

УДК 539.3:534:532.5

05.13.18 – Математическое моделирование, численные методы и комплексы программ (технические науки)

**ИССЛЕДОВАНИЕ КАЧЕСТВЕННЫХ ПОКАЗАТЕЛЕЙ РЕЗУЛЬТАТОВ ИЗУЧЕНИЯ БАЗОВЫХ ПРЕДМЕТОВ ВЫПУСКНИКАМИ ШКОЛ КАК ОСНОВНОГО УСЛОВИЯ УСПЕШНОГО ОБУЧЕНИЯ В ВУЗЕ**

Анищик Татьяна Алексеевна  
старший преподаватель  
РИНЦ SPIN-код: 7310-5179  
tanja63@mail.ru

Коблянский Владимир Сергеевич  
студент факультета Прикладной информатики  
vovakob2@mail.ru

*Кубанский государственный аграрный университет имени И.Т. Трубилина, Краснодар, Россия*

Успешное обучение в вузе предполагает высокий уровень интеллектуального развития абитуриентов, эрудированности и развитого мышления, поэтому отбор абитуриентов, наилучшим образом подготовленных к обучению, является одним из условий получения выпускниками вуза высокого качества знаний, умений и навыков. Современное динамично развивающееся общество не может не предъявлять повышенных требований к качеству подготовки выпускников вузов. Весомым наукометрическим показателем из набора критериев оценки вузов России в категории «Условия для получения качественного образования» является «Средний балл ЕГЭ студентов», зачисленных по общему конкурсу на очную форму обучения по программам бакалавриата и специалитета. Отрадно, что для зачисления абитуриентов в аграрные вузы на все направления обучения год от года повышают проходной балл, однако, высокие баллы ЕГЭ не могут гарантировать качества подготовки выпускников школ. В статье проведен анализ эмпирических данных на однородность, вариационную изменчивость и характер связи. Обоснован выбор модели данных, методов проведения исследования и программных средств. Исследована взаимосвязь результатов сдачи ЕГЭ по математике и информатике до и после этапа зачисления; представлены различия в показателях абитуриентов и студентов, выбравших одно из инженерных направлений обучения. Обоснован выбор метода регрессионно-корреляционного анализа. Рассмотрены варианты выполнения расчетов средствами табличного процессора MS Excel. Построены уравнения регрессии и установлено их качество по величинам средних абсолютных ошибок аппроксимации (MAPE) для каждой модели. Определены значения и выявлена

UDC 539.3:534:532.5

05.13.18 – Mathematical modeling, numerical methods and software packages (technical sciences)

**RESEARCH OF QUALITATIVE INDICATORS OF THE RESULTS OF STUDYING BASIC SUBJECTS BY SCHOOL GRADUATES AS THE MAIN CONDITION FOR SUCCESSFUL EDUCATION IN HIGHER EDUCATION**

Anishchik Tatyana Alekseevna  
senior lecturer  
RSCI SPIN-code: 7310-5179  
tanja63@mail.ru

Koblyanski Vladimir Sergeevich  
student of Applied Informatics  
vovakob2@mail.ru  
*Kuban state agrarian University named after I. T. Trubilin, Krasnodar, Russia*

Successful training at university requires a high level of intelligence, erudition and developed thinking of applicants, so the selection of the most prepared applicants is one of the conditions for getting high-quality knowledge, skills and abilities by graduates of the University. Modern dynamically developing society cannot but impose increased requirements to the quality of training of University graduates. A significant scientometric factor from the set of criteria for evaluating Russian universities in the category "The conditions for obtaining quality education" is the "average exam score of students" who are enrolled in the General competition for full-time training in programs of bachelor and specialist. It is great that the passing score for enrolling in agricultural universities for any field of study is increased year by year, however, high exam scores cannot guarantee the good quality of knowledge of school graduates. The article analyzes the empirical data on homogeneity, variational variability and the nature of the relationship. The choice of data model, research methods and software tools is justified. The relationship between the results of passing the unified state exam in mathematics and computer science before and after the enrollment stage is studied; differences in the results of applicants and students who chose one of the engineering areas of study are presented. The choice of the method of regression and correlation analysis is justified. Options for performing calculations using the MS Excel table processor are considered. Regression equations are constructed and their quality is determined by the values of average absolute approximation errors (MAPE) for each model. The article identifies values and the statistical significance of the most important factors of models: regression coefficient; values of the sample Pearson correlation coefficient; its significance by the

статистическая значимость наиболее важных показателей моделей: коэффициента регрессии; величины выборочного коэффициента корреляции Пирсона, установлена его значимость по t-критерию Стьюдента, проведена оценка тесноты связи; величины коэффициента детерминации

student t-criterion is established, the evaluation of correlation and the coefficient of determination was done as well

Ключевые слова: АБИТУРИЕНТ, ОБУЧЕНИЕ, ОБРАЗОВАНИЕ, ЕДИНЫЙ ГОСУДАРСТВЕННЫЙ ЭКЗАМЕН, РЕЙТИНГ ВУЗА, ОДНОРОДНОСТЬ, КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

Keywords: APPLICANT, TRAINING, EDUCATION, UNIFIED STATE EXAM, UNIVERSITY RATING, UNIFORMITY, CORRELATION AND REGRESSION ANALYSIS

DOI: <http://dx.doi.org/10.21515/1990-4665-161-015>

## Введение

Успешное обучение в вузе предполагает высокий уровень интеллектуального развития абитуриентов, эрудированности и развитого мышления, поэтому отбор абитуриентов, наилучшим образом подготовленных к обучению, является одним из условий получения выпускниками вуза высокого качества знаний, формирования умений и навыков [6]. Компетентностная модель российского образования рассчитана на существенное увеличение объема самостоятельной познавательной активности студентов, должным образом подготовленных к такой деятельности. Современное динамично развивающееся общество не может не предъявлять повышенных требований к качеству подготовки выпускников вузов. Результат успешного внедрения инновационных проектов в агропромышленный комплекс связан, в том числе, с уровнем квалификации агроинженерных кадров, а значит, возрастает объем и уровень требований к базовому инженерному образованию [5].

Весомым наукометрическим показателем из набора критериев оценки вузов России в категории «Условия для получения качественного образования» является «Средний балл ЕГЭ студентов», зачисленных по общему конкурсу на очную форму обучения по программам бакалавриата и специалитета [10]. Отрадно, что для зачисления абитуриентов в аграрные вузы на инженерные направления обучения год от года повышается проходной

балл, однако высокие баллы ЕГЭ не гарантируют качества подготовки выпускников школ. Об этом свидетельствует не уменьшающееся количество отчисляемых студентов из вузов по причинам неуспеваемости. Устранение пробелов школьного образования в вузе, например, по информатике, приводит к дублированию школьного курса и к потере значительной доли учебного процесса [4].

В разное время проблемами успешности обучения в вузе занимались видные педагоги и психологи, например: Ю. К. Бабанский, И. Я. Лернер, В. В. Краевский, В. В. Давыдов, Г. И. Щукин. Выявлено, что успешность обучения студентов в вузе – это сложное комплексное понятие, определяемое внутренними (психологические) и внешними (социальные и педагогические) факторами. На успешность обучения студентов в высших учебных заведениях влияют многие факторы: материальное положение; состояние здоровья; возраст; семейное положение; уровень довузовской подготовки; владение навыками самоорганизации, планирования и контроля своей деятельности (прежде всего учебной); мотивы выбора вуза; адекватность исходных представлений о специфике вузовского обучения; форма обучения (очная, вечерняя, заочная, дистанционная и др.); плата за обучение и ее величина; организация учебного процесса в вузе; материальная база вуза; уровень квалификации преподавателей и обслуживающего персонала; престижность вуза и, наконец, индивидуальные психологические особенности студентов [8]. В настоящем уровне довузовской подготовки определяется результатами ЕГЭ по базовым (профильным) предметам, под которыми будем понимать те из них, которые определяют выбор абитуриентами направления обучения.

Исследованиям подготовленности первокурсников к освоению определенных дисциплин в вузе, связанных с результатами ЕГЭ, посвящено много исследовательских работ. Например, И. А. Ермакова, В. А. Гоголин и Л. Ю. Ефремова анализируют степень подготовки обучаемых по

определенному предмету [13]; исследуют связь результатов ЕГЭ: с успеваемостью студентов на первом курсе Т. Е. Хавенсон и А. А. Соловьева [15], с успеваемостью студентов на протяжении всех лет обучения Н. А. Чернышова [18], с уровнем интеллекта студентов А. В. Капцов и Е. И. Колесникова [14]. В чем состоят качественные различия показателей ЕГЭ абитуриентов и первокурсников? – Представляется целесообразным проведение анализа качественных показателей результатов ЕГЭ по базовым дисциплинам абитуриентов и студентов в одном из лучших аграрных вузов страны. Для этого потребуются описание, вычисление и сравнение параметров моделей эмпирических данных до и после этапа зачисления.

Задачами исследования являются:

1. Выявить однородность подготовленных данных.
2. Проверить выборку на нормальность распределения.
3. Установить форму связи эмпирических данных и уточнить способ выбора соответствующей модели.
4. Получить уравнение парной регрессии, если связь будет установлена.
5. Определить значение параметров для выбранной формы модели.
6. Осуществить прогнозирование по выбранной модели.
7. Оценить среднюю абсолютную ошибку аппроксимации для модели.

#### **Подготовка исходных данных**

Исследование проводилось в Кубанском государственном аграрном университете имени И. Т. Трубилина, входящем в тройку лидеров в 2018-2019 гг. среди 49 вузов России сельскохозяйственного направления [11]. Генеральной совокупностью исследования являются данные случайной выборки абитуриентов, подававших документы в приемную комиссию на направление подготовки 09.03.03 – Прикладная информатика, бакалавриат, очную форму обучения на бюджетные места в 2018 году. Объем первой

выборочной совокупности, элементы которой будут исследоваться, составили данные опроса 98 абитуриентов об итогах сдачи ЕГЭ двух профильных дисциплин – математики, информатики и ИКТ (таблица 1). При интерпретации и представлении результатов ЕГЭ важно понимать, что 100-балльные шкалы по разным учебным предметам по сути своей не совпадают и баллы (максимальные, минимальные, средние и т. д.) нельзя сравнивать напрямую ни между предметами, ни по одному и тому же предмету в динамике по годам [9].

Таблица 1 – Результаты ЕГЭ по базовым предметам у абитуриентов

№	x	y	№	x	y	№	x	y	№	x	y	№	x	y	№	x	y	№	x	y
1	70	62	15	56	53	29	70	61	43	70	75	57	45	51	71	56	79	85	68	84
2	74	79	16	68	64	30	72	70	44	50	70	58	50	51	72	56	61	86	65	61
3	33	51	17	72	62	31	72	77	45	39	66	59	62	68	73	70	73	87	56	61
4	70	51	18	56	62	32	62	53	46	62	61	60	70	66	74	56	62	88	45	62
5	50	62	19	62	61	33	50	51	47	68	62	61	33	42	75	39	53	89	62	68
6	33	59	20	70	73	34	45	46	48	39	48	62	72	70	76	62	53	90	50	42
7	50	48	21	50	66	35	68	75	49	62	51	63	45	46	77	70	64	91	39	61
8	68	83	22	56	51	36	50	62	50	55	55	64	68	50	78	68	64	92	33	51
9	74	77	23	39	42	37	62	61	51	50	57	65	68	79	79	53	67	93	50	42
10	74	73	24	56	74	38	56	72	52	56	57	66	56	50	80	56	58	94	39	59
11	50	72	25	68	77	39	50	53	53	56	61	67	74	55	81	39	53	95	66	75
12	39	61	26	68	73	40	68	55	54	33	48	68	39	42	82	62	70	96	69	77
13	45	64	27	50	59	41	62	77	55	45	61	69	74	61	83	62	59	97	58	63
14	50	55	28	68	72	42	62	59	56	50	62	70	68	57	84	50	61	98	51	58

Введены обозначения предметов: x – Математика, y – Информатика и ИКТ.

Второй выборочной совокупностью являются данные 23 первокурсников (таблица 2), отобранных из прежнего списка респондентов согласно критериям зачисления [19], с проходным баллом, равным 220.

Таблица 2 – Результаты ЕГЭ по базовым предметам у первокурсников

№	x	y	№	x	y	№	x	y	№	x	y
1	74	79	7	68	73	13	72	70	19	70	66
2	68	83	8	68	72	14	68	79	20	69	77
3	74	77	9	72	70	15	70	73	21	74	61
4	74	73	10	72	77	16	68	84	22	56	79
5	70	73	11	68	75	17	66	75			
6	68	77	12	70	75	18	62	77			

Введены обозначения предметов: x – Математика, y – Информатика и ИКТ.

### Обоснование выбора метода решения проблемы

Исследование зависимостей значений двух переменных можно выполнить с использованием метода парного корреляционно-регрессионного анализа. Основными задачами парного корреляционного анализа являются: отбор факторов, оказывающих наиболее существенное влияние на результативный признак, на основании измерения степени связности между явлениями; обнаружение неизвестных причинных связей. В регрессионном анализе основными задачами считаются: генерация гипотезы по установлению формы зависимости между переменными одним из выбранных способов; получение и исследование уравнения регрессии; оценка и прогнозирование значений результатов.

### **Обоснование выбора программного средства**

На современном рынке программного обеспечения ПК представлено около одной тысячи компьютерных программ для статистической обработки данных. К ним можно отнести статистические пакеты: универсальные (*Statistica, SPSS*), профессиональные (*BMDP, SAS*), специализированные (*DATASCOPE, BioStat*) и системы автоматизированного проектирования (*Mathcad*). Программы подобного рода реализуют сложные статистические методы, некоторые из них могут работать с большими данными, что особенно актуально. Однако работа с такими программами потребует от пользователя достаточных знаний и умений, т. е. обладания определенными компетенциями.

Статистическую обработку данных можно реализовать в пользовательских программах, написанных на языках программирования высокого уровня и встроенных в интегрированные системы программирования, что предоставляет возможность рассматривать разные варианты решения задачи и приводит к формированию умений выбора программных продуктов и выработке навыков обработки данных в разных средах [1, с. 58; 2, с. 30].

Выбор программы для проведения расчетов – табличного процессора *MS Excel 2010* является давно обоснованным и связан с некоторыми основными его преимуществами:

- высокой степенью распространенности и доступности версий офисных программ;
- наличием удобного графического интерфейса и простотой в освоении программы;
- возможностью подключения встроенного в *MS Excel* пакета «Анализ данных» (*Файл – Параметры – Настройки – Управление – Настройки MS Excel – Перейти – Пакет анализа*). В результате доступна команда *Данные – Анализ данных*, позволяющая существенно облегчить проведение более сложной статистической обработки данных;
- графическим представлением результатов обработки данных и сохранением их вместе с расчетными таблицами в одном файле;
- возможностью импортирования результатов обработки данных и описанием итогов исследования в текстовом редакторе.

### **Реализация поставленных задач средствами *MS Excel***

#### *1. Проверка данных выборки на однородность*

Объем первой выборки данных достаточно большой, поэтому корреляционная связь, если будет установлена, должна достаточно полно проявиться и, как следствие, произойдет сглаживание влияния других факторов. Одним из требований корреляционно-регрессионного анализа к исследуемым данным является их однородность относительно распределения около среднего уровня. Критериями однородности данных для показателей считаются величина среднеквадратического отклонения ( $\sigma$ ) и коэффициент вариации ( $V$ ), которые можно вычислить следующими способами:

1. С использованием расчетных формул [3, с. 95]:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad V = \frac{\sigma}{x} \cdot 100,$$

где  $n$  – количество элементов выборки,

$x_i$  – элементы,

$\bar{x}$  – среднеарифметическое значение элементов.

2. С использованием функций СТАНДОТКЛОН.Г(диапазон) и СРЗНАЧ(диапазон).

Значение коэффициента вариации находится в прямой зависимости с величиной разброса данных: чем он больше, тем больше разброс (таблица 3).

Таблица 3 – Показатели однородности данных

Показатель	Среднеарифметическое значение	Среднеквадратическое отклонение	Коэффициент вариации, %
x	56,9081633	11,68121	20,52643
y	61,4387755	10,07225562	16,39397194

В соответствии с полученными коэффициентами вариации по фактору  $x$  наблюдается значительная вариация (таблица 4). Следовательно, совокупность данных однородна и для ее исследования могут использоваться корреляционно-регрессионный метод.

Если значение вариации превышает 33%, то исследуемые данные считаются неоднородными и их необходимо исключить из выборки.

Таблица 4 – Оценка изменчивости вариационного ряда

Диапазон вариации	< 10%	10-20%;	20- 33%	> 33%
Характер изменчивости	Однородные данные			Неоднородные данные
	незначительный	средний	значительный	

### 2. Проверка выборки на нормальность распределения

Нормальное распределение выборки является условием применимости большей части статистических методов, поэтому начальным этапом анализа данных считается выдвижение гипотезы о нормальности их распределения. Существует несколько способов выявления нормальности



распределения, например, с построением графика функции распределения для нормального закона; с применением критерия нормальности, например, Дарбина, хи-квадрат, Колмогорова-Смирнова; с применением встроенных программных средств расчета показателей и построения графиков, например, *Statistica*. Воспользуемся способом с применением критерия асимметрии и эксцесса. Для этого необходимо рассчитать показатели описательной статистики (таблица 5).

Таблица 5 – Показатели описательной статистики

Показатель	x	y
Коэффициент асимметричности (скос)	-0,368439379	0,109813
Фактическое значение <i>t</i> -критерия асимметрии	-1,51165	0,450546673
Коэффициент эксцесса	-0,888088213	-0,5791
Фактическое значение <i>t</i> -критерия эксцесса	-1,83979	-1,19968899

Коэффициенты асимметрии ( $A_s$ ) и эксцесса ( $E_x$ ) можно рассчитать, воспользовавшись функциями СКОС(диапазон) и ЭКСЦЕСС(диапазон).

Значения коэффициентов асимметрии и эксцесса, недостаточно близкие к нулю, указывают, что распределение данных около средних величин не совсем симметрично. Отрицательная асимметрия ( $x$ ) означает, что преобладают данные с большими значениями и правый хвост распределения короче левого; отрицательный эксцесс – кривая распределения является плосковершинной, т. е. данные не густо сгруппированы около средней.

Ошибки показателей асимметрии ( $S_{A_s}$ ) и эксцесса ( $S_{E_x}$ ) вычисляются по формулам:

$$S_{A_s} = \sqrt{\frac{6}{N+3}} = 0,243733339; \quad S_{E_x} = 2\sqrt{\frac{6}{N+5}} = 0,482710792,$$

где  $N$  – размерность выборки.

Фактическим значением *t*-критерия ( $t_\phi$ ) коэффициентов асимметрии и эксцесса является частное от деления показателей асимметрии и эксцесса на их ошибки соответственно. Табличное значение ( $t_T$ ) определяется из

таблицы Стьюдента, расположенное на пересечении строки, равной числу степеней свободы ( $df = N - 1$ ), и колонки, содержащей уровень значимости ( $p$ ): при  $df = 100$  (ближайшее к  $98 - 1 = 97$ ) на уровне значимости:  $p = 0,05$  (допускаемая вероятность ошибок 5%). Полученное фактическое значение необходимо сравнить с табличным значением  $t_T \approx 1,984$ , если  $t_\phi < t_T$ , то распределение признается нормальным. Результат сравнения фактического и табличного значений для асимметрии:  $-1,51165 < 1,984$  показывают незначительность асимметрии кривой распределения. Для показателей эксцесса:  $-1,83979 < 1,984$  подтверждается гипотеза нормальность распределения данных. Представленные в таблице 5 фактические значения  $t$ -критерия асимметрии и эксцесса меньше трех, что говорит о несущественности значений асимметрии и эксцесса, следовательно, изучаемая информация соответствует закону нормального распределения и ее можно использовать для корреляционного анализа [16]. График нормального распределения выборки можно получить, выбрав команду *Данные – Анализ данных – Регрессия – График нормальной вероятности*.

### 3. Определение вида связи и выбор формы модели

Сформулируем гипотезу о наличии линейной формы связи парных данных. Наиболее востребованным способом выбора модели из существующих (аналитический, графический и экспериментальный) является графический, предусматривающий построение диаграммы рассеяния:

1. Занести эмпирические данные в электронную таблицу.
2. Построить диаграмму рассеяния, в которой координатами точек являются значения двух переменных (*Вставка – Точечная диаграмма*).
3. Построить линию тренда (*Макет – Линия тренда – Линейное приближение*).
4. Провести анализ, установить характер связи данных и выбрать модель.

При увеличении значения переменной  $x$  происходит увеличение значения переменной  $y$ , что свидетельствует о согласованности изменения значений двух переменных на диаграмме рассеяния (рисунок 1).

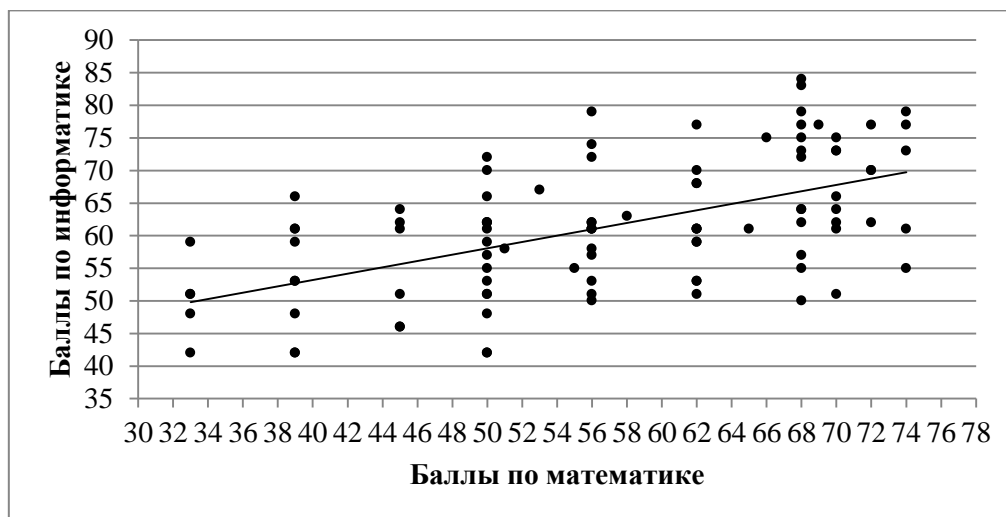


Рисунок 1 – Диаграмма рассеяния данных таблицы 1

#### 4. Получение уравнения парной регрессии

Предполагаем наличие линейной положительной корреляции, которой соответствует модель вида:  $y = ax + b + e$ ,

где  $x$  – фактор,

$y$  – отклик,

$a, b$  – коэффициенты,

$e$  – случайная ошибка (погрешность), возникшая, например, по причине неправильного описания структуры модели, либо из-за ошибок измерения.

Гипотеза принимается, т. е. для проведения исследования зависимости между переменными будет использована парная линейная регрессия. Предполагаемый характер связи будет справедлив и для данных из таблицы 2 (модель 2), поскольку данные являются подмножеством данных таблицы 1 (модель 1).

#### 5. Определение значений параметров для выбранной формы модели

Для определения значений параметров для линейной формы модели необходимо:

- вычислить коэффициенты уравнения линейной регрессии и выяснить их статистическую значимость;
- выяснить статистическую значимость коэффициента регрессии;
- вычислить выборочный коэффициент корреляции Пирсона, установить его значимость и оценить тесноту связи;
- вычислить коэффициент детерминации и выяснить его эконометрический смысл;

Вычисление коэффициентов уравнения линейной регрессии осуществляется следующими способами:

1. С использованием расчетных формул метода наименьших квадратов:

$$a = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{N \sum_{i=1}^N x_i^2 - \left( \sum_{i=1}^N x_i \right)^2} \quad b = \frac{\sum_{i=1}^N y_i \sum_{i=1}^N x_i^2 - \sum_{i=1}^N x_i y_i \sum_{i=1}^N x_i}{N \sum_{i=1}^N x_i^2 - \left( \sum_{i=1}^N x_i \right)^2}$$

где  $a, b$  – коэффициенты,

$x_i, y_i$  – исходные значения,

$N$  – размерность выборки.

2. С использованием команды *Данные – Анализ данных – Регрессия*. Результаты вычислений находятся в итоговой таблице на пересечении ячеек *Коэффициенты и Y-пересечение* и *Коэффициенты – Переменная X1*.

Получили уравнение регрессии  $y \approx 0,65x + 16,77$  для модели 1 и для модели 2:  $y \approx -0,3x + 91,72$ . В обоих уравнениях  $b > 0$ , поэтому относительное изменение отклика происходит медленнее, чем изменение фактора.

Статистическая значимость коэффициентов уравнения линейной регрессии. В модели 1 коэффициент  $b \approx 16,77$  показывает, каким будет зна-

чение  $y$  при  $x = 0$ , т. е. на величину  $y$  влияют и другие незначительные факторы, не описанные в модели. Коэффициент  $a \approx 0,65$  показывает степень влияния переменной  $x$  на  $y$ .

В модели 2 коэффициент  $b \approx 91,72$  показывает, каким будет значение  $y$  при  $x = 0$ , т. е. на величину анализируемого параметра  $y$  влияют другие значительные факторы, не описанные в модели. Коэффициент  $a \approx -0,3$  показывает степень влияния переменной  $x$  на  $y$ , причем знак « $-$ » указывает на отрицательное влияние.

Определение статистической значимости коэффициента регрессии (a) позволяет судить о среднем изменении отклика при изменении фактора на одну единицу. С увеличением результата ЕГЭ по математике у абитуриентов на 1 балл, значение результата по информатике возрастет в среднем на 0,65 баллов; у студентов значение результата по информатике уменьшится в среднем на 0,3 балла.

Вычисление выборочного коэффициента корреляции Пирсона можно осуществить следующими способами:

1. С использованием расчетной формулы [12, с. 84]:

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

где  $r_{xy}$  – коэффициент корреляции,

$N$  – размерность выборки,

$x_i$  и  $y_i$  – исходные пары данных,  $i = 1, \dots, N$ ,

$\bar{x}$ ,  $\bar{y}$  – среднеарифметические значения элементов  $x_i$  и  $y_i$ .

2. С использованием функции КОРРЕЛ(диапазон X, диапазон Y).

3. С использованием команды *Данные – Анализ данных – Корреляция*. В итоговой таблице на пересечении ячеек *Столбец 1* и *Столбец 2* найти значение.

В результате проведенных вычислений получили значения коэффициентов корреляции для модели 1:  $r_{xy} \approx 0,563381$  и для модели 2:  $r_{xy} \approx 0,375639$

Оценка тесноты корреляционной связи. Округленная до сотых величина обоих коэффициентов корреляции 0,56 и 0,38 положительная. Следовательно, между данными связь прямая. Значение коэффициента корреляции для таблицы 1 находится в интервале (0,5; 0,7), поэтому характер тесноты связи по шкале Чеддока определен как «заметный», а для таблицы 2 – в интервале (0,3; 0,5) как «умеренный» (таблица 6).

Таблица 6 – Качественная оценка тесноты корреляционной связи

Диапазон	0,1 – 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9 – 0,99
Характер тесноты связи	Слабая	Умеренная	Заметная	Высокая	Весьма высокая

Установление статистической значимости выборочных коэффициентов корреляции по  $t$ -критерию Стьюдента. Выдвигается гипотеза об отсутствии влияния фактора на отклик. Фактическое значение  $t$ -критерия вычисляется следующими способами:

1. С помощью расчетной формулы:

$$t_{\text{факт}} = \frac{|r| \sqrt{N-2}}{\sqrt{1-r^2}},$$

где  $t_{\text{факт}}$  – фактическим значение  $t$ -критерия,

$r$  – коэффициент корреляции,

$N$  – размерность выборки.

2. С использованием команды *Данные – Анализ данных – Регрессия*. Искомое значение находится в итоговой таблице на пересечении ячеек *Коэффициенты –  $t$ -статистика* и *Переменная X1*.

Для модели 1 определяем  $t_{\text{факт}} = \frac{0,563381\sqrt{98-2}}{\sqrt{1-0,317398}} = 6,681186$ .

По таблице критических точек распределения Стьюдента находим значение критерия:  $t_{\text{табл}} = 1,984$  для уровня значимости 0,05 с ближайшим значением 100 к величине числа степеней свободы:  $k = N - 2 = 98 - 2 = 96$ .

Сравним значения коэффициентов:  $6,681186 > 1,98$ . Коэффициент корреляции значимо отличается от нуля и  $t_{\text{факт}} > t_{\text{табл}}$  следовательно, статистически значим. Влияние фактора на отклик обнаружено, поэтому гипотеза отклоняется.

Для модели 2 определяем  $t_{\text{факт}} = \frac{0,375639\sqrt{22-2}}{\sqrt{1-0,141105}} = -1,812659$ .

Определяем значение критерия:  $t_{\text{табл}} = 2,086$  для уровня значимости 0,05 с числом степеней свободы:  $k = N - 2 = 22 - 2 = 20$  в таблице Стьюдента.

Сравним значения коэффициентов:  $0,375639324 < 2,086$ . Получили, что  $t_{\text{факт}} < t_{\text{табл}}$ , следовательно гипотеза принимается, фактор на отклик не оказывает существенного влияния.

Вычисление коэффициента детерминации парной линейной регрессии (%) осуществляется путем возведения в квадрат коэффициента корреляции и умножения результата на 100. В результате получим для модели 1:

$$R \approx 0,56^2 \cdot 100 \approx 0,317398 \cdot 100 \approx 32\%,$$

для модели 2:  $R \approx 0,38^2 \cdot 100 \approx 0,141105 \cdot 100 \approx 14\%$ .

Эконометрический смысл коэффициента детерминации состоит в том, что 32% результата по информатике у абитуриентов определяется их достижениями в области математики, а 68% связаны с другими факторами влияния. У студентов только 14% результата по информатике объясняется достижениями в области математики и 86% – влиянием других факторов.

### 6. Оценка средней абсолютной ошибки аппроксимации для каждой модели

Вычисление средней абсолютной ошибки прогноза (*MAPE* – *Mean Absolute Percent Error*) показывает, на сколько процентов в среднем прогноз отклоняется от факта и осуществляется по формуле:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{e_i}{y_i} \right| \cdot 100,$$

где  $N$  – размерность выборки,

$e_i$  – величина погрешности, равная  $y_i - ux_i$ ,

$y_i$  – исходные значения,

$y$  – новые значения, полученные из уравнения  $y = ax + b$ .

Для модели 1  $MAPE \approx 11,7\%$ , для модели 2:  $MAPE \approx 8\%$ . При прогнозе по модели 1 ошибка в среднем будет составлять 11,7%, а по модели 2 – 8%. Это означает, что модель 2 описана с меньшими погрешностями, т. е. является более точной. Считается, что если  $MAPE < 15\%$ , то результат является приемлемым (достоверным) и данное уравнение можно использовать в качестве регрессии.

### 7. Оценка полученных результатов и их обобщение

1. Выполнено исследование исходных данных на однородность, произведена оценка вариационной изменчивости.

2. Выявлена нормальность распределения данных с применением критерия асимметрии и эксцесса, определена значимость результатов по  $t$ -критерию Стьюдента.

3. Произведен анализ эмпирических данных, установлен характер их связи, выбрана линейная форма модели.

4. Обоснован выбор метода исследования – регрессионно-корреляционный анализ и программное средство – табличный процессор *MS Excel*.



5. Рассмотрены несколько вариантов выполнения расчетов и представления результатов средствами табличного процессора *MS Excel*.

6. Выявлена статистическая значимость коэффициента регрессии: с увеличением результата ЕГЭ по математике у абитуриентов на 1 балл, значение результата по информатике возрастет в среднем на 0,65 баллов; у студентов значение результата по информатике уменьшится в среднем на 0,3 балла.

7. Вычислена величина выборочного коэффициента корреляции Пирсона и установлена его значимость по  $t$ -критерию Стьюдента. Проведена оценка тесноты связи по таблице Чеддока.

8. Вычислена величина коэффициента детерминации, которая показывает: сколько процентов отклика объясняется с помощью фактора, включенного в модель: у абитуриентов 32% результата по информатике объясняется результатами по математике, у студентов – 14%.

9. Установлено качество уравнений регрессии по величинам средних абсолютных ошибок аппроксимации (*MAPE*) для каждой модели.

10. Выполнено прогнозирование по выбранной модели для данных таблицы 1. Значение коэффициента детерминации для абитуриентов – 32%, что превышает 30%, и означает: прогнозирование по исследуемой модели целесообразно. Прогнозирование для модели 2 не требуется.

### **Выводы**

Исследование подтвердило гипотезу о прямой, заметной корреляционной связи эмпирических данных модели 1 и умеренной корреляционной связи данных модели 2, правильно выбранной линейной модели. В результате проведения анализа показателей построенных моделей выяснено, что первокурсники имеют преимущество, по сравнению с абитуриентами, не только в суммарном исчислении баллов, но и в качественных характеристиках результатов ЕГЭ по базовым предметам, что может являться одним из условий успешного обучения в вузе на базе активной умственной дея-

тельности. Активность умственной деятельности студентов проявляется, например, в готовности к осознанному и систематичному выполнению учебных заданий; в стремлении к повышению интеллектуального уровня и способности к самостоятельной работе. Успешное обучение в вузе, по мнению автора, связано с занятием каждодневной, трудоемкой и мотивированной учебной деятельностью [7, 17].

Опыт выполнения исследования, полученный в данной работе, можно применить к результатам ЕГЭ по двум профильным предметам в учебном заведении любого уровня образования.

### Литература

1. Анищик Т. А. Алгоритмизация и программирование на языке Паскаль : учеб. пособие / Т. А. Анищик. – Краснодар : КубГАУ, 2016. – 160 с.
2. Анищик Т. А. Методические аспекты преподавания алгоритмизации и программирования на агроинженерных направлениях обучения / Т. А. Анищик, В. С. Маций // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар : КубГАУ, 2017. – № 08 (132). С. 23–34.
3. Анищик Т. А. Основы алгоритмического программирования на языке Паскаль : учеб. пособие / Т. А. Анищик. – Краснодар : КубГАУ, 2017. – 183 с.
4. Анищик Т. А. Проблемы преподавания информационных технологий в аграрном вузе и подходы к их решению / Т. А. Анищик // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар : КубГАУ, 2017. – № 07 (131). С. 644–658.
5. Анищик Т. А. Проблемы формирования информационно-коммуникационной компетентности современного агроинженера на занятиях по информатике / Т. А. Анищик, Е. В. Фешина // Наука XXI века: проблемы, перспективы и актуальные вопросы развития общества, образования и науки: сб. науч. тр. по матер. Межвуз. осенней науч. конф. – Майкоп : Изд-во «Магарин О. Г.», 2018. С. 23–29.
6. Анищик Т. А. Об алгоритме усвоения знаний, умений и навыков, реализуемом информационными технологиями / Т. А. Анищик // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар : КубГАУ, 2008. – № 05 (039). С. 222–229.
7. Анищик Т. А. Рабочая тетрадь по математическим и логическим основам информатики [Текст] : учеб. -метод. пособие / Т. А. Анищик, А. С. Креймер – Краснодар : КубГАУ, 2005. – 96 с.
8. Арсентьева М. В. Психологические факторы успешного обучения студентов в вузе // Известия ТулГУ. Технические науки. 2012. №11-2.
9. Боченков С. А., Вальдман И. А. Вопросы интерпретации и представление результатов ЕГЭ: проблемы и возможные решения. URL:

[https://vo.hse.ru/data/2014/08/04/1314337503/2013-3\\_Vochenkov\\_Valdman.pdf](https://vo.hse.ru/data/2014/08/04/1314337503/2013-3_Vochenkov_Valdman.pdf) (дата обращения: 10.08.2020).

10. Рейтинги лучших вузов и школ России 2019 // URL: [https://raex-a.ru/media/uploads/bulletins/pdf/2019\\_vuz\\_bul.pdf](https://raex-a.ru/media/uploads/bulletins/pdf/2019_vuz_bul.pdf) (дата обращения: 01.04.2020).

11. Рейтинг востребованности вузов в РФ – 2019 // URL: <https://na.ria.ru/20200324/1569014904.html> (дата обращения: 05.04.2020).

12. Карманова А. В. Высшая математика [Текст] : учеб. пособие / А. В. Карманова, Н. А. Соловьева. – Краснодар : КубГАУ, 2018. – 96 с.

13. Ермакова И. А. Анализ степени подготовленности первокурсников к обучению математике в техническом вузе / Ермакова И. А., Гоголин В. А., Ефремова Л. Ю. // Наука и образование: современные тренды. Чебоксары : «Итерактив плюс», 2015, № 3 (9), с. 288–296.

14. Капцов А. В. Колесникова Е. И. Взаимосвязь уровня интеллекта студентов и результатов сдачи ЕГЭ // Вестник Самарской гуманитарной академии. Серия: Психология. 2009. № 2.

15. Хавенсон Т. Е., Соловьева А. А. Связь результатов Единого государственного экзамена и успеваемости в вузе // Вопросы образования. 2014. №1.

16. Хромцова Л. С. Корреляционно-регрессионный анализ основных показателей нефтедобывающей промышленности // Экономический анализ: теория и практика. 2007. №7.

17. Фешина Е. В. Экономическая необходимость решения проблемы активизации мыслительной деятельности у студентов инженерных направлений обучения и подходы к их решению / Е. В. Фешина, Т. А. Анищик // Вестник Академии знаний, 2018. № 6 (29). – С. 259–264.

18. Чернышова Н. А. Связь результатов ЕГЭ и академических успехов студентов в сельскохозяйственном вузе / Н. А. Чернышова // Вестник Нижегородского университета им. Н. И. Лобачевского. Серия: Социальные науки, 2017, № 1 (45), с. 171–177.

19. Информация о приеме в 2020 году / Бакалавриат, специалитет, магистратура / Проходной балл 2018 г. и 2019 г.

URL: <https://kubsau.ru/upload/iblock/180/180cdaeabb76ce6f9f6a9e140546f766.pdf> (дата обращения: 10.08.2020).

## References

1. Anishhik T. A. Algoritmizacija i programmirovanie na jazyke Paskal' : ucheb. posobie / T. A. Anishhik. – Krasnodar : KubGAU, 2016. – 160 s.

2. Anishhik T. A. Metodicheskie aspekty prepodavaniya algoritmizacii i programmirovaniya na agroinzhenernyh napravlenijah obuchenija / T. A. Anishhik, V. S. Macij // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar : KubGAU, 2017. – № 08 (132). S. 23–34.

3. Anishhik T. A. Osnovy algoritmicheskogo programmirovaniya na jazyke Paskal' : ucheb. posobie / T. A. Anishhik. – Krasnodar : KubGAU, 2017. – 183 s.

4. Anishhik T. A. Problemy prepodavaniya informacionnyh tehnologij v agrarnom vuze i podhody k ih resheniju / T. A. Anishhik // Politematicheskij setевой jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar : KubGAU, 2017. – № 07 (131). S. 644–658.

5. Anishhik T. A. Problemy formirovaniya informacionno-kommunikacionnoj kompetentnosti sovremennogo agroinzhenera na zanjatijah po informatike / T. A. Anishhik, E. V. Feshina // Nauka XXI veka: problemy, perspektivy i aktual'nye voprosy razvitija ob-

shhestva, obrazovaniya i nauki: sb. nauch. tr. po mater. Mezhvuz. osennej nauch. konf. – Majkop : Izd-vo «Magarin O. G.», 2018. S. 23–29.

6. Anishhik T. A. Ob algoritme usvoeniya znanij, umenij i navykov, realizuemom informacionnymi tehnologijami / T. A. Anishhik // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar : KubGAU, 2008. – № 05 (039). S. 222–229.

7. Anishhik T. A. Rabochaja tetrad' po matematicheskim i logicheskim osnovam informa-tiki [Tekst] : ucheb. -metod. posobie / T. A. Anishhik, A. S. Krejmer – Krasnodar : KubGAU, 2005. – 96 s.

8. Arsent'eva M. V. Psihologicheskie faktory uspešnogo obuchenija studentov v vuze // Izvestija TulGU. Tehniceskie nauki. 2012. №11-2.

9. Bochenkov S. A., Val'dman I. A. Voprosy interpretacii i predstavlenie rezul'tatov EGJe: problemy i vozmozhnye reshenija. URL: [https://vo.hse.ru/data/2014/08/04/1314337503/2013-3\\_Bochenkov\\_Valdman.pdf](https://vo.hse.ru/data/2014/08/04/1314337503/2013-3_Bochenkov_Valdman.pdf) (data obrashhenija: 10.08.2020).

10. Rejtingi luchshih vuzov i shkol Rossii 2019 // URL: [https://raex-a.ru/media/uploads/bulletins/pdf/2019\\_vuz\\_bul.pdf](https://raex-a.ru/media/uploads/bulletins/pdf/2019_vuz_bul.pdf) (data obrashhenija: 01.04.2020).

11. Rejting vostrebovannosti vuzov v RF – 2019 //

URL: <https://na.ria.ru/20200324/1569014904.html> (data obrashhenija: 05.04.2020).

12. Karmanova A. V. Vysshaja matematika [Tekst] : ucheb. posobie / A. V. Karmanova, N. A. Solov'eva. – Krasnodar : KubGAU, 2018. – 96 s.

13. Ermakova I. A. Analiz stepeni podgotovlennosti pervokursnikov k obucheniju matematike v tehničeskom vuze / Ermakova I. A., Gogolin V. A., Efremova L. Ju. // Nauka i obrazovanie: sovremennye trendy. Cheboksary : «Iteraktiv pljus», 2015, № 3 (9), s. 288–296.

14. Kapcov A. V. Kolesnikova E. I. Vzaimosvjaz' urovnja intellekta studentov i rezul'tatov sdachi EGJe // Vestnik Samarskoj gumanitarnej akademii. Serija: Psihologija. 2009. № 2.

15. Havenson T. E., Solov'eva A. A. Svjaz' rezul'tatov Edinogo gosudarstvennogo jek-zamena i uspevaemosti v vuze // Voprosy obrazovaniya. 2014. №1.

16. Hromcova L. S. Korreljacionno-regressionnyj analiz osnovnyh pokazatelej neftedobывajushhej promyshlennosti // Jekonomicheskij analiz: teorija i praktika. 2007. №7.

17. Feshina E. V. Jekonomicheskaja neobhodimost' reshenija problemy aktivizacii myslitel'noj dejatel'nosti u studentov inženernyh napravlenij obuchenija i podhody k ih resheniju / E. V. Feshina, T. A. Anishhik // Vestnik Akademii znanij, 2018. № 6 (29). – S. 259–264.

18. Chernyshova N. A. Svjaz' rezul'tatov EGJe i akademicheskikh uspehov studentov v sel'skohozjajstvennom vuze / N. A. Chernyshova // Vestnik Nizhegorodskogo universiteta im. N. I. Lobachevskogo. Serija: Social'nye nauki, 2017, № 1 (45), s. 171–177.

19. Informacija o prieme v 2020 godu / Bakalavriat, specialitet, magistratura / Prohodnoj ball 2018 g. i 2019 g.

URL: <https://kubsau.ru/upload/iblock/180/180cdaeabb76ce6f9f6a9e140546f766.pdf> (data obrashhenija: 10.08.2020).