

# **МОДЕЛИРОВАНИЕ ВЗАИМОСВЯЗЕЙ ДОЛЕЙ УЧАСТНИКОВ ЕДИНОГО ГОСУДАРСТВЕННОГО ЭКЗАМЕНА ПО ПРОФИЛЬНОЙ МАТЕМАТИКЕ И ЕГО РЕПЕТИЦИЙ В ЗАВИСИМОСТИ ОТ КОЛИЧЕСТВА ПРАВИЛЬНЫХ ОТВЕТОВ**

**Ж. Н. Зенкова**

*Томский государственный университет, Россия  
Российская академия народного хозяйства  
и государственной службы, Москва  
E-mail: zhanna.zenkova@mail.tsu.ru*

В работе найдены нелинейные модели взаимосвязи долей участников Единого Государственного Экзамена (ЕГЭ), профильная математика, и его репетиций, ответивших на одинаковое количество вопросов. При анализе использовались данные о результатах ЕГЭ и его репетиций в Томской области за 2016-2020 гг. Обнаружены два кластера, характеризующие результаты: в первом кластере оказались участники, ответившие верно на 1-8 вопросов ЕГЭ, во втором – все остальные. Для каждого кластера построены параболические модели, проходящие через точку начала координат. Выявлен нетипичный характер модели 2020 гг., который объясняется влиянием пандемии COVID-19. Построена усреднённая модель взаимосвязи, показывающая, что доли участников, ответивших на 1-8 баллов на репетиции, на ЕГЭ снижаются, в то время как доли ответивших на 9-32 вопроса увеличиваются, более того, доля ответивших на 15-32 вопроса практически удваиваются.

## **MODELING THE RELATIONSHIP BETWEEN THE SHARES OF UNIFIED STATE EXAM PARTICIPANTS IN SPECIALIZED MATHEMATICS AND ITS REHEARSALS DEPENDING ON THE NUMBER OF ANSWERS**

**Z. N. Zenkova**

In the paper, the nonlinear models of relationship between the shares of participants of Unified State Exam (USE) in specialized mathematics and its rehearsals depending on the number of answers are considered. The modeling is based on the data collected in Tomsk region in 2016-2020. We divided the data on two clusters: the first unites the participants with 1-8 correct answers on USE, the second contains the rest. For each cluster we found the parabolic models passing through the origin. Due to the untypical model of 2020, we found the influence of COVID-19 pandemic. The generalized average model of the results' relationship is built. It is shown that the shares of participants with 1-8 correct answers (rehearsals) are decreased on USE but for 9-32 correct answers are increased, moreover, for 15-32 correct answers it is increased almost twice.

В работе рассматривались данные о результатах репетиционных экзаменов и ЕГЭ по профильной математике по Томской области за 2016-2020 гг. [1-4]. Анализировалась связь между распределениями долей участников по количеству вопросов, на которые были даны ответы, на репетиционном экзамене ЕГЭ и ЕГЭ. В табл. 1 приведены объемы генеральных совокупностей ( $N$ ). Заметим, что количество участников репетиций обычно больше, чем число участников ЕГЭ, тогда как число тех, кто ответил на 1 вопрос и более на репетициях

всегда меньше, чем на ЕГЭ. Это позволяет сделать вывод о том, что многие плохо подготовленные участники после репетиции ЕГЭ по математике не участвуют в ЕГЭ [3-8].

Таблица 1

**Число сдававших ЕГЭ и репетицию ЕГЭ, профильная математика, Томская обл.**

Год		2016	2017	2018	2019	2020
Все сдававшие	Реп. ЕГЭ	4000	3500	3561	2803	3334
	ЕГЭ	3478	3406	3390	3111	2859
Набравшие больше нуля баллов	Реп. ЕГЭ	2895	2852	2588	2503	2466
	ЕГЭ	3466	3402	3371	3101	2856

На рис. 1 приведены графики распределений участников, получивших более нуля баллов, в долях. На рис. 2 приведены графики зависимостей по годам, где по оси абсцисс отображена доля участников, ответивших правильно на определенное количество вопросов на репетиционном экзамене, а по оси ординат – доля участников, давших ответ на такое же количество вопросов на ЕГЭ (на рисунках это количество указано возле соответствующих точек).

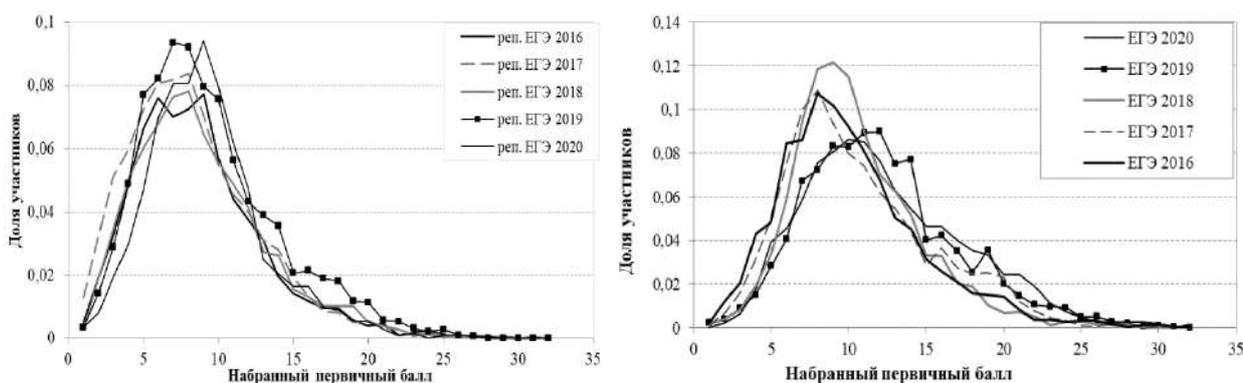


Рис. 1. Распределения долей по количеству ответов на репетициях ЕГЭ (слева) и на ЕГЭ (справа)

Для построения моделей зависимостей данные были разбиты на два кластера: первый – с низкими результатами – от 1 до 8 отвеченных вопросов (здесь результаты репетиционного экзамена были обычно лучше, чем результаты ЕГЭ), второй – от 9 до 32 вопросов. Наиболее адекватно данные описывают полиномиальные модели второго порядка следующего вида: для первого кластера:

$$y=ax^2, \tag{1}$$

для второго кластера:

$$y=bx^2+cx, \tag{2}$$

где  $x$  – доля участников, ответивших правильно на заданное количество вопросов на репетиционном экзамене,  $y$  – доля школьников, ответивших правильно на это же количество вопросов на ЕГЭ,  $a$ ,  $b$  и  $c$  – найденные методом наименьших квадратов коэффициенты модели – приведены в таблице 2, как и коэффициенты детерминации  $R^2$ , которые говорят о высоком качестве полученных моделей [9].

Нормальность остатков моделей проверялась критерием Шапиро-Уилка

[10-11], достигнутый уровень значимости  $p$ -value указан в таблице 2 и позволяет во всех случаях принять гипотезу о нормальности на уровне значимости  $\alpha = 0,03$ . Точность моделирования характеризует средняя абсолютная относительная ошибка аппроксимации  $A$ , которая определяется формулой:

$$A = \frac{1}{N} \sum \frac{|y_{\text{факт}} - y_{\text{модель}}|}{y_{\text{факт}}} \cdot 100\%.$$

Значения  $A$  также отображены в табл. 2. Заметим, что ошибки моделей варьируются в пределах 10,7-35,9%. Это объясняется тем, что рассматривался универсальный класс моделей, функциональный вид которых не зависел от года.

Таким образом, полученные модели позволяют сделать вывод о том, что доли школьников, ответивших на одинаковое количество вопросов на репетиции и ЕГЭ, связаны между собой нелинейно, при этом для первого кластера низкие результаты (ответили на 1-5 вопросов) по сравнению с репетиционным экзаменом на ЕГЭ снижаются очень существенно, и чем хуже результат, тем больше это снижение. Для тех, кто ответил на 6-8 вопросов разница в изменении долей не столь существенная. Модели второго кластера говорят о том, что доли школьников, ответивших на 15-32 вопроса, практически удваиваются, а для 9-14 вопросов изменения в долях уже не такие радикальные. На рис. 2 приведены графики исходных данных и полученных моделей.

Таблица 2

**Значения коэффициентов моделей, коэффициентов детерминации  $R^2$ , достигнутых уровней значимости  $p$ -value критерия Шапиро-Уилка, ошибок  $A$**

Год	Первый кластер				Второй кластер				
	$a$	$R^2$	$p$ -value	$A$ , %	$b$	$c$	$R^2$	$p$ -value	$A$ , %
2016	16,39	0,89	0,08	23,4	- 12,55	2,30	0,99	0,17	27,7
2017	12,75	0,91	0,21	22,8	- 13,56	2,23	0,96	0,17	27,1
2018	15,25	0,90	0,35	27,5	- 11,33	2,00	0,99	0,03	18,3
2019	9,17	0,92	0,65	28,6	-16,61	2,63	0,97	0,20	35,9
2020	13,42	0,96	0,28	10,7	<b>-24,25</b>	<b>3,04</b>	0,92	0,14	30,9

Заметим, что модель 2020 г. (второй кластер) существенно отличается от остальных, что можно объяснить прямым влиянием пандемии COVID-19 [2].

В целях прогнозирования будущих результатов ЕГЭ [3, 7] по результатам репетиционного экзамена рассмотрим усреднённые модели по каждому кластеру, исключив 2020 г. во втором кластере как выброс. Получим, что первый кластер может быть описан моделью следующего вида:

$$y = 13,396 x^2, \quad (5)$$

для второго кластера модель определяется формулой:

$$y = -13,513x^2 + 2,290x \quad (6)$$

На рис. 2 представлена итоговая модель, которая может быть использована при прогнозировании долей школьников, которые ответят на определенное количество вопросов на ЕГЭ, по итогам репетиционного экзамена. Заметим, что для

первого кластера (ответили на 1-8 вопросов) на участке  $x > 0,07465$  (точка пересечения уравнения (2) с прямой  $y=x$ ) модель прогнозирует, что соответствующая заданному количеству вопросов доля на ЕГЭ будет больше, чем на репетиционном экзамене, для второго (ответили на 9-32 вопроса) – на участке  $x < 0,09547$  (пересечение (6) с прямой  $y=x$ ). Заметим, что этот переход в основном касается модальных значений, т.е., тех участников, которые ответили на 7-10 вопросов.

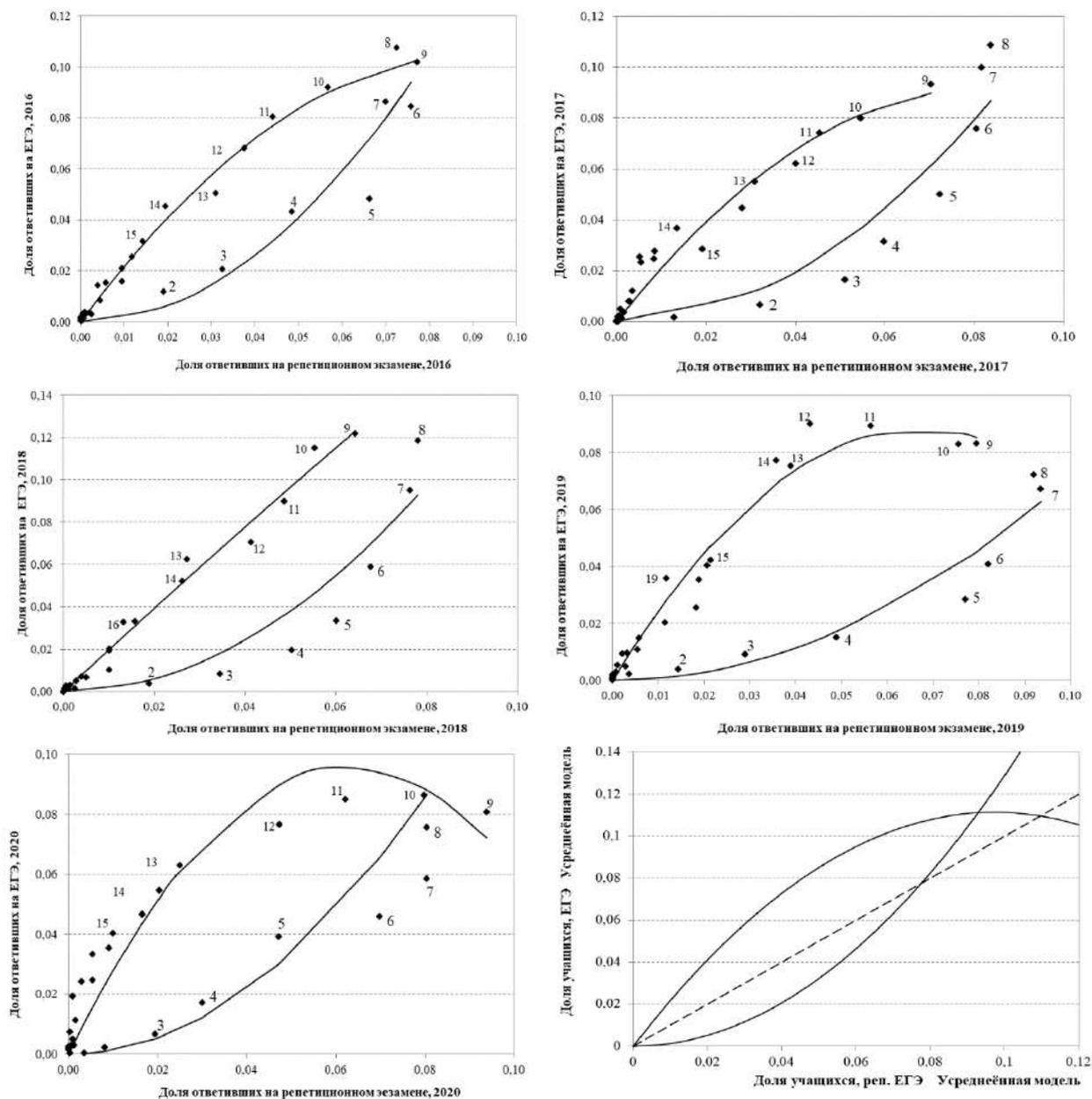


Рис. 2. Связь долей по количеству отвеченных вопросов на репетиционном экзамене и ЕГЭ и их модели, 2016-2020 гг., а также обобщающая усреднённая модель. Прямая  $y=x$  отображена в виде пунктирной линии

В итоге, если на репетиционном экзамене на три вопроса ответили, например, 2% участников, то на ЕГЭ следует ожидать, что на три вопроса ответят уже

$y = 13,396 \cdot 0,0004 = 0,00536$ , т.е. практически 0,54% (т.е., почти 4-кратное снижение). Если же на репетиционном экзамене 5% ответили на 15 вопросов, то на ЕГЭ следует ожидать  $y = -13,513 \cdot 0,05^2 + 2,290 \cdot 0,05 = 0,0807$ , т.е., практически 8,1% (увеличение более чем в полтора раза).

Таким образом, найдены модели взаимосвязи долей участников ЕГЭ (профильная математика) и его репетиций, ответивших на одинаковое количество вопросов. Выявлен нетипичный характер модели 2020 г., который объясняется влиянием пандемии COVID-19. Построена усреднённая модель взаимосвязи долей, показывающая, что доли ответивших на 1-8 баллов на репетиции на ЕГЭ снижаются, в то время как доли ответивших на 9-32 вопроса увеличиваются, более того, доля ответивших на 15-32 вопроса существенно увеличивается, т.е. в среднем, те, кто был хорошо подготовлен к репетиционному экзамену, перед ЕГЭ усиленно готовились и показали существенно лучший результат.

#### СПИСОК ЛИТЕРАТУРЫ

1. *Benks E. A., Ilyukhin B. V., Serbina N. P.* External barriers affecting the education success in Russian schools // Proceedings of 16th International Technology, Education and Development Conf. 2022. P. 10501-10505.
2. *Benks E. A., Ilyukhin B. V., Serbina N. P.* Identification of the influence of unforeseen external influences (COVID-19 pandemic) on educational outcomes in the general education system // Proceedings of 15th International Conf. of Education, Research and Innovation. 2022. P. 8214.
3. *Katsman Y. Y., Lepustin A. V., Ilyukhin B. V., Lepustina E. V., Zenkova Z. Z.* The stochastic model of the impact of context factors to educational results of Tomsk school graduates // Proceeding of 2016 IEEE Global Engineering Education Conference. 2016. P. 767-771.
4. *Khudorozhkov I., Ilyukhin B. V., Benks E. A., Serbina N. P.* Assessment of the impact of resource management mechanisms on the educational results of the school // Proceedings of 14th International Conference On Education And New Learning Technologies. 2022. P. 7504-7510.
5. *Багирова И. Х., Зенкова Ж. Н.* Моделирование готовности студентов к профессиональной деятельности с учетом мнений работодателей // Экономика и предпринимательство. 2016. № 8 (73). С. 719-726.
6. *Илюхин Б. В., Бенкс Е. А., Сербина Н. П., Гусякова В. И.* Анализ внешнего контекста общеобразовательных организаций, значимо влияющего на образовательные результаты, в региональных системах управления образованием // Перспективы науки и образования. 2021. № 5 (53). С. 495-511.
7. *Илюхин Б. В., Савиных Г. П., Зенкова Ж. Н., Сербина Н. П., Кабанова Т. В.* Использование результатов государственной итоговой аттестации в организации обучения школьников точным наукам // Перспективы науки и образования. 2022. № 5 (59). С. 219-232.
8. *Савиных Г. П., Илюхин Б. В., Машарова Т. В.* О состоянии внутренних систем оценки качества образования в школах России // Перспективы науки и образования. 2021. № 2 (50). С. 501-512.
9. Мхитарян В. С., Архипова М. Ю., Балаш В. А. и др. Эконометрика : учебник / М. : Проспект. 2014, 380 с.
10. *Shapiro S. S., Wilk M. B.* An analysis of variance test for normality (complete samples) // *Biometrika*. 1965. Vol. 52. No 3 (4). P. 591-611.
11. *Кобзарь А. И.* Прикладная математическая статистика для инженеров и научных работников / М. : ФИЗМАТЛИТ, 2006. 816 с.