

В.В. Братищенко

*Байкальский государственный университет,
г. Иркутск, Российская Федерация*

ПОШАГОВАЯ МОДЕЛЬ С ЛАТЕНТНЫМИ ПАРАМЕТРАМИ ДЛЯ ОЦЕНОК СТУДЕНТОВ

Аннотация. В статье рассматриваются недостатки традиционных подходов статистической обработки оценок промежуточных аттестаций студентов. Предлагается модель, в которой получение оценки k в порядковой шкале связывается с успешным выполнением k шагов аттестации. По аналогии с Item Response Theory, применяемой для обработки результатов тестирования, вероятность успешного выполнения шага определяется латентными параметрами — уровнем подготовленности студента и трудностью шага. Предложены методы определения латентных параметров по полученным оценкам и статистические процедуры проверки адекватности модели. Приведены данные обработки массива оценок. Итоги обработки подтверждают возможность применения предложенной модели для более точного оценивания подготовленности студентов и трудности аттестаций.

Ключевые слова. Статистическая обработка оценок, Item Response Theory, модель с латентными параметрами, адекватность модели, Infit-статистика, Outfit-статистика, коэффициент детерминации.

Информация о статье. Дата поступления: 11 ноября 2021 г.

V.V. Bratishchenko

*Baikal State University,
Irkutsk, Russian Federation*

STEP-BY-STEP MODEL WITH LATENT PARAMETERS FOR STUDENT ASSESSMENTS

Abstract. The article discusses the disadvantages of traditional approaches to statistical processing of assessments of intermediate attestations of students. We proposed a model in which the obtaining of the k grade on an ordinal scale is associated with the successful completion of k certification steps. By analogy with Item Response Theory used for processing test results the probability of a successful step is determined by latent parameters — the student's ability and the difficulty of the step. Methods for determining latent parameters from the estimates obtained and statistical procedures for checking the adequacy of the model are proposed. The data of processing the array of estimates are presented. The processing results confirmed the possibility of using the proposed model for a more accurate assessment of students' ability and the difficulty of attestations.

Keywords. Statistical processing of assessments, Item Response Theory, model with latent parameters, model adequacy, Infit statistics, Outfit statistics, coefficient of determination.

Article info. Received 11 November 2021.

Статистическая обработка оценок промежуточных аттестаций применяется для управления качеством обучения в вузах. В работах [1–4] изучается влияние на оценки разных факторов (ре-

зультаты ЕГЭ), пол, источники финансирования обучения, проживание в общежитии), но при этом исключается зависимость оценок от дисциплин и преподавателей. Однако, такие зависимости несомненно есть [5; 6], и их изучение позволяет выявить дисциплины со значимыми отклонениями в распределении оценок. Конечно, такие отклонения могут быть вызваны объективными причинами, но чаще всего они связаны с недостатками в системе преподавания и методике оценивания.

Применение средних значений и других статистических характеристик к оценкам не вполне корректно, так как оценки измеряются в порядковых, а не метрических шкалах [7]. Более обоснованным является подход, в котором изучаются распределения вероятностей значений оценок. В данной работе предлагается модель оценки, использующая для описания вероятностей значений латентные параметры подготовленности студента и трудности испытания по аналогии с параметрами, применяемыми в Item Response Theory [8] для анализа оценок тестирования.

Чем выше подготовленность студента и ниже трудность испытания, тем выше вероятность хорошей оценки. Для дихотомических оценок («зачтено» или «не зачтено») модель выглядит наиболее просто. Для полиномических оценок применяются порядковые шкалы (далее используется шкала $\{0, 1, \dots, k\}$) и связывают вероятности более высокой оценки с дополнительными параметрами.

Предлагается следующая модель, описывающая получение оценки в виде последовательности шагов. В случае «неудачи» на l -м шаге ($l = 1, \dots, k$) с вероятностью $1 - p_l$ оценка принимается равной $l - 1$, в случае «успеха» с вероятностью p_l испытание продолжается на следующем шаге. Предполагается независимость шагов. Успешно выполнив все шаги обучающийся получает максимальный балл k . Случайная величина X — итоговая оценка будет иметь следующее распределение вероятностей

$$P\{X = x\} = (1 - p_{x+1}) \prod_{l=0}^x p_l, \quad x = 0, \dots, k,$$

где для удобства полагается $p_0 = 1, p_{k+1} = 0$.

Математическое ожидание X можно вычислить в виде следующей суммы

$$M[X] = \sum_{h=1}^k \prod_{l=1}^h p_l.$$

Основную проблему подобных моделей составляет оценка параметров. Для применения байесовского подхода к оценке вероятностей «успеха» на каждом шаге необходимо предложить некото-

рое априорное распределение, которое каждая новая оценка будет переводить в апостериорное. Для того чтобы априорное и апостериорное распределения отличались только параметрами, применяют так называемые сопряженные распределения [9]. Для бернуллиевского распределения сопряженным является бета-распределение.

$$f(x) = \frac{x^{\theta-1}(1-x)^{\delta-1}}{B(\theta, \delta)},$$

с математическим ожиданием

$$M[X] = \frac{\theta}{\theta + \delta}$$

и дисперсией

$$D[X] = \frac{\theta\delta}{(\theta + \delta)^2(\theta + \delta + 1)}.$$

Свойства бета-распределения позволяют интерпретировать θ как параметр подготовленности студента, а δ как параметр трудности испытания.

На рис.1 представлены графики изменения вероятностей P_0 — «неявка», P_1 — «неудовлетворительно», P_2 — «удовлетворительно», P_3 — «хорошо», P_4 — «отлично» при изменении параметра θ подготовленности, если параметры трудности соответствующих четырех шагов экзамена равны 1; 1,5; 4; 12,5. Вероятность успеха на каждом шаге принимается равной математическому ожиданию бета-распределения.

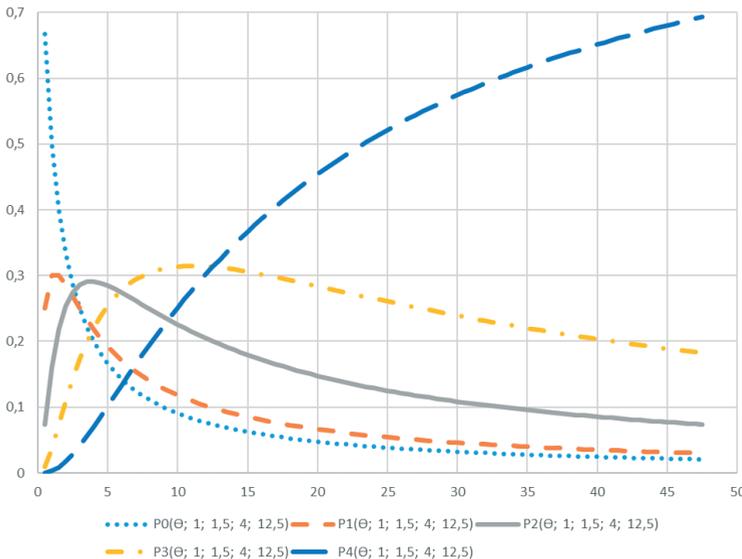


Рис. 1. Вероятности оценок в зависимости от параметра θ

Влияние параметра подготовленности на распределение оценки при фиксированных параметрах шагов представлено на рис. 2. Как и следовало ожидать, большее значение подготовленности увеличивает вероятности высоких оценок.

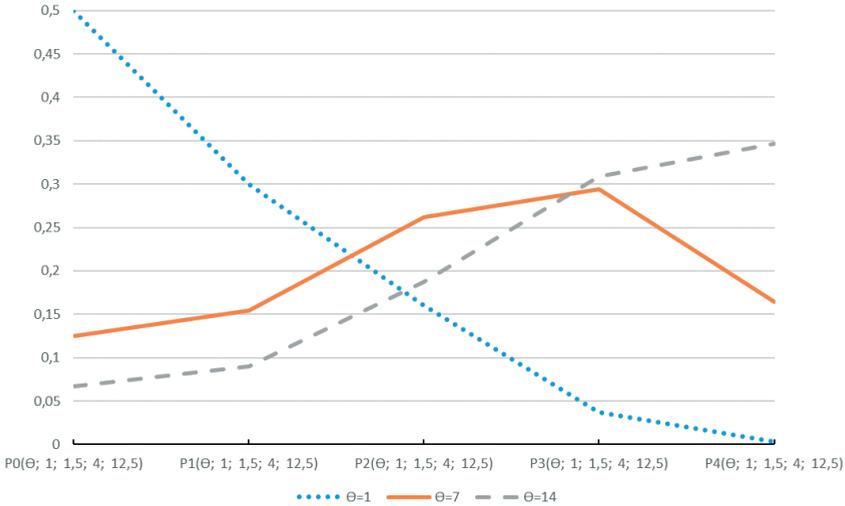


Рис. 2. Распределения вероятностей оценок студентов с уровнями подготовленности $\theta = 1$, $\theta = 7$, $\theta = 14$

Особенностью предлагаемой модели является, то что линейное увеличение ее латентных параметров в c раз не меняет математические ожидания параметров, но почти в c раз уменьшает их дисперсии.

Для получения байесовских оценок предположим, что бернуллиевская случайная величина X , параметр которой имеет бета-распределение, получает в результате испытания значение a . В этом случае апостериорная плотность

$$f(x|X=a) = \frac{f(x)P\{X=a|x\}}{P\{X=a\}} = \frac{x^{\theta-1}(1-x)^{\delta-1}x^a(1-x)^{1-a}B(\theta, \delta)}{B(\theta, \delta)B(\theta+a, \delta+1-a)} = \frac{x^{\theta+a-1}(1-x)^{\delta+1-a-1}}{B(\theta+a, \delta+1-a)}$$

будет иметь бета-распределение с параметрами $\theta + a$ и $\delta + 1 - a$, т.е. «успех» увеличивает параметр подготовленности, а «неудача» — параметр трудности.

Предлагается следующая модель оценок. Пусть имеется набор оценок, где оценка x_{ij} получена i -м обучающимся на j -м испытании, θ_i — «подготовленность» обучающегося, δ_{jl} — «трудность» l -го шага j -го испытания, n — количество обучающихся, m — коли-

чество испытаний. Тогда вероятность успеха i -го обучающегося на l -м шаге j -го испытания будет равна

$$P_{ijl} = \frac{\theta_i}{\theta_i + \delta_{jl}},$$

а вероятность оценки

$$P\{X_{ij} = x_{ij}\} = \left(1 - \frac{\theta_i}{\theta_i + \delta_{j(x_{ij}+1)}}\right) \prod_{l=0}^{x_{ij}} \frac{\theta_i}{\theta_i + \delta_{jl}},$$

$$x_{ij} = 0, \dots, k,$$

$$\delta_{j0} = 0, \delta_{j(k+1)} = \infty$$

будет зависеть от соответствующих латентных параметров подготовленности обучающегося и трудностей шагов испытания.

Кроме этого, предположим, что бернуллиевская случайная величина (индикатор успеха) на l -м шаге j -го испытания i -го обучающегося определяется априорным бета-распределением вероятности успеха с параметрами θ_i и δ_{jl} . Событие $X_{ij} = x_{ij}$ означает, что шаги $1, \dots, x_{ij}$ завершились успехом, а шаг $x_{ij} + 1$ – неудачей.

Соответственно апостериорные параметры θ_i и $\delta_{j(x_{ij}+1)}$ изменятся на $\theta_i + x_{ij}$ и $\delta_{j(x_{ij}+1)} + 1$. Таким образом, по итогам сдачи нескольких экзаменов параметр подготовленности увеличивается на сумму оценок, а параметры трудности увеличиваются для шагов с номерами большими чем полученный балл. Максимальная оценка не увеличивает трудности шагов, зато дает наибольшее увеличение параметра подготовленности.

Байесовский подход, в данном случае, оправдывает применение сумм оценок в качестве оценки подготовленности обучающегося. Однако, такому применению препятствуют следующие аргументы. Во-первых, величина суммы зависит от количества оценок – одинаковые суммы можно набрать, получив одну оценку «отлично» или 4 оценки «неудовлетворительно» (в приведенных числовых экспериментах шкала: «неявка», «неудовлетворительно», «удовлетворительно», «хорошо», «отлично» приводилась к шкале 0, ..., 4). Применение средних уменьшает этот недостаток, но не исключает его совсем. Во-вторых, оценка «неудовлетворительно» увеличивает подготовленность студента, что не вполне обосновано.

Кроме байесовского подхода, для оценивания латентных параметров используются численные методы. Одним из способов определения латентных параметров модели по наблюдениям x_{ij} оценок является метод моментов [8]. В соответствии с методом моментов латентные параметры определяются по уравнениям

$$\sum_{i=1}^n M[X_{ij}] = \sum_{i=1}^n x_{ij}, \quad j = 1, \dots, m, \quad \sum_{j=1}^m M[X_{ij}] = \sum_{j=1}^m x_{ij}, \quad i = 1, \dots, n,$$

где $M[X_{ij}] = \sum_{h=1}^k \prod_{l=1}^h \frac{\theta_i}{\theta_i + \delta_{jl}}$ – математическое ожидание оценки.

Для решения уравнений предлагается использовать метод Ньютона. В этом методе итерационная формула

$$x^{(s+1)} = x^{(s)} - \frac{f(x^{(k)})}{f'(x^{(k)})}$$

использует производные по параметрам

$$\frac{\partial}{\partial \theta_i} \left(\sum_{j=1}^m \sum_{h=1}^k \prod_{l=1}^h \frac{\theta_i}{\theta_i + \delta_{jl}} - \sum_{j=1}^m x_{ij} \right) = \sum_{j=1}^m \sum_{h=1}^k \left(\prod_{l=1}^h \frac{1}{\theta_i + \delta_{jl}} \right) \left(\sum_{l=1}^h \frac{\delta_{jl}}{\theta_i + \delta_{jl}} \right)$$

$$\frac{\partial}{\partial \delta_{jv}} \left(\sum_{i=1}^n \sum_{h=1}^k \prod_{l=1}^h \frac{\theta_i}{\theta_i + \delta_{jl}} - \sum_{i=1}^n x_{ij} \right) = - \sum_{i=1}^n \sum_{h=v}^k \left(\prod_{l=1}^h \frac{\theta_i}{\theta_i + \delta_{jl}} \right) \frac{1}{\theta_i + \delta_{jv}}.$$

В итоге получаем формулы

$$\theta_i^{(s+1)} = \theta_i^{(s)} - \frac{\sum_{j=1}^m \sum_{h=1}^k \prod_{l=1}^h \frac{\theta_i^{(s)}}{\theta_i^{(s)} + \delta_{jl}^{(s)}} - \sum_{j=1}^m x_{ij}}{\sum_{j=1}^m \sum_{h=1}^k \left(\prod_{l=1}^h \frac{1}{\theta_i^{(s)} + \delta_{jl}^{(s)}} \right) \left(\sum_{l=1}^h \frac{\delta_{jl}^{(s)}}{\theta_i^{(s)} + \delta_{jl}^{(s)}} \right)},$$

$$\delta_{jv}^{(s+1)} = \delta_{jv}^{(s)} + \frac{\sum_{i=1}^n \sum_{h=1}^k \prod_{l=1}^h \frac{\theta_i^{(s)}}{\theta_i^{(s)} + \delta_{jl}^{(s)}} - \sum_{i=1}^n x_{ij}}{\sum_{i=1}^n \sum_{h=v}^k \left(\prod_{l=1}^h \frac{\theta_i^{(s)}}{\theta_i^{(s)} + \delta_{jl}^{(s)}} \right) \frac{1}{\theta_i^{(s)} + \delta_{jv}^{(s)}}}$$

для определения оценок латентных параметров численными методами.

Для вычисления начальных значений параметров можно применить байесовские оценки

$$\theta_i^{(0)} = \sum_{j=1}^m x_{ij}$$

$$\delta_{jv}^{(0)} = \sum_{i=1}^n \begin{cases} 1, & v \geq x_{ij} + 1, \\ 0, & v < x_{ij} + 1. \end{cases}$$

В процессе вычисления нужно учитывать неотрицательность латентных параметров и возможность неограниченного роста параметров для случаев, когда все оценки обучающегося или экзамена максимальны — для круглых отличников или экзаменов со всеми пятерками.

На рис. 3 представлено уменьшение суммы квадратов отклонений для решения системы уравнений численными методами при увеличении количества итераций.

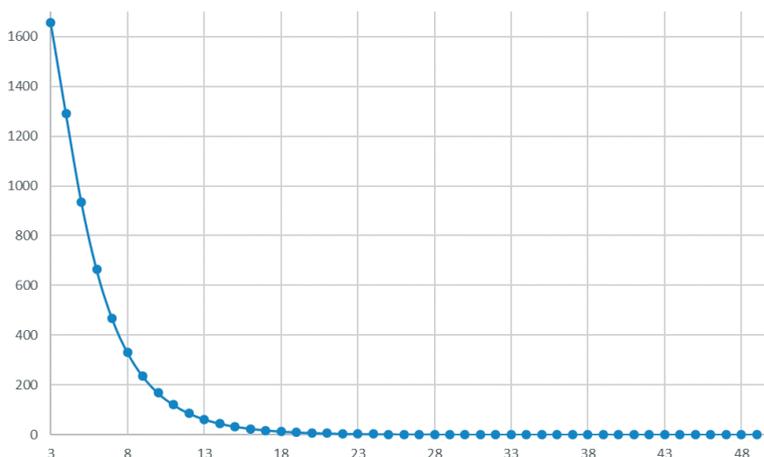


Рис. 3. Изменение суммы квадратов отклонений при решении системы уравнений численными методами

Значения параметров трудностей, использованных для получения графиков на рис. 1 и 2 примерно соответствуют средним значениям по всем экзаменам, полученным в результате обработки оценок группы из 23 студентов по 1037 оценкам 72 аттестаций.

Дисперсионный анализ позволяет оценить влияния зависимостей оценок от студентов. Для этого сравниваются усредненные выборочные дисперсии по студентам

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

с межгрупповой дисперсией

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

В случае отсутствия влияния 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 $nm - n$. Как и следовало ожидать, проверка подтвердила предположение о наличии влияния студента на оценки. Аналогично была подтверждена зависимость оценок от экзамена.

По такой же схеме были проверены гипотезы о независимости остатков $x'_{ij} = x_{ij} - M[X_{ij}]$ от студентов и экзаменов. Вычисления подтвердили эту гипотезу с доверительной вероятностью близкой к единице.

Еще одна традиционная проверка моделей с латентными параметрами использует коэффициенты корреляции. Коэффициент корреляции можно вычислять между рядом оценок и рядом соответствующих математических ожиданий. Для выбранного массива оценок коэффициент корреляции составил 0,74, что свидетельствует о значимой статистической связи предсказанных моделью значений с наблюдениями. Коэффициенты корреляции соответствующих рядов одного студента или одного экзамена дают более пеструю картину. Для 3 студентов из 23 и 22 экзаменов из 72 коэффициент корреляции оказывается меньше 0,5. Такие значимые отклонения свидетельствуют о нестандартных профилях соответствующих студентов и экзаменов.

При обработке результатов тестирования применяют статистики Infit и Outfit [8]. Статистики определяются по стандартизованным отклонениям

$$z_{ij} = \frac{x_{ij} - M[X_{ij}]}{\sqrt{D[X_{ij}]}}$$

Статистика Outfit определяется для студента

$$U_i = \frac{1}{m} \sum_{j=1}^m z_{ij}^2$$

или для экзамена

$$U'_j = \frac{1}{n} \sum_{i=1}^n z_{ij}^2$$

и является средним квадратов стандартизованных отклонений и случайной величиной, распределенной по закону хи-квадрат и деленной на число степеней свободы. Считается [8], что данная

статистика чувствительна к выбросам. Поэтому наряду с ней используют статистику Infit для студента

$$V_i = \sum_{j=1}^m z_{ij}^2 / \sum_{j=1}^m D[X_{ij}]$$

или для задания

$$V'_i = \sum_{i=1}^n z_{ij}^2 / \sum_{i=1}^n D[X_{ij}]$$

Эта статистика учитывает отклонение с весом равным доле дисперсии отклонения в сумме дисперсий всех отклонений. Взвешивание уменьшает влияние менее информативных ответов с малой дисперсией. Считается, что статистика Infit имеет Хи-квадрат распределение, деленное на число степеней свободы, равное сумме дисперсий. В таблице приведены рекомендации использования статистик по итогам тестирования

Правила интерпретации значений Infit и Outfit статистик

Значение	Интерпретация
< 0,5	Малопродуктивное задание. Может ошибочно породить ощущение высокой надежности заданий
0,5–1,5	Задание может быть использовано для измерения
1,5–2,0	Задание малопродуктивно для измерения, но может быть использовано без редактирования.
> 2	Задание нарушает систему измерений. Допустимо 1–2 таких задания

По итогам вычисления статистики Outfit в первый интервал попали 34 экзамена, во второй — 36, в третий — 2. Для студентов эта статистика распределила студентов следующим образом: в первый интервал попало 6 студентов, во второй — 16, в третий — 1. Данные статистики можно использовать для выделения экзаменов с методиками оценивания, отличающимися от среднестатистических.

Коэффициент детерминации

$$R^2 = 1 - \frac{D[X|\theta, \delta]}{D[X]} = 1 - \sum (x_{ij} - M[X_{ij}])^2 / \sum (x_{ij} - \bar{x})^2$$

определяет «долю изменчивости», которую описывает модель. Для предложенной модели и массива оценок коэффициент детерминации составил 0,548. Если в качестве математических ожиданий взять средние по студентам, то он будет равен 0,220, если средние по оценкам, то 0,287.

Предложенная модель оказывается существенно точнее усредненных оценок. Практическое ее использование целесообразно для выделения экзаменов, отличающихся по трудностям шагов от средних значений по всем экзаменам, с целью более детального анализа и улучшения методик оценивания.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Сосницкий В.Н. Вероятностный поход к анализу успеваемости студентов / В.Н. Сосницкий, Н.И. Потанин // *Фундаментальные исследования*. — 2014. — № 8. — С. 734–738.
2. Лебедева Т.В. Статистический анализ факторов, влияющих на успеваемость студентов российских вузов / Т.В. Лебедева, А.П. Цыпин, В.С. Сидоренко // *Интеллект. Инновации. Инвестиции*. — 2016. — № 9. — С. 55–58.
3. Галимова Н.С. Построение многофакторной модели успеваемости студента / Н.С. Галимова, Л.Р. Загитова // *Международный научно-исследовательский журнал*. — 2020. — № 6-3 (96). — С. 31–36.
4. Русаков С.В. Нейросетевая модель прогнозирования группы риска по успеваемости студентов первого курса / С.В. Русаков, О.Л. Русакова, К.А. Посохина // *Современные информационные технологии и ИТ-образование*. — 2018. — № 4. — С. 815–822.
5. Гранков М.В. Анализ и кластеризация основных факторов, влияющих на успеваемость учебных групп вуза / М.В. Гранков, В.М. Аль-Габри, М.Ю. Горлова // *Инженерный вестник Дона*. — 2016. — № 4 (43). — С. 57.
6. Братищенко В.В. Статистический анализ экзаменационных оценок / В.В. Братищенко // *Известия Иркутской государственной экономической академии (Байкальский государственный университет экономики и права) (электронный журнал)*. — 2011. — № 3. — URL: <http://brj-bguer.ru/reader/article.aspx?id=8014>.
7. Шафоростова Е.Н. Проблемы внедрения информационной системы контроля качества обучения студентов / Е.Н. Шафоростова, Т.И. Лазарева // *Вестник Белгородского государственного технологического университета им. В.Г. Шухова*. — 2010. — № 3. — С. 173–176.
8. Нейман Ю.М. Введение в теорию моделирования и параметризации педагогических тестов / Ю.М. Нейман, В.А. Хлебников. — Москва : Прометей, 2000. — 168 с.
9. Айвазян С.А. Байесовский подход в эконометрическом анализе / С.А. Айвазян // *Прикладная эконометрика*. — 2008. — № 1. — С. 93–130.

REFERENCES

1. Sosnitskiy V.N., Potanin N.I. A probabilistic approach to analysis of the student performance. *Fundamental'nye issledovaniya = Fundamental research*, 2014, no. 8, pp. 734–738. (In Russian).
2. Lebedeva T.V., Tsy-pin A.P., Sidorenko V.S. Statistical analysis of factors influencing on students progress in russian universities. *Intellekt. Innovatsii. Investitsii = Intelligence. Innovations. Investments*. 2016, no. 9, pp. 55–58. (In Russian).
3. Galimova N.S., Zagitova L.R. Construction of a multi-factor model of the students' academic performance. *Mezhdunarodnyi nauchno-issledovatel'skii zhurnal = International Research Journal*, 2020, no. 6-3, pp. 31–36. (In Russian).
4. Rusakov S.V., Rusakova O.L., Posokhina K.A. Neural network model of predicting the risk group for the accession of students of the first course. *Sovremennye informatsionnye tekhnologii i IT-obrazovanie = Modern Information Technology and IT-education*, 2018, no. 4, pp. 815–822. (In Russian).

5. Grankov M.V., Al-Gabri W.M., Gorlova M.U. Analysis and clustering of basic factors affect on academic performance of university groups. *Inzhenernyy vestnik Dona = Engineering journal of Don*, 2016, no. 4, pp. 57. (In Russian).

6. Bratishenko V.V. Statistic analysis of examination grades. *Izvestiya Irkutskoy gosudarstvennoy ekonomicheskoy akademii (Baykalskiy gosudarstvennyy universitet ekonomiki i prava) = Izvestiya of Irkutsk State Economics Academy (Baikal State University of Economics and Law)*, 2011, no. 3. Available at: <http://brj-bguerp.ru/reader/article.aspx?id=8014>. (In Russian).

7. Shaforostova E.N., Lazareva T.I. Problems of introducing an information system for monitoring the quality of student learning. *Vestnik Belgorodskogo gosudarstvennogo tekhnologicheskogo universiteta im. V.G. Shukhova = Bulletin of bstu named after V.G. Shukhov*, 2010, no. 3, pp. 173–176. (In Russian).

8. Neiman Yu.M., Khlebnikov V.A. Introduction to the theory of modeling and parametrization of pedagogical tests. Moscow, Prometei Publ., 2000. 168 p.

9. Aivazian S. Bayesian methods in econometrics. *Prikladnaya ekonometrika = Applied econometrics*, 2008, no. 1, pp. 93–130. (In Russian).

ИНФОРМАЦИЯ ОБ АВТОРЕ

Братищенко Владимир Владимирович — кандидат физико-математических наук, доцент, кафедра математических методов и цифровых технологий, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: vvb@bgu.ru.

INFORMATION ABOUT THE AUTHOR

Vladimir V. Bratishenko — PhD in Physics and Mathematics, Department of Mathematical Methods and Digital Technologies, Baikal State University, Irkutsk, Russian Federation, e-mail: vvb@bgu.ru.

ДЛЯ ЦИТИРОВАНИЯ

Братищенко В.В. Пошаговая модель с латентными параметрами для оценок студентов / В.В. Братищенко. — DOI: 10.17150/2713-1734.2021.8(3).188-198 // *System Analysis & Mathematical Modeling*. — 2021. — Т. 3, № 3. — С. 188–198.

FOR CITATION

Bratishenko V.V. Step-by-Step Model with Latent Parameters for Student Assessments. *System Analysis & Mathematical Modeling*, 2021, vol. 3, no. 3, pp. 188–198. DOI: 10.17150/2713-1734.2021.8(3).188-198. (In Russian).