УДК 378:001.891

В.В. Братищенко

МОДЕЛЬ СТОБАЛЛЬНЫХ ОЦЕНОК С ЛАТЕНТНЫМИ ПАРАМЕТРАМИ НА ОСНОВЕ РАСПРЕДЕЛЕНИЯ ГАУССА

Педагогические измерения тесно связаны с технологией тестирования, как наиболее удовлетворяющей формальным требованиям валидности, надежности и достоверности. Оценки преподавателей являются наиболее распространенными измерителями знаний, умений и навыков обучающихся, но не все учебные результаты могут быть измерены тестами. Обосновывается актуальность статистических исследований оценок. Формулируются требования к модели оценок. Предлагается модель с латентными параметрами для стобалльных оценок. Разработан алгоритм оценки параметров модели по данным наблюдений. Предложены статистические процедуры оценки адекватности модели, использующие дисперсионный анализ, коэффициенты корреляции и критерий хи-квадрат.

Ключевые слова: стобалльная оценка; модель оценки; латентный параметр; модель Раша; текущая успеваемость.

V.V. Bratischenko

MODEL OF 100-GRADES MARK WITH LATENT PARAMETERS BASED ON GAUSSIAN DISTRIBUTION

Pedagogical measurements are closely related to testing technology as the most satisfying formal requirements of validity, reliability and reliability. Teacher assessments are the most common measures of learners' knowledge, skills and abilities, but not all learning outcomes can be measured by tests. The relevance of marks statistical research is justified. Requirements for a model of marks are formulated. A model with latent parameters for hundred grades marks is proposed. An algorithm for estimating model parameters from observational data has been developed. Statistical procedures for assessing the adequacy of the model using dispersion analysis, correlation coefficients and the chi-square test are proposed.

Keywords: 100-grades mark; model of marks; latent parameter; Rasch model; current performance.

Проблемам педагогических измерений посвящено большое количество научных работ [1–3]. Особенно актуальным становится разработка методик измерения сформированности компетенций [5–9]. В основном педагогические измерения связывают с технологией тестирования, как наиболее удовлетворяющей формальным требованиям валидности, надежности и достоверности. Однако, с одной стороны, не все учебные результаты могут быть измерены тестами, с другой, оценки преподавателей являются наиболее распространенными измерителями знаний, умений и навыков обучающихся. Статистические методы исследования оценок преподавателей позволяют вычислять характеристики, полезные для организации процесса оценивания и изучения качеств измерительных материалов.

В качестве объекта исследований выбраны стобалльные оценки, широко применяемые в дополнение к традиционным. Общепризнано, что традиционная шкала включающая четыре градации: «неудовлетворительно», «удовлетворительно», «хорошо» и «отлично» – является слишком грубой. Переход на стобалльную шкалу позволяет сделать оценки более гибкими и точными. Необходимо учитывать то, что шкалы оценок не являются метрическими. Это приводит к тому, что оценки некорректно интерпретировать как числа и некорректно использовать числовые характеристики, вычисляемые по массиву оценок. Еще одним фактором, препятствующим применению традиционных статистических методов, является различные интерпретации шкал разными преподавателями. Оценки одного преподавателя могут быть значительно смещены относительно оценок другого преподавателя. Все эти особенности должны быть учтены в модели оценок.

Стобалльные оценки складываются под влиянием многих факторов. Следовательно, должны хорошо описываться нормальным распределением. Однако, область значений нормальных случайных величин включает все значения от $-\infty$ до $+\infty$. Для установления соответствия стобалльной величины нормальной случайной величине можно предложить логистическое преобразование – сигмоиду. Пусть x – стобальная оценка, y – вещественная переменная, связанная с ней соотношением

$$y = ln\left(\frac{x+1}{101-x}\right).$$

Величина 101 выбрана для исключения бесконечных значений, если оценка равна 100 баллам.

В современных методах статистической обработки результатов тестирования [4], основанных на идеях Раша применяются латентные переменные, характеризующие подготовленность обучаемого и трудность задания. Предлагается распространить такое описание на стобалльные оценки. Предположим, что y_{ij} – реализация нормальной случайной величины Y_{ij} с математическим ожиданием $\theta_i - \delta_j$, где θ_i – латентный параметр подготовленности $i -$ го студента $i = 1, \dots, n$, δ_j – латентный параметр трудности $j -$ го задания $j = 1, \dots, m$, σ – среднеквадратическое отклонение. Случайные величины Y_{ij} соответствующие оценкам студентов будем считать независимыми в совокупности. Используя логарифм функции правдоподобия

$$\ln(L) = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) + \frac{\left(\theta_i-\delta_j-y_{ij}\right)^2}{2\sigma^2}\right),$$

можно найти условия достижения максимума правдоподобия

$$\frac{\partial(\ln(L))}{\partial\theta_i} = \sum_{j=1}^{m}\left(\frac{\theta_i-\delta_j-y_{ij}}{\sigma^2}\right) = 0 \text{ или } m\theta_i = \sum_{j=1}^{m}\left(\delta_j + y_{ij}\right),$$

$$\frac{\partial\left(ln(L)\right)}{\partial\delta_j} = \sum_{i=1}^{n}\left(\frac{\theta_i-\delta_j-y_{ij}}{\sigma^2}\right) = 0 \text{ или } n\delta_j = \sum_{i=1}^{n}\left(\theta_i - y_{ij}\right).$$

Данная система уравнений вырождена, вероятность зависит от разности $\theta_i - \delta_j$ латентных параметров. Поэтому для получения одного из решений можно дополнительно принять $\sum_{j=1}^{m}\delta_j$=0. В этом случае

$$\theta_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij},\ \delta_j = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} - \frac{1}{n}\sum_{i=1}^{n} y_{ij}.$$

Однако, при обработке оценок (особенно, оценок текущей успеваемости) достаточно часто встречаются неполные наборы оценок, если некоторые студенты не выполняют какие-либо учебные задания. Для описания неполного набора оценок введем индикаторы

$$I_{ij} = \begin{cases} 1, \text{если есть оценка } i - \text{го студента за } j - \text{е задание}, \\ 0, \text{если нет оценки } i - \text{го студента за } j - \text{е задание}. \end{cases}$$

В этом случае уравнения

$$\theta_i \left(\sum_{j=1}^{m} I_{ij} \right) = \sum_{j=1}^{m} I_{ij}(\delta_j + y_{ij})$$

$$\delta_j \left(\sum_{i=1}^{m} I_{ij} \right) = \sum_{i=1}^{n} I_{ij}(\theta_i - y_{ij})$$

по-прежнему образуют вырожденную систему. Матрица данной системы имеет блочную структуру с диагональными матрицами в диагональных блоках, причем диагональный элемент равен сумме элементов строки смежного блока, и матрица имеет, таким образом, доминирующую диагональ. Такую систему уравнений удобно решать приближенными методами. Приближенное решение на k-й итерации вычисляется по формулам, полученных методом касательных,

$$\theta_i^k = \frac{\sum_{j=1}^{m} I_{ij}(\delta_j^{k-1} + y_{ij})}{\sum_{j=1}^{m} I_{ij}},$$

$$\delta_j^k = \frac{\sum_{i=1}^{n} I_{ij}\left(\theta_i^{k-1} - y_{ij}\right)}{\sum_{i=1}^{n} I_{ij}}$$

при начальных условиях:

$$\theta_i^0 = \frac{\sum_{j=1}^{m} I_{ij} y_{ij}}{\sum_{j=1}^{m} I_{ij}},$$

$$\delta_j^0 = -\frac{\sum_{i=1}^{n} I_{ij} y_{ij}}{\sum_{i=1}^{n} I_{ij}}.$$

При условиях вариативности оценок вычислительный процесс быстро сходится. Для проверки адекватности модели применим дисперсионный анализ. По причинам неполных данных суммы будут вычисляться по следующим формулам:

$$s_t = \sum_{i,j=1}^{n,m} I_{ij}\left(y_{ij} - y_{**}\right)^2,\ y_{**} = \frac{1}{N}\sum_{i,j=1}^{n,m} I_{ij} y_{ij},\ N = \sum_{i,j=1}^{n,m} I_{ij},$$

$$s_b = \sum_{i,j=1}^{n,m} I_{ij}\left(y_{i*} - y_{**}\right)^2,\ y_{i*} = \sum_{j=1}^{m} I_{ij} y_{ij} / \sum_{j=1}^{m} I_{ij},$$

$$s_w = \sum_{i,j=1}^{n,m} I_{ij}\left(y_{ij} - y_{i*}\right)^2$$

со степенями свободы $N-1$, $n-1$, $N-n$. Далее по критерию Фишера

$$F = \frac{s_b}{n-1} \Big/ \frac{s_w}{N-n}$$

со степенями свободы $n-1$, $N-n$ проверяется гипотеза о равенстве средних по группам студентов.

Исходным статистическим материалом для исследования выбраны оценки текущей успеваемости студентов БГУ. Вычисления проводились для группы из 16 студентов, 988 оценок по 83 заданиям. Для проверки преобразованных оценок y_{ij} значение критерия Фишера составило 6,406 со степенями свободы 15 и 972. Гипотеза о равенстве средних отвергается, так как доверительная вероятность, соответствующая значению критерия практически равна 0.

По точно такой же схеме проверялись остатки:

$$z_{ij} = y_{ij} - M[Y_{ij}] = y_{ij} - (\theta_i - \delta_j).$$

В данном случае проверка по критерию Фишера дала противоположные результаты – гипотеза о равенстве средних подтверждается с почти единичной вероятностью.

Проверка влияния заданий на средние оценки дала такие же результаты. Это подтверждает предположение, что модель с латентными параметрами точнее описывает процесс формирования оценок.

Проверки по критерию Фишера не дают полной гарантии адекватности модели. В частности, остатки z_{ij} не прошли проверку на нормальность распределения (рис. 1).

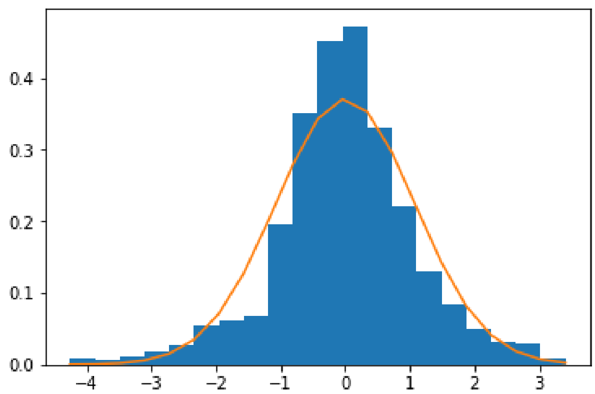

Рис. 1. Гистограмма остатков и плотность нормального распределения

Еще одна проверка, традиционно применяемая при обработке результатов тестирования – это вычисление суммы квадратов нормированных остатков

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

которые при выполнении приведенных выше предположений будут иметь распределение хи-квадрат с степенями свободы $\sum_{i=1}^{n} I_{ij}$ и $\sum_{j=1}^{m} I_{ij}$. Для проверки студентов уровень значимости колеблется от 0,256 до 0,724. Это означает, что оценки параметров подготовленности прошли проверку по критерию хи-квадрат. Для параметров трудности заданий оказалось, что 17 из них не прошли проверку по критерию хи-квадрат по уровню значимости 0,05.

Проверка согласованности уровней подготовленности студентов $\theta_1, \dots, \theta_n$ (трудности заданий $\delta_1, \dots, \delta_m$) с рядами оценок $y_{1j}, \dots, y_{nj}$ $(y_{i1}, \dots, y_{im})$ может быть проведена с использованием соответствующих коэффициентов корреляции. Вычисленные коэффициенты корреляции трудностей заданий и оценок студентов достаточно однородно лежат в интервале от –0,9 до –0,72. Отрицательные значения отражают очевидное соотношение: чем труднее задание, тем ниже оценки. В целом такая согласованность свидетельствует, что набор заданий в целом достаточно адекватно оценивает знания студентов.

Коэффициенты корреляции подготовленностей студентов и набора оценок некоторого задания характеризуют само задание. Наблюдаемый разброс коэффициентов (см. рис. 2) свидетельствует о существенной неоднородности заданий. Прежде всего в девяти заданиях все студенты получили одинаковые оценки (выброс в точке ноль на рис. 2). Для таких случаев вычисление коэффициента корреляции невозможно. Для двадцати одного задания коэффициент корреляции оказался меньше 0,5. Все это говорит о том, что для данных заданий плохо определена процедура оценивания. Для получения статистически обоснованных результатов следует исключить из статистической обработки оценки таких заданий.

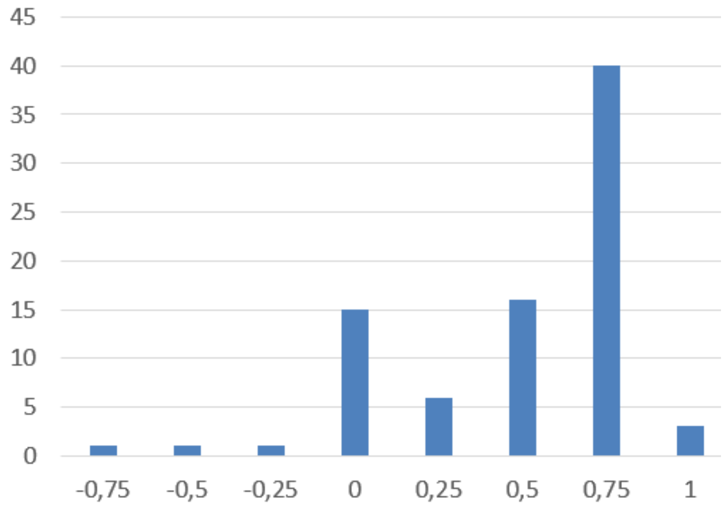

Рис. 2. Распределение коэффициентов корреляции для заданий

Предложенная модель предоставляет приближенное описание для процедур оценивания. Достоинством модели является возможность объединения шкал и методик оценивания разных преподавателей для получения статистически обоснованных оценок подготовленности студентов. Модель можно применять для выявления заданий, для которых процедура оценивания определена недостаточно корректно.

Список использованной литературы

1. Челышкова М.Б. Теория и практика конструирования педагогических тестов : учеб. пособие / М.Б. Челышкова. – М. : Логос, 2002. – 432 с.

2. Гуськова М.В. Этапы развития эвалюации в образовании [Электронный ресурс] / М.В. Гуськова, В.И. Звонников // Экономика образования. – 2011. – № 4. – Режим доступа: https://cyberleninka.ru/article/n/etapy-razvitiya-evalyuatsii-v-obrazovanii.

3. Богоудинова Р.З. Основные подходы к оцениванию результатов образовательной деятельности [Электронный ресурс] / Р.З. Богоудинова // Вестник Казанского технологического университета. – 2011. – № 22. – Режим доступа:

https://cyberleninka.ru/article/n/osnovnye-podhody-k-otsenivaniyu-rezultatov-obrazovatelnoy-deyatelnosti

4. Нейман Ю.М. Введение в теорию моделирования и параметризации педагогических тестов / Ю.М. Нейман, В.А. Хлебников. – М. : Прометей, 2000. – 168 с.

5. Озерникова Т.Г. Качество образования – приоритет развития университета / Т.Г. Озерникова, Т.А. Бутакова // Известия Иркутской государственной экономической академии. – 2015. – Т. 25, № 2. – С. 196–205. – DOI: 10.17150/1993-3541.2015.25(2).196-205.

6. Родионов А.В. Модификация рейтинговой параметрической модели оценки латентных факторов для измерения уровня сформированности компетенций / А.В. Родионов // Известия Иркутской государственной экономической академии. – 2014. – № 6 (98). – С. 168–174. – DOI: 10.17150/1993-3541.2014.24(6).168-174.

7. Кешиков К.А. Компетентностный подход и методы оценки компетенций в современном высшем образовании / К.А. Кешиков // Перспективы развития информационных технологий. – 2016. – № 29. – С. 115–122.

8. Пахаруков А.А. Проблемы применения компетентностного подхода при разработке и реализации основных образовательных программ высшего профессионального образования / А.А. Пахаруков, В.Н. Белоусов // Вестник Иркутского государственного технического университета. – 2013. – № 7 (78). – С. 298–303.

9. Пашин В.И. Проблемы оценки знаний учащихся [Электронный ресурс] / В.И. Пашин // Наука, техника и образование. – 2015. – № 9 (15). – Режим доступа: https://cyberleninka.ru/article/n/problemy-otsenki-znaniy-uchaschihsya.

Информация об авторе

Братищенко Владимир Владимирович – кандидат физико-математических наук, доцент кафедры математики и информатики Байкальского государственного университета, 664003, г. Иркутск, ул. Ленина, 11, e-mail: vbrat56@mail.ru.

Author

Bratischenko Vladimir Vladimirovich – PhD in Physics and Mathematical Sciences, Associate Professor of Mathematics and Informatics Department, Baikal State University, 664003, Irkutsk, 11 Lenina St., e-mail: vbrat56@mail.ru.