УДК 65.011, 519.256, 004.043

ОЦЕНКА ЧИСЛА ПОБЕДИТЕЛЕЙ ЗАКЛЮЧИТЕЛЬНЫХ ЭТАПОВ ОЛИМПИАД

В.Н. Васильев, Т.В. Харченко, В.В. Клименко

Описан математический подход оценивания результатов участников массовых олимпиад, приведены результаты численных расчетов, демонстрирующие полученные оценки. Предлагаемый подход может применяться методическими комиссиями и жюри массовых олимпиад при разработке заданий этапов олимпиад, определении численности участников заключительных этапов и количества победителей и призеров. В рамках математического подхода оценивания результатов участников массовых олимпиад используются данные об индексах решаемости заданий отборочного этапа, распределении участников отборочного этапа по типам решенных задач и набранным баллам.

Ключевые слова: математическая статистика, олимпиады школьников, критерии оценки результатов.

Введение

Ежегодно в цикле олимпиад по информатике, проводимых ГОУ ВПО «СПбГУ ИТМО», принимает участие порядка 4 000 школьников 11 классов из всех федеральных округов Российской Федерации. Олимпиады проводятся в два этапа: отборочный (дистанционный) и заключительный (очный). Отборочный этап разбит на три тура и длится около четырех месяцев. В ходе отборочного этапа участникам предлагаются задачи всех разделов информатики, соответствующие школьной программе. К заключительному этапу допускаются участники, набравшие проходной балл, который устанавливается жюри олимпиады.

Методическая комиссия олимпиады разрабатывает задачи отборочного и заключительного этапов, руководствуясь утвержденными организационным комитетом Критериями определения победителей и призеров. Так, в соответствии с Критериями, победителем становится участник, решивший обе творческие задачи по технологиям программирования при условии правильного решения не менее 9 из 10 задач по общим вопросам информатики и информационно-коммуникативных технологий (ИКТ) уровней воспроизведения и применения. Призером становится участник, который решил 9 из 10 задач по общим вопросам ИКТ и не справился или не решал творческие задачи. Таким образом, можно утверждать, что призер олимпиады уверенно владеет материалом информатики и ИКТ на уровнях воспроизведения и применения, но еще не достигает творческого уровня владения материалом. Методическая комиссия и жюри олимпиады определяют баллы за задачи заключительного этапа с учетом описанных выше Критериев. При определении победителей и призеров этапа олимпиады необходимо учитывать дополнительно следующие требования: количество победителей и призеров не должно превосходить 45%, а количество победителей не должно превосходить 10% от числа участников этапа.

В настоящее время актуальной является задача определения проходного балла p в заключительный этап олимпиады, поскольку среди не допущенных в заключительный этап могли быть те, кто смог бы справиться с предложенными задачами и войти в состав победителей и призеров. Для решения этой задачи требуется оценить вероятности решения задач участниками заключительного этапа и соответственно оценить численность победителей и призеров заключительного этапа в зависимости от их результатов в отборочном этапе, предполагаемого состава и уровня сложности заданий заключительного этапа.

Оценка числа участников

В цикле олимпиад 2009–2010 учебного года к заключительному этапу были допущены участники отборочного этапа, набравшие 23 и более баллов из 90 возможных баллов. Таким образом, проходной балл p=23. Количество участников заключительного этапа – 931 школьник 11 классов. Фактически к участию в заключительном этапе допущено менее трети участников отборочного этапа. На рис. 1 представлено распределение участников по набранным баллам по итогам *отборочного* этапа среди школьников выпускных классов в цикле олимпиад по информатике 2009–2010 учебного года.

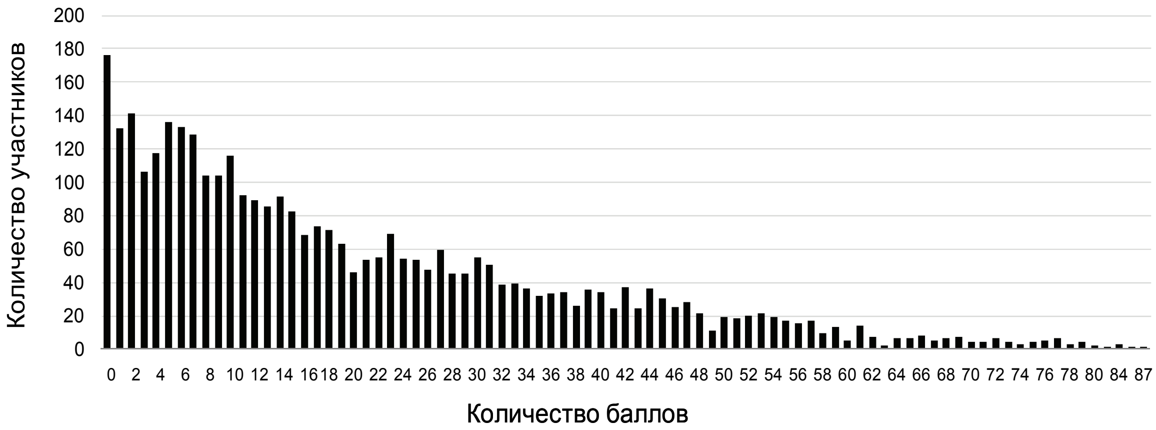

Рис. 1. Гистограмма распределения участников отборочного этапа по набранным баллам

Каждой задаче заключительного этапа соответствует по теме и уровню сложности группа заданий отборочного этапа. По этой группе заданий отборочного этапа рассчитан средний индекс решаемости. Проведено сопоставление индексов решаемости заданий заключительного этапа с усредненными индексами решаемости соответствующих им групп заданий отборочного этапа. Полученные результаты приведены в табл. 1.

№ задания	1	2	3	4	5	6	7	8	9	10	11	12
Заключительный	45,86	43,39	35,23	44,68	44,36	68,85	29,43	41,03	40,28	53,38	26,42	27,71
Отборочный	29,43	39,42	33,30	33,51	8,92	40,49	29,32	34,59	32,01	46,62	15,79	17,08

Таблица 1. Индексы решаемости заданий (%) по этапам олимпиады

Как видно из табл. 1, не существует прямой зависимости между индексом решаемости задач отборочного этапа и индексом решаемости соответствующих типов задач заключительного этапа. Для построения оценки количества победителей и призеров заключительного этапа в зависимости от значения проходного балла при p<23 необходимо определить принципы назначения баллов за задачи заключительного этапа. Рассматривается несколько подходов: экспертная оценка (вариант “A”), оценка по индексу решаемости соответствующих заданий в отборочном этапе (вариант “B”), оценка по индексу решаемости задач очного этапа (вариант “C”).

Номер задачи	1	2	3	4	5	6	7	8	9	10	11	12
	Количество баллов											
вариант “A”	1	2	1	1	2	2	2	1	2	1	3	3
вариант “B”	5	3	4	4	10	2	5	4	4	1	8	8
вариант “C”	5	6	7	6	7	1	9	6	7	4	10	10

Таблица 2. Распределение баллов по задачам

С использованием данных табл. 2 и фактических результатов участников построены распределения участников заключительного этапа по набранным баллам для вариантов “A” и “B” распределения баллов по задачам (см. рис. 2, 3). Рассмотрение варианта “C” детально не проводится в связи с тем, что этот вариант представляет собой апостериорную оценку уровня сложности заданий заключительного этапа.

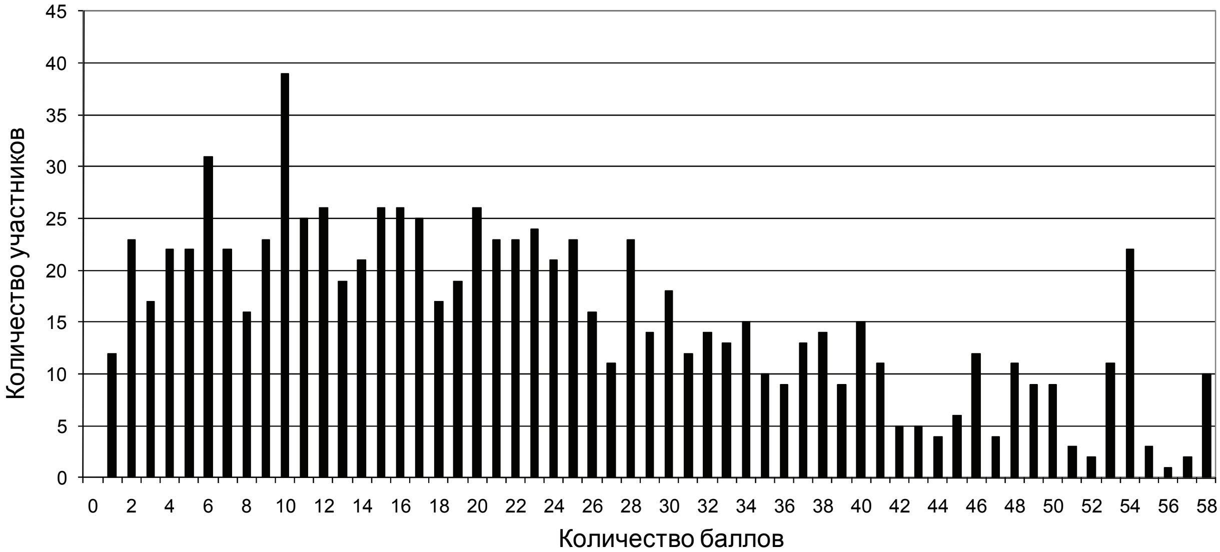

Рис. 2. Гистограмма распределения участников заключительного этапа по количеству набранных баллов. Количество участников – 931. Вариант “B” распределения баллов по задачам (см. табл. 2)

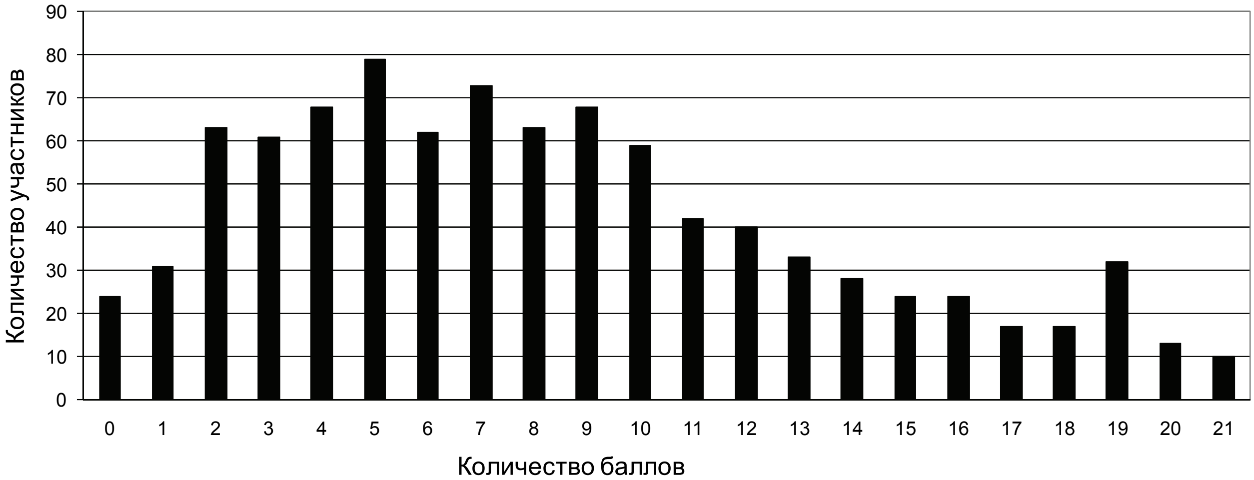

Рис. 3. Гистограмма распределения участников заключительного этапа по количеству набранных баллов. Количество участников – 931. Вариант “A” распределения баллов по задачам (см. табл. 2)

Кроме того, проведена оценка «погрешности» вариантов “A”, “B” и “C” с точки зрения критериев определения победителей и призеров. Погрешность возникает, когда победитель или призер определяется по числу набранных баллов, а не по составу решенных заданий, и заключается в том, что участник фактически решил набор заданий, удовлетворяющий критерию «победитель», но по набранным баллам отнесен к «призерам», и наоборот.

	Погрешность: победитель – призер	Общая погрешность
вариант “A”	89 участников	144 частника
вариант “B”	35 участников	52 участника
вариант “C”	5 участников	22 участника

Таблица 3. Погрешности методов в определении состава групп призеров и победителей

Из двух вариантов “A” и “B” предпочтительным является вариант “B”, поскольку он дает меньшую величину погрешности по сравнению с вариантом “A” (см. табл. 3).

Для получения оценки количества победителей и призеров заключительного этапа олимпиады в зависимости от проходного балла при $p < 23$ используется метод экстраполяции (см. [4]).

Пусть K – максимальное количество баллов отборочного этапа, N – максимальное количество баллов заключительного этапа. Вводятся случайные величины X и $Y^{(n)}$, $0 \le n \le N$. Значения x_k случайной величины X – число участников, набравших k баллов в отборочном этапе. Значения $y_k^{(n)}$ случайной величины $Y^{(n)}$ – число участников, набравших k баллов в отборочном этапе и n баллов в заключительном этапе. Таким образом, рассматриваются значения ($y_k^{(n)}$, x_k) двумерной случайной величины ($Y^{(n)}$, X). Для оценки числа победителей и призеров заключительного этапа в зависимости от проходного балла при $p < 23$ рассмотрим линейную регрессию случайной величины $Y^{(n)}$ на X (см. [4, 5]). Используем уравнение линейной регрессии в виде [1]

$$y = m_Y + \frac{\rho_{XY}\sigma_Y}{\sigma_X}\left(x - m_X\right), \tag{1}$$

где m_X, m_Y – выборочные средние значения случайных величин X и Y, σ_X, σ_Y – выборочные среднеквадратичные отклонения случайных величин X и Y, ρ_{XY} – выборочный коэффициент корреляции случайных величин X и Y. Уравнение (1) дает оценку значений случайной величины, определяемых теоретическим уравнением регрессии, наилучшим образом в смысле принципа наименьших квадратов [1, 2]. Подобно формуле (1), выражения для частных уравнений эмпирической линейной регрессии для оценки значений $Y^{(n)}$ имеют вид

$$y_k^{(n)} = m_{Y^{(n)}} + \frac{\rho_{XY}^{(n)}\sigma_{Y^{(n)}}}{\sigma_X}\left(x_k - m_X\right), \tag{2}$$

где $m_X, m_{Y^{(n)}}$ – выборочные средние значения случайных величин X и $Y^{(n)}$, $\sigma_X, \sigma_{Y^{(n)}}$ – выборочные среднеквадратичные отклонения случайных величин X и $Y^{(n)}$, где $\rho_{XY}^{(n)}$ – выборочный частный корреляционный коэффициент случайных величин X и $Y^{(n)}$ при фиксированном n. Таким образом, получаем случайную величину $R = \{\rho_{XY}^{(n)},\ n=1,\ldots,M\}$ из значений частных корреляционных коэффициентов со средним значением $E(R) = 0{,}603$ и дисперсией $D(R) = 0{,}453$. Используя тот факт, что величина $t = \dfrac{Y - \overline{Y}}{s_Y}$ (величина s_Y пропорциональна σ_Y) имеет t-распределение Стьюдента с 60 степенями свободы [2], проведены расчеты величин d_p 95-процентных доверительных интервалов для оценки значений случайной величины $Y^{(n)}$ в интервале $10<p<50$ значений проходного балла, полученных из уравнения (2). Полученные значения величин d_p для отдельных p приведены в табл. 4.

Проходной балл, p	50	45	40	35	30	25	20	15	10
Величина, d_p	2,24	2,67	2,87	3,09	3,38	3,46	3,52	3,62	3,86

Таблица 4. Размеры доверительных интервалов при некоторых значениях проходного балла

На основании расчитанных значений случайных величин $Y^{(n)}$ построено семейство $\{\Psi_p\}$ распределений числа участников по набранным баллам в заключительном этапе в зависимости от проходного балла p. По каждому из распределений Ψ_p случайной величины $Y^{(n)}$ рассчитываются величины λ_p, μ_p, λ_{10}^p, μ_{45}^p. Здесь p – проходной балл в заключительный этап, λ_p – количество победителей этапа, μ_p – число победителей и призеров этапа, λ_{10}^p – количество участников, составляющих 10% от числа участников заключительного этапа, μ_{45}^p – количество участников, составляющих 45% от числа участников заключительного этапа.

Анализ полученных результатов

Проведенные расчеты значений случайной величины $Y^{(n)}$ позволяют с учетом результатов отборочного этапа – индексов решаемости заданий и распределения участников по набранным баллам – оценить при $p < 23$ количество победителей заключительного этапа с учетом критериев определения победителей и призеров, а также проанализировать выполнение условий $\lambda_p < \lambda_{10}^p$ и $\mu_p < \mu_{45}^p$ в зависимости от значений проходного балла $p < 23$.

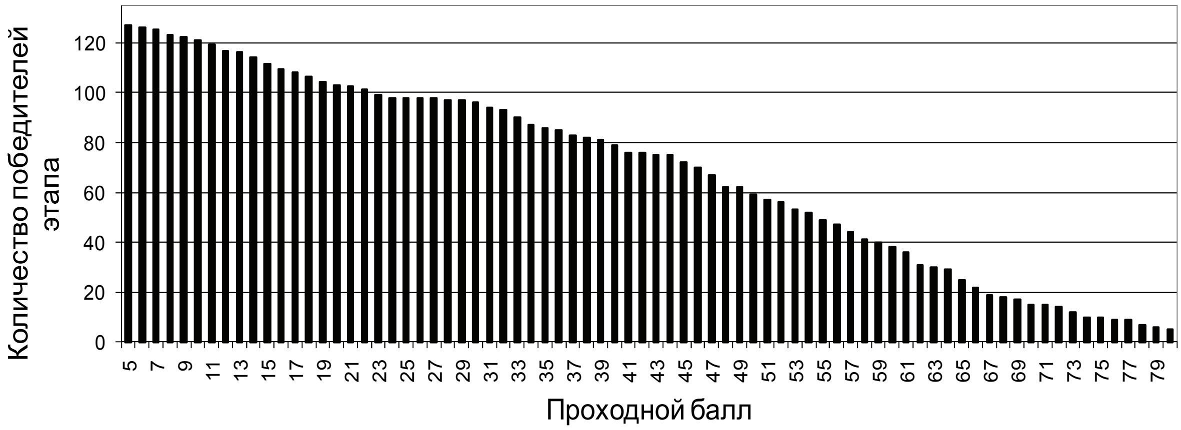

Рис. 4. Рассчитанная численность победителей от величины проходного балла

Рис. 4 показывает равномерный рост численности победителей заключительного этапа при уменьшении значения проходного балла.

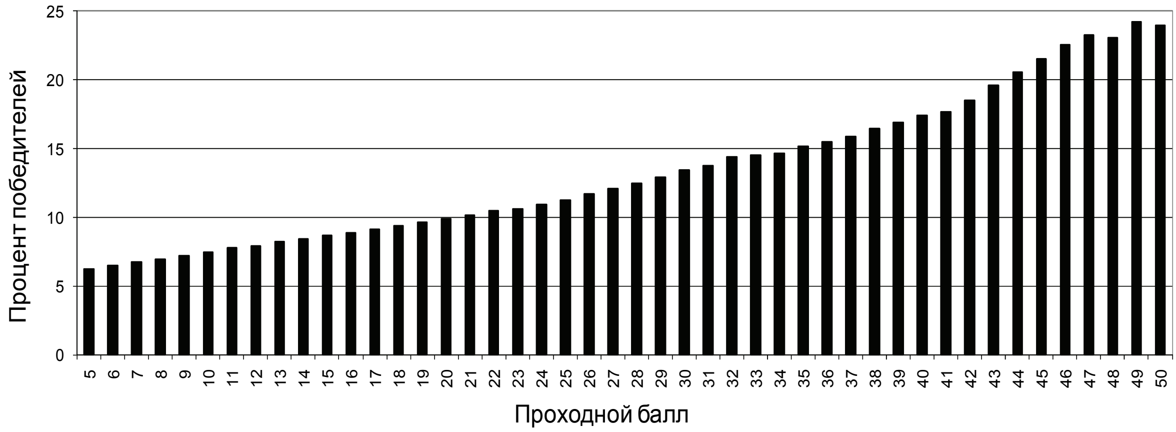

Рис. 5. Процент победителей этапа

Гистограмма на рис. 5 иллюстрирует рассчитанные значения процента победителей заключительного этапа от значения проходного балла, полученные с использованием распределения участников отборочного этапа по баллам (см. рис. 1), индексов решаемости заданий (см. табл. 1) и варианта “B” распределения баллов по задачам (см. табл. 3). Как следует из рис. 5, при значении проходного балла 21 и выше число победителей превышает 10%. Данный факт необходимо принимать во внимание при определении проходного балла в заключительный этап. Можно показать, что для проходного балла, меньшего 37, общее число победителей и призеров не превосходит 45% от числа участников этапа. Таким образом, если проходной балл $p \geq 21$, то условия $\lambda_p < \lambda_{10}^p$ и $\mu_p < \mu_{45}^p$ выполняются.

Заключение

В работе средствами математической статистики построены оценки числа победителей заключительного этапа олимпиады в зависимости от значения проходного балла.

Показано, что существует зависимость распределения участников заключительного этапа по набранным баллам от распределения участников отборочного этапа по набранным баллам и дополнительного параметра – проходного балла в заключительный этап.

Предполагается использовать описанный в статье математический подход в цикле олимпиад 2010–2011 учебного года при определении проходного балла в заключительный этап и определении баллов за правильно решенные задания в заключительном этапе.

Литература

1. Крамер Г. Математические методы статистики. – М.: Мир, 1975. – 468 с.
2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичная обработка данных.– М.: Финансы и статистика, 1983. – 471 с.
3. Беспалько В.П. Образование и обучение с участием компьютеров (педагогика третьего тысячелетия). – М.: Изд-во Московского психолого-социального института, 2002. – 234 с.
4. Круг Г.К., Сосулин Ю.А., Фатуев В.А. Планирование эксперимента в задачах идентефикации и экстраполяции, М.: Наука, 1977. – 207 с.
5. Elfving G. Optimum allocation in linear regression theory // Ann. Math. Statist. – 1952. – № 23. – P. 255–262.

Васильев Владимир Николаевич – Санкт-Петербургский государственный университет информационных технологий, механики и оптики, доктор технических наук, профессор, ректор, vasilev@mail.ifmo.ru

Харченко Татьяна Владимировна – Санкт-Петербургский государственный университет информационных технологий, механики и оптики, ст. преподаватель, kharchenko@mail.ifmo.ru

Клименко Виктор Владимирович – Санкт-Петербургский государственный университет информационных технологий, механики и оптики, магистр математики, shurup239@mail.com