ВЕСТНИК ЮГОРСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
2023 г. Выпуск 2 (69). С. 9–16

УДК 519.718.7
DOI: 10.18822/byusu20230209-16

О НЕКОТОРЫХ ПОДХОДАХ ПРЕОБРАЗОВАНИЯ ТЕСТОВЫХ БАЛЛОВ В ОТМЕТКИ

Коротков Михаил Геннадьевич
кандидат физико-математических наук,
доцент Инженерной школы цифровых технологий,
Югорский государственный университет
Ханты-Мансийск, Россия
E-mail: mkorotkov@rambler.ru

Петров Алексей Аверьянович
кандидат физико-математических наук,
доцент Инженерной школы цифровых технологий,
Югорский государственный университет
Ханты-Мансийск, Россия
E-mail: averpetrov@mail.ru

Предмет исследования: компьютерное тестирование для оценки результатов освоения учебной дисциплины в высшем учебном заведении.

Цель исследования: проблема субъективности определения пороговых значений при переводе тестовых баллов (от 0 до 100) в привычную пятибалльную шкалу отметок.

Методы и объекты исследования: предлагается применение методов математической статистики для обработки данных результатов тестирования при определении пороговых значений тестовых баллов.

Основные результаты исследования: рассмотрены и оценены подходы, применяемые в настоящее время на практике. Предложены варианты альтернативных подходов решения проблемы.

Ключевые слова: обработка статистических данных, тестирование, тестовый балл, шкала отметок, распределение по отметкам.

ABOUT SOME APPROACHES FOR CONVERTING TEST SCORES INTO GRADES

Mikhail G. Korotkov
Candidate of Physical and Mathematical Sciences
Associate Professor of the Engineering School of Digital Technologies,
Yugra State University
Khanty-Mansiysk, Russia
E-mail: mkorotkov@rambler.ru

Aleksey A. Petrov
Candidate of Physical and Mathematical Sciences
Associate Professor of the Engineering School of Digital Technologies,
Yugra State University
Khanty-Mansiysk, Russia
E-mail: averpetrov@mail.ru

Subject of research: Computer testing for evaluating the results of mastering an academic discipline in a higher educational institution.

Purpose of research: The problem of subjectivity in determining threshold values when translating test scores (from 0 to 100) into the usual five-point scale of marks.

Methods and objects of research: It is proposed to use methods of mathematical statistics for processing test results data in determining the threshold values of test scores.

Main results of research: The approaches currently used in practice are considered and evaluated. Variants of alternative approaches to solving the problem are proposed.

Keywords: processing of statistical data, testing, test score, grade scale, grade distribution.

Введение

Распространение компьютерного тестирования для оценки результатов освоения учебной дисциплины в высшем учебном заведении зачастую приводит к необходимости перевода тестовых баллов (от 0 до 100) в привычную пятибалльную шкалу отметок. В соответствии с устоявшейся на практике терминологией в статье термины «отметка» и «оценка» будут использоваться как синонимы, несмотря на их принципиальное отличие друг от друга.

Пятибалльная шкала отметок является порядковой, в то время как тестовые баллы относятся, в общем случае, к интервальной шкале. В случае результатов прохождения конкретного теста ее можно рассматривать как шкалу отношений.

Перевод баллов тестирования в отметки происходит либо с использованием таблицы или номограммы соответствия некоторого диапазона тестовых баллов отметкам, либо применяется некоторое математическое выражение, позволяющее определить отметку [1, 2, 3].

В первом случае определение интервалов, соответствующих отметкам, представляется и является субъективным и не учитывает множество скрытых факторов.

Применение математических выражений для указанных выше целей тоже не лишено недостатков. Например, в [2] предлагается соотношение вида

$$Y = 3{,}3 \cdot lg\left(\frac{100}{100-X}\right), \tag{1}$$

где Y – отметка по пятибалльной шкале, X- тестовые баллы.

Здесь представлена нелинейная зависимость величины отметки от тестового балла. Если даже не учитывать проблему с областями определений и значений в (1), просматривается завышенный уровень требований к знаниям испытуемых. Например, отметке «3» соответствует значение $X = 75$, что, очевидно, сильно завышено. Кроме того, здесь не учитывается сложность конкретного теста.

В качестве еще одного варианта зависимости отметки от тестового балла в этих работах рассматривается зависимость вида

$$Y = aX^n + b, \tag{2}$$

где a, b, n – коэффициенты, подлежащие определению.

Из них нас будет интересовать коэффициент n, определяющий тип зависимости.

При $n = 1$ мы получаем линейную зависимость, $n < 1$соответствует сублинейной зависимости, $n > 1$ дает надлинейную зависимость.

В случае линейной зависимости наблюдается прямая пропорциональная зависимость между оценками и индивидуальным тестовым баллом. Это самая простая зависимость, но на практике она, как правило, не используется.

Сублинейные зависимости означают, что при переходе к отметкам больший интерес представляет возможность лучшей классификации в области низких баллов для сложных тестов. Надлинейные зависимости предпочтительнее применять, в области высоких оценок для простых тестов.

Даже в случае применения математических выражений при переводе тестовых баллов в отметки, очевидно, сохраняется проблема субъективности выбора типа зависимости и значений используемых параметров, не говоря о признании теста простым или сложным.

Результаты и обсуждение

Постановка задачи

В данной работе авторами предпринята попытка уменьшить субъективный фактор при переводе тестовых баллов в отметки. Кроме этого, сравнить полученные результаты с ситуацией привлечения мнения экспертов по формированию таблиц перевода баллов тестирования в отметки пятибалльной шкалы.

В дальнейшем будем предполагать, что, рассматривая большое количество тестируемых, значение тестового балла можно считать непрерывной случайной величиной.

Шкала отметок, являясь порядковой, разбивает тестируемых на непересекающиеся группы, внутри каждой из них распределение тестового балла близко к нормальномузакону. Это можно считать следствием ЦПТ [4], если разбиение на эти группы проведено корректно.

Вышеизложенное позволяет предположить возможность выделять интервалы тестовых баллов, соответствующих определенным отметкам поэтапно. Например, сначала выделить интервал, соответствующий отметке «2» (что иногда можно рассматривать как отметку «не зачтено»). Повторяя эту процедуру, можно выявить интервалы тестовых баллов для остальных отметок.

Пусть из k (количество тестируемых) студентов, k_1человек получили отметку «2» (неудовлетворительно), k_2 – остальные отметки.

Предполагая нормальность распределения тестовых баллов внутри этих двух групп, обозначим μ_1, σ_1 и μ_2, σ_2 – математические ожидания и среднеквадратические отклонения. Тогда:

$$f_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \tag{3}$$

$$f_2(x) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{\frac{(x-\mu_2)^2}{2\sigma_2^2}} \tag{4}$$

плотности распределения, а

$$F_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{(x-\mu_1)^2}{2\sigma_1^2}} dx, \tag{5}$$

$$F_2(x) = \frac{1}{\sigma_2\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{(x-\mu_2)^2}{2\sigma_2^2}} dx \tag{6}$$

функции распределения для каждой из этих групп, соответственно [5].

Разобьем диапазон баллов тестирования [0, 100] на N промежутков одинаковой длины.

Обозначим $Fact_i$ фактическую частоту попадания в -ый промежуток, а $Teor1_i$, $Teor1_i$ – теоретические частоты для первой и второй групп в отдельности.

$$Teor1_i = k_1\big(F_1(x_i) - F_1(x_{i-1})\big) \tag{7}$$

$$Teor2_i = k_2\big(F_2(x_i) - F_2(x_{i-1})\big) \tag{8}$$

Тогда теоретическая частота попадания в -ый промежуток $Teor_i = Teor1_i + Teor2_i$.

Введем в рассмотрение квадратичный функционал

$$\Phi = \sum_{i=1}^{N}(Fact_i - Teor_i)^2 \tag{9}$$

Значения величин $k_1, k_2, \mu_1, \mu_2, \sigma_1, \sigma_2$ находим, минимизируя функционал (9)[6], учитывая их неотрицательность, требования к математическим ожиданиям $\mu_1 < \mu_2 \leq 100$ и ограничения на среднеквадратические отклонения с учетом «правила трех сигм» для попадания теоретических баллов в диапазон от 0 до 100.

Граничное значение тестового балла, разделяющее отметку «2» от остальных находим как локальный минимум функции

$$h(x) = k_1 f_1(x) + k_2 f_2(x). \tag{10}$$

Проиллюстрируем изложенное выше на примере реального тестирования в ЮГУ (по естественно-научной дисциплине), в котором $k = 130, N = 10$.

Фактическое распределение результатов тестирования представлено на Рис. 1, а в Таб. 1 – значения статистик.

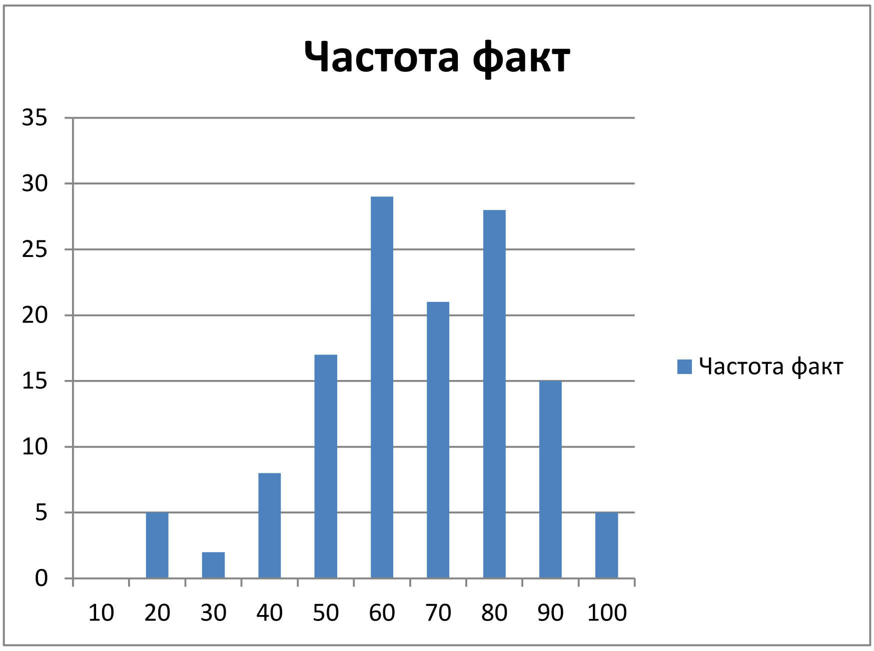

Рисунок 1 – Фактическое распределение результатов тестирования

Таблица 1

Среднее	61,67
Медиана	62,96
Стандартное отклонение	18,03
Асимметричность	-0,46
Минимум	12,96
Максимум	96,30
Счет	130

Наблюдаемая существенная отрицательная асимметричность позволяет предположить наличие в выборке группы студентов, получивших низкие тестовые баллы. Этот же факт наблюдается и на Рис. 1.

Теоретические и фактические частоты представлены на Рис. 2. Теоретические частоты получены в результате минимизации функционала (9).

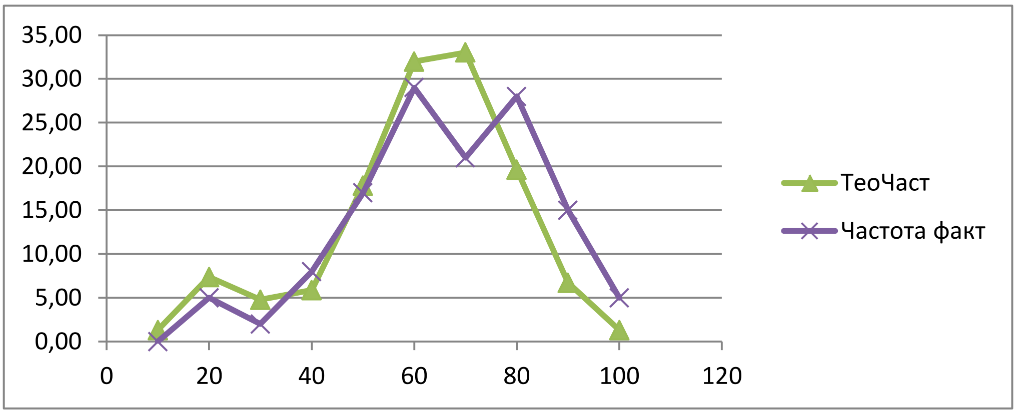

Рисунок 2 – Теоретические и фактические частоты

Граничное значение тестового балла, разделяющее отметку «2» от остальных, находим как локальный минимум функции (10). В нашем примере пороговое значение отметок «3» и выше равно 29,07 (смотри Рис. 3). При необходимости пороговое значение округляется.

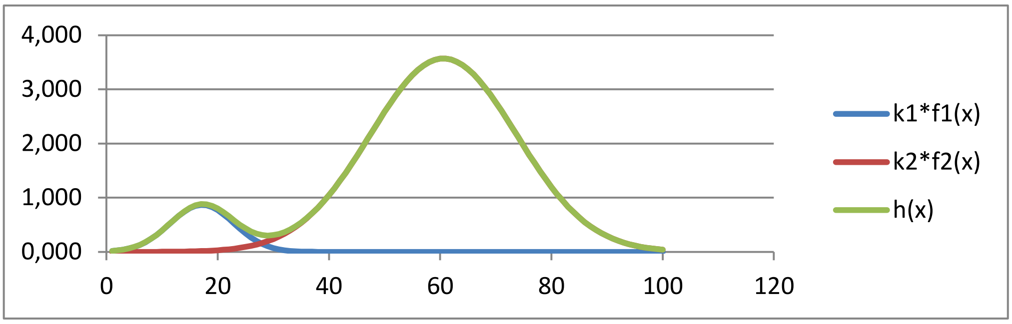

Рисунок 3

В итоге количество студентов, получивших отметку «2», оказалось равным 7. Распределение остальных студентов представлено на Рис. 4, а в Таб. 2 – значения соответствующих статистик.

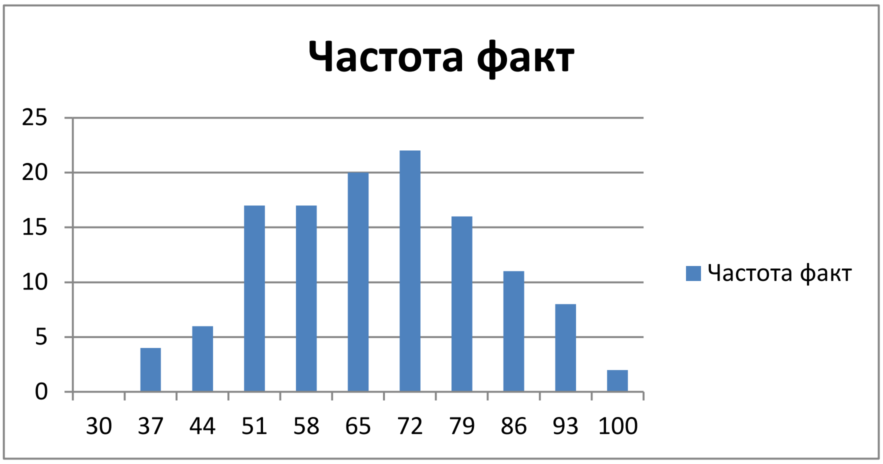

Рисунок 4

Таблица 2

Среднее	64,09
Медиана	62,96
Стандартное отклонение	15,27
Асимметричность	0,01
Минимум	30,56
Максимум	96,30
Счет	123

Повторяя процедуру выделения пороговых значений тестовых баллов сначала для отметки «4», а затем для отметки «5», были найдены значения 64,59 и 80,14, соответственно. В результате применения рассматриваемого подхода в данном примере получено следующее распределение студентов по отметкам: Рис. 5, Таб. 3.

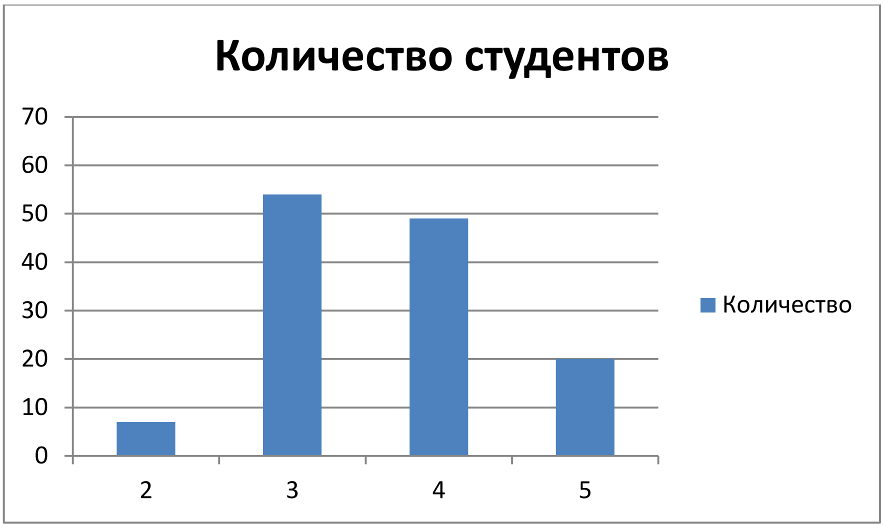

Рисунок 5

Таблица 3

Отметка	Количество студентов
2	7
3	54
4	49
5	20

Чтобы сравнить полученные результаты с ситуацией привлечения экспертного мнения по формированию таблиц перевода баллов данного теста в отметки пятибалльной шкалы, в Таб. 4 представлены пороговые значения, предложенные шестью экспертами.

Таблица 4

Отметка	Эксперт А	Эксперт Б	Эксперт В	Эксперт Г	Эксперт Д	Эксперт Е
3	45	40	50	55	60	45
4	60	60	70	75	75	70
5	75	70	90	90	90	85

Как видно из таблицы, субъективные мнения экспертов о сложности теста и требованиях к освоению учебной дисциплины существенно различаются, что вполне ожидаемо. Распределения студентов по отметкам в зависимости от эксперта представлены на Рис. 6, а на Рис. 7 – распределение при пороговых значениях, усредненных по экспертам. Усреднение проводилось посредством нахождения простого среднего арифметического, что не могло учесть различия в квалификации привлеченных экспертов.

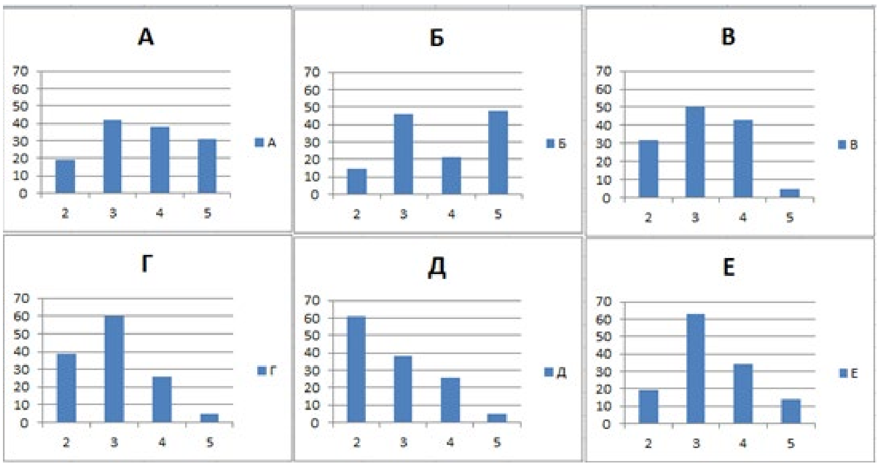

Рисунок 6

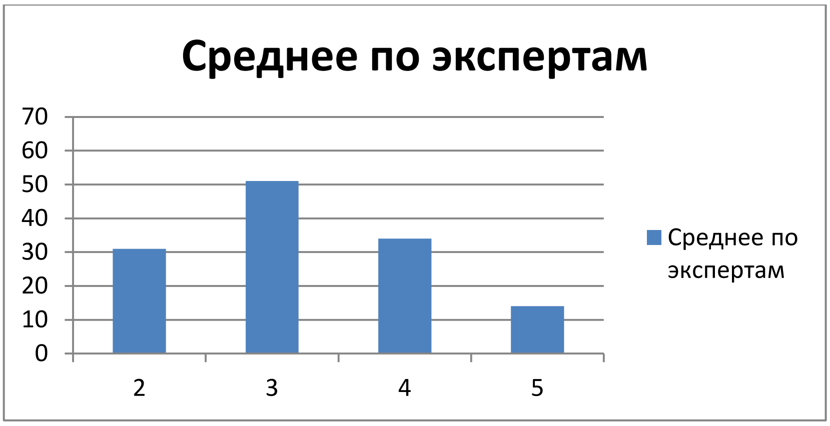

Рисунок 7

Все распределения студентов по отметкам, полученные с учетом мнения экспертов, включая осредненное, статистически (при 1% уровне значимости) отличаются друг от друга. И нет объективных оснований предпочесть мнение одного эксперта или выбрать осредненное.

Заключение и выводы

Рассмотренная в работе проблема перехода от тестовых баллов, полученных студентом в результате прохождения теста по учебной дисциплине, к пятибалльной шкале отметок имеет несколько аспектов. Перечислим некоторые, лежащие на поверхности. С одной стороны, правила перевода тестовых баллов в отметки должны быть сформулированы до начала прохождения теста. В этом случае невозможно избавиться от крайней субъективности, формулируя их как жесткие или в форме "рекомендаций". С другой стороны, попытка уменьшить влияние субъективного фактора в этом вопросе, предложенная в данной работе, предполагает многократное использование конкретного теста с предварительным его прохождением для накопления достаточной статистики. И только после этого проведения тестирования "на отметку" с четкими правилами. Но и в этом случае проблема субъективности не решается в полной мере. Например, остается открытым вопрос выбора количества интервалов N. Вероятно, можно пытаться решить его, применяя инструменты математической статистики.

По мнению авторов, возможен несколько иной подход к компьютерному тестированию, как инструменту проверки уровня освоения учебной дисциплины, неявно включающий в себя элементы классического экзамена. Он заключается в использовании многоуровневого тестирования. Тестируемый должен последовательно пройти задания уровня отметки "3", уровня "4" и уровня "5", причем, переход к следующему уровню становится возможным только при условии получения "зачета" по предыдущему. Тестовые задания различных уровней должны не пересекаться, различаться по типу и возрастать по степени сложности при возрастании уровня отметки. Наиболее напрашивающимся такой подход представляется при изучении естественно-научных дисциплин. Например, к уровню "3" можно отнести задания теста, относящиеся к определениям изучаемых в дисциплине объектов, терминологии, понятийному аппарату. Без прохождения этого уровня бессмысленно говорить о более сложных вещах. На уровне "4" задания относятся к свойствам изучаемых объектов и их взаимодействию друг с другом. Уровень "5" предполагает применение знаний при решении практических и ситуационных задач.

Многоуровневый подход, по-видимому, исключит ситуацию, когда в рамках одного теста студент может, например, верно сформулировать условие монотонности дифференцируемой функции, дав перед этим неверные определения понятиям "функция", "монотонность", "дифференцируемость".

Создание многоуровневого теста для проверки уровня освоения учебной дисциплины технически, очевидно, более трудоемко по сравнению с "обычным" тестом. Тем не менее, определение порогового значения для получения «зачета» в рамках одного уровня является более простой задачей, чем выделение уровней отметок "3", "4", "5".

Литература

1. Дубас, В. Об оценивании знаний при программированном контроле / В. Дубас. – Текст : непосредственный // Физика в школе. – 1990. – № 3. – С. 83.

2. Молибог, А. Г. Программированное обучение (вопросы научной организации педагогического труда) / А. Г. Молибог. М., Высшая школа, 1967. – 243 с. – Текст : непосредственный.

3. Звонников, В. И. Контроль качества обучения при аттестации (компетентностный подход) / В. И. Звонников, М. Б. Челышкова. – М. : Логос, 2009. – 272 с. – Текст : непосредственный.

4. Чистяков, В. П. Курс теории вероятностей / В. П. Чистяков. – М.: Наука, 1987. – Текст : непосредственный.

5. Ивченко, Г. И. Математическая статистика : учеб. / Г. И. Ивченко, Ю. И. Медведев. – М. : Книжный дом «ЛИБРОКОМ», 2014. – 352 с. – Текст : непосредственный.

6. Шукаев, Д. Н. Прикладные методы оптимизации : учеб. / Д. Н. Шукаев. – М. : Издательский дом Академии Естествознания, 2017. – Текст : непосредственный.