УДК 378:001.891

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ОЦЕНОК ПАРАМЕТРОВ МОДЕЛИ РАША, ПОЛУЧЕННЫХ МЕТОДАМИ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ И НАИМЕНЬШИХ КВАДРАТОВ

А.А. Маслак, С.И. Моисеев, С.А. Осипов

В имитационном эксперименте получены и исследованы статистические характеристики оценок параметров модели Раша, полученных методом максимального правдоподобия и методом наименьших квадратов. Проведен сравнительный анализ этих двух наборов оценок, который показал, что оценки параметров модели на основе метода наименьших квадратов ближе к сгенерированным данным. Рассмотрена специфика оценок, получаемых этими методами. Определено влияние пропусков данных на точность вычисления параметров дихотомической модели Раша.

Ключевые слова: латентная переменная, измерение, модель Раша, точность оценок, метод максимального правдоподобия, метод наименьших квадратов.

ВВЕДЕНИЕ

Специфика управления в социальных системах обусловлена тем, что многие переменные являются латентными (скрытыми). Примерами латентных переменных служат уровень жизни населения, качество образования, личностные качества и многие другие.

Непосредственно измерить латентную переменную нельзя, латентная переменная проявляется через так называемые индикаторные переменные. Например, латентная переменная «уровень жизни населения» согласно информации, ежегодно публикуемой Росстатом, характеризуется 36-ю разнородными индикаторными переменными, которые измеряются в рублях, процентах, квадратных метрах, килограммах и других единицах. Этот набор индикаторных переменных является операциональным определением латентной переменной. Примерами индикаторных переменных служат средняя зарплата в регионе, процент безработицы, площадь жилищ, потребление мяса, яиц, приходящихся в среднем на одного жителя и др. Фактически латентная переменная «уровень жизни населения» представляет собой системный показатель, который необходим для сравнения и мониторинга регионов, а также и формирования оптимальной политики управления социальной сферой в целом.

В рамках теории измерения латентных переменных на основе операционального определения осуществляется оценивание латентной переменной на интервальной шкале. Эта теория приобретает все большую популярность при исследовании социальных и экономических систем. Ядро теории измерения латентных переменных составляет семейство моделей Раша, в которых индикаторные переменные измеряются на шкале Лайкерта [1—4].

Однако многие латентные переменные определяются через непрерывные количественные индикаторные переменные [5, 6]. Примером служит рассмотренная выше латентная переменная «уровень жизни населения». Для того, чтобы воспользоваться теорией измерения латентных переменных, обычно применяется категоризация индикаторных переменных [6—8]. Однако при категоризации (сегментировании области значения переменных) часть информации теряется, что отмечено в работах [9—11]. Показано, что точность измерения латентной переменной существенно зависит от способа дискретизации количественной индикаторной переменной, в том числе от числа категорий [12—14]. В известных работах точность оценок рассматривается только в случае использования категоризованных индикаторных переменных [2, 15—18]. Поэтому представляет интерес исследование непосредственного использования непрерывных количественных индикаторных переменных в моделях Раша. Одним из методов, позволяющих выполнить такое исследование, является метод наименьших квадратов (МНК), который в настоящей работе выбран благодаря доступности соответствующего программного обеспечения.

1. ЦЕЛЬ ИССЛЕДОВАНИЯ

В большинстве пакетов для вычисления параметров модели Раша (CONQUEST, FACETS, JMETRIK, MINISTEP, PARSCALE, QUEST, RUMM, WINMIRA, WINSTEPS и др.) применяются разновидности метода максимального правдоподобия (ММП) и парных сравнений [6]. Наиболее известны диалоговые системы RUMM и WINSTEPS в различных модификациях [6]. Отличительная особенность этих пакетов заключается в том, что в случае адекватности данных модели измерения, сумма значений индикаторных переменных (тестовый балл) является достаточной статистикой. Это означает, что испытуемые, набравшие один и тот же тестовый балл (независимо от того, на какие задания они ответили правильно) получают одну и ту же оценку на шкале латентной переменной. Это принципиально важно для получения оценок параметров модели на линейной шкале.

В данной работе показано, что МНК позволяет описать результаты тестирования с большей точностью, чем ММП. Однако принципиальный недостаток МНК заключается в том, что испытуемые с одним и тем же тестовым баллом могут получить различные оценки уровня их подготовленности. Это означает, что при использовании МНК тестовый балл не является достаточной статистикой, что противоречит основам теории измерения латентных переменных.

Цель данной работы заключается в сравнительном анализе оценок параметров модели Раша, полученных методом максимального правдоподобия и методом наименьших квадратов. Для ее достижения необходимо решить следующие задачи:

— определить, в какой степени оценки, получаемые МНК, нарушают принцип достаточности тестового балла (испытуемые, имеющие один и тот же тестовый балл, должны иметь одну и ту же оценку на шкале латентной переменной);

— выполнить сравнительный анализ точности вычисления параметров модели Раша двумя методами;

— оценить влияние пропусков данных на точность вычисления параметров модели Раша двумя методами.

В качестве модели исследования выбрана дихотомическая модель Раша для измерения латентных переменных [3]. Акцент на точности измерения латентной переменной объясняется практической важностью результатов измерения [16—18].

В основе модели Раша лежит логистическая функция, позволяющая найти вероятность правильного ответа i-го испытуемого с уровнем подготовленности β_i на j-е задание с трудностью δ_j

$$p_{ij} = \frac{e^{\beta_i - \delta_j}}{1 + e^{\beta_i - \delta_j}}. \qquad (1)$$

В классической дихотомической модели Раша элементы x_{ij} матрицы данных принимают значения:

$x_{ij} = 1$, если i-й испытуемый правильно выполнил j-е задание;

$x_{ij} = 0$, если i-й испытуемый неправильно выполнил j-е задание.

Если i-й испытуемый не решал j-е задание, то значение индикаторной переменной «равно» пробелу.

2. ДАННЫЕ

Исследование проводилось на основе имитационного эксперимента, поскольку модель измерения нелинейная, и аналитические методы исследования неэффективны [6, 14]. Для решения поставленных задач использовалась следующая схема. Уровень подготовленности испытуемых (объекты) и трудность тестовых заданий (индикаторы) варьируется в одном и том же диапазоне от −4,0 до +4,0 логит. Число тестовых заданий выбрано равным 30, равномерно распределенных в диапазоне от −4,0 до +4,0 логит. Подготовленность испытуемых варьировалась на 20 уровнях в двукратной повторности (всего 40 уровней), равномерно распределенных в том же диапазоне. Двукратное повторение уровней подготовленности необходимо для оценки точности их измерения.

Данные эксперимента были сгенерированы в соответствии с дихотомической моделью Раша, т. е. данные априорно соответствовали модели измерения [6, 14].

Доля пропусков (пропущенных данных в матрице) была выбрана равной 10 %. Координаты элементов с пропущенными данными выбирались в матрице случайным образом.

Матрица сгенерированных данных без пропусков представлена в табл. 1.

Таблица 1

Матрица сгенерированных данных на основе дихотомической модели Раша

Данные	−4,00	−3,72	−3,45	−3,17	−2,90	−2,62	−2,35	−2,07	−1,79	−1,52	−1,24	−0,97	−0,69	−0,41	−0,14	0,14	0,41	0,69	0,97	1,24	1,52	1,79	2,07	2,35	2,62	2,90	3,17	3,45	3,72	4,00
−4,00	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−4,00	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−3,58	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−3,58	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−3,16	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−3,16	1	1	1	1	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−2,74	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−2,74	1	1	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−2,32	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
−2,32	1	1	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−1,90	1	0	1	1	1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
−1,90	1	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−1,47	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
−1,47	1	1	1	1	1	0	1	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−1,05	1	1	1	1	1	1	1	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−1,05	1	1	1	1	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
−0,63	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
−0,63	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
−0,21	1	1	1	1	1	0	1	1	1	0	0	1	1	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0
−0,21	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0,21	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	1	0	0	0	0	0
0,21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	0	1	0	0
0,63	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	0	0	0	1	0	0	0	0	0	0
0,63	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	1	0	1	0	0	0	0	0	0	0
1,05	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0
1,05	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0
1,47	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	1	1	0	0	1	0	0	0	0	0	0
1,47	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	0	0	1	0	0	0	0
1,90	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0
1,90	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0
2,32	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1
2,32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1	1	1	1	0	1	0	0	0
2,74	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
2,74	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	0	1
3,16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0
3,16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0
3,58	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0
3,58	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
4,00	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	0	1
4,00	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0

3. МЕТОД

Решение сформулированных в § 1 задач осуществлялось в рамках теории измерения латентных переменных. Для получения оценок латентной переменной использовалась диалоговая система ИЛП (измерение латентных переменных), разработанная в лаборатории объективных измерений Кубанского государственного университета [19, 20]. В этой системе оценки вычисляются на основе метода максимального правдоподобия. Оценки методом МНК получены в среде MS Excel [10, 11].

3.1. Оценка латентных переменных методом максимального правдоподобия

Для вычисления оценок параметров модели Раша наиболее часто применяют метод максимального правдоподобия, предложенный Р. Фишером. Этот метод реализован в диалоговой системе ИЛП.

Вероятностная модель выполнения M заданий для i-го испытуемого представляется в виде функции правдоподобия дискретной случайной величины $\vec{x}_i = \{x_{i1}, x_{i2}, x_{i3}, ..., x_{iM}\}$

$$L_i\{\vec{x}_i \mid \beta_i\} = \prod_{j=1}^{M} P_{ij}^{x_{ij}} Q_{ij}^{1-x_{ij}}, \qquad (2)$$

где P_{ij} — вероятность правильного выполнения i-м испытуемым j-го задания; Q_{ij} — вероятность неправильного выполнения i-м испытуемым j-го задания теста (2).

Все задания теста локально независимые, т. е. при данном уровне знаний ответ на каждое задание теста не зависит от результатов выполнения остальных его заданий. Значение β_i, при котором функция правдоподобия достигает максимума, принимают в качестве объективной оценки β_i и называют оценкой наибольшего правдоподобия. Так как функции L_i и $\ln L_i$ достигают максимума при одном и том же значении β_i, то удобно ввести в рассмотрение логарифмическую функцию правдоподобия

$$\ln L_i\{\vec{x}_i \mid \beta_i\} = \sum_{j=1}^{M} \{x_{ij}\ln P_{ij} + (1 - x_{ij})\ln Q_{ij}\}.$$

Аналогичная функция

$$\ln L_j\{\vec{x}_j \mid \delta_j\} = \sum_{i=1}^{N} \{x_{ij}\ln P_{ij} + (1 - x_{ij})\ln Q_{ij}\}$$

составляется для вычисления оценок наибольшего правдоподобия параметра δ_j, $j = 1, 2, , M$.

Система уравнений для определения оценок уровня знаний в группе из N испытуемых и уровня трудности M заданий теста имеет вид:

$$\begin{cases} \dfrac{\partial \ln L_i\{\vec{x}_i|\beta_i\}}{\partial \beta_i} = 0, \\ \dfrac{\partial \ln L_j\{\vec{x}_j|\delta_j\}}{\partial \delta_j} = 0, \end{cases}$$

где $i = 1, 2, ..., N$, $j = 1, 2, ..., M$.

3.2. Оценка латентных переменных методом наименьших квадратов

При использовании МНК параметры β_i и δ_j модели измерения латентных переменных выбираются так, чтобы сумма квадратов отклонений вероятностей, полученных на основе модели Раша (1), от эмпирических данных x_{ij} была минимальной.

Таким образом, задача оценивания параметров модели сводится к минимизации суммы квадратов отклонений:

$$S(\beta_i, \delta_j) = \sum_{i=1}^{N}\sum_{j=1}^{M} (p_{ij} - x_{ij})^2 = $$
$$= \sum_{i=1}^{N}\sum_{j=1}^{M} \left(\frac{e^{\beta_i - \delta_j}}{1 + e^{\beta_i - \delta_j}} - x_{ij}\right)^2 \to \min.$$

Основное преимущество МНК состоит в том, что вместо дихотомических данных x_{ij} можно использовать нечеткое множество x_{ij}, значения элементов которого варьируются непрерывно от 0 до 1. Это позволит применять модель для измерения латентных переменных, индикаторные переменные которых оцениваются на непрерывных количественных шкалах.

Еще одно немаловажное преимущество этой модели заключается в том, что предлагаемый метод оценивания параметров модели значительно расширяет инструментальные возможности для получения оценок латентной переменной. Так, при применении метода максимального правдоподобия необходимо специализированное программное обеспечение (ИЛП, RUMM, WINSTEP), которого нет в свободном доступе [21, 22]. Применение же МНК представляет собой классическую задачу нелинейного программирования. Эту задачу можно решить в среде MS Excel, более доступной, чем специализированное программное обеспечение.

4. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Оценки латентной переменной, полученные с помощью ММП и МНК, приведены в табл. 2, где β_i^* — уровень подготовленности i-го испытуемого, который использовался при генерировании матрицы результатов тестирования; $\hat{\beta}_i$ — оценка уровня подготовленности i-го испытуемого; Δ_i — абсолютное отклонение оценки уровня подготовленности от истинного значения, используемого при генерировании; SE — среднеквадратическая ошибка вычисления оценки уровня подготовленности испытуемого.

Таблица 2

Оценки латентной переменной методами МП и МНК без пропусков и с пропусками данных

№ п/п	β_i^*	Тестовый балл	ММП (без пропусков)			МНК (без пропусков)			ММП (с пропусками)			МНК (с пропусками)		
			$\hat{\beta}_i$	Δ_i	SE	$\hat{\beta}_i$	Δ_i	SE	$\hat{\beta}_i$	Δ_i	SE	$\hat{\beta}_i$	Δ_i	SE
1	−4,00	4	−3,77	0,23	0,70	−3,64	0,36	0,68	−4,02	0,02	0,74	−2,35	1,65	0,59
2	−4,00	3	−4,27	0,27	0,76	−3,80	0,20	0,69	−4,55	0,55	0,79	−2,61	1,39	0,60
3	−3,58	5	−3,31	0,27	0,67	−2,92	0,66	0,63	−3,43	0,15	0,70	−2,22	1,36	0,58
4	−3,58	3	−4,27	0,69	0,76	−3,67	0,09	0,68	−5,08	1,50	0,84	−2,86	0,71	0,62
5	−3,16	5	−3,31	0,16	0,67	−2,87	0,29	0,63	−4,08	0,92	0,75	−2,61	0,55	0,60
6	−3,16	8	−2,13	1,03	0,60	−1,98	1,18	0,58	−1,87	1,29	0,61	−0,80	2,36	0,53
7	−2,74	7	−2,50	0,24	0,62	−2,26	0,48	0,60	−2,39	0,35	0,64	−1,70	1,03	0,56
8	−2,74	6	−2,89	0,16	0,64	−2,46	0,27	0,61	−3,37	0,63	0,70	−2,09	0,65	0,57
9	−2,32	8	−2,13	0,18	0,60	−2,16	0,16	0,59	−2,13	0,19	0,62	−1,45	0,87	0,55
10	−2,32	6	−2,89	0,58	0,64	−2,54	0,22	0,61	−3,02	0,70	0,68	−2,22	0,10	0,58
11	−1,90	8	−2,13	0,24	0,60	−2,18	0,29	0,59	−1,79	0,11	0,60	−1,32	0,58	0,54
12	−1,90	5	−3,31	1,42	0,67	−3,12	1,23	0,64	−3,50	1,60	0,71	−2,09	0,20	0,57
13	−1,47	9	−1,78	0,31	0,59	−1,99	0,51	0,59	−2,13	0,66	0,62	−1,32	0,16	0,54
14	−1,47	10	−1,45	0,03	0,58	−1,19	0,28	0,55	−1,09	0,39	0,57	−0,80	0,67	0,53
15	−1,05	11	−1,13	0,07	0,56	−0,99	0,06	0,55	−1,12	0,07	0,57	−0,80	0,25	0,53
16	−1,05	8	−2,13	1,08	0,60	−1,77	0,71	0,58	−2,42	1,37	0,64	−1,83	0,78	0,56
17	−0,63	15	0,07	0,70	0,54	0,07	0,70	0,53	0,10	0,73	0,55	0,10	0,74	0,52
18	−0,63	14	−0,22	0,41	0,54	−0,09	0,54	0,53	0,36	0,99	0,55	0,49	1,12	0,53
19	−0,21	14	−0,22	0,01	0,54	−0,30	0,09	0,53	−0,19	0,02	0,55	0,10	0,31	0,52
20	−0,21	16	0,36	0,57	0,54	0,37	0,58	0,53	0,17	0,38	0,55	−0,03	0,19	0,52
21	0,21	17	0,65	0,43	0,54	0,98	0,77	0,53	0,82	0,61	0,55	0,49	0,28	0,53
22	0,21	19	1,23	1,02	0,55	1,02	0,81	0,53	1,28	1,06	0,56	0,75	0,54	0,53
23	0,63	18	0,94	0,30	0,54	1,00	0,37	0,53	0,95	0,32	0,55	0,75	0,12	0,53
24	0,63	18	0,94	0,30	0,54	1,03	0,39	0,53	1,19	0,56	0,56	0,88	0,25	0,54
25	1,05	19	1,23	0,18	0,55	1,22	0,16	0,54	1,16	0,11	0,56	0,62	0,43	0,53
26	1,05	18	0,94	0,12	0,54	1,02	0,03	0,53	1,48	0,43	0,56	0,75	0,30	0,53
27	1,47	19	1,23	0,24	0,55	1,13	0,35	0,53	1,65	0,18	0,57	1,14	0,34	0,54
28	1,47	20	1,53	0,06	0,56	1,36	0,12	0,54	1,63	0,15	0,57	1,52	0,05	0,55
29	1,90	20	1,53	0,36	0,56	1,37	0,52	0,54	1,57	0,33	0,57	1,14	0,76	0,54
30	1,90	20	1,53	0,36	0,56	1,27	0,63	0,54	1,61	0,28	0,57	1,26	0,63	0,54
31	2,32	26	3,69	1,37	0,68	4,09	1,78	0,73	3,16	0,84	0,64	2,43	0,11	0,58
32	2,32	22	2,17	0,15	0,58	2,13	0,18	0,56	2,49	0,17	0,61	1,91	0,41	0,56
33	2,74	25	3,26	0,53	0,65	2,84	0,11	0,60	3,41	0,67	0,66	2,43	0,31	0,58
34	2,74	25	3,26	0,53	0,65	3,17	0,43	0,62	3,85	1,11	0,69	2,55	0,18	0,59
35	3,16	28	4,69	1,53	0,84	4,09	0,94	0,73	4,63	1,48	0,78	3,07	0,09	0,62
36	3,16	26	3,69	0,53	0,68	3,76	0,61	0,68	3,53	0,38	0,67	2,43	0,73	0,58
37	3,58	27	4,15	0,57	0,74	3,55	0,03	0,66	4,38	0,80	0,75	3,07	0,51	0,62
38	3,58	26	3,69	0,11	0,68	3,15	0,43	0,62	3,88	0,30	0,70	2,81	0,77	0,60
39	4,00	26	3,69	0,32	0,68	3,14	0,86	0,62	3,88	0,12	0,70	2,68	1,32	0,60
40	4,00	28	4,69	0,69	0,84	4,03	0,03	0,72	4,90	0,90	0,82	2,94	1,06	0,61

Таблица 3

Коэффициенты корреляции тестового балла с оценками ММП и МНК без пропусков и с пропусками данных

	Тестовый балл	ММП (без пропусков)	ММП (с пропусками)	МНК (без пропусков)	МНК (с пропусками)
Тестовый балл	1,000	0,998	0,989	0,996	0,994
ММП (без пропусков)	0,998	1,000	0,995	0,996	0,992
ММП (с пропусками)	0,989	0,995	1,000	0,989	0,994
МНК (без пропусков)	0,996	0,996	0,989	1,000	0,987
МНК (с пропусками)	0,994	0,992	0,994	0,987	1,000

4.1. Практическая оценка достаточности тестового балла по МНК

Достаточность тестового балла с практической точки зрения определяется на основе корреляции тестового балла с оценками латентной переменной. Результаты корреляционного анализа приведены в табл. 3.

Коэффициенты корреляции показывают, что все виды оценок уровня подготовленности (ММП и МНК, без пропусков и при наличии пропусков) достаточно хорошо коррелируют с тестовым баллом. Таким образом, полученные результаты свидетельствуют о том, что для получения оценок латентной переменной можно применять МНК. Однако необходимо рассмотреть и точность получаемых оценок.

4.2. Сравнение абсолютной погрешности вычисления оценок латентной переменной, полученных по ММП и МНК

Абсолютная погрешность вычисления оценок латентной переменной для i-го испытуемого определяется на основе их близости к соответствующим значениям в имитационном эксперименте: $\Delta_i = |\hat{\beta}_i - \beta_i^*|$, $\hat{\beta}_i$ — оценка латентной переменной для i-го испытуемого; β_i^* — истинное значение латентной переменной для i-го испытуемого, которое использовалось для генерирования матрицы тестирования.

Для определения близости оценок параметров модели, полученных обоими методами, к истинным значениям проведен трехфакторный дисперсионный анализ (табл. 4). Анализировалось влияние факторов:

— фактор A — метод вычисления оценок, варьируется на двух уровнях (ММП и МНК);

— фактор B — варьируется на двух уровнях — отсутствие или наличие (10 %) пропусков в матрице данных;

— фактор C — число уровней, на которых варьируется латентная переменная (в данном эксперименте 20).

Проинтерпретируем результаты, приведенные в табл. 4.

Фактор A незначим — в среднем по всем уровням остальных факторов нет статистически значимого различия по абсолютной ошибке вычисления оценок латентной переменной между обоими методами; эмпирический уровень значимости $p = 0{,}780$, что значительно превышает номинальный уровень значимости, равный 0,05. Средние значения абсолютных ошибок по обоим методам приведены в табл. 5.

Таблица 4

Результаты дисперсионного анализа абсолютной погрешности оценок уровня подготовленности испытуемых

Источник дисперсии	Сумма квадратов	Степени свободы	Средний квадрат	F-статистика	p
A	0,023	1	0,023	0,079	0,780
B	0,025	1	0,025	0,086	0,771
C	938,880	19	49,415	167,319	Менее 0,001
AB	0,003	1	0,003	0,011	0,918
AC	14,862	19	0,782	2,649	0,001
BC	3,304	19	0,174	0,589	0,904
ABC	5,689	19	0,299	1,014	0,455
Ошибка	23,627	80	0,295	—	—
Всего	986,413	159	—	—	—

Таблица 5

Абсолютные ошибки измерений

Метод измерения	Среднее значение	Объем выборки	Стандартная ошибка	95%-й доверительный интервал	
				Нижняя граница	Верхняя граница
ММП	0,140	80	0,066	0,010	0,269
МНК	0,164	80	0,066	0,034	0,293

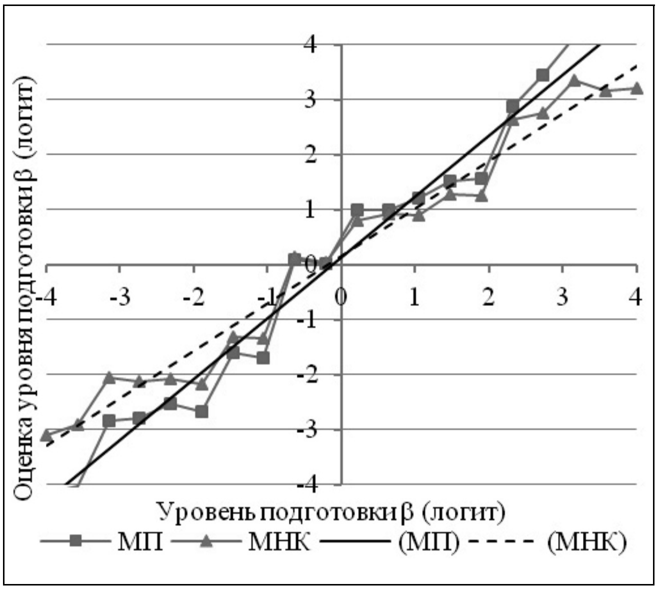

Рис. 1. Эффект взаимодействия факторов *A* и *C*

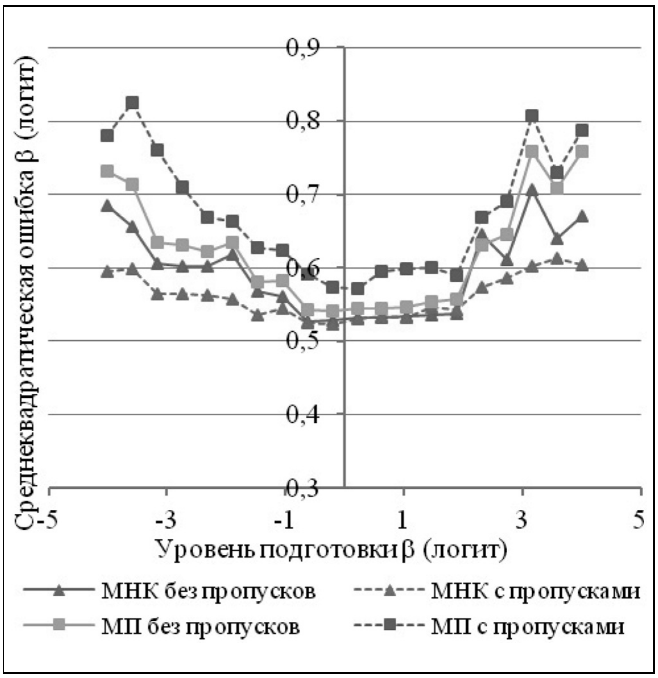

Рис. 2. Среднеквадратическая ошибка измерения (*SE*) уровня подготовленности испытуемых в зависимости от метода получения оценок параметров модели Раша и наличия пропусков

Фактор *B* незначим, это свидетельствует о том, что в среднем по всем уровням остальных факторов наличие пропусков, в объеме 10 % от общего числа данных, статистически значимо не влияет на точность вычисления оценок ($p = 0,771$). Абсолютные ошибки измерения в зависимости от наличия пропусков в среднем для уровней всех остальных факторов приведены в табл. 6.

Фактор *C* значим ($p < 0,001$). Это означает, что абсолютная ошибка измерения латентной переменной в среднем для всех уровней факторов *A* и *B* зависит от местоположения латентной переменной. Поскольку есть значимые взаимодействия факторов, информации о значимости факторов *A* и *C* недостаточно. Необходимо рассмотреть взаимодействие факторов.

Из всех взаимодействий значимо только взаимодействие факторов *A* и *C* (на очень высоком уровне значимости $p < 0,001$. Это означает, что абсолютная ошибка измерения латентной переменной зависит как от ее местоположения на шкале, так и от применяемого метода. Этот эффект иллюстрируется на рис. 1, где по оси абсцисс отложены моделируемые значения уровня подготовленности испытуемых, а по оси ординат полученные оценки этих уровней. Для большей наглядности на этом рисунке представлены оценки точности вычисления латентной переменной не только в зависимости от ее местоположения на шкале, но и от метода вычисления оценок и наличия пропусков данных. Как видно из рис. 1, по краям шкалы отклонения оценок от истинных (моделируемых) значений больше, чем в середине шкалы. Кроме того, отклонения оценок, полученных по ММП и МНК различны, об этом и свидетельствует взаимодействие факторов.

Поскольку фактор *A* является незначимым, то различием в разности оценок по ММП и МНК можно пренебречь.

Важно сравнение рассматриваемых методов по точности вычисления самих оценок уровня подготовленности испытуемых.

Таблица 6

Абсолютные ошибки измерений в зависимости от наличия пропусков

Наличие/отсутствие пропусков	Среднее значение	Объем выборки	Стандартная ошибка	95%-й доверительный интервал	
				Нижняя граница	Верхняя граница
Без пропусков	0,139	80	0,061	0,018	0,260
С пропусками	0,164	80	0,061	0,043	0,285

4.3. Сравнение оценок латентной переменной по точности их вычисления

Точность вычисления оценок важна при практическом применении результатов измерения. Точность оценок латентной переменной для дихотомической модели Раша вычисляется как

$$SE_i = 1/\sqrt{\sum_{j=1}^{M} P_{ij}(1 - P_{ij})},$$

где P_{ij} — вероятность правильного выполнения i-м испытуемым j-го задания.

В целях наглядности на рис. 2 представлены оценки точности вычисления уровня подготовленности испытуемых по обоим методам при наличии и отсутствии пропусков.

Для анализа точности вычисления оценок (SE) обоими методами проведен дисперсионный анализ на основе полной модели трехфакторного эксперимента (табл. 7). Анализируется влияние тех же факторов, что и в предыдущем случае, однако здесь фактор A значимый ($p < 0{,}001$). Это означает, что точность вычисления оценок зависит от метода (табл. 8). Как и следовало ожидать, в среднем по уровням всех остальных факторов точность вычисления оценок по МНК меньше, чем по ММП.

Фактор C также значимый ($p < 0{,}001$). Это хорошо видно на рис. 2, где среднеквадратическая ошибка оценки существенно зависит от значения латентной переменной. Среднеквадратическая ошибка наименьшая в середине интервала варьирования и увеличивается по мере удаления от него.

Таблица 7

Результаты дисперсионного анализа точности вычисления оценок уровня подготовленности испытуемых

Источник дисперсии	Сумма квадратов	Степени свободы	Средний квадрат	F-статистика	p
A	0,195	1	0,195	102,834	Менее 0,001
B	0,003	1	0,003	1,458	0,231
C	0,557	19	0,029	15,482	Менее 0,001
AB	0,069	1	0,069	36,434	Менее 0,001
AC	0,055	19	0,003	1,539	0,094
BC	0,015	19	0,001	0,412	0,984
ABC	0,020	19	0,001	0,559	0,924
Ошибка	0,152	80	0,002	—	—
Всего	1,066	159	—	—	—

Таблица 8

Абсолютные ошибки измерений

Метод измерения	Среднее значение	Объем выборки	Стандартная ошибка	95%-й доверительный интервал	
				Нижняя граница	Верхняя граница
ММП	0,648	80	0,005	0,639	0,658
МНК	0,578	80	0,005	0,569	0,588

Таблица 9

Абсолютные ошибки измерений в зависимости от наличия пропусков

Наличие/отсутствие пропусков	Среднее значение	Объем выборки	Стандартная ошибка	95%-й доверительный интервал	
				Нижняя граница	Верхняя граница
Без пропусков	0,609	80	0,005	0,600	0,619
С пропусками	0,617	80	0,005	0,608	0,627

Фактор B незначимый ($p < 0{,}231$), но для наглядности в табл. 9 приведены среднеквадратические ошибки без пропусков и при наличии пропусков данных в среднем для всех уровней остальных факторов. Видно, что при наличии пропусков среднеквадратическая ошибка (SE) несколько больше.

Представляет интерес эффект значимого взаимодействия факторов A и B, т. е. зависимости среднеквадратической ошибки (SE) от метода вычисления оценок и наличия пропущенных данных ($p < 0{,}001$). Он объясняется тем, что среднеквадратическая ошибка вычисления оценок по МНК всегда меньше, чем по ММП. Это вызвано тем, что при использовании МНК нет ограничения на то, что испытуемые с одинаковым тестовым баллом должны получить одинаковые оценки. Поэтому оценки получаются более точными.

5. ОБСУЖДЕНИЕ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Проведенное исследование показывает, что для вычисления оценок параметров дихотомической модели Раша можно применять МНК. Для выбранной схемы эксперимента различие между оценками, полученными двумя методами (ММП и МНК) практически несущественно.

Дальнейшее исследование применимости МНК целесообразно вести по двум направлениям:

— определить границы применимости МНК в зависимости об числа испытуемых, числа индикаторов (тестовых заданий), объема пропущенных данных и ограничений на оценки МНК;

— второе, наверное, наиболее важное направление, это сравнение оценок, полученных по ММП и МНК, для политомических данных. Это позволит определить возможность получения достаточно точных оценок параметров модели Раша для непрерывных количественных индикаторов.

ЗАКЛЮЧЕНИЕ

В рамках имитационного моделирования сгенерирована матрица данных на основе дихотомической модели Раша, по которым вычислены параметры модели Раша методом максимального правдоподобия и методом наименьших квадратов.

Рассмотрена специфика оценок, получаемых этими методами. Как и следовало ожидать, оценки параметров модели на основе МНК ближе к сгенерированным данным, поскольку нет ограничения на то, что оценки испытуемых должны быть одни и те же в случае одинакового тестового балла.

Определено влияние пропуска данных в сгенерированных данных на точность вычисления оценок параметров модели Раша.

ЛИТЕРАТУРА

1. *Rasch* Models. Foundations, Recent Developments and Applications / Eds. G.H. Fischer, I.W. Molenaar. — N.-Y.: Springer, 1997.
2. *Wilson M.* Constructing measures: an item response modeling approach. — Mahwah, New Jersey: Lawrence Erlbaum associates, 2005. — 228 p.
3. *Летова Л.В., Маслак А.А., Осипов С.А.* Семейство моделей Раша для объективного измерения латентных переменных // Информатизация образования и науки. — 2013. — № 4. — С. 131—141.
4. *Маслак А.А., Поздняков С.А.* Модель Раша для проверки качества метода измерения толерантности // Социология: методология, методы, математическое моделирование. — 2008. — Т. 26. — С. 87—104.
5. *Маслак А.А., Поздняков С.А.* Методика измерения и мониторинга уровня жизни населения в субъектах Южного федерального округа Российской Федерации // Вестник Воронежского гос. техн. ун-та. — 2008. — Т. 4. — № 10. — С. 159—171.
6. *Маслак А.А.* Измерение латентных переменных в социальных системах. — Славянск-на-Кубани: Изд. центр КубГУ, 2012. — 432 с.
7. *Maslak A.A., Anisimova T.S., Osipov S.A., Karabatsos G.* Measuring and Comparing Higher Education Quality Between Countries Worldwide // Journal of Applied Measurement. — 2005. — Vol. 6, N 4. —P. 432—442.
8. *Маслак А.А., Поздняков С.А., Данилов А.А.* Измерение уровня развития инфраструктуры сферы образования в субъектах РФ // Высшее образование в России. — 2008. — № 2. — С. 102—108.
9. *Linacre J.M.* Understanding Rasch measurement: estimation methods for Rasch Measures // Journal of Outcome Measurement. — 1999 N 3. — P. 381—405.
10. *Баркалов С.А., Моисеев С.И., Кочерга Н.С., Соловьева Е.В.* Математические модели подготовки и проверки качества освоения компетенций в образовательном процессе // Открытое образование. — 2014. — № 2. — С. 9—16.
11. *Баркалов С.А., Моисеев С.И., Соловьева Е.В.* Применение метода наименьших квадратов при оценке латентных переменных методом Раша // Научный вестник Воронежского ГАСУ / Сер. Управление строительством. — 2014. — № 1 (6). — С. 98—100.
12. *Данилов А.А., Маслак А.А.* Исследование точности измерения латентной переменной в зависимости от числа градаций индикаторных переменных // Вестник Воронежского гос. техн. ун-та. — 2009. — Т. 5, № 11. — С. 106—114.
13. *Maslak A.A., Osipov S.A., Goncharova T.N.* Investigation of measurement precision of latent variables in education // Образование и наука. — 2014. — № 7 (116). — С. 36—47.
14. *Летова Л.В.* Точность измерения латентных переменных // Дистанционное и виртуальное обучение. — 2013. — № 12. — С. 75—88.
15. *Chen S., Hsiao C., Wang L.* Measurement Errors and Censored Structural Latent Variables Models // Econometric Theory. — 2012. — N 28 (03). — P. 696—703.
16. *Елисеев И.Н.* Исследование погрешности расчета трудности заданий теста на основе моделирования дихотомической матрицы ответов // Педагогическая информатика. — 2011. — № 4. — С. 92—101.
17. *Ефремов Н.Ф.* Формирование и оценивание компетенций в образовании: монография. — Ростов-на-Дону: Аркол, 2010. — 386 с.
18. *Звонников В.И., Челышкова М.Б.* Современные средства оценивания результатов обучения: учебное пособие.— М.: Изд. центр «Академия», 2007. — 223 с.
19. *Осипов С.А.* Вычисление оценок модели Раша методом наибольшего правдоподобия // Оценка эффективности образовательных инноваций и технологий: Тезисы докладов и выступлений Третьей Всерос. науч.-практ. конф. (27—28 июня 2001 г). Славянск-на-Кубани, 2001. — С. 26—30.
20. *Маслак А.А., Осипов С.А.* Измерение латентных переменных / Свидетельство о гос. регистрации программ для ЭВМ № 2013618487. Дата гос. регистрации в Реестре программ для ЭВМ 10 сентября 2013 г.
21. *Getting* Started RUMM 2010. Rasch Unidimensional Measurement Models. — Perth: RUMM Laboratory Ltd, 2001. — 87 p.
22. *Smith R.M.* Rasch Measurement Models: Interpreting WINSTEPS/BIGSTEPS and Facets Output. — Gainesville, Florida: JAM Press, 1999. — 68 p.

Статья представлена к публикации членом редколлегии А.С. Манделем.

Маслак Анатолий Андреевич — д-р техн. наук, зав. лабораторией, филиал Кубанского гос. ун-та, г. Славянск-на-Кубани,
✉ anatoliy_maslak@mail.ru,

Моисеев Сергей Игоревич — канд. физ.-мат. наук, доцент, Институт менеджмента, маркетинга и финансов, г. Воронеж,
✉ mail@moiseevs.ru,

Осипов Сергей Александрович — канд. техн. наук, доцент, филиал Кубанского гос. ун-та, г. Славянск-на-Кубани,
✉ slnosa@mail.ru.