УДК 378

В. В. Братищенко*

ПОЛИТОМИЧЕСКАЯ БИНОМИАЛЬНАЯ МОДЕЛЬ ОЦЕНОК

В работе предлагается модель педагогических измерений, аналогичная модели Раша. Предложенную модель можно применять, если педагогические измерения имеют больше чем две градации и формируются в виде суммарной оценки знаний. Приведены уравнения для поиска параметров модели. Предложены оценки точности определения параметров, полученные с помощью неравенства Рао-Крамера. Проверка адекватности модели выполнена на основе дисперсионного анализа и изучения коэффициентов корреляции. Кроме этого, применялись традиционные для модели Раша статистика согласия и взвешенная общая статистика согласия. Обсуждаются результаты применения модели к экзаменационным оценкам студентов. Проведенные вычисления демонстрируют согласие наблюдений с моделью с высоким уровнем доверительной вероятности.

Ключевые слова: Политомическая биномиальная модель, модель Раша-Андрича, оценки латентных параметров, адекватность, дисперсия оценок параметров, неравенство Рао-Крамера, статистика согласия, взвешенная общая статистика согласия.

Для выдающихся идей характерным является порождение множества новых моделей. Так произошло с моделью Георга Раша (Georg Rasch) [1]. Модель Раша применяется для дихотомических тестов, в которых возможны два варианта ответа: «правильно» — 1 или «неправильно» — 0. Вероятность

$$P_{ij} = \frac{\exp(\theta_i - \delta_j)}{1 + \exp(\theta_i - \delta_j)}$$

правильного ответа i-го студента на j-е задание зависит от подготовленности θ_i тестируемого и трудности δ_j задания.

Достаточно часто ответ на задание имеет не две, а большее количество категорий, из которых выбирается одна. В таких случаях применяют политомические модели. Одна из таких моделей была предложена Андричем [2]. В этой модели вероятность P_{ijk} того, что i-й студент в j-ом задании выберет k-ю категорию ответа, зависит от подготовленности θ_i тестируемого, трудности δ_j

* Братищенко Владимир Владимирович — кандидат физико-математических наук, доцент, заведующий кафедрой информатики и кибернетики, Байкальский государственный университет экономики и права, г. Иркутск, vbrat56@mail.ru.

задания и уровня ϕ_k трудности категории – порога Раша-Андрича. Для задания вероятностей используют соотношения $\ln\left(P_{ijk} / P_{ij(k-1)}\right) = \theta_i - \delta_j - \phi_k$. При возрастающих значениях ϕ_k уровней трудности категорий выбор $k+1$-й категории труднее, чем выбор k-й категории.

В модели Раша-Андрича количество параметров увеличивается на количество категорий без единицы, т.е. предполагается, что все k-е категории разных заданий имеют одинаковую трудность ϕ_k. Для надежной оценки параметров модели необходимо увеличивать объем наблюдений (оценок).

В данной работе предлагается биномиальная политомическая модель Раша для обработки результатов измерений, формирование которых аналогично сумме дихотомических измерений, и которые принимают значения 0, 1, …., l. Для описания вероятностей оценки X_{ij} i-го студента ($i = 1,...,n$) на j-ом задании ($j = 1,..., m$) предлагается использовать биномиальное распределение

$$P\left\{X_{ij} = k\right\} = C_l^k p_{ij}{}^k q_{ij}{}^{l-k},\ k \in \left\{0,1,...,l\right\}, \tag{1}$$

$$p_{ij} = \frac{e^{\theta_i}}{e^{\theta_i} + e^{\delta_j}},$$

$$q_{ij} = 1 - p_{ij} = \frac{e^{\delta_j}}{e^{\theta_i} + e^{\delta_j}},$$

где как и в модели Раша, θ_i — параметр подготовленности i-го студента, δ_j — параметр трудности j-го задания. Случайные величины $\left\{X_{ij}\right\}$ будем полагать независимыми в совокупности.

Количество параметров в биномиальной политомической модели меньше, чем в модели Раша-Андрича. Это означает, что точность оценивания параметров биномиальной модели (при условии ее адекватности) будет выше при одном и том же объеме выборки.

Для определения параметров $\theta_1,...,\theta_n$ и $\delta_1,...,\delta_m$ по известным оценкам $\left\|x_{ij}\right\|$ можно использовать метод моментов. Приравнивая суммы оценок соответствующим математическим ожиданиям, получаем систему уравнений

$$\begin{aligned} b_i &= \sum_{j=1}^{m} x_{ij} = l\sum_{j=1}^{m} p_{ij}, \quad i = 1,...,n, \\ c_j &= \sum_{i=1}^{n} x_{ij} = l\sum_{i=1}^{n} p_{ij}, \quad j = 1,...,m. \end{aligned} \tag{2}$$

Аналогичные уравнения дает и метод максимального правдоподобия. Логарифм функции правдоподобия

$$\ln(L) = \ln\left(\prod_{i,j=1}^{n,m} C_l^{x_{ij}} \left(\frac{e^{\theta_i}}{e^{\theta_i} + e^{\delta_j}} \right)^{x_{ij}} \left(\frac{e^{\delta_j}}{e^{\theta_i} + e^{\delta_j}} \right)^{l-x_{ij}} \right) =$$

$$= \sum_{i,j=1}^{n,m} \ln\left(C_l^{x_{ij}}\right) + \left(l - x_{ij}\right)\left(\delta_j - \theta_i\right) - l\ln\left(1 + \exp\left(\delta_j - \theta_i\right)\right)$$

достигает максимума при условиях

$$\frac{\partial \ln(L)}{\partial \theta_i} = \sum_{j=1}^{m} x_{ij} - l\frac{e^{\theta_i}}{e^{\theta_i} + e^{\delta_j}} = 0\,,$$

$$\frac{\partial \ln(L)}{\partial \delta_j} = \sum_{i=1}^{n} l\frac{e^{\theta_i}}{e^{\theta_i} + e^{\delta_j}} - x_{ij} = 0\,,$$

которые совпадают с системой (2). Решение данной системы находится численными методами.

Предложенная модель применялась для обработки экзаменационных оценок студентов. Значения оценок являются традиционными для четырехбалльной шкалы ($l = 3$) российской высшей школы: «неудовлетворительно» — 0, «удовлетворительно» — 1, «хорошо» — 2, «отлично» — 3. На рис.1 и 2 приведены гистограммы параметров «подготовленности» студентов и «сложности» экзаменов, рассчитанные по 4619 оценкам 151 студента по 150 экзаменам.

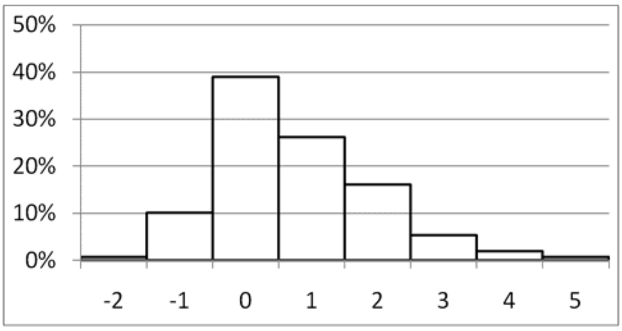

Рис.1. Распределение значений $\theta_1,\ldots,\theta_n$ «подготовленности» студентов

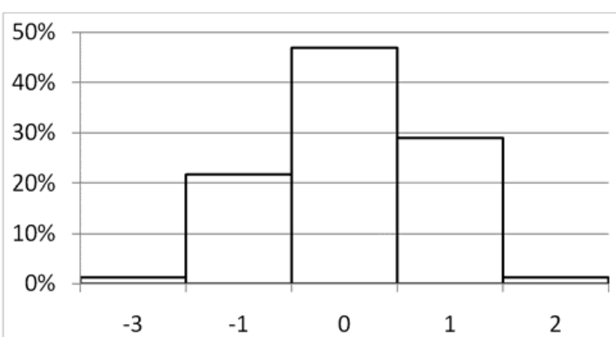

Рис. 2. Распределение значений $\delta_1,\ldots,\delta_m$ «сложности» экзамена

Применение параметрической модели экзаменационных оценок позволяет выявить ненормально «легкие» или «сложные» экзамены для того, чтобы помочь преподавателям скорректировать методики преподавания и оценивания знаний. Кроме этого, модель можно применять для прогнозирования итогов успеваемости.

Для оценки точности определения параметров «подготовленности» студентов и «сложности» экзаменов можно воспользоваться неравенством Рао-Крамера [3]. Для параметров модели информационная матрица Фишера будет иметь достаточно сложную структуру. Для простоты, предположим, что все параметры, кроме θ_i, известны и совпадают с решением уравнений (2). В этом случае количество информации Фишера

$$I(\theta_i;\{X_{ij}\}) = M\left[\left(\frac{d\ln(L)}{d\theta_i}\right)^2\right] = M\left[\left(\sum_{j=1}^{m}\left(X_{ij} - l\frac{e^{\theta_i}}{e^{\theta_i}+e^{\delta_j}}\right)\right)^2\right] =$$

$$= M\left[\left(\sum_{j=1}^{m}\left(X_{ij} - M[X_{ij}]\right)\right)^2\right] = \sum_{j=1}^{m} D[X_{ij}]$$

будет совпадать с суммой дисперсий оценок i-го студента. Аналогично для параметра δ_j количество информации

$$I(\delta_j;\{X_{ij}\}) = \sum_{i=1}^{n} D[X_{ij}]$$

будет совпадать с суммой дисперсий оценок j-го экзамена. Применяя неравенство Рао-Крамера, получаем неравенства

$$D[\hat{\theta}_i] \geq \frac{1}{\sum_{j=1}^{m} D[X_{ij}]}, \quad D[\hat{\delta}_j] \geq \frac{1}{\sum_{i=1}^{m} D[X_{ij}]},$$

которые определяют нижнюю границу дисперсий соответствующих оценок. Подставляя в неравенства дисперсию биномиального распределения (1) и оценки $\hat{\theta}_1,\ldots,\hat{\theta}_n$, $\hat{\delta}_1,\ldots,\hat{\delta}_m$ параметров вместо истинных значений получаем следующие граничные значения

$$D[\hat{\theta}_i] \geq \frac{1}{l\sum_{j=1}^{m}\frac{e^{\hat{\theta}_i-\hat{\delta}_j}}{\left(1+e^{\hat{\theta}_i-\hat{\delta}_j}\right)^2}}, \quad D[\hat{\delta}_j] \geq \frac{1}{l\sum_{i=1}^{m}\frac{e^{\hat{\theta}_i-\hat{\delta}_j}}{\left(1+e^{\hat{\theta}_i-\hat{\delta}_j}\right)^2}}, \qquad (3)$$

которые можно использовать для построения интервальных оценок, с учетом сходимости распределения оценок максимального правдоподобия к нормальному.

На рис. 3 представлен график нижней границы среднеквадратичного отклонения оценки параметра подготовленности студента для $l = 3$, в зависимости от величины параметра для разных количеств оценок. При построении графика использовался ряд $\hat{\delta}_1, \dots, \hat{\delta}_m$ параметров сложности, соответствующий нормальному распределению с нулевым математическим ожиданием и среднеквадратическим отклонением равным 0,78 (это оценка среднеквадратичного отклонения параметра сложности экзаменов, полученного по набору экзаменационных оценок).

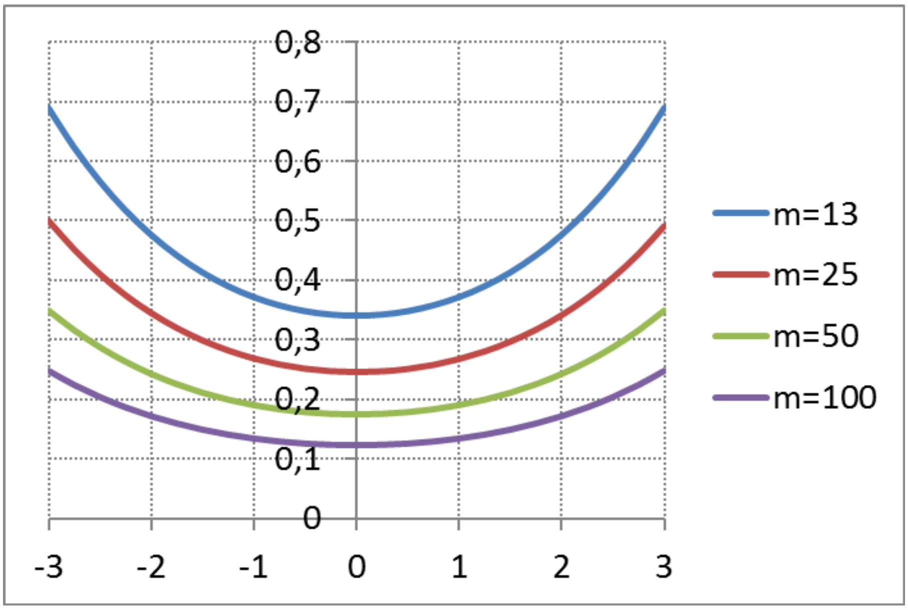

Рис. 3. Нижняя граница среднеквадратичного отклонения оценки параметра подготовленности студента

В соответствии с неравенством (3) Рао-Крамера на дисперсию оценки параметра подготовленности студента (трудности экзамена) влияют только его оценки. Таким образом, чем больше оценок студента (экзамена), тем выше точность определения соответствующего параметра.

Применение параметрической модели экзаменационных оценок позволяет выявить ненормально «легкие» или «сложные» экзамены для того, чтобы помочь преподавателям скорректировать методики преподавания и оценивания знаний. Кроме этого, модель можно применять для прогнозирования итогов успеваемости.

Для проверки адекватности политомической биномиальной модели можно применить дисперсионный анализ [4]. Для оценки влияния подготовленности студента сравниваются усредненные выборочные дисперсии по студентам

$$M_2 = \frac{1}{n(m-1)} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(x_{ij} - \bar{x}_{i*} \right)^2, \quad \bar{x}_{i*} = \frac{1}{m} \sum_{j=1}^{m} x_{ij}$$

с межгрупповой (факторной) дисперсией

$$M_1 = \frac{m}{n-1} \sum_{i=1}^{n} \left(\bar{x}_{i*} - \bar{x}_{**} \right)^2, \quad \bar{x}_{**} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij} .$$

В случае отсутствия влияния подготовленности студента на оценку, M_1 и M_2 являются разными оценками дисперсии одной и той же случайной величины. Статистика $F = M_1 / M_2$, при условии одинакового нормального распределения и независимости вариаций среди оценок, будет иметь распределение Фишера со степенями свободы $n-1$ и $mn-n$.

Для анализируемого набора экзаменационных оценок межгрупповая дисперсия $M_1 = 56{,}36$, характеризующая влияние фактора – подготовленности студента, оказалась значительно больше выборочной дисперсии $M_2 = 0{,}33$, которая объясняется случайной составляющей наблюдений. Как и следовало ожидать, при таком соотношении дисперсий гипотеза об отсутствии влияния подготовленности студента отвергается с практически нулевым уровнем значимости.

Если вместо оценок изучать отклонения

$$y_{ij} = x_{ij} - M\left[X_{ij}\right] = x_{ij} - l \frac{e^{\theta_i}}{e^{\theta_i} + e^{\delta_j}}$$

от математических ожиданий оценок, то дисперсионный анализ приводит к противоположным результатам. Межгрупповая дисперсия $M_1 = 0{,}00000027$ уменьшилась почти до нуля. Средние отклонений, вычисленные по множеству оценок каждого студента, оказались практически нулевыми. Поэтому гипотеза об отсутствии влияния подготовленности студента на отклонения от математических ожиданий оценок принимается практически с единичной доверительной вероятностью.

Аналогичные результаты дает проверка влияния на оценки фактора сложности экзамена. Статистика F подтверждает влияние сложности на экзамены и отвергает влияние сложности на отклонения оценок от своих математических ожиданий. Это свидетельствует о том, что предложенная модель достаточно правдоподобно объясняет разброс экзаменационных оценок.

Одним из вариантов проверки адекватности модели является изучение коэффициентов корреляции [5]. Оценка

$$r_j = \frac{\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_{i*}\right)\left(\bar{x}_{i*} - \bar{x}_{**}\right)}{\sqrt{\sum_{j=1}^{m}\left(x_{ij} - \bar{x}_{i*}\right)^2 \sum_{j=1}^{m}\left(\bar{x}_{i*} - \bar{x}_{**}\right)^2}}$$

коэффициента корреляции ряда $\left(x_{1j},\ldots,x_{nj}\right)$ оценок j-го задания и ряда $\left(\bar{x}_{1*},\ldots,\bar{x}_{n*}\right)$ средних оценок студентов позволяет оценить согласованность их результатов. Теоретический максимум +1, достигаемый в случае линейной зависимости рядов, на практике оказывается значительно меньше.

Другими вариантами могут быть коэффициенты r'_j корреляции ряда $\left(x_{1j},\ldots,x_{nj}\right)$ оценок и уровней подготовленности $\left(\theta_1,\ldots,\theta_n\right)$. Если вместо ряда $\left(\theta_1,\ldots,\theta_n\right)$ взять ряд $\left(p_{1j},\ldots,p_{nj}\right)$ вероятностей, то из полученного коэффициента r''_j корреляции будет исключена нелинейная зависимость вероятности от латентных параметров и коэффициент будет показывать меру линейной зависимости случайных оценок и их математических ожиданий. Для упомянутого выше набора экзаменационных оценок r'_j колеблется от 0,646 до 0,950, а r''_j — от 0,641 до 0,995. Совсем отсутствуют отрицательные значения, свидетельствующие о серьезных отклонениях оценок от ожидаемых значений.

Проверка монотонности заключается в изучении значений среднего уровня $\bar{\theta}_j(k)$ подготовленности студентов, получивших одинаковую оценку k в j-м задании. Проверка должна подтвердить возрастание среднего уровня подготовленности при увеличении оценки за задание $\bar{\theta}_j(0) < \bar{\theta}_j(1) < \ldots < \bar{\theta}_j(l)$. Для изучаемого множества экзаменационных оценок приведенное выше условие «чем выше оценка, тем выше уровень подготовленности» выполняется для 83 экзаменов из 85.

Традиционно для проверки модели применяют общую статистику согласия (MNSQ Outfit) и взвешенную общую статистику согласия (MNSQ Infit) [6]. В этих статистиках используются стандартизованные отклонения

$$z_{ij} = \left(x_{ij} - M\left[X_{ij}\right]\right) \Big/ \sqrt{D[X_{ij}]} = \left(x_{ij} - l\,\frac{e^{\theta_i}}{e^{\theta_i} + e^{\delta_j}}\right) \Big/ \sqrt{q_{ij}}$$

$$q_{ij} = l\,\frac{e^{\theta_i} e^{\delta_j}}{\left(e^{\theta_i} + e^{\delta_j}\right)^2}$$

с нулевыми математическими ожиданиями и единичными дисперсиями. Общие статистики согласия

$$U_i = \sum_{j=1}^{m} \frac{z_{ij}^2}{m}, \quad U_j = \sum_{i=1}^{n} \frac{z_{ij}^2}{n}$$

имеют распределения хи-квадрат с n и m степенями свободы, деленные на количество степеней свободы. Математическое ожидание статистик равно единице. Взвешенные общие статистики согласия

$$V_i = \sum_{j=1}^{m} z_{ij}^2 \bigg/ \sum_{j=1}^{m} q_{ij}, \quad V_j = \sum_{i=1}^{n} z_{ij}^2 \bigg/ \sum_{i=1}^{n} q_{ij}$$

также имеют распределения хи-квадрат с числом степеней свободы равным сумме среднеквадратичных отклонений, деленных на число степеней свободы. Значения статистик для экзаменационных оценок приведены в таблице 1. Изучение статистик согласия демонстрирует высокую доверительную вероятность практически для всех экзаменов и студентов, за небольшим количеством исключений, связанных с высокой долей одинаковых оценок.

Значения статистик согласия для экзаменационных оценок студентов

Статистика	Минимальное значение	Максимальное значение	Среднее значение
Общая статистика согласия для студентов U_i	0,117	1,456	0,599
Доверительная вероятность для U_i	0,036	1	0,856
Взвешенная общая статистика согласия для студентов V_i	0,0002	1,449	0,624
Доверительная вероятность для V_i	0,105	0,999998	0,78
Общая статистика согласия для экзаменов U_j	1,533664	0,258	0,645
Доверительная вероятность для U_j	0,0168	1	0,891
Взвешенная общая статистика согласия для экзаменов V_j	0,258	1,07	0,622
Доверительная вероятность для V_j	0,294	1	0,832

Приведенные результаты статистических проверок предложенной модели оценок подтверждают возможность ее применения для изучения экзаменационных оценок студентов.

Список использованной литературы

1. Rasch G. Probabilistic models for some intelligence and attainment tests / G. Rasch. — Chicago: The University of Chicago Press, 1980.

2. Andrich D. A. rating formulation for ordered response categories. / D. A. Andrich // Psychometrika, 43, 357–74.

3. Крамер Г. Математические методы статистики / Г. Крамер. — М. : Мир, 1975. — 648 с.

4. Хьютсон А. Дисперсионный анализ / А. Хьютсон. — М. : Статистика, 1971. — 88с.

5. Linacre J. M. Correlations: point-biserial, point-measure, residual / J. M. Linacre. URL : http://www.winsteps.com/winman/correlations.htm.

6. Wright B. D., Stone, M. H. Measurement essentials. 2nd Edition / B. D. Wright, M. H. Stone. — Wilmington, Delaware: Wide Range, 1999.