1) или «решка» (0) – никогда не равен своей теоретической вероятности (0,5).

Как же тогда применить модель? Известно, что статистические закономерности проявляются тем полнее и точнее, чем больше объём данных. Отдельные, единичные явления (например, угадывание ответа) содержат в себе элемент случайного. При анализе большого количества подобных фактов случайные отклонения от основной статистической закономерности взаимно поглощаются, причём тем больше, чем большее число фактов рассматривается. Именно поэтому тест, как правило, состоит из нескольких десятков тестовых заданий, что позволяет применить для обработки результатов тестирования аппарат теории вероятностей и математической статистики.

Обычно анализируется не результат выполнения теста отдельным испытуемым, а весь массив результатов выполнения m тестовых заданий n испытуемыми. Данные могут быть представлены в виде матрицы, столбцы которой – результаты выполнения заданий, а строки – результаты тестируемых (табл.1).

Таблица 1

		Тестовые задания						
		1	2	j	...	m
Тестируемые	1	1	1	0	0	0
	2	1	1	...	0	0
	1	0	...	1	0	0
	...	1	1	...	1	1	1	0
	i	1	...	1	0	0	0	0
	1	1	1	0	1	0
	n	1	1	1	1	1	0	1

По модели Г.Раша вероятность правильного ответа i -го испытуемого на j -е задание равна:

$$P_{ij} = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}}, \quad (5)$$

где θ_i – уровень подготовленности i -го испытуемого; β_j – уровень трудности j -го задания.

Вероятность неправильного ответа найдем как вероятность противоположного события⁶

$$\begin{aligned}\bar{P}_{ij} &= 1 - P_{ij} = 1 - \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} = \frac{1 + e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} - \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \\ \bar{P}_{ij} &= \frac{1}{1 + e^{\theta_i - \beta_j}}\end{aligned}\quad (6)$$

Из уравнений (5) и (6) следует, что вероятность результата x при решении i -ым испытуемым j -го задания

$$P_{ij}(x) = \frac{e^{x_{ij}(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} \quad (7)$$

Нетрудно заметить, что при $x = 1$ (правильный ответ) выражение (7) преобразуется в формулу (5), а при $x = 0$ (неправильный ответ) – в формулу (6).

Считается, что результат выполнения каждого отдельного тестового задания в исходной матрице не зависит от результатов выполнения других тестовых заданий. Тогда вероятность полученных при тестировании результатов $\mathbf{P}(x)_{ij}$ (т.е. вероятность всех входящих в матрицу значений x) согласно теоремы умножения вероятностей независимых событий⁷ может быть найдена как произведение вероятностей выполнения каждого задания каждым тестируемым:

$$\mathbf{P}(x)_{ij} = \prod_{i=1}^n \prod_{j=1}^m \frac{e^{x_{ij}(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} \quad (8)$$

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

$$\mathbf{P}(x)_{ij} = \frac{e^{\sum_{i=1}^n \sum_{j=1}^m x_{ij}(\theta_i - \beta_j)}}{\prod_{i=1}^n \prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} \quad (9)$$

Введём обозначения:

R_i – сумма результатов выполнения тестовых заданий i -м тестируемым

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

⁶ Два события называются противоположными, если появление одного из них равносильно не появлению другого.

⁷ Вентцель Е.С. Теория вероятностей. – М.: Высшая школа, 2001. – 576 с.

s_j – сумма результатов выполнения j -го задания всеми тестируемыми

$$s_i = \sum_{j=1}^n x_{ij} \quad (11)$$

Используя (10) и (11), заменим

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

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

Таким образом, уравнение (9) принимает вид

$$P(x)_{ij} = \frac{e^{\sum_{i=1}^n R_i \theta_i - \sum_{j=1}^m s_j \beta_j}}{\prod_{i=1}^n \prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} \quad (12)$$

Уравнение (12) является достаточно важным, так как оно обосновывает следующие положения⁸:

- для определения уровней подготовленности испытуемых и уровней трудности заданий достаточно знать величины R_i и s_j (сумму результатов выполнения тестовых заданий i -м тестируемым и сумму результатов выполнения j -го задания всеми тестируемыми);
- уровни подготовленности тестируемых и уровни трудности заданий могут оцениваться независимо друг от друга;
- независимость трудности заданий от подготовленности тестируемых позволяет обойтись без шкалирования уровней трудности заданий по отношению к уровню трудности образцового (базового) задания.

Педагогические измерения с использованием модели Раша основано на подборе таких значений уровней подготовленности тестируемых θ_i и уровней трудности заданий β_j , при которых результаты конкретного тестирования и математическая модель наилучшим образом совпадают.

Решение этой задачи во многом зависит от того, что считать «наилучшим», т.е. от принятого критерия оптимальности. Обычно критерий оптимальности находят по методу максимального (наибольшего) правдоподобия, хотя может быть принят метод

⁸ Wright B.D., Stone M.H. Best Test Design. - Chicago: MESA PRESS. 1979.

моментов или метод наименьших квадратов⁹.

В общем виде метод максимума правдоподобия можно сформулировать так: наилучшим описанием явления считается то, которое дает наибольшую вероятность получить в результате измерений именно те значения, которые и были фактически получены¹⁰. В формальной записи этот метод может быть представлен в виде максимума произведения вероятностей всех наблюдаемых независимых событий¹¹:

$$p_1 \cdot p_2 \cdot \dots \cdot p_N = \prod_{i=1}^N p_i \rightarrow \max \quad (13)$$

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

$$P(x)_{ij} = \prod_{i=1}^n \prod_{j=1}^m \frac{e^{x_{ij}(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} = \frac{e^{\sum_{i=1}^n R_i \theta_i - \sum_{j=1}^m S_j \beta_j}}{\prod_{i=1}^n \prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} \rightarrow \max \quad (14)$$

Существует два основных подхода к обработке результатов тестирования с использованием полученных формул.

Первый подход основан на непосредственном применении критерия оптимальности (14). Этот подход известные в области педагогических измерений американские теоретики Б.Д.Райт и М.Х.Стоун¹² сочли идеальным, поскольку он позволяет достичь максимально возможной точности. Однако ввиду очень большого объёма вычислений метод может быть реализован только при помощи численных методов, компьютерной техники и соответствующего программного обеспечения (подробно этот подход описан в работе¹³). Определённым недостатком данного подхода является непрозрачность процедуры оценивания. Программа, выполняющая обработку результатов тестирования, в глазах педагога представляет собой «чёрный ящик»¹⁴:

- известны входные данные (матрица ответов),
- на выходе получаем значения уровней подготовленности тестируемых и трудности

⁹ Linacre J.M. User's Guide to WINSTEPS – Rasch-Model Computer Programs. Program Manual. – Chicago, 2007. 385p.

¹⁰ Львовский Б.Н. Статистические методы построения эмпирических формул. - М.: Высшая школа, 1988.

¹¹ Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. - М.: ЮНИТИ, 1998.

¹² Wright B.D., Stone M.H. Best Test Design. - Chicago: MESA PRESS. 1979.

¹³ Деменчёнок О.Г. Подбор параметров модели педагогических измерений.// Педагогические измерения, 2008, №1.

¹⁴ Чёрный ящик – объект изучения, внутреннее устройство которого либо неизвестно, либо слишком сложно для того, чтобы можно было по свойствам его составных частей (элементов) и структуре связей между ними делать выводы о поведении объекта; метод исследования таких объектов. (Большая советская энциклопедия, электронная версия. – М.: Большая Российская энциклопедия, 2002).

заданий,

- как именно производится оценка – непонятно (перепроверить результаты вручную невозможно). Непрозрачность процедуры оценивания может вызвать недоверие, как педагога, так и обучаемых, а равно и к этому методу оценивания.

Второй подход основан на упрощении обработки результатов тестирования за счёт введения дополнительных допущений (предположений). Благодаря этим допущениям становится возможной обработка результатов тестирования вручную, без компьютера. Целесообразность этого Б.Д.Райт и М.Х.Стоун в 1979 году мотивировали непрактичностью использования математической модели ввиду больших затрат машинного времени компьютера.

В настоящее время такой довод неактуален: матрица ответов десятков испытуемых на десятки тестовых заданий в соответствии с моделью Раша на обычном персональном компьютере обрабатывается около секунды. Читатель может лично убедиться в этом, скачав бесплатную компьютерную программу Estimate2PL (сайт www.asksystem.narod.ru). Кроме того, упрощения неизбежно снижают точность педагогического измерения. Однако упрощенный вариант вычислений ценен тем, что может быть выполнен без специализированных программ. Тем самым процедура оценивания становится более понятной и обретает прозрачность – любое заинтересованное лицо получает возможность проверки результатов. Таким образом, упрощенный вариант вычислений помогает педагогу понять и принять IRT.

Простая процедура оценивания PROX

Внедрение модели Раша в практику тестирования началось в 60-е годы прошлого столетия. В то время компьютеров было мало, их вычислительные ресурсы в тысячи раз уступали возможностям современных компьютеров. Поэтому пользовались популярностью различные варианты упрощенной обработки результатов тестирования.

В качестве примера можно привести алгоритм PROX, который адаптирован для проведения вычислений без использования компьютера. Этот алгоритм для модели Раша предложен Лесли Коэном (Leslie Cohen)¹⁵. PROX (от англ. Normal Approximation Estimation Algorithm – алгоритм оценки путём нормального приближения) базируется на предположении о нормальности распределения уровней подготовленности (среднее значение $\bar{\theta}$, стандартное отклонение σ_{θ}) и уровней трудности (среднее значение $\bar{\beta}$, стандартное отклонение σ_{β}).

¹⁵ Cohen L. Approximate Expressions for Parameter Estimates in the Rasch Model, The British Journal of Mathematical and Statistical Psychology, 1979, №32, p.113-120.

PROX основан на близости модели Раша и функции распределения стандартной нормальной случайной величины:

$$\frac{e^{\theta-\beta}}{1+e^{\theta-\beta}} \approx F\left(\frac{\theta-\beta}{1,7}\right)$$

Просуммируем результаты выполнения j -го задания всеми тестируемыми

$$S_j = \sum_{i=1}^n x_{ij} = \sum_{i=1}^n P_{ij} = \sum_{i=1}^n \frac{e^{\theta_j - \beta_i}}{1 + e^{\theta_j - \beta_i}} \approx \sum_{i=1}^n F\left(\frac{\theta_j - \beta_i}{1,7}\right) \quad (15)$$

Считая, что уровни трудности заданий подчиняются нормальному распределению, запишем (15) в виде¹⁶:

$$S_j = \sum_{i=1}^n P_{ij} = n \int_{-\infty}^{+\infty} F\left(\frac{\theta - \beta_i}{1,7}\right) \frac{d}{d\theta} \left[F\left(\frac{\theta - \bar{\theta}}{\sigma_\theta}\right) \right] d\theta$$

Исходя из свойств функции распределения стандартной нормальной случайной величины

$$\int_{-\infty}^{+\infty} F(a+bt)F'(t) dt = F\left(\frac{a}{\sqrt{1+b^2}}\right)$$

запишем

$$S_j = \sum_{i=1}^n P_{ij} = n \cdot F\left(\frac{\bar{\theta} - \beta_j}{\sqrt{\sigma_\theta^2 + 1,7^2}}\right).$$

Перенесём n в левую часть уравнения и проведём обратную замену функции распределения стандартной нормальной случайной величины на модель Раша:

$$\begin{aligned} \frac{S_j}{n} &= \frac{1}{1 + e^{-\frac{\bar{\theta} - \beta_j}{\sqrt{\sigma_\theta^2 + 2,9}} \cdot 1,7}} = \frac{1}{1 + e^{-\frac{\bar{\theta} - \beta_j}{\sqrt{1 + \sigma_\theta^2 / 2,9}}}}, \\ \frac{n}{S_j} &= 1 + e^{-\frac{\bar{\theta} - \beta_j}{\sqrt{1 + \sigma_\theta^2 / 2,9}}}, \\ e^{-\frac{\bar{\theta} - \beta_j}{\sqrt{1 + \sigma_\theta^2 / 2,9}}} &= \frac{n - S_j}{S_j}, \end{aligned}$$

¹⁶ Linacre J.M. PROX with missing data, or known item or person measures./Rasch Measurement Transactions, 1994, 8:3 p.378.

$$-\frac{\bar{\theta} - \beta_j}{\sqrt{1 + \frac{\sigma_\theta^2}{2,9}}} = \ln \frac{n - S_j}{S_j}.$$

Окончательно для уровня трудности j -го задания получаем:

$$\beta_j = \bar{\theta} + \ln \frac{n - S_j}{S_j} \sqrt{1 + \frac{\sigma_\theta^2}{2,9}}. \quad (16)$$

Аналогичный ход преобразований уравнения (10) приводит к следующей зависимости для уровня подготовленности i -го тестируемого:

$$\theta_i = \bar{\beta} + \ln \frac{R_i}{m - R_i} \sqrt{1 + \frac{\sigma_\beta^2}{2,9}}. \quad (17)$$

Уравнения (16) и (17) являются ключевыми в алгоритме PROX.

Практические расчёты с использованием PROX

Сначала определяются начальные значения уровней подготовки

$$\theta_i^0 = \ln \frac{p_i}{q_i} = \ln \frac{R_i}{m - R_i}, \quad (18)$$

где p_i и q_i – доля правильных и неправильных ответов i -го тестируемого;
и начальные значения логитов трудности заданий

$$\beta_j^0 = \ln \frac{p_j}{q_j} = \ln \frac{n - s_j}{s_j}, \quad (19)$$

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

Примечание: при делении на ноль получается неопределенность, а логарифм нуля не существует. Следовательно, и числитель, и знаменатель выражений (18) и (19) должен быть ненулевым. Поэтому в рамках модели Раша не могут быть достаточно точно оценены испытуемые, все ответы которых правильны или неправильны, а также задания, на которые даны только правильные или только неправильные ответы. В таких случаях может быть принята следующая интерпретация:

- если все ответы испытуемого правильны, то уровень его подготовленности превышает уровень трудности любого из выполненных им заданий;
- если все ответы испытуемого неправильны, то уровень его подготовленности ниже уровня трудности любого из выполненных им заданий;
- если все ответы на тестовое задание правильны, то уровень его трудности меньше уровня подготовленности любого из испытуемых;

- если все ответы на тестовое задание неправильны, то уровень его трудности больше уровня подготовленности любого из испытуемых.

Далее находим приближенные значения поправочных коэффициентов X и Y

$$X = \sqrt{1 + \frac{\sigma_{\beta}^2}{2,9}} \approx \sqrt{\frac{1 + U / 2,9}{1 - UV / 8,35}}, \quad (20)$$

$$Y = \sqrt{1 + \frac{\sigma_{\theta}^2}{2,9}} \approx \sqrt{\frac{1 + V / 2,9}{1 - UV / 8,35}}, \quad (21)$$

где V и U – дисперсии по выборкам уровней подготовленности тестируемых и уровней сложности заданий

$$V = \frac{\sum_{i=1}^n (\theta_i^0)^2 - n \bar{\theta}^2}{n - 1}, \quad (22)$$

$$U = \frac{\sum_{j=1}^m (\beta_j^0)^2 - m \bar{\beta}^2}{m - 1}, \quad (23)$$

Нетрудно заметить, что константа $8,35 = 1,7^4$, что также связано с аппроксимацией модели Раша функцией распределения стандартной нормальной случайной величины.

Производим перерасчет полученных ранее начальных значений уровней подготовленности и уровней трудности заданий

$$\theta_i = \bar{\beta} + X \theta_i^0, \quad (24)$$

$$\beta_j = \bar{\theta} + Y \beta_j^0. \quad (25)$$

Для оценки стандартных ошибок измерения предлагаются формулы:

$$\sigma_{\theta_i} = X \sqrt{\frac{m}{R_i(m - R_i)}}, \quad (26)$$

$$\sigma_{\beta_j} = Y \sqrt{\frac{n}{s_j(n - s_j)}}. \quad (27)$$

Чем меньше стандартные ошибки, тем точнее результат измерения.

Имеются модификации алгоритма PROX для других моделей IRT, например, для модели с несколькими градациями оценок выполнения тестовых заданий¹⁷.

Очевидно, что результат оценивания по алгоритму PROX тем точнее, чем ближе к истине предположения о нормальности распределения уровня подготовленности ис-

¹⁷ Wright B.D., Masters G.N. Rating Scale Analysis: Rasch Measurement. Chicago: Mesa Press, 1982. – 204 p.

пытуемых и уровня трудности заданий.

Реализация PROX средствами электронной таблицы Microsoft Excel

Хотя оценивание по алгоритму PROX можно выполнить вручную, это слишком трудоёмко и ненадёжно. Гораздо проще выполнить расчёты в электронной таблице Microsoft Excel. Кроме того, при изменении матрицы ответов новые значения уровней подготовленности тестируемых и трудности заданий будут получены автоматически. Реализация PROX в Microsoft Excel позволяет многократно снизить трудоёмкость работы, перенести акцент с выполнения вычислений на анализ результатов.

Для примера взята задача подбора параметров модели Г.Раша из работы Б.Д.Райта и М.Х.Стоуна¹⁸ (табл.2). Исходные данные – это результаты выполнения 14 тестовых заданий 34 испытуемыми. Результаты по некоторым заданиям 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.

Таблица 2

Результаты выполнения теста

		Тестовые задания															R	θ
		4	5	7	6	9	8	10	11	13	12	14	15	16	17			
Тестируемые	25	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2	-3,8	
	4	1	0	1	0	1	0	0	0	0	0	0	0	0	0	3	-2,7	
	33	1	0	1	0	0	0	1	0	0	0	0	0	0	0	3	-2,7	
	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	4	-1,9	
	27	1	1	1	1	0	0	0	0	0	0	0	0	0	0	4	-1,9	
	11	0	1	1	1	1	1	0	0	0	0	0	0	0	0	5	-1,2	
	12	1	1	1	0	1	0	1	0	0	0	0	0	0	0	5	-1,2	
	17	1	0	1	1	1	1	1	0	0	0	0	0	0	0	6	-0,6	
	19	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	-0,6	
	30	1	1	1	1	1	1	0	0	0	0	0	0	0	0	6	-0,6	
	2	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0	
	3	1	1	1	1	1	1	1	0	0	1	0	0	0	0	8	0	
	5	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0	
	6	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0	
	8	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0	
	9	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0	
	13	1	1	0	0	1	1	1	1	0	1	0	0	0	0	7	0	
	16	1	1	1	1	1	1	0	1	0	0	0	0	0	0	7	0	
	26	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0	
	28	1	1	1	1	1	1	0	1	0	0	0	0	0	0	7	0	
29	1	1	0	1	1	0	1	1	0	0	1	0	0	0	7	0		
31	1	1	1	1	1	1	1	0	0	0	0	0	0	0	7	0		

¹⁸ Wright B.D., Stone M.H. Best Test Design. - Chicago: MESA PRESS. 1979.

	10	1	1	1	1	1	1	1	1	0	0	0	0	0	0	8	0,6
	18	1	1	1	1	1	1	1	0	1	0	0	0	0	0	8	0,6
	14	1	1	1	1	1	1	1	1	0	0	0	0	0	0	8	0,6
	32	1	1	1	1	1	1	1	1	0	0	0	0	0	0	8	0,6
	20	1	1	1	1	1	1	1	0	1	0	0	0	0	0	8	0,6
	21	1	1	1	1	1	1	1	1	1	0	0	0	0	0	9	1,2
	22	1	1	1	1	1	1	1	1	0	1	0	0	0	0	9	1,2
	23	1	1	1	1	1	1	1	0	1	0	1	0	0	0	9	1,2
	34	1	1	1	1	1	1	1	0	0	1	1	0	0	0	9	1,9
	15	1	1	1	1	1	1	1	1	1	1	0	0	0	0	10	1,9
	7	1	1	1	1	1	1	1	1	1	1	0	1	0	0	11	2,7
	24	1	1	1	1	1	1	1	1	1	0	0	0	1	1	11	2,7
	<i>S</i>	32	31	31	30	30	27	25	12	7	6	3	1	1	1		
	<i>β</i>	-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		

В той же работе Б.Д.Райта и М.Х.Стоуна опубликованы результаты подбора параметров θ и β модели Г.Раша по алгоритму PROX, которые также приведены в табл.2.

Опишем технологию реализации PROX в электронной таблице Microsoft Excel.

Сначала вводим данные табл.2 в клетки B2:O35 (рис.3).

	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

Рис.3. Ввод исходных данных (курсивом выделены номера тестируемых и заданий)

Находим сумму баллов каждого тестируемого R : 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.

По формулам (18) и (19) рассчитываем начальные значения параметров модели θ^0 и β^0 :

- для уровней подготовленности испытуемых в Q2 вводим =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).

По формулам (22) и (23) определяем V и U – дисперсии уровней подготовленности тестируемых и уровней трудности заданий:

- для дисперсии уровней подготовленности в любую свободную клетку (например, B40) вводим =ДИСП(Q2:Q35);
- для дисперсии трудности заданий в клетку B41 вводим =ДИСП(B37:O37).

Далее по формулам (20) и (21) находим значения коэффициентов X и Y :

- для коэффициента X в клетку B42 вводим **=КОРЕНЬ((1+B41/2,89)/(1-B40*B41/8,35));**
- для коэффициента Y в клетку B43 вводим **=КОРЕНЬ((1+B40/2,89)/(1-B40*B41/8,35)).**

По формулам (24) и (25) производим перерасчет полученных ранее значений уровней подготовки и уровней сложности заданий:

- для уровней подготовленности испытуемых в R2 вводим формулу **=СРЗНАЧ(\$B\$37:\$O\$37)+\$B\$42*Q2** и копируем в клетки R3:R35 (символом \$ обозначается абсолютная ссылка, которая не меняется при копировании формулы);
- для уровней трудности заданий в B38 вводим **=СРЗНАЧ(\$Q\$2:\$Q\$35)+\$B\$43*B37** и копируем в C38:O38.

Найдём стандартные ошибки значений уровней подготовки и уровней трудности заданий формулам (26) и (27):

- для ошибок уровней подготовленности испытуемых в S2 вводим формулу **=\$B\$42*КОРЕНЬ(14/(P2*(14-P2)))** и копируем в клетки S3:S35;
- для ошибок уровней трудности заданий в B39 вводим **=\$B\$43*КОРЕНЬ(34/(B36*(34-B36)))** и копируем в C39:O39.

Полученные результаты полностью соответствуют результатам обработки данных табл.2 по алгоритму PROX, которые получены Б.Д.Райтом и М.Х.Стоуном¹⁹, что подтверждает корректность описанной реализации PROX средствами Microsoft Excel.

Проверка нормальности распределения

Базовым предположением, которое делается при обработке результатов тестирования по алгоритму PROX, является гипотеза о том, что уровни подготовленности тестируемых и уровни трудности заданий подчиняются нормальному закону распределения. Такие допущения часто используются в статистическом анализе. Однако справедливость предположения о нормальности распределения рекомендуется проверять.

Один из критериев проверки нормальности распределения основан на анализе асимметрии A и эксцесса распределения E ²⁰:

$$A = \frac{m_3}{\sigma^3} = \frac{\sum (x_i - \bar{x})^3}{n \cdot \sigma^3},$$

¹⁹ Wright B.D., Stone M.H. Best Test Design. - Chicago: MESA PRESS. 1979.

²⁰ Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высшее Образование, 2008.

$$E = \frac{m_4}{\sigma^4} - 3 = \frac{\sum (x_i - \bar{x})^4}{n \cdot \sigma^4} - 3,$$

где $\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$, m_3 и m_4 – центральные эмпирические моменты третьего и четвёртого порядков.

Асимметрия является мерой несимметричности распределения данной величины. Асимметрия положительна, если правый конец графика плотности вероятностей распределения длиннее левого, и отрицательна в противном случае. Если распределение симметрично относительно математического ожидания, то его коэффициент асимметрии равен нулю.

Эксцесс распределения показывает остроту пика кривой по сравнению с нормальным законом: положительная величина означает более острый пик, отрицательная – менее. Для нормального распределения $A = E = 0$.

Значения коэффициента асимметрии и эксцесса распределения используют для оценки нормальности закона распределения через вспомогательные коэффициенты a_3 и a_4 , которые зависят только от количества значений и определяются по формулам²¹:

$$a_3 = \sqrt{D(A)} = \sqrt{\frac{6 \cdot (n-1)}{(n+1) \cdot (n+3)}},$$

$$a_4 = \sqrt{D(E)} = \sqrt{\frac{24 \cdot (n-2) \cdot (n-3)}{(n+1)^2 \cdot (n+3) \cdot (n+5)}},$$

где $D(A)$ и $D(E)$ – оценка дисперсии асимметрии и эксцесса распределения.

Считается, что случайная величина распределена по нормальному закону, если выполняются условия, являющиеся следствием из нормального закона распределения:

$$|A| \leq 3a_3,$$

$$|E| \leq 5a_4.$$

В случае выполнения обоих условий отличие A и E от нуля считается недостоверным, что позволяет принять гипотезу о нормальном распределении.

Проверка нормальности распределения для рассмотренного выше примера проводилась в электронной таблице Microsoft Excel, имеющей встроенные функции для нахождения асимметрии и эксцесса распределения – СКОС() и ЭКСЦЕСС(). Результаты расчета (табл.3) доказывают некорректность гипотезы о нормальном распределении.

²¹ Минько А.А. Статистический анализ в MS Excel. - М.: Вильямс, 2004. – 448 с.

Таблица 3

Проверяемая величина	Асимметрия A	Коэффициент a_3	Условие $ A \leq 3a_3$	Экцесс E	Коэффициент a_4	Условие $ E \leq 5a_4$	Нормальность распределения
Уровень подготовленности θ	-0,48	0,39	$0,48 \leq 1,2$ выполняется	0,78	0,17	$0,78 > 0,6$ не выполняется	не подтверждается
Уровень трудности β	0,24	0,55	$0,24 \leq 1,7$ выполняется	-1,7	0,21	$1,7 > 1,0$ не выполняется	не подтверждается

На рис.4 и 5 представлены гистограммы²² уровня подготовленности испытуемых и уровня трудности заданий.

Для большей наглядности на гистограммы пунктирными линиями нанесены графики функции нормального распределения, математическое ожидание и дисперсия которого соответствуют отображаемому гистограммой набору данных. Значительное расхождение графиков функции нормального распределения и гистограмм убедительно иллюстрирует несостоятельность гипотезы о нормальном распределении уровня подготовленности испытуемых и уровня трудности заданий.

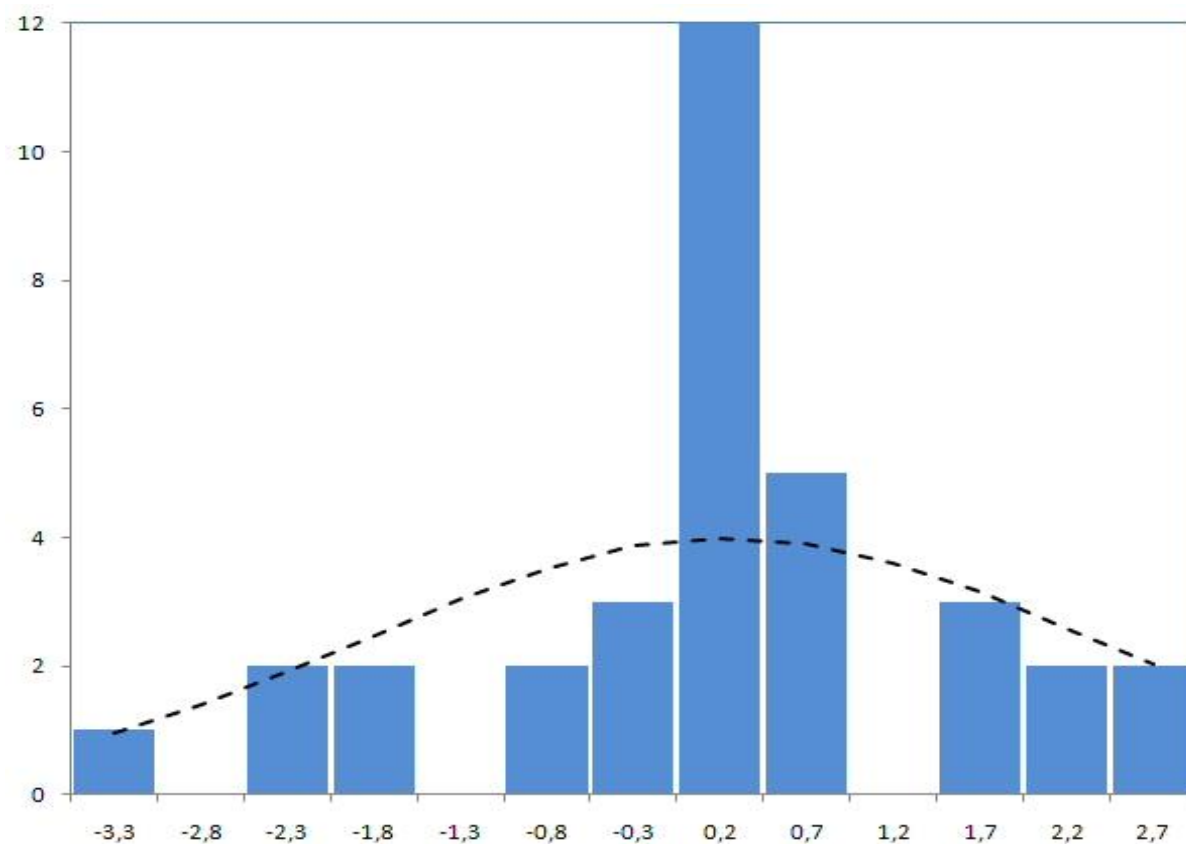


Рис.4. Гистограмма уровня подготовленности испытуемых

²² Гистограмма (от греч. histos – столб и gramma – черта) — графическое изображение частоты попадания элементов выборки в соответствующие интервалы данных.

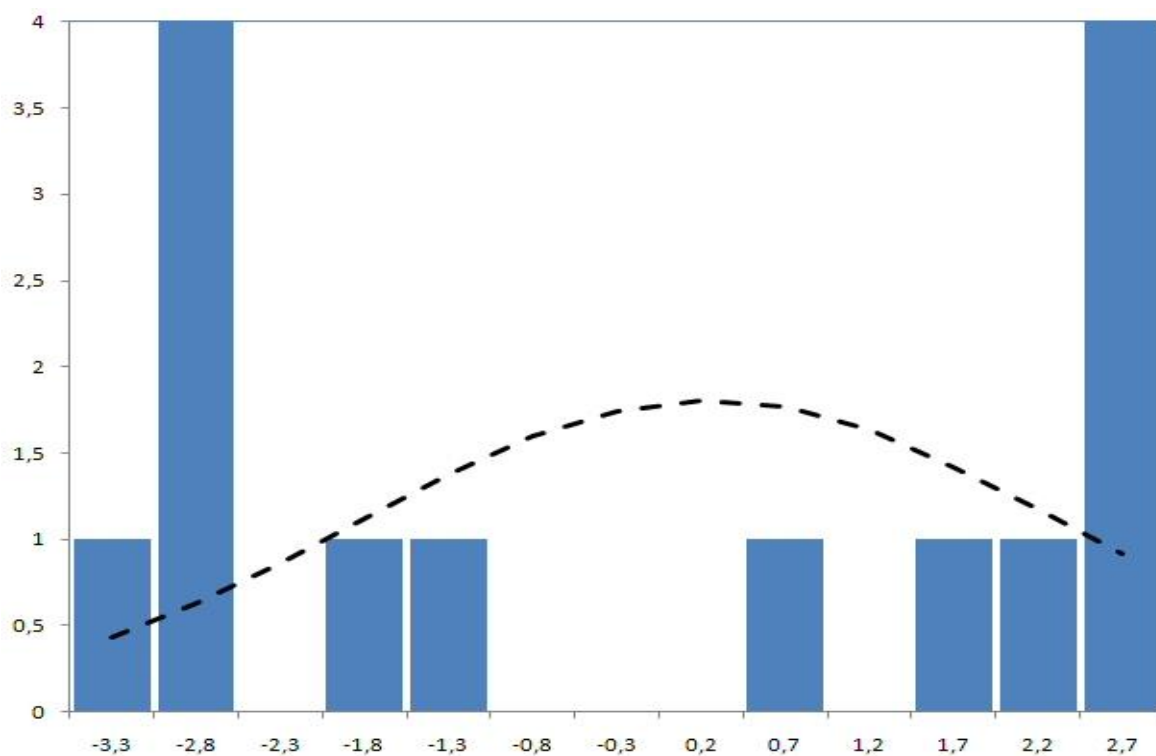


Рис.5. Гистограмма уровня трудности заданий

В рассмотренном примере не подтверждены базовые предположения PROX о нормальности распределения уровней подготовленности и уровней трудности, что:

- 1) ставит под сомнение правомерность использования алгоритма PROX для обработки данных этого примера;
- 2) увеличивает ошибку педагогического измерения.

Снижение точности измерений – это плата за упрощение решения. Сам автор алгоритма PROX Лесли Козн назвал свою работу «Приближенные выражения для оценки параметров модели Раша»²³, подчёркивая тем самым невысокую точность результатов контроля знаний.

Хотя упрощенное решение задачи определения уровня подготовленности тестируемого по процедуре PROX может быть найдено без использования компьютера, такой подход в настоящее время не рационален. Во-первых, исчезла необходимость упрощения решения – компьютеры доступны, и использование в решении моделей IRT не создает никаких проблем современным компьютерам. Во-вторых, снижается точность педагогических измерений – реальное распределение уровней подготовленности и уровней трудности всегда будет отличаться от идеального нормального распределения. В-третьих, ручной счет (даже с помощью калькулятора) слишком трудоёмок и ненаде-

²³ Cohen L. Approximate Expressions for Parameter Estimates in the Rasch Model, The British Journal of Mathematical and Statistical Psychology, 1979, №32, p.113-120.

жен – требуется провести большой объем вычислений, при котором ошибки, просчеты, неверная запись результатов практически неизбежны. Алгоритм PROX может найти применение только для предварительной оценки параметров модели Раша, например, для нахождения начальной точки численного поиска решения. Так, в программе Winsteps Д.М.Линека начальные значения параметров модели улучшаются с помощью алгоритма PROX, затем для поиска решения используется один из алгоритмов численного поиска – JMLE или XMLE²⁴.

Таким образом, на современном этапе развития педагогических измерений алгоритм PROX может рассматриваться исключительно как средство вспомогательного назначения для грубой оценки параметров модели Раша.

Автор полагает, что эффективное решение задачи определения уровня подготовленности должно основываться на непосредственном использовании выбранной математической модели и численных методов без дополнительных упрощений, ведущих к снижению точности педагогического измерения. Представляется важной задачей разработка адаптированной для практического применения в оценочной деятельности преподавателя компьютерной реализации моделей IRT, в первую очередь модели Г.Раша.

²⁴ Linacre J.M. User's Guide to WINSTEPS – Rasch-Model Computer Programs. Program Manual. – Chicago, 2007. 385p.