ЕГЭ: как учесть погрешность баллов и снизить ошибку педагогического оценивания

Олег Гениевич Деменчёнок,
начальник кафедры математики и информатики Восточно-Сибирского института МВД России, доцент, кандидат технических наук

• тест • тестовый балл • ЕГЭ, погрешность измерения • математические модели педагогических измерений • Partial Credit Model •

Проблемная ситуация

Результаты измерений из-за погрешностей всегда несколько отличаются от истинного значения измеряемой величины. Анализ погрешностей педагогических измерений приобретает особую актуальность ввиду косвенного характера измерения — уровни подготовленности обучаемых и параметры тестовых заданий определяются расчётным путём на основе выбранной модели измерения — и невозможности проверки полученных результатов. Поэтому учёт допущенной погрешности является необходимым условием корректности педагогического измерения.

На необходимость учёта погрешности измерений, проведения экспертизы качества контрольных материалов единого государственного экзамена неоднократно указывали многие авторы[1]. Однако в свидетельство о результатах ЕГЭ выставляются окончательные баллы по общеобразовательным дисциплинам, без указания погрешности измерения. Это правомерно только в том случае, когда погрешность измерения не превышает 0,5 балла (т.е. погрешность менее половины цены деления шкалы)[2].

В противном случае следует либо явно указать погрешность измерения (например, 76 ± 12 баллов), либо увеличить цену деления шкалы (например, до двадцати баллов: 0, 20, 40 … 80, 100 баллов). К сожалению, организаторы ЕГЭ не придают гласности отчёты с анализом погрешности результатов ЕГЭ. Независимые эксперты не могут провести такой анализ из-за завесы секретности, окутавшей ЕГЭ.

Практически вся значимая информация о ЕГЭ включена в перечень сведений конфиденциального характера, в том числе:

- критерии оценивания по предметам;
- оригиналы бланков ЕГЭ, их электронные изображения и ксерокопии;
- экзаменационные материалы, протоколы проверки заданий с развёрнутыми ответами;
- итоговые протоколы;
- аналитические и информационные материалы;
- свидетельства о результатах ЕГЭ выпускников — участников ЕГЭ;
- бланки апелляций по результатам ЕГЭ, решения конфликтной комиссии;
- резервные копии баз данных ЕГЭ.

Доступ к указанной информации ограничен в соответствии с законодательством Российской Федерации. Эти материалы запрещается публиковать в открытой печати, переписке, использовать в передачах по радио и телевидению, в публичных выступлениях. Гриф конфиденциальности делает процедуру ЕГЭ абсолютно непрозрачной, что снижает степень доверия к результатам ЕГЭ.

Однако конфиденциальность сведений ЕГЭ не является преградой для анализа погреш-

[1] **Аванесов В.С.** Являются ли КИМы ЕГЭ методом педагогических измерений? // Педагогические измерения. № 3. 2007. С. 3–26; **Хлебников В.** Краткий анализ технологии и результатов единого государственного экзамена (ЕГЭ) // Педагогические измерения. № 4. 2008. С. 25–40.

[2] **Тейлор Дж.** Введение в теорию ошибок: Пер. с англ. М.: Мир, 1985. 272 с.

ности самой применяемой технологии. Эту погрешность можно найти, исходя из количества и структуры заданий ЕГЭ, а также математической модели. Такое исследование способно «пролить свет» на реальные возможности ЕГЭ как инструмента педагогического измерения.

Примерная структура контрольных материалов ЕГЭ

В качестве примерной структуры экзамена ЕГЭ принят демонстрационный вариант контрольных материалов 2010 года по математике, подготовленный Федеральным институтом педагогических измерений и утверждённый директором этого института А.Г. Ершовым 21.07.2009 года[3] (сайт официального информационного портала ЕГЭ, http://www1.ege.edu.ru/images/stories/demo_2010/ma_demo_2010.pdf).

Экзаменационная работа состоит из двух частей, включающих 18 заданий.

Часть 1-я включает 12 заданий с кратким ответом (B1 — B12) базового уровня по материалу курса математики. За правильный ответ на задание части 1-й ставится один балл, а за неверный ответ или отсутствие ответа — 0 баллов.

Часть 2-я (С) содержит шесть сравнительно трудных заданий по материалу школьного курса математики. Максимальная оценка за задания второй части:

• за первое и второе задание — 2 балла (за частично правильный ответ — 1 балл);
• за третье и четвёртое задание — 3 балла (частично правильный ответ — от 1 до 2-х баллов);
• за пятое и шестое задание — 4 балла (частично правильный ответ — от 1 до 3-х баллов).

Полное и безошибочное выполнение всех заданий этого варианта ЕГЭ соответствует 30 баллам.

Надо отметить, что варианты контрольных материалов и по некоторым другим общеобразовательным дисциплинам обладают подобной структурой.

Параметры оценки погрешности измерений

В соответствии с примерной структурой контрольных материалов ЕГЭ для проведения пробных расчётов примем модель метода измерения (теста), включающего 18 заданий, из которых:
— 12 дихотомических заданий (т.е. заданий, оцениваемых 0 или 1 баллом);
— 6 политомических заданий (максимальный балл за первое и второе задание — 2 балла, за третье и четвёртое — 3 балла, пятое и шестое — 4 балла).

К сожалению, нет никаких сведений о возможном распределении уровней трудности заданий. Автор полагает, что истинное распределение становится известным только после проведения ЕГЭ, так как для определения уровней трудности заданий нужна достаточно представительная статистика ответов школьников, а получить такие данные до проведения ЕГЭ трудно ввиду конфиденциальности заданий. Поэтому в первом приближении примем распределение уровней трудности заданий β для пробных расчётов, близких к равномерному:
— 12 дихотомических заданий равномерно распределены по уровню трудности от —5 до 2 (шаг равен 7/11, т.е. $\beta_1 = -5$; $\beta_2 = -4{,}36$; $\beta_3 = -3{,}73$ … $\beta_{11} = 1{,}36$; $\beta_{12} = 2$);
— 6 политомических задания имеют уровень трудности первого шага $\beta_{j\,1}=1$ (где j = 13, 14 … 18 — порядковый номер задания) и максимальный уровень трудности, равный 5. Остальные уровни трудности этих заданий равномерно распределены от 1 до 5, что соответствует следующим значениям уровней трудности:
• $\beta_{13\,1}=1$; $\beta_{13\,2}=5$;
• $\beta_{14\,1}=1$; $\beta_{14\,2}=5$;
• $\beta_{15\,1}=1$; $\beta_{15\,2}=3$; $\beta_{15\,3}=5$;
• $\beta_{16\,1}=1$; $\beta_{16\,2}=3$; $\beta_{16\,3}=5$;
• $\beta_{17\,1}=1$; $\beta_{17\,2}=2{,}33$; $\beta_{17\,3}=3{,}67$; $\beta_{17\,4}=5$;
• $\beta_{18\,1}=1$; $\beta_{18\,2}=2{,}33$; $\beta_{18\,3}=3{,}67$; $\beta_{18\,4}=5$.

Математическая модель

Согласно находящейся на сайте официального информационного портала ЕГЭ (www1.ege.edu.ru) инфор-

[3] Демонстрационный вариант контрольных измерительных материалов единого государственного экзамена 2010 года по математике. М.: Федеральное государственное научное учреждение «Федеральный институт педагогических измерений», 2009. 18 с.

мации, результаты выполнения заданий обрабатываются в рамках модификации модели Раша с произвольными промежуточными категориями выполнения тестового задания, известной в англоязычной литературе как Partial Credit Model.

PCM — это наиболее известная математическая модель педагогических измерений, с градацией степени правильности ответа (т.е. с возможностью учёта частично правильных ответов)[4]:

$$\pi_{ij\tilde{o}} = \frac{e^{\sum_{k=0}^{x}(\theta_i - \beta_{jk})}}{\sum_{l=0}^{x_{\max j}} e^{\sum_{k=0}^{l}(\theta_i - \beta_{jk})}}, \tag{1}$$

где π_{ijx} — вероятность достижения результата x_{ij} (т.е. того, что тестируемый i выполнит ровно x шагов и получит x баллов в задании j); x=0, 1… x_{ij} …x_{maxj} — количество шагов; x_{maxj} — максимально возможное количество баллов за задание j;

$$\beta_{j0} = 0, \ \sum_{n=0}^{0}(\theta_i - \beta_{j0}) = 0.$$

Например, для задания, максимально оцениваемого двумя баллами, вероятности получения одного и двух баллов соответственно равны:

$$\pi_1 = \frac{e^{\theta-\beta_1}}{1 + e^{\theta-\beta_1} + e^{\theta-\beta_1} \cdot e^{\theta-\beta_2}}, \tag{2}$$

$$\pi_2 = \frac{e^{\theta-\beta_1} \cdot e^{\theta-\beta_2}}{1 + e^{\theta-\beta_1} + e^{\theta-\beta_1} \cdot e^{\theta-\beta_2}}. \tag{3}$$

Если максимальная оценка задания равна трём баллам, имеем:

$$\pi_1 = \frac{e^{\theta-\beta_1}}{1 + e^{\theta-\beta_1} + e^{\theta-\beta_1} \cdot e^{\theta-\beta_2} + e^{\theta-\beta_1} \cdot e^{\theta-\beta_2} \cdot e^{\theta-\beta_3}}, \tag{4}$$

$$\pi_2 = \frac{e^{\theta-\beta_1} \cdot e^{\theta-\beta_2}}{1 + e^{\theta-\beta_1} + e^{\theta-\beta_1} \cdot e^{\theta-\beta_2} + e^{\theta-\beta_1} \cdot e^{\theta-\beta_2} \cdot e^{\theta-\beta_3}}. \tag{5}$$

$$\pi_3 = \frac{e^{\theta-\beta_1} \cdot e^{\theta-\beta_2} \cdot e^{\theta-\beta_3}}{1 + e^{\theta-\beta_1} + e^{\theta-\beta_1} \cdot e^{\theta-\beta_2} + e^{\theta-\beta_1} \cdot e^{\theta-\beta_2} \cdot e^{\theta-\beta_3}}. \tag{6}$$

Аналогичным образом можно применить уравнение Partial Credit Model (1) для анализа заданий с большим количеством градаций степени правильности ответа.

Для оценки стандартной ошибки измерения уровня подготовленности испытуемого i используется формула[5]:

$$\sigma_{\theta i} = \frac{1}{\sqrt{\sum_{j=1}^{m}\left(\sum_{l=1}^{x_{\max j}} l^2 \cdot \pi_{ijl} - \left(\sum_{l=1}^{x_{\max j}} l \cdot \pi_{ijl}\right)^2\right)}}, \tag{7}$$

где m — число тестовых заданий.

[4] **Wright B.D., Masters G.N.** Rating Scale Analysis: Rasch Measurement. Chicago: Mesa Press, 1982. 204 p.

[5] Там же.

Для заданий, максимально оцениваемых двумя баллами, уравнение (7) принимает вид:

$$\sigma_{\theta i} = \frac{1}{\sqrt{\sum_{j=1}^{m}\left(\pi_{ij1} + 4\pi_{ij2} - \left(\pi_{ij1} + 2\pi_{ij2}\right)^2\right)}}. \quad (8)$$

Запишем уравнение (7) для заданий, максимально оцениваемых тремя и четырьмя баллами:

$$\sigma_{\theta i} = \frac{1}{\sqrt{\sum_{j=1}^{m}\left(\pi_{ij1} + 4\pi_{ij2} + 9\pi_{ij3} - \left(\pi_{ij1} + 2\pi_{ij2} + 3\pi_{ij3}\right)^2\right)}}, \quad (9)$$

$$\sigma_{\theta i} = \frac{1}{\sqrt{\sum_{j=1}^{m}\left(\pi_{ij1} + 4\pi_{ij2} + 9\pi_{ij3} + 16\pi_{ij4} - \left(\pi_{ij1} + 2\pi_{ij2} + 3\pi_{ij3} + 4\pi_{ij4}\right)^2\right)}}. \quad (10)$$

Оценки стандартной ошибки измерения уровня подготовленности испытуемого для анализа заданий с большим количеством градаций степени правильности ответа находятся аналогичным образом.

Для дихотомических заданий ($x_{max\,j}$ = 1) формула (7) существенно упрощается:

$$\sigma_{\theta i} = \frac{1}{\sqrt{\sum_{j=1}^{m}\left(\pi_{ij1} - \pi_{ij1}^{2}\right)}} = \frac{1}{\sqrt{\sum_{j=1}^{m}\pi_{ij1}(1-\pi_{ij1})}} = \frac{1}{\sqrt{\sum_{j=1}^{m}P_{ij}(1-P_{ij})}}, \quad (11)$$

где $\pi_{ij1} = P_{ij}$ — вероятность получения испытуемым *i* одного балла в задании *j*, что для дихотомического задания соответствует вероятности правильного ответа испытуемого *i* на задание *j*.

Вероятность правильного ответа для дихотомического задания может быть найдена по модели Георга Раша[6]:

$$P_{ij} = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} = \frac{1}{1 + e^{-(\theta_i - \beta_j)}}. \quad (12)$$

Расчёт стандартной ошибки измерения уровня подготовленности

Все расчёты выполним в электронной таблице Microsoft Excel. Это позволит любому желающему перепроверить полученные результаты или провести оценку погрешности измерений для другой структуры контрольных измерительных материалов ЕГЭ.

Сначала введём исходные данные (см. рис. 1) — идентификаторы заданий (строка 2) и испытуемых (столбец А), используя для простоты нумерацию. Примем, что уровни подготовленности выпускников θ меняются от — 5 до 5 с шагом 0,5 (столбец В). Уровни трудности 12 дихотомических заданий равномерно распределены от — 5 до 2 (шаг равен 7/11. Т.е., β_1=-5; β_2=-4,36 … β_{12}=2). Для политомических заданий

[6] **Rasch G.** Probabilistic models for some intelligence and attainment tests. — Copenhagen, Denmark: Danish Institute for Educational Research, 1960.

введём все принятые ранее уровни трудности отдельных шагов, например: значения $\beta_{15\ 1}=1$; $\beta_{15\ 2}=3$; $\beta_{15\ 3}=5$ поместим в ячейки S3, T3 и U3 (рис.1).

	A	B	C	D	E	…	S	T	U	…
2			Задание 1	Задание 2	Задание 3	…	Задание 15			…
3			-5	-4,36	-3,73	…	1	3	5	…
4	Выпускник 1	-5								
5	Выпускник 2	-4,5								
6	Выпускник 3	-4								
…	…	…								
24	Выпускник 21	5								

Рис.1. Ввод исходных данных

Далее рассчитаем вероятности правильных ответов. Для этого готовим форму (рис.2), аналогичную форме ввода данных, и вводим расчётные формулы. Для дихотомических заданий в ячейку C29 запишем формулу (12) в соответствии с правилами Microsoft Excel =1/(1+EXP(C$3-$B4)) и скопируем содержимое C29 в диапазон ячеек C29:N49.

Для политомических заданий расчётные формулы составляются на основе уравнений (1–6). Так, для 15 задания, которое оценивается тремя баллами, в ячейки S29, T29 и U29 запишем:

- =EXP($B4-S$3)/(1+EXP($B4-S$3)+EXP($B4-S$3)*EXP($B4-T$3)+EXP($B4-S$3)*EXP($B4-T$3)*EXP($B4-U$3))
- =EXP($B4-S$3)*EXP($B4-T$3)/(1+EXP($B4-S$3)+EXP($B4-S$3)*EXP($B4-T$3)+EXP($B4-S$3)*EXP($B4-T$3)*EXP($B4-U$3))
- =EXP($B4-S$3)*EXP($B4-T$3)*EXP($B4-U$3)/(1+EXP($B4-S$3)+EXP($B4-S$3)*EXP($B4-T$3)+EXP($B4-S$3)*EXP($B4-T$3)*EXP($B4-U$3))

Затем скопируем содержимое S29:U29 в диапазон S29:U49.

Аналогично вводятся формулы для других политомических заданий.

	A	B	C	D	E	…	S	T	U	…
27			Задание 1	Задание 2	Задание 3	…	Задание 15			…
28			π_1	π_1	π_1	…	π_1	π_2	π_3	…
29	Выпускник 1		0,5	0,35	0,22	…	0	0	0	…
30	Выпускник 2		0,62	0,47	0,32	…	0	0	0	…
31	Выпускник 3		0,73	0,59	0,43	…	0	0	0	…
…	…	…	…	…	…	…	…	…	…	…
49	Выпускник 21		1	1	1	…	0,1	0,5	0,5	…

Рис.2. Расчёт вероятностей правильных ответов

Осталось найти погрешности измерений. Для этого запишем входящие в уравнение (7) слагаемые в диапазон C4:AC24. Так, в ячейку C4 введём =C29*(1 — C29), что соответствует выражению $P_{11}(1 — P_{11})$, затем скопируем содержимое C4 в диапазон дихотомических заданий C4:N24.

Слагаемые уравнения (7) для политомических заданий составляются на основе уравнений (8–10). Например, для 15 задания в ячейку S4 запишем =S29+4*T29+9*U29-(S29+2*T29+3*U29) 2 и скопируем полученную формулу в диапазон S4:S24. Слагаемые остальных заданий вводятся аналогично.

Теперь рассчитаем стандартную ошибку измерения уровня подготовленности, для чего в ячейку AJ4 введём =1/СУММ(C4:AM4) 0,5 и скопируем эту формулу в диапазон AJ4:AJ24. Результаты свидетельствуют о нелинейном характере зависимости стандартной ошибки измерения σ от уровня подготовленности θ (рис.3). Стандартная ошибка минимальна при уровне подготовленности, равном 2 логитам $\sigma(2)=0,52$; максимальна при $\theta = — 5$, достигая значения $\sigma(— 5)=1,04$.

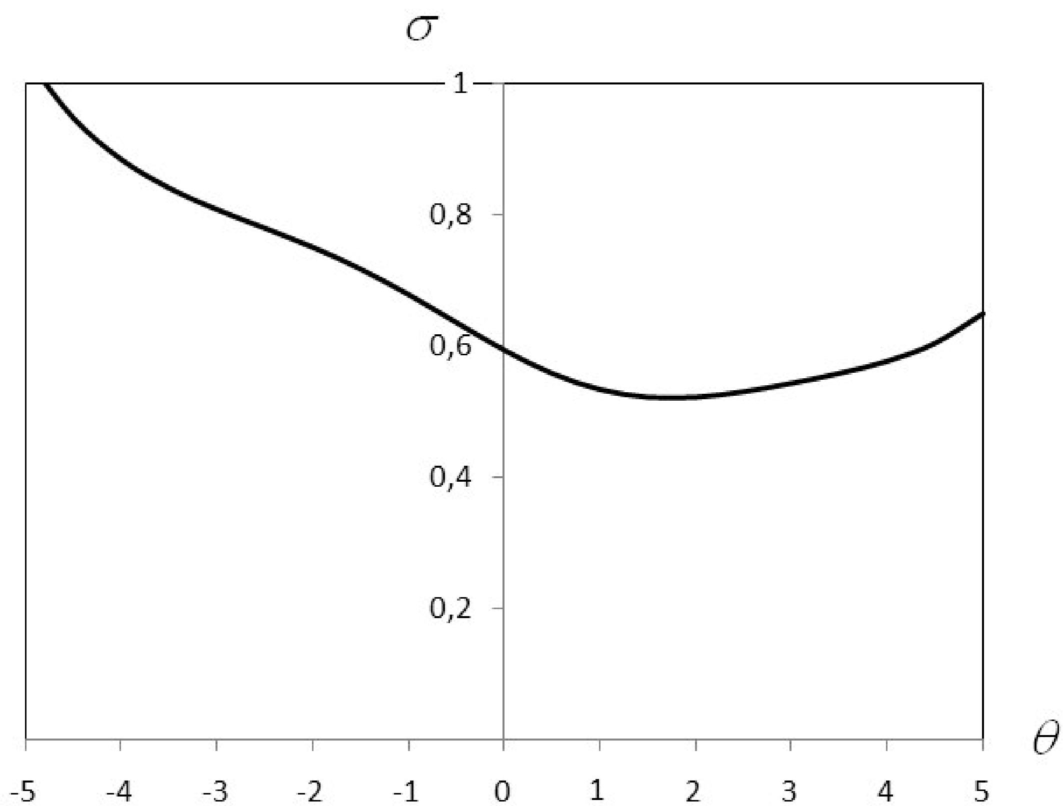

Рис.3. Зависимость стандартной ошибки измерения от уровня подготовленности

Оценка погрешности измерений в баллах

Результаты выполнения выпускником заданий ЕГЭ переводятся в стобалльную шкалу следующим образом[7]. Если участник ЕГЭ не выполнил ни одного задания и получил ноль первичных баллов, он получает ноль баллов. Если участник ЕГЭ верно выполнил все задания и получил максимально возможный первичный балл, он получает сто баллов. Баллы остальных участников ЕГЭ вычисляются с помощью линейного преобразования:

[7] Методика шкалирования результатов ЕГЭ в 2008 году. М.: Федеральная служба по надзору в сфере образования и науки, 2008. 2 с.; http://www1.ege.edu.ru/content/view/431/166/

$$t=\begin{cases}0 & \text{ïðè}\quad \theta<\theta_{\min}\\ \text{ОКРУГЛ}\left(\dfrac{6\cdot\theta_{\max}+8\ \ \cdot\theta-94\cdot\theta_{\min}}{\theta_{\max}-\theta_{\min}}\right) & \text{ïðè}\ \ \theta_{\min}\le\theta\le\theta_{\max}\ ,\\ 100 & \text{ïðè}\quad \theta>\theta_{\max}\end{cases}\qquad(13)$$

где t — балл, θ — оценка уровня подготовленности участника ЕГЭ в логитах, θ_{min} — оценка в логитах, соответствующая одному исходному баллу, θ_{max} — оценка в логитах, соответствующая исходному баллу, на единицу меньшему максимально возможного.

Нетрудно заметить, что по уравнению (13) интервалу уровней подготовленности θ_{min} … θ_{max} соответствует диапазон от 6 до 94 баллов ЕГЭ:

$$t(\theta_{\min})=\text{ОКРУГЛ}\left(\frac{6\cdot\theta_{\max}+8\ \ \cdot\theta_{\min}-94\cdot\theta_{\min}}{\theta_{\max}-\theta_{\min}}\right)=\frac{6(\theta_{\max}-\theta_{\min})}{\theta_{\max}-\theta_{\min}}=6,$$

$$t(\theta_{\max})=\text{ОКРУГЛ}\left(\frac{6\cdot\theta_{\max}+8\ \ \cdot\theta_{\max}-94\cdot\theta_{\min}}{\theta_{\max}-\theta_{\min}}\right)=\frac{94(\theta_{\max}-\theta_{\min})}{\theta_{\max}-\theta_{\min}}=94\,.$$

Таким образом, баллы ЕГЭ могут принимать одно из 90 значений: 0; 6; 7; 8 … 93; 94; 100.

Коэффициент линейного преобразования (коэффициент масштабирования оценки в логитах) по формуле (13) равен:

$$a=\frac{88}{\theta_{\max}-\theta_{\min}}\,.\qquad(14)$$

Тогда стандартная ошибка измерения балла ЕГЭ может быть найдена из выражения:

$$\sigma_t=a\cdot\sigma=\frac{88\sigma}{\theta_{\max}-\theta_{\min}}\,.\qquad(15)$$

Результат такого преобразования показан на рис. 4. Форма графика совпадает с формой линии графика рис.3, что объясняется линейностью преобразования. Стандартная ошибка измерения балла ЕГЭ изменяется в пределах от 4,6 до 9,2 балла.

Результаты теоретического исследования хорошо согласуются с немногочисленными опубликованными данными. Так, в работе В. Хлебникова[8] приведена зависимость стандартной ошибки измерения от баллов ЕГЭ 2004 года по русскому языку (рис.5). Разумеется, полного совпадения графиков рис.4 и рис.5 нет — его и не может быть, поскольку это стандартные ошибки измерений, выполненных по разным общеобразовательным дисциплинам и разным наборам заданий ЕГЭ. Однако числовые значения достаточно близки, что косвенно подтверждает корректность предлагаемой методики оценки погрешности результатов ЕГЭ.

Подбор уровней трудности для набора заданий, обеспечивающего минимальную стандартную ошибку измерения

Самый точный измерительный инструмент даёт минимальную ошибку при измерении любого объекта. Следовательно, лучшим следует признать тот набор заданий, при котором стандартная ошибка минимальна для испытуемого с любым уровнем подготов-

[8] **Хлебников В.** Краткий анализ технологии и результатов единого государственного экзамена (ЕГЭ) // Педагогические измерения. 2008. № 4. С. 25–40.

ленности. Формально это условие можно записать в виде:

$$\max\left(\sigma_i(\theta_i, \beta_1 \ldots \beta_{m\, x_{\max}})\right) \to \min \quad \text{при} \quad i = 1 \ldots n\,. \qquad (16)$$

Попробуем найти этот идеальный вариант. Для этого в ячейку AI26 введём формулу =МАКС(AI4:AI24), которая находит максимальное значение стандартной ошибки измерения σ. Далее следует подобрать уровни трудности всех заданий таким образом, чтобы минимизировать значение AI26. Встроенное в электронную таблицу Microsoft Excel средство *Поиск решения* легко справляется с подобными задачами. В окне *Поиска решения* указываем целевую ячейку AI26, диапазон изменяемых ячеек C3:AF3 (в этих ячейках хранятся значения уровней трудности всех заданий) и отмечаем направление поиска, *равное минимальному значению* (рис.6).

Кроме того, ограничим пределы изменения уровней трудности заданий интервалом [— 5; 5], а для политомических заданий укажем, что уровень трудности каждого последующего шага должен быть больше уровня трудности шага предыдущего. Для увеличения точности решения можно с помощью кнопки *Параметры* увеличить принятое по умолчанию предельное число итераций и уменьшить относительную погрешность численного решения.

Результаты поиска решения сведены в таблицу 1. С помощью такого набора заданий можно практически выровнять ошибку для всего интервала измерения: стандартная ошибка измерения находится в сравнительно узком интервале значений $\sigma = 0{,}62 \ldots 0{,}66$, что соответствует стандартной ошибке тестового балла $\sigma_t = 5{,}5 \ldots 5{,}8$ баллов (рис.7).

Очевидно, что реальное распределение уровней трудности заданий будет отличаться от идеального с точки зрения точности варианта, так как:

• уровни трудности заданий определяются после проведения ЕГЭ. На этапе формирования наборов заданий эти уровни неизвестны, поэтому целенаправленный

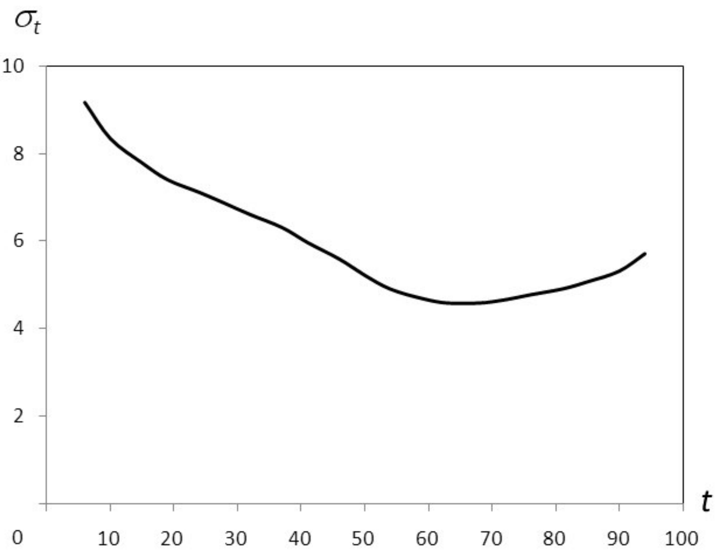

Рис. 4. Зависимость стандартной ошибки измерения от балла ЕГЭ

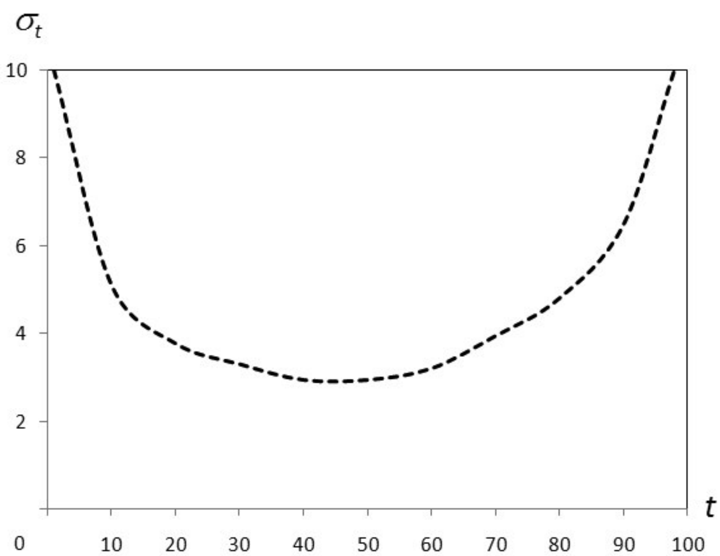

Рис. 5. Стандартная ошибка измерения ЕГЭ 2004 года по русскому языку (по данным В.А. Хлебникова)

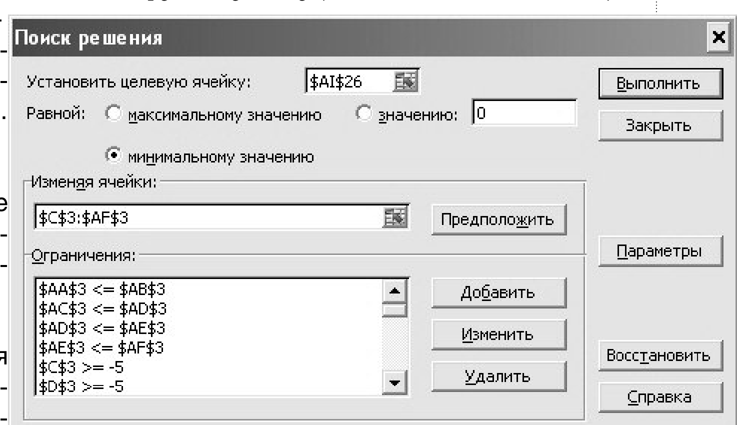

Рис. 6. Ввод параметров поиска решения

Таблица 1

Уровни трудности заданий, обеспечивающие минимизацию ошибки измерения

№ задания	Уровень трудности	№ задания	Уровень трудности	№ задания	Уровень трудности
1	-4,47	2	-4,45	3	-4,46
4	-4,46	5	-4,45	6	-4,43
7	-4,46	8	-4,35	9	-4,49
10	-2,18	11	0,95	12	1,39
13	$\beta_{13\,1}$=-1,51 $\beta_{13\,2}$=3,78	14	$\beta_{14\,1}$=-1,51 $\beta_{14\,2}$=3,78	15	$\beta_{15\,1}$=-1,25 $\beta_{15\,2}$=1,81 $\beta_{15\,3}$=4,20
16	$\beta_{16\,1}$=-1,25 $\beta_{16\,2}$=1,81 $\beta_{16\,3}$=4,20	17	$\beta_{17\,1}$=-0,55 $\beta_{17\,2}$=1,78 $\beta_{17\,3}$=4,78 $\beta_{17\,4}$=4,83	18	$\beta_{18\,1}$=-0,55 $\beta_{18\,2}$=1,78 $\beta_{18\,3}$=4,78 $\beta_{18\,4}$=4,83

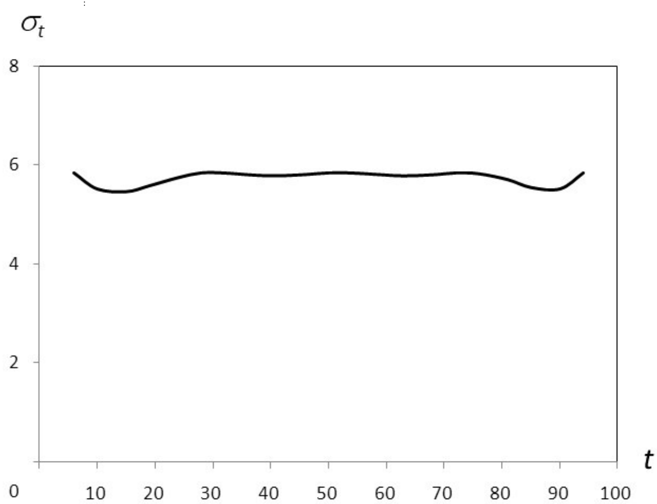

Рис.7. Результаты минимизации стандартной ошибки тестового балла

подбор заданий по уровню трудности невозможен;

- даже при наличии достаточной статистики ответов на политомические задания представляется весьма проблематичным подбор нужного количества таких заданий с уровнями трудности каждого шага, полностью совпадающими со значениями, найденными при минимизации стандартной ошибки измерения.

Поэтому найденная минимальная стандартная ошибка тестового балла σ = 5,5…5,8 — это предельная точность, которую теоретически можно достичь при заданном количестве и структуре заданий ЕГЭ, а также математической модели оценивания. Рост реальной стандартной ошибки тестового балла может быть обусловлен:

- неоптимальностью распределения уровней трудности заданий реального варианта ЕГЭ;
- неравноценностью (непараллельностью) заданий отдельных вариантов ЕГЭ;
- субъективностью оценивания экспертами заданий части С. Так, считается допустимым расхождение оценок экспертов за задание части С, не превышающее одного балла. Для рассматриваемого варианта ЕГЭ часть С содержит шесть заданий, поэтому допустимое расхождение может достигнуть шести баллов.

Оценка минимальной погрешности измерения

Стандартная ошибка тестового балла σ_t характеризует случайную ошибку педагогического измерения. Оценка погрешности измерения (ошибки нахождения балла ЕГЭ) связана со стандартной ошибкой зависимостью[9]:

$$\Delta t = \varepsilon \cdot \sigma_t \qquad (17)$$

где ε — аргумент функции Лапласа, при котором она равна половине выбранного значения вероятности α (табличная величина, например: *α = 0,68* соответствует *ε = 1,0*; α

[9] **Айвазян С.А., Мхитарян В.С.** Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998.

$= 0,90$ соответствует $\varepsilon = 1,65$; $\alpha = 0,997$ соответствует $\varepsilon = 3,0$ и т.д.).

Балл абитуриента по ЕГЭ должен рассматриваться не как конкретное числовое значение, а как интервал вида $t \pm \Delta t$. Например, в рассмотренном выше идеальном с точки зрения точности варианте тестовому баллу $t = 63$ соответствует стандартная ошибка Δt=5,78. Это означает:

— с вероятностью 68% балл по ЕГЭ этого выпускника находится в интервале $t = 63 \pm 1{\cdot}5,78$ (или 57,2 … 68,8);
— с вероятностью 90% $t = 63 \pm 1,65{\cdot}5,78 = 63 \pm 9,5$ (или 53,5 … 72,5);
— с вероятностью 99,7% $t = 63 \pm 3{\cdot}5,78$ (или 45,6 … 80,4).

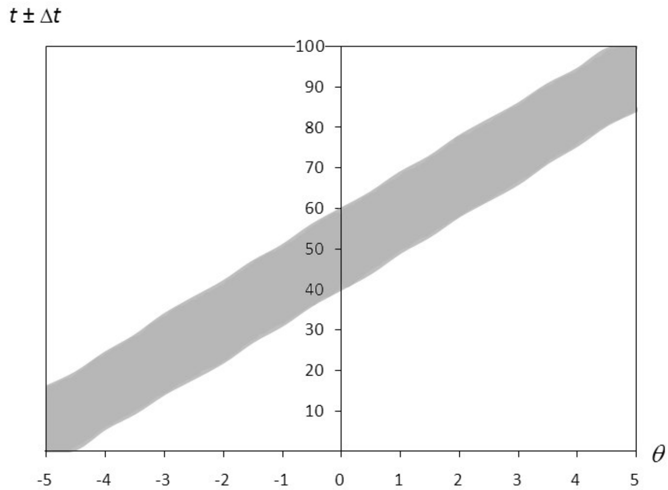

Рис. 8. Зависимость доверительного интервала балла ЕГЭ от уровня подготовленности выпускника при $\alpha = 0,90$

Для наглядности зависимость доверительного интервала балла ЕГЭ от уровня подготовленности абитуриента при $\alpha = 0,90$ приведена на рис. 8. Баллу ЕГЭ соответствует не линия, а полоса шириной 19 баллов. Такова наилучшая теоретически достижимая точность при использовании технологии ЕГЭ — ошибка плюс-минус 9,5 балла. При этом результат, равный 59 баллам, нельзя отличить от 50 баллов (или от 68 баллов), так как различие результатов меньше погрешности измерений.

Практическая пригодность балла ЕГЭ

Сейчас основным назначением баллов ЕГЭ стало обеспечение справедливого конкурса абитуриентов при зачислении в вузы.

Приёмные комиссии вузов учитывают баллы по общеобразовательным предметам, которые выставляются в свидетельствах о результатах ЕГЭ без указания погрешности измерения. Погрешность определения балла не учитывается, и преимущество даже в один балл оказывается решающим для зачисления в вуз, что существенно влияет на судьбу абитуриента.

При этом реальная погрешность тестового балла многократно превышает неявно заявляемую ошибку ±0,5 балла. Для рассмотренного демонстрационного варианта контрольных материалов единого государственного экзамена 2010 года по математике нижний предел стандартной ошибки балла для всего интервала измерения равен 5,5 балла. Соответствующий нижний предел погрешности ЕГЭ при доверительной вероятности 90% равен 9,1 балла.

Высокая погрешность баллов ЕГЭ в определённой степени нивелирует различие результатов выпускников. Например, при указанной погрешности нет оснований признавать различие между 60 и 69 баллами, поскольку различие меньше погрешности измерения.

Таким образом, баллы ЕГЭ ограниченно пригодны для обеспечения конкурса абитуриентов при зачислении в вузы.

Выводы

1. Разработанная методика позволяет оценить нижний предел погрешности баллов ЕГЭ на основе анализа количества и структуры заданий ЕГЭ, а также математической модели Partial Credit Model.
2. Реальная погрешность баллов ЕГЭ многократно превышает неявно заявляемую ошибку ±0,5 балла.
3. Без учёта погрешности измерения тестовый балл ЕГЭ слабо пригоден для обеспечения справедливого конкурса абитуриентов при зачислении в вузы.
4. Для полного решения задач ЕГЭ необходима оценка погрешности баллов каждого выпускника, а также оптимизация количества и структуры заданий с целью снижения ошибки педагогического оценивания. ❑