

Статистический анализ факторов академической успеваемости (на примере обучающихся экономического факультета ФГБОУ ВО «ВГУ»)

П. А. Канапухин¹, В. В. Коротких², С. С. Щекунских³

^{1, 2, 3} Воронежский государственный университет, Университетская пл., 1,
394018, Воронеж, Российская Федерация

Для цитирования: Канапухин П. А., Коротких В. В., Щекунских С. С. Статистический анализ факторов академической успеваемости (на примере обучающихся экономического факультета ФГБОУ ВО «ВГУ») // Вестник Воронежского государственного университета. Серия: Экономика и управление. 2020. № 2. С. 27–44. DOI: 10.17308/econ.2020.2/2899

Предмет. Образовательные процессы являются одной из конституирующих составляющих жизни современного социума. При этом процессы высшего образования играют важную роль в достижении как индивидуальной, так и коллективной эффективности и поэтому подвергаются тщательному государственному и общественному мониторингу.

Цели. Выявление статистически значимых детерминант академической успеваемости обучающихся с целью повышения качества учебного процесса.

Метод. В процессе исследования использовались параметрические и непараметрические методы анализа данных и машинного обучения. Информационную базу исследования сформировали данные об учебной деятельности, успеваемости, баллах на Едином государственном экзамене, а также обобщенные адресные данные обучающихся в бакалавриате по направлениям «Экономика», «Менеджмент» 2013–2015 гг. набора.

Результаты. В эмпирической части исследования протестированы несколько рабочих гипотез о связи академической успеваемости обучающихся с их домашним регионом, полом, источником финансирования учебной деятельности, а также результатами вступительных испытаний. Также протестирована гипотеза о взаимосвязи преждевременного завершения учебной деятельности обучающихся по причине неуспеваемости и их низким вступительным баллом. Мы установили, что стресс, связанный со сменой места жительства в период учебы, почти не сказывается на показателях успеваемости. Обучающиеся девушки демонстрируют более высокие результаты на обоих направлениях подготовки. Девушки, обучающиеся за счет федерального бюджета, характеризуются большей мотивацией на высокую успеваемость по сравнению с обучающимися за счет собственных средств. Статистически значимых линейных связей между вступительными баллами и успеваемостью обучающихся не выявлено. Поскольку обучающиеся, завершающие учебную деятельность по причине неуспеваемости, поступили преимущественно с низкими суммарными баллами ЕГЭ, имеет смысл значительно усилить фундаментальную подготовку, а именно аудиторную нагрузку по дисциплинам базовой части учебных планов, развивать конвергентные образовательные направления, а также укреплять междисциплинарные связи и практики изучения логики познания в целях сглаживания неоднородности в формировании компетенций у обучающихся.

Ключевые слова: стратегия, ресурсы, капитал, доход, рынок.

Введение

Высшее образование в современном мире – одна из важнейших составляющих жизни социума. Высшее образование играет ключевую роль как в индивидуальной, так и в коллективной эффективности.

Использование аналитического инструментария в исследовании успеваемости обучаю-

щихся нацелено на выявление статистически значимых детерминант успеваемости и определение методов воздействия на них. Данная проблематика рассматривается, например, такими учеными, как Т. Ю. Быстрова и др. [7], Е. Г. Антоненков и Ю. А. Крюков [6]. Однако в научной среде нет единого представления ни о пространстве детерминант успеваемости обучающихся, ни о методах его анализа.

Очевидно, что возможность использования баллов на ЕГЭ в качестве предикторов успеваемости обучающихся в университете широко изучается российскими учеными: Г. П. Мартыновым [14], Е. В. Кондратьевой и О. Ю. Кондратьевой [11], И. П. Степановой и др. [17]. Е. В. Тарасова [18] исследует особенности влияния климато-экологических условий, в частности дефицита солнечной инсоляции в зимний период года, на успешность умственной работы и эффективность обучения. М. В. Гранков и В. М. Аль-Габри [8] предлагают в качестве факторов успеваемости обучающихся рассматривать порядковый номер экзамена в сессию и учитывать специфику конкретных экзаменаторов на основе имеющейся статистики. В исследовании И. В. Никулиной и А. Э. Снежковой [15] изучается роль мотивационных переменных в эффективности учебной деятельности. Е. С. Шмарихина [20] для целей анализа факторов успеваемости обучающихся предлагает рассматривать систему из 27 показателей внутренних (личностных факторов) и внешних (факторов среды) показателей. В. В. Кузнецов и др. [13] исследовали взаимосвязь самооценки состояния здоровья и уровня заболеваемости с академической успеваемостью. Е. Е. Котова и А. С. Писарев [12] изучали возможность прогнозирования успеваемости обучающихся с учетом их индивидуальных параметров когнитивно-стилевого потенциала (познавательные-мыслительные стратегии обучающихся в процессе учебной деятельности).

Целью данной научной работы является исследование факторов успеваемости с помощью вероятностно-статистического аппарата. В ходе проведения исследования мы будем рассматривать следующие **рабочие гипотезы**.

1. Обучающиеся, прибывающие из различных регионов, не различаются по уровню академической успеваемости в университете.

2. Обучающиеся девушки и юноши не различаются по уровню академической успеваемости в университете.

3. Источник финансирования не оказывает влияния на мотивацию обучающегося к высокому уровню академической успеваемости.

4. Источник финансирования в равной степени оказывает влияние на мотивацию к высокому уровню академической успеваемости обучающихся девушек и юношей.

5. Завершение учебной деятельности по причине неуспеваемости не может быть обу-

словлено низким вступительным баллом обучающегося.

6. Вступительные баллы обучающихся не могут рассматриваться значимым предиктором академической успеваемости.

Методология

Проверку выдвинутых гипотез о различии целесообразно осуществить средствами дисперсионного анализа. Зависимыми переменными в дисперсионном анализе будут выступать средний балл обучающегося (гипотезы 1–4), суммарный балл ЕГЭ (вступительный экзамен) (гипотезы 5, 6), независимыми – домашний регион обучающегося, пол обучающегося, источник финансирования. Проверка гипотез сводится к задаче одномерного дисперсионного анализа. Достоверность статистического вывода по результатам параметрического дисперсионного анализа требует обязательного соблюдения требований о нормальности распределения зависимой переменной в сравниваемых группах и однородности дисперсий в группах по всем градациям номинальных переменных. Гипотезу о соответствии нормальному закону распределения анализируемого признака будем тестировать по критерию Шапиро – Уилка (Shapiro – Wilk Test) [3] или критерию отклонения распределения случайных величин от нормального (Probability Plot Correlation Coefficient Test for Normality) [4] в случае большой размерности сравниваемых групп. Гипотезу об однородности дисперсий анализируемого признака можно тестировать, используя критерий Левене (Levene Test) [1] или критерий Бартлетта (Bartlett Test) [2]. При нарушении любого из требований следует обратиться к непараметрическому дисперсионному анализу в версии Краскела – Уоллиса (Kruskal – Wallis Test) [5]. В этом случае нам потребуются решать совокупность отдельных задач одномерного однофакторного дисперсионного анализа.

Набор данных

Основными данными для проведения статистического анализа образовательного процесса выступают данные об успеваемости обучающихся очного отделения бакалавриата по направлениям «Экономика» и «Менеджмент» на экономическом факультете Воронежского государственного университета различных годов поступления по соответствующим предметам и соответствующим годам (табл. 1, 2).

Т а б л и ц а 1

Краткая характеристика набора данных «Экономика»

Характеристика	Год набора		
	2013	2014	2015
Зачислено обучающихся на первый курс, чел.	290	281	176
Успешно завершили обучение, чел.	254	224	131
Минимальный балл по сумме трех ЕГЭ	130	80	126
Максимальный балл по сумме трех ЕГЭ	278	289	272
Средний балл	3,5	3,18	3,05

Т а б л и ц а 2

Краткая характеристика набора данных «Менеджмент»

Характеристика	Год набора		
	2013	2014	2015
Зачислено обучающихся на первый курс, чел.	113	98	90
Успешно завершили обучение, чел.	95	76	72
Минимальный балл по сумме трех ЕГЭ	128	120	128
Максимальный балл по сумме трех ЕГЭ	274	257	256
Средний балл	3,55	3,15	3,48

Каждый обучающийся в обоих наборах данных имеет уникальный закодированный номер и описывается следующими номинативными переменными:

– завершение учебной деятельности (переменная «**out**» с двумя градациями «успешно закончил обучение», «отчислен»);

– пол обучающегося (переменная «**sex**» с двумя градациями);

– источник финансирования (переменная «**fund**» с двумя градациями «договор» – у обучающихся на договорной основе, «бюджет» – у остальных);

– год зачисления (переменная «**year**» с соответствующими градациями);

– домашний регион обучающегося (переменная «**geo**» со следующими градациями: «Воронеж» – у обучающихся из Воронежа, «ВоронежОбл» – у обучающихся из Воронежской области, «ЛипецкОбл» – у обучающихся из Липецка и Липецкой области, «БелгородОбл» – у обучающихся из Белгорода и Белгородской области, «ТамбовОбл» – у обучающихся из Тамбова и Тамбовской области, «Другое» – у всех остальных);

и несколькими количественными:

– результаты вступительных испытаний по математике ЕГЭ, баллы (переменная «**exam_math**»);

– результаты вступительных испытаний по обществознанию ЕГЭ, баллы (переменная «**exam_soc**»);

– результаты вступительных испытаний по русскому языку ЕГЭ, баллы (переменная «**exam_rus**»);

– средний балл обучающегося (переменная «**univav**»).

По значениям номинативных переменных в каждом наборе данных могут быть идентифицированы соответствующие группы обучающихся.

Данные предоставлены управлением информатизации и компьютерных технологий ФГБОУ ВО «Воронежский государственный университет». Мы констатируем отсутствие нарушений условий согласия на обработку и публикацию персональных данных обучающихся, персональные данные в явном виде нигде не используются.

Результаты эмпирического исследования

Результаты по набору данных «Экономика»

В табл. 3 приведены результаты параметрического дисперсионного анализа по набору данных «Экономика». В первых трех строках

Параметрический дисперсионный анализ

Фактор	Df	Sum Sq	Mean Sq	F value	P-value
geo	5	4,42	0,883	4,108	0,0011
sex	1	15,36	15,361	71,433	0,0002
fund	1	4,99	4,987	23,19	0,0001
Residuals	601	129,24	0,215	–	–

Результаты тестирования распределения исследуемого признака в сравниваемых группах на соответствие нормальному закону

Фактор	Градации фактора	Объем группы	P-value
geo	Воронеж	220	0,0002
	ВоронежОбл	156	0,0028
	ЛипецкОбл	36	0,2260
	БелгородОбл	30	0,1720
	ТамбовОбл	17	0,3940
	Другое	150	0,0004
sex	Девушки	422	0,0001
	Юноши	187	0,0001
fund	Договор	527	0,0001
	Бюджет	82	0,0001

представлены главные эффекты переменных «geo», «sex» и «fund», суммы квадратов отклонений (Sum Sq), средние суммы квадратов отклонений (Mean Sq) с указанием значения статистики Фишера (F value) и наблюдаемого р-уровня значимости (P-value). В последней строке таблицы приводятся суммы квадратов отклонений (Sum Sq), средние суммы квадратов отклонений (Mean Sq) для остатков или внутригруппового разброса. Как видим из таблицы результатов дисперсионного анализа, значимые различия выявлены по всем переменным.

Предварять содержательную интерпретацию результатов параметрического дисперсионного анализа должна процедура установления их достоверности по критерию Шапиро – Уилка (табл. 4) и критерию Левене.

Исходя из приведенных р-уровней значимости, требование по нормальности распределения не выполняется в группах, полученных по градациям всех рассматриваемых факторов (р-уровень значимости < 0,05). В этом случае уже нет необходимости в проведении тестирования на гомогенность. Для получения достоверных результатов мы будем использовать непараметрическую процедуру дисперсионного анализа Краскела – Уоллиса для **тестирования рабочих гипотез 1–3**.

Содержательная интерпретация результатов

На рис. 1 изображены распределения успеваемости обучающихся в группах, сформированных по градациям фактора «домашний регион».

Поскольку требования параметрического дисперсионного анализа не были выполнены, для получения достоверных результатов воспользуемся аналогичными непараметрическими процедурами (табл. 5).

Приведенный в тесте Краскела – Уоллиса р-уровень значимости (0,0039) указывает на необходимость отклонения нулевой гипотезы об отсутствии различий в успеваемости обучающихся в зависимости от домашнего региона. Обнаружены значимые различия между рассматриваемыми группами.

Однако выявление различий не является окончательным результатом анализа, поскольку в случае более двух градаций фактора он предполагает дальнейшее раскрытие соотношений между средними рангами. Для выявления того, какие именно группы отличаются друг от друга, воспользуемся процедурой апостериорных сравнений, представляющих собой попарные сравнения исследуемых групп для обнаружения различий между ними (табл. 6).

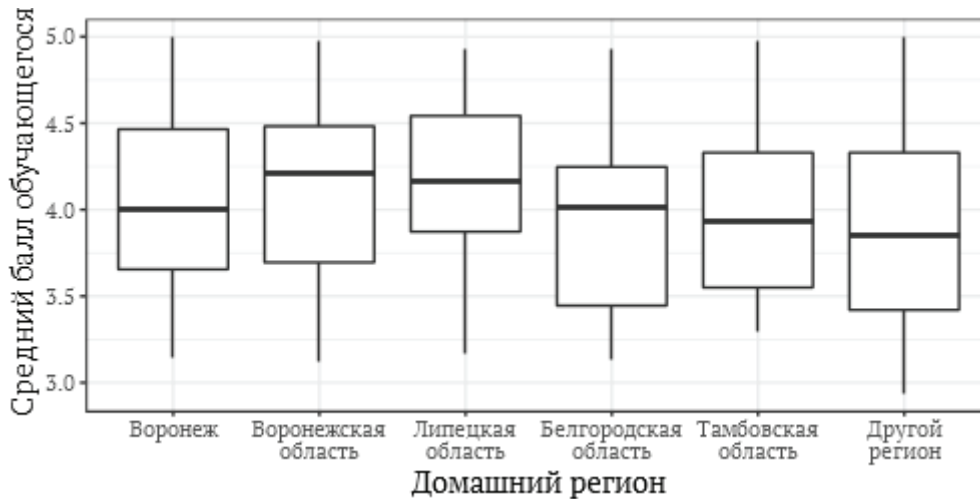


Рис. 1. Диаграмма размаха успеваемости обучающихся по группам

Т а б л и ц а 5

Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	df	P-value
Успеваемость	Домашний регион	17,28	5	0,0039

Т а б л и ц а 6

Апостериорные сравнения групп по критерию Тьюки – Крамера

Градации фактора	Средний ранг по группе	Однородные группы
ЛипецкОбл	351,3472	AB
ВоронежОбл	330,3301	B
Воронеж	315,8341	B
ТамбовОбл	282,2059	AB
БелгородОбл	267,2667	AB
Другое	261,7733	A

Во втором столбце табл. 6 приведены средние ранги исследуемого признака в группах, выделенных по градациям фактора «домашний регион». Несмотря на то что группа обучающихся из Липецка и Липецкой области имеет наивысший ранг по успеваемости, значимых различий у этой группы не выявлено. В третьем столбце приведены данные об однородности сравниваемых групп. Наличие у двух групп хотя бы одного одинакового символа говорит об отсутствии различий между ними по значениям исследуемого признака на пятипроцентном уровне значимости.

Вывод по фактору «домашний регион»: по результатам апостериорных сравнений можно констатировать наличие значимых различий между группами обучающихся бакалавриата по

направлению «Экономика» из Воронежа и Воронежской области и группой из «других» регионов.

На рис. 2 приведены распределения успеваемости обучающихся в группах, сформированных по градациям фактора «пол». Обратимся к результатам непараметрического дисперсионного анализа (табл. 7).

Проведенный тест Краскела – Уоллиса позволяет отклонить нулевую гипотезу об отсутствии различий в успеваемости юношей и девушек (p-уровень значимости менее 1 %). Поскольку фактор имеет только две градации, нет необходимости проведения апостериорных сравнений.

Вывод по фактору «пол»: успеваемость девушек, обучающихся в бакалавриате по направлению «Экономика», значимо превосходит успеваемость юношей.

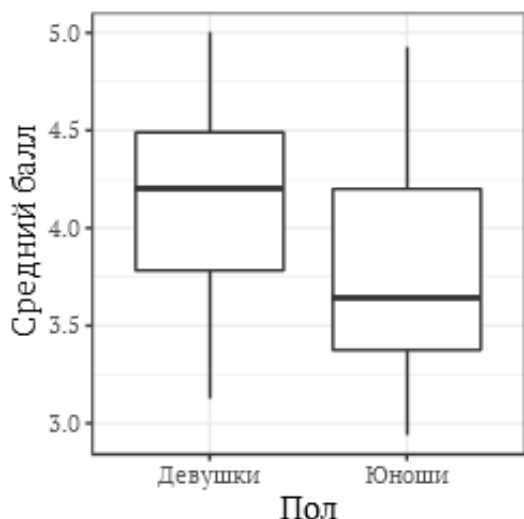


Рис. 2. Диаграмма размаха успеваемости обучающихся по группам

Т а б л и ц а 7
Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	df	P-value
Успеваемость	Пол	60,324	1	8,048e-15

На рис. 3 представлены распределения успеваемости обучающихся в группах, сформированных по градациям фактора «источник финансирования». По результатам процедуры Краскела – Уоллиса (табл. 8) мы получили достаточные основания для отклонения нулевой гипотезы об отсутствии различий в успеваемости обучающихся за счет собственных средств и за счет бюджетных средств на р-уровне значимости 1 %.

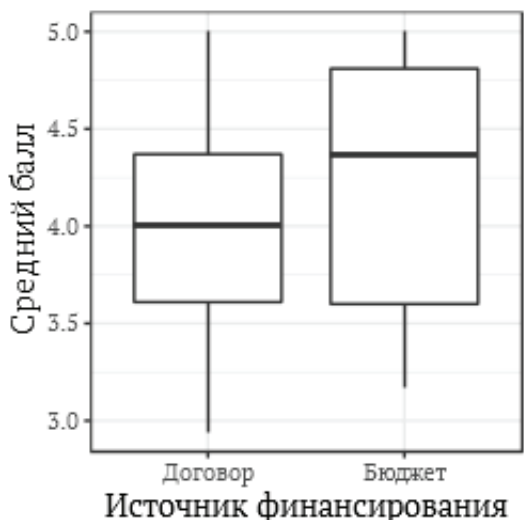


Рис. 3. Диаграмма размаха успеваемости обучающихся по группам

Т а б л и ц а 8
Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	df	P-value
Успеваемость	Источник финансирования	8,852	1	0,0029

Вывод по фактору «источник финансирования»: успеваемость обучающихся за собственные средства значительно уступает успеваемости обучающихся за счет бюджетных средств.

Интересным представляется исследовать, как взаимодействуют факторы «пол» и «источник финансирования» (рабочая гипотеза 4). Рассмотрим распределения исследуемого признака по соответствующим группам (рис. 4).

Тест Краскела – Уоллиса (табл. 8) показал, что группы обучающихся, сформированные по всем комбинациям градаций факторов «пол» и «источник финансирования», имеют значимые различия на уровне значимости 1 %, что указывает на необходимость отклонения нулевой гипотезы об отсутствии различий в исследуемых группах. Поскольку мы имеем четыре комбинации градаций факторов, проведем процедуру апостериорных сравнений (табл. 9).

Самую высокую успеваемость имеют девушки, обучающиеся за счет средств федерального бюджета. Эта группа значительно отличается от остальных. Несколько меньшую успеваемость демонстрируют девушки, обучающиеся за счет собственных средств. Такая группа также значительно отличается от остальных. Группы юношей оказались однородными.

Вывод по взаимодействию факторов «пол», «финансирование»: по результатам апостериорных сравнений можно констатировать наличие значимых различий в успеваемости между группами девушек и между группами девушек и юношей.

Результаты по набору данных «Менеджмент»

По результатам параметрического дисперсионного анализа (табл. 10) можно было бы говорить о значимых различиях успеваемости на всех уровнях номинативных переменных. Однако такой вывод не являлся бы достовер-

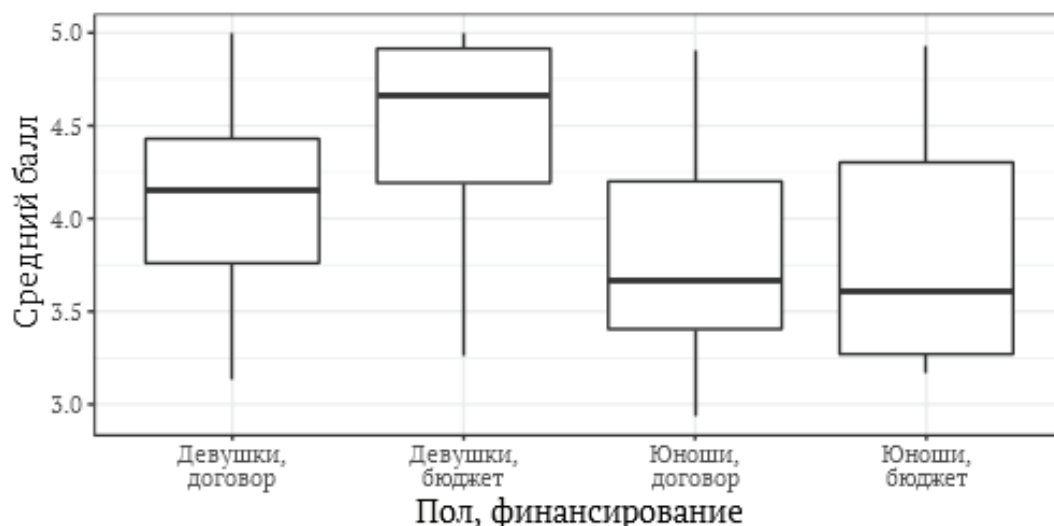


Рис. 4. Диаграмма размаха успеваемости обучающихся по группам

Т а б л и ц а 8

Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	df	P-value
Успеваемость	Пол, источник финансирования	81,942	3	2,2e-16

Т а б л и ц а 9

Апостериорные сравнения групп по критерию Тьюки – Крамера

Градации фактора	Средний ранг по группе	Однородные группы
Девушки, бюджет	450,3600	B
Девушки, договор	327,2769	A
Юноши, договор	223,0871	C
Юноши, бюджет	215,6719	C

Т а б л и ц а 10

Параметрический дисперсионный анализ

Фактор	Df	Sum Sq	Mean Sq	F value	P-value
geo	5	3,19	0,637	2,468	0,0334
sex	1	14,03	14,025	54,323	0,0001
fund	1	7,24	7,237	28,031	0,0001
Residuals	235	60,67	0,285	-	-

ным, поскольку распределения исследуемого признака значимо отличаются от нормального почти на всех градациях факторов (табл. 11). Для тестирования рабочих гипотез 1–3 на наборе данных «Менеджмент» воспользуемся непараметрическим вариантом дисперсионного анализа Краскела – Уоллиса.

Содержательная интерпретация результатов

Распределения успеваемости обучающихся бакалавриата по направлению «Менеджмент» в группах, сформированных по градациям фактора «домашний регион», имеют следующий вид (рис. 5).

Результаты тестирования распределения исследуемого признака в сравниваемых группах на соответствие нормальному закону

Фактор	Градации фактора	Объем группы	P-value
geo	Воронеж	113	0,0001
	ВоронежОбл	39	0,0706
	ЛипецкОбл	25	0,1320
	БелгородОбл	7	0,6450
	ТамбовОбл	4	0,6550
	Другое	55	0,0026
sex	Девушки	151	0,0001
	Юноши	92	0,0009
fund	Договор	196	0,0001
	Бюджет	47	0,0001

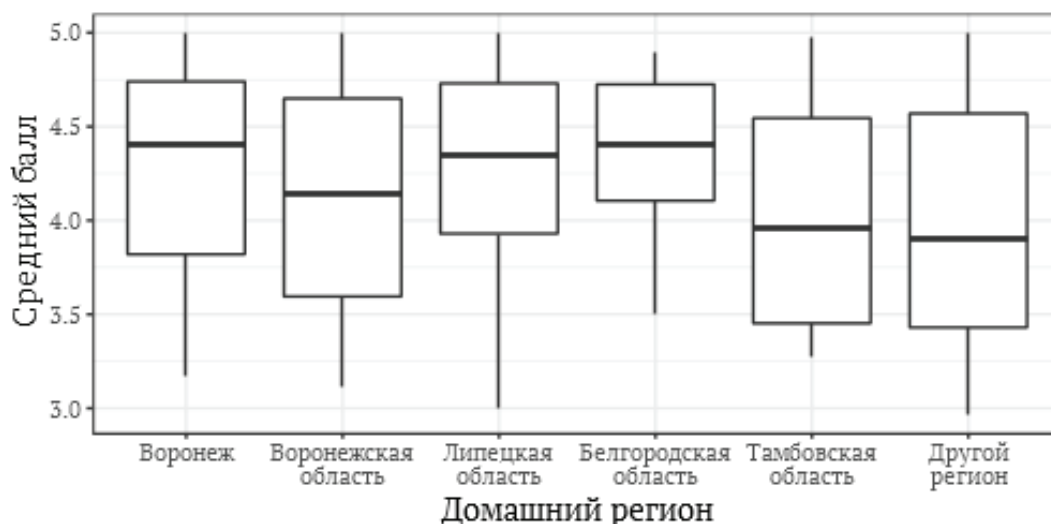


Рис. 5. Диаграмма размаха успеваемости обучающихся по группам

Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	df	P-value
Успеваемость	Домашний регион	8,38	5	0,1365

По результатам реализации непараметрической процедуры дисперсионного анализа (табл. 12) можно констатировать отсутствие различий в успеваемости обучающихся в зависимости от домашнего региона.

Вывод по фактору «домашний регион»: обучающиеся бакалавриата по направлению «Менеджмент» демонстрируют однородный уровень успеваемости вне зависимости от домашнего региона.

На рис. 6 приведены распределения успеваемости обучающихся в группах, сформированных по градациям фактора «пол».

Тест Краскела – Уоллиса (табл. 13) позволяет отклонить нулевую гипотезу об отсутствии различий в успеваемости юношей и девушек (p-уровень значимости менее 1 %).

Вывод по фактору «пол»: успеваемость девушек, обучающихся в бакалавриате по направлению «Менеджмент», значимо превосходит успеваемость юношей.

На рис. 7 приведены распределения успеваемости обучающихся в группах, сформированных по градациям фактора «источник финансирования».

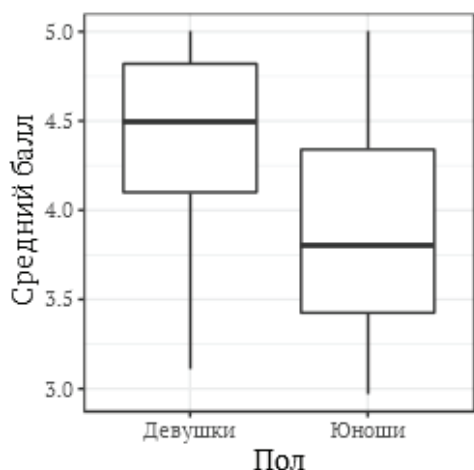


Рис. 6. Диаграмма размаха успеваемости обучающихся по группам

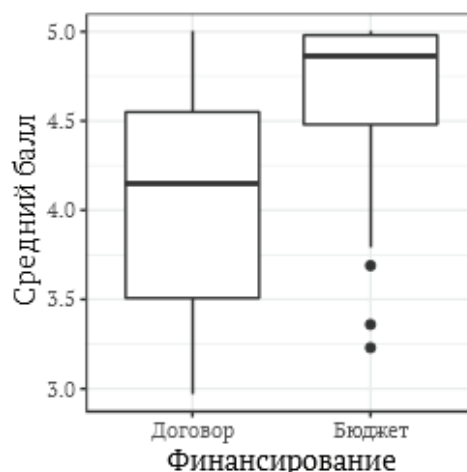


Рис. 7. Диаграмма размаха успеваемости обучающихся по группам

Таблица 13

Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	df	P-value
Успеваемость	Пол	34,194	1	4,989e-09

Таблица 14

Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	df	P-value
Успеваемость	Источник финансирования	42,691	1	6,409e-11

По результатам дисперсионного анализа (табл. 14) мы получили достаточные основания для отклонения нулевой гипотезы об отсутствии различий в успеваемости обучающихся за счет собственных средств и за счет бюджетных на p -уровне значимости 1 %.

Вывод по фактору «источник финансирования»: успеваемость обучающихся за соб-

ственные средства значительно уступает успеваемости обучающихся за счет бюджетных средств.

Исследуем взаимодействие факторов «пол» и «источник финансирования» на анализируемой выборке. Рассмотрим распределения исследуемого признака по соответствующим группам (рис. 8).

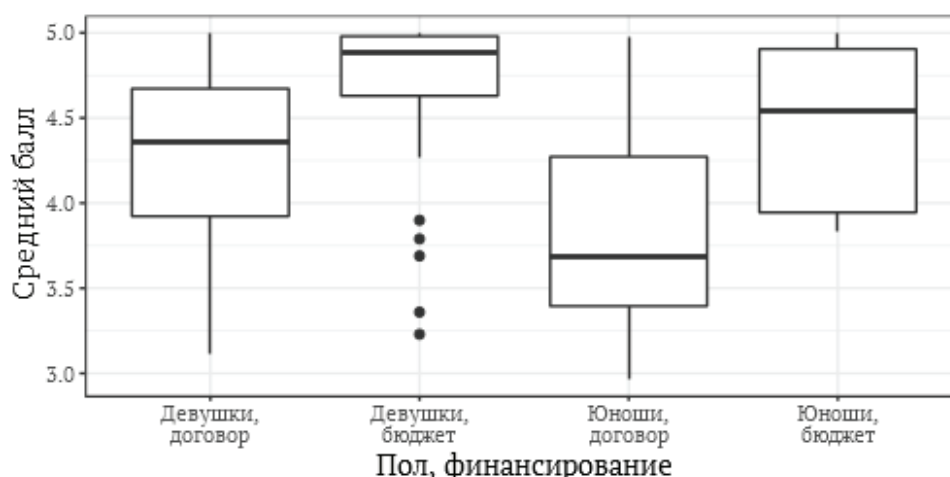


Рис. 8. Диаграмма размаха успеваемости обучающихся по группам

Таблица 15

Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	df	P-value
Успеваемость	Пол, источник финансирования	63,842	3	8.874e-14

Анализ выявил различия между рассматриваемыми группами. Вероятность получить такие или более ярко выраженные различия в условиях, когда верна нулевая гипотеза, составляет менее 1 %. Это указывает на необходимость отклонения нулевой гипотезы об отсутствии различий в исследуемых группах. Проведем процедуру апостериорных сравнений и выясним, между какими именно группами обнаружены значимые различия (табл. 16).

Т а б л и ц а 16
Апостериорные сравнения групп по критерию Тьюки – Крамера

Градации фактора	Средний ранг по группе	Однородные группы
Девушки, бюджет	187,5384	B
Юноши, бюджет	155,9375	AB
Девушки, договор	126,9241	A
Юноши, договор	81,7738	C

Самый высокий уровень успеваемости демонстрируют девушки, обучающиеся за счет средств федерального бюджета. Девушки, обучающиеся за собственные средства, имеют значимо более низкую успеваемость. В отличие от результата на наборе данных «Экономика» группы юношей значимо отличаются по успеваемости, значительно более высокую успеваемость имеют обучающиеся за средства федерального бюджета.

Вывод по взаимодействию факторов «пол», «финансирование»: по результатам апостериорных сравнений можно констатировать наличие значимых различий как внутри групп девушек, так и внутри групп юношей.

Анализ эффективности учебной деятельности обучающихся

Объединим два имеющихся набора данных и протестируем нулевую гипотезу об отсутствии различий суммарных баллов ЕГЭ в группах обучающихся, сформированных по градациям номинативной переменной «завершение учебной деятельности» (рис. 9).

По причине того, что распределение исследуемого признака значимо отклоняется от нормального в группе отчисленных обучающихся (табл. 17), используем непараметрическую процедуру дисперсионного анализа (табл. 18).

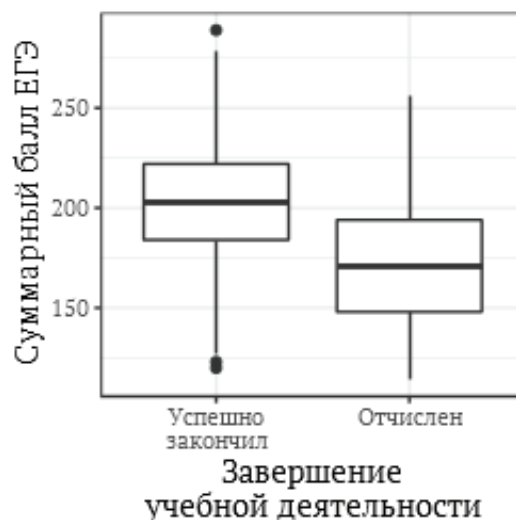


Рис. 9. Диаграмма размаха суммарного балла ЕГЭ по группам

Т а б л и ц а 17
Результаты тестирования распределения исследуемого признака в сравниваемых группах на соответствие нормальному закону

Фактор	Градации фактора	Объем группы	P-value
Out	Успешно закончил обучение	665	0,584
	Отчислен	164	0,021

Т а б л и ц а 18
Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	df	P-value
Суммарный балл ЕГЭ	Завершение учебной деятельности	113,590	1	2.2e-16

Вывод по фактору «завершение учебной деятельности»: обучающиеся бакалавриата по направлениям «Экономика» и «Менеджмент», успешно завершающие учебную деятельность, имели статистически значимо более высокий суммарный балл ЕГЭ.

Ранее мы установили, что среди обучающихся, успешно заканчивающих обучение, более высокую успеваемость демонстрируют те, кто финансируется за счет средств федерального бюджета. Предположим, что мотивация обучающихся вне зависимости от способа завершения учебной деятельности различна для разных источников финанси-

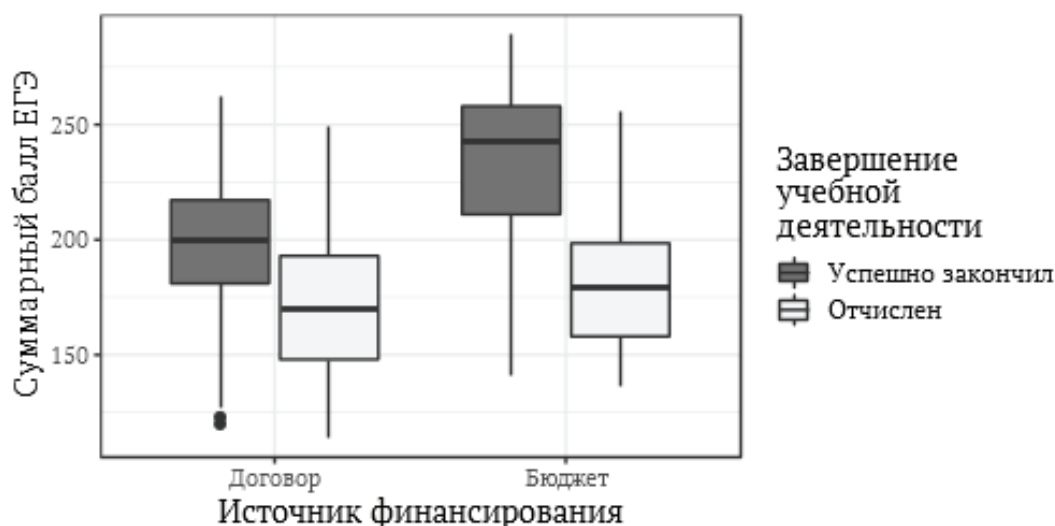


Рис. 10. Диаграмма размаха суммарного балла ЕГЭ по группам

Т а б л и ц а 19

Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	Df	P-value
Суммарный балл ЕГЭ	Завершение учебной деятельности, источник финансирования	179,77	3	2.2e-16

Т а б л и ц а 20

Апостериорные сравнения групп по критерию Тьюки – Крамера

Градации фактора	Средний ранг по группе	Однородные группы
Успешно закончил, бюджет	753,2287	B
Успешно закончил, договор	599,1750	A
Отчислен, договор	98,5000	C
Отчислен, бюджет	98,5000	C

ния. Добавим в анализ фактор «источник финансирования» и визуализируем распределения исследуемого признака (рис. 10). По графику видно, что распределения значительно отличаются от нормального, поэтому воспользуемся непараметрическим анализом.

По результатам (табл. 19) можно заключить, что в рассматриваемых группах обучающихся обнаружены значимые различия, поскольку вероятность получить такие или более ярко выраженные различия при условии, что верна нулевая гипотеза об отсутствии взаимосвязи суммарных баллов ЕГЭ, источника финансирования и завершения учебной деятельности обучающихся, составляет менее 1 %. Нулевая гипотеза отклоняется. Проведем процедуру апостериорных сравнений (табл. 20).

Вывод по факторам «завершение учебной деятельности», «источник финансирования»: группы обучающихся, отчисленных за неуспеваемость, статистически однородны по суммарному баллу ЕГЭ вне зависимости от источника финансирования.

В заключение эмпирического раздела исследования мы попытаемся ответить на вопрос о взаимосвязи успеваемости обучающихся в университете и их суммарного балла ЕГЭ. Визуализация соответствующих данных представлена на рис. 11. Мы можем говорить лишь о слабой положительной линейной взаимосвязи, поскольку обе регрессионные прямые имеют положительный наклон, но большая часть точек расположена достаточно далеко от соответствующих прямых. Суммарный балл ЕГЭ не дает исчерпывающего объяснения успеваемости

обучающихся университета. Можно говорить о том, что успеваемость при обучении в университете отлична от успеваемости в школе.

Интересным представляется изучение вопроса о том, одинаково ли изменяется успеваемость у обучающихся на различных направлениях подготовки по сравнению с их начальным уровнем успеваемости (суммарным баллом ЕГЭ). Рисунок 12-1 показывает, что группы обучающихся, поступивших на оба направления подготовки, являются однородными. Однако итоговая успеваемость по на-

правлениям стала значимо различаться, причем обучающиеся по направлению «Менеджмент» демонстрируют более высокую успеваемость (рис. 12-2). Надежность выводов обеспечена непараметрическим дисперсионным анализом (табл. 21).

Выявленная разница по большей части обусловлена тем, что девушки, обучающиеся на направлении «Менеджмент», демонстрируют более высокую успеваемость по сравнению с девушками с направления «Экономика» (рис. 13, табл. 22).

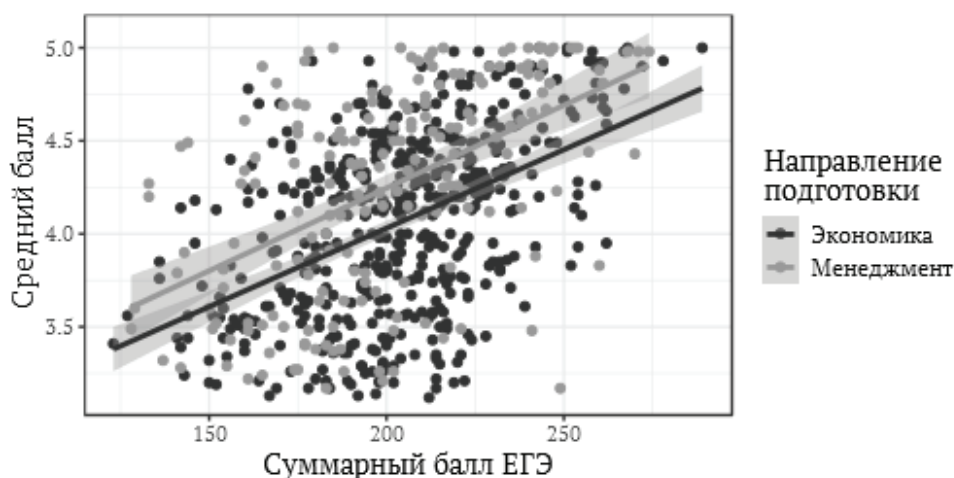


Рис. 11. Зависимость успеваемости обучения в университете от суммарного балла ЕГЭ

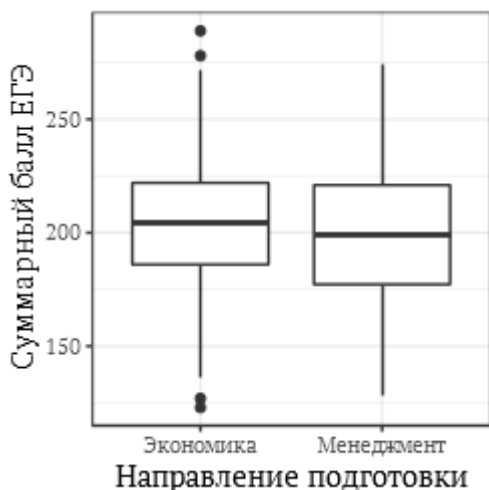


Рис. 12-1. Диаграмма размаха суммарного балла ЕГЭ по группам

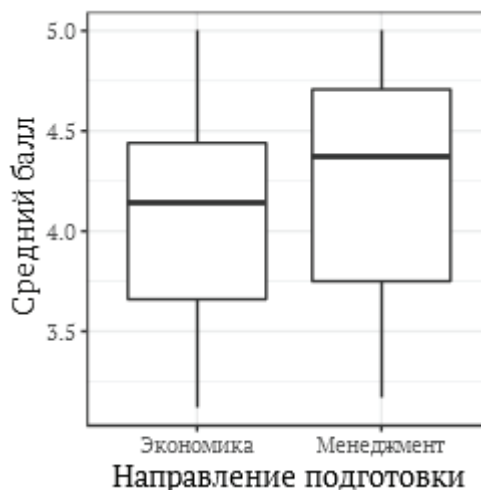


Рис. 12-2. Диаграмма размаха среднего балла по группам

Т а б л и ц а 21

Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	df	P-value
Суммарный балл ЕГЭ	Направление подготовки	2,7552	1	0,0969
Средний балл	Направление подготовки	17,639	1	2,67e-05

Непараметрический однофакторный дисперсионный анализ

Зависимая переменная	Фактор	Test statistic	df	P-value
Суммарный балл ЕГЭ	Направление подготовки, пол	10,766	3	0,0136

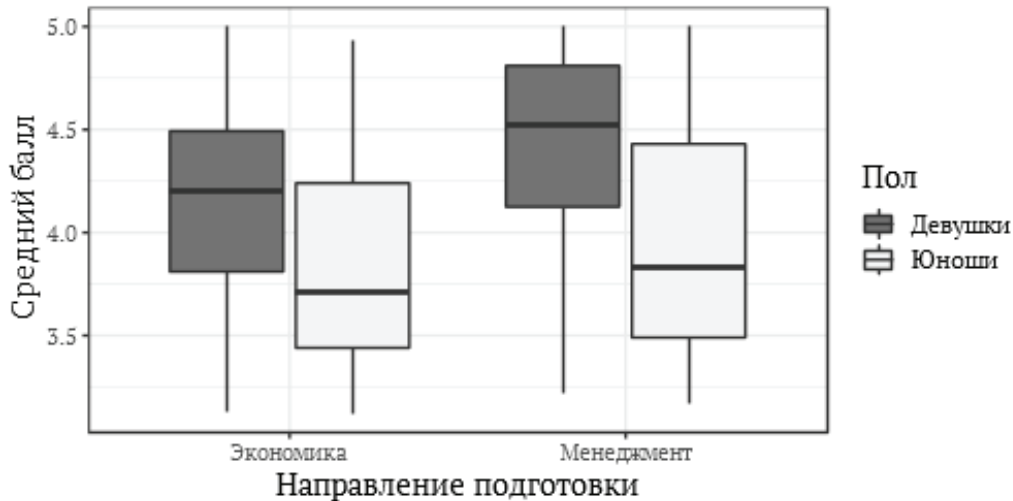


Рис. 13. Диаграмма размаха среднего балла по группам

Заключение

В работе представлены результаты статистического анализа учебной деятельности бакалавров экономического факультета ФГБОУ ВО «Воронежский государственный университет». Информационную базу исследования сформировали данные об учебной деятельности, успеваемости, вступительных баллах по ЕГЭ, а также обобщенные адресные данные. Мы рассматривали два набора данных, соответствующих обучающимся по направлениям «Экономика», «Менеджмент».

В качестве основных факторов успеваемости обучающегося мы рассмотрели его домашний регион, пол, источник финансирования учебной деятельности, а также результаты вступительных испытаний.

Стресс, связанный со сменой места жительства в период учебы, почти не оказывает влияния как на когнитивные способности обучающихся, так и на показатели успеваемости. Различия в успеваемости обучающихся, в отличие от домашнего региона, выявлены только на направлении «Экономика». Более высокую успеваемость продемонстрировали обучающиеся из Воронежа, Воронежской и Липецкой областей. При проведении довузовской работы на эти регионы следует обращать особое внимание.

Обучающиеся девушки демонстрируют более высокие результаты на обоих направлени-

ях подготовки. Для девушек, обучающихся за счет федерального бюджета, характерно наличие познавательной энергии, силы воли и большая мотивация на высокую успеваемость. Обучающимся за счет собственных средств, в частности юношам, свойственно «клиповое» мышление, может не хватать навыков и технологий самостоятельной работы, хотя не стоит списывать со счетов факты недисциплинированности и неисполнительности. (Наличие недисциплинированности и неисполнительности мы полагаем, исходя из эвристических соображений, хотя в действительности их можно было бы установить по показателям посещаемости обучающихся или по показателям активности в случае дистанционных форм.)

Обнаруженные нами закономерности могут указывать на неявный социальный заказ обучающихся на *получение именно фундаментального университетского образования*.

Нам не удалось выявить статистически значимые линейные связи между вступительными баллами и успеваемостью обучающихся, но особое внимание следует уделить тому факту, что обучающиеся, завершающие учебную деятельность по причине неуспеваемости, поступили преимущественно с низкими суммарными баллами ЕГЭ. Низкая результативность обучающихся на вступительных испытаниях не должна становиться помехой в учеб-

ном процессе. Перед университетским образованием стоит задача подготовки профессионалов высокого уровня независимо от их начальных запаса знаний и когнитивных способностей. В современных условиях перманентных изменений в сфере образования и трансформации требований к специалистам одним из важных ориентиров при достижении социальной миссии университетов является целенаправленное сглаживание неоднородности в формировании компетенций обучающихся, а именно формировании не только «широких», но и в первую очередь «глубоких» компетенций, т. е. системы компетенций.

В соответствии с концепцией классического университетского образования, суть фундаментального, следует:

Библиографический список

1. *Levene H. Robust Tests for Equality of Variances / H. Levene // Contributions to Probability and Statistics. – Palo Alto, Calif. : Stanford University Press, 1960. – P. 278–292.*

2. *Bartlett M. S. Properties of Sufficiency of Statistical Tests / M. S. Bartlett // Proc. Roy. Soc. – 1937. – A 160. – P. 268–287.*

3. *Shapiro S. S. An Analysis of Variance Test for Normality / S. S. Shapiro, M. B. Wilk // Biometrika. – 1965. – Vol. 52, no. 3. – P. 591–611.*

4. *Filliben J. J. The Probability Plot Correlation Coefficient Test for Normality / J. J. Filliben // Technometrics. – 1975. – Vol. 17, no. 1. – P. 111–117.*

5. *Kruskal W. H. Use of Ranks in One-criterion Variance Analysis / W. H. Kruskal, W. A. Wallis // Journal of the American Statistical Association. – 1952. – Vol. 47, no. 260. – P. 583–621.*

6. *Антоненков Е. Г. Интеллектуальный анализ персональных предпочтений студентов в ходе подготовки кадров в высшей школе / Е. Г. Антоненков, Ю. А. Крюков // Современная наука : актуальные проблемы теории и практики. Серия : Естественные и технические науки. – 2019. – № 11-2. – С. 30–34.*

7. *Быстрова Т. Ю. Учебная аналитика MOOK как инструмент прогнозирования успешности обучающихся / Т. Ю. Быстрова, В. А. Ларионова, Е. В. Синицын, А. В. Толмачев // Вопросы образования. – 2018. – № 4. – С. 139–166.*

8. *Гранков М. В. Регрессионная модель успеваемости учебных групп вуза / М. В. Гранков, В. М. Аль-Габри // Инженерный вестник Дона. – 2017. – № 1 (44). – С. 46.*

9. *Ендовицкий Д. А. Московский международный рейтинг «Три миссии университета» в анализе состояния и перспектив развития вузов /*

– значительно усилить фундаментальную подготовку, а именно аудиторную нагрузку по дисциплинам базовой части учебных планов;

– совершенствовать методические аспекты образовательного процесса, использование которых позволит преподнести материал в наиболее доступной для обучающихся форме;

– развивать конвергентные образовательные направления;

– укреплять междисциплинарные связи и практики изучения логики познания.

Конфликт интересов

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Д. А. Ендовицкий // Вестник Воронежского государственного университета. Серия : Проблемы высшего образования. – 2019. – № 1. – С. 5–11.

10. *Ендовицкий Д. А. Конкурентоспособность российских университетов в глобальной системе высшего образования : количественный анализ / Д. А. Ендовицкий, В. В. Коротких, М. В. Воронова // Высшее образование в России. – 2020. – Т. 29, № 2. – С. 9–26.*

11. *Кондратьева Е. В. Анализ эффективности прогнозирования успеваемости на основе fuzzy logic / Е. В. Кондратьева, О. Ю. Кондратьева // Математическое и компьютерное моделирование в экономике, страховании и управлении рисками. – 2018. – № 3. – С. 79–82.*

12. *Котова Е. Е. Автоматизация прогнозирования результатов обучения студентов / Е. Е. Котова, А. С. Писарев // Известия СПбГЭТУ ЛЭТИ. – 2019. – № 5. – С. 31–39.*

13. *Кузнецов В. В. Взаимосвязь самооценки состояния здоровья и уровня заболеваемости с академической успеваемостью у студентов старших курсов медицинских специальностей с учетом влияния социально-экономических и демографических характеристик / В. В. Кузнецов, Р. А. Байрамов, Е. А. Смирнов, Е. К. Косилова, К. В. Косилов // Медицинский альманах. – 2019. – № 5-6 (61). – С. 10–15.*

14. *Мартынов Г. П. Статистический анализ зависимостей успеваемости обучающихся от сопутствующих факторов / Г. П. Мартынов // Актуальные вопросы образования. – 2019. – Т. 3. – С. 62–66.*

15. *Никулина И. В. Исследование уровня сформированности мотивации учебной деятельности обучающихся / И. В. Никулина, А. Э. Снежкова // Вестник Самарского университета. Серия : Исто-*

рия, педагогика, филология. – 2019. – Т. 25, № 3. – С. 89–94.

16. *Садовничий В. А.* Евразийская миссия университетов / В. А. Садовничий // Три миссии университета : образование, наука, общество / ред. кол.: В. А. Садовничий [и др.]. – Москва : Макс Пресс, 2019. – С. 7–19.

17. *Степанова И. П.* Прогнозирование успешности обучения по химическим дисциплинам по результатам ЕГЭ в медицинском вузе / И. П. Степанова, И. В. Ганзина, О. В. Атавина, Т. В. Постнова, В. В. Мугак // Международный журнал экспериментального образования. – 2018. – № 11. – С. 17–22.

18. *Тарасова Е. В.* Анализ влияния факторов окружающей среды на годовую динамику успеваемости обучающихся северного вуза / Е. В. Тарасова // Бюллетень Северного государственного медицинского университета. – 2018. – № 1 (40). – С. 13–14.

19. *Темняткина О. В.* Модели оценки эффективности работы педагогов, используемые в зарубежных странах / О. В. Темняткина, Д. В. Токменинова // Перспективы науки и образования. – 2019. – № 3 (39). – С. 489–499.

20. *Шмарихина Е. С.* Исследование факторов успеваемости обучающихся / Е. С. Шмарихина // Вестник НГУЭУ. – 2018. – № 3. – С. 130–143.

Канапухин Павел Анатольевич, доктор экономических наук, доцент, заведующий кафедрой экономики, маркетинга и коммерции, Воронежский государственный университет, Воронеж, Российская Федерация

E-mail: kanapukhin@econ.vsu.ru

ORCID ID: 0000-0002-2236-4871

Коротких Вячеслав Владимирович, кандидат экономических наук, доцент кафедры информационных технологий и математических методов в экономике, Воронеж-

ский государственный университет, Воронеж, Российская Федерация

E-mail: korotkikh@econ.vsu.ru

ORCID ID: 0000-0001-9029-7466

Щекунских Светлана Станиславовна, кандидат физико-математических наук, доцент кафедры информационных технологий и математических методов в экономике, Воронежский государственный университет, Воронеж, Российская Федерация

E-mail: shchekunskikh@econ.vsu.ru

ORCID ID: 0000-0003-0830-4554

Поступила в редакцию 15.03.2020

Подписана в печать 12.04.2020

Statistical analysis of academic performance (by an example from the faculty of economics of Voronezh State University)

P. A. Kanapukhin¹, V. V. Korotkikh^{2✉}, S. S. Shchekunskikh³

^{1, 2, 3} Voronezh State University, 1, University sq., 394018, Voronezh, Russian Federation

Cite as: Kanapukhin, P. A., Korotkikh, V. V., Shchekunskikh, S. S. (2020). Statistical analysis of academic performance (by an example from the faculty of economics of Voronezh State University). *Proceedings of Voronezh State University. Series: Economics and Management*. 2, 27-44. (In Russ., abstract in Eng.) DOI: 10.17308/econ.2020.2/2899

Importance. We live in a world in which educational processes are one of the constituent components of the modern society life. It is a well-known fact that higher education processes play an important role in achieving both individual and collective effectiveness. Therefore, they are subject to careful state and public monitoring.

Objectives. Identification of statistically significant determinants of students' academic performance and in order to improve the quality of the educational process.

Methods. During the research, we used parametric and nonparametric methods of data analysis and machine learning. We formed the research information base using data on educational activities, academic performance, points on the Unified State Exam, as well as generalized address data of students in the bachelor's degree in the areas of "Economics", "Management" 2013-2015 recruitment.

Results. In the empirical part of the study, we tested several working hypotheses about the relationship of student academic performance with home region, gender, source of funding for educational activities, and the results of entrance tests. We also tested the hypothesis about the relationship between the completion of the educational activities of students due to academic failure and their low entrance score. We found that school-related stress has little to no impact on academic performance. Female students show better results in both areas of training. Girls who study at the expense of the federal budget demonstrate greater motivation for high academic performance compared to students at their own expense. There were no statistically significant linear relationships between the entrance scores and student performance. It would be unfair not to mention that fact that students completing their educational activities due to academic failure entered mainly with low total USE scores. Therefore it makes sense to significantly strengthen fundamental training, namely the classroom load in the disciplines of the basic part of the curriculum, develop convergent educational areas, and also strengthen interdisciplinary ties and the practice of studying the logic of cognition in order to smooth out the heterogeneity in the formation of competencies among students.

Key words: higher education, economic analysis, academic performance.

Conflict of Interest

The authors declare that there are no obvious or potential conflicts of interest associated with the publication of this article.

References

1. Levene, H. (1960) Robust tests for equality of variances. In Olkin, I. (ed.) *Contributions to Probability and Statistics*. Palo Alto, California, Stanford University Press, pp. 278-292.
2. Bartlett, M. S. (1937) Properties of Sufficiency of Statistical Tests. *Proc. Roy. Soc. A* 160, 268-287.
3. Shapiro, S. S. & Wilk, M. B. (1965) An Analysis of Variance Test for Normality. *Biometrika*. 52(3), 591-611.
4. Filliben, J. J. (1975) The Probability Plot Correlation Coefficient Test for Normality. *Technometrics*. 17(1), 111-117.
5. Kruskal, W. H. & Wallis, W. A. (1952) Use of Ranks in One-criterion Variance Analysis. *Journal of the American Statistical Association*. 47(260), 583-621.
6. Antonenkov, E. & Kryukov, Y. (2019) Intellectual analysis of personal preferences of students in the course of training in higher education. *Modern Science: Actual problems of theory and practice. Series: Natural and Technical Sciences*. 11(2), 30-34. (In Russ.)
7. Bystrova, T., Larionova, V., Sinitsyn, E. & Tolmachev, A. (2018) Learning Analytics in Massive Open Online Courses as a Tool for Predicting Learner Performance. *Educational Studies Moscow*. 4, 139-166. (In Russ.)
8. Grankov, M. & Al-Gabri, W. (2017) Regression model of the performance of students groups in higher education institution. *Engineering journal of Don*. 1(44), 46. (In Russ.)
9. Endovitsky, D. (2019). [The Moscow International Ranking "Three University Missions" as an Item for the Analysis of the Current State and Development Prospects of Universities]. *Proceedings of Voronezh State University. Series: Problems of Higher Education*. 1, 5-11. (In Russ.)
10. Endovitsky, D., Korotkikh, V., & Voronova, M. (2020). Competitiveness of Russian Universities in the Global System of Higher Education: Quantitative Analysis. *Vysshee obrazovanie v Rossii [Higher Education in Russia]*. 29(2), 9-26. (In Russ., abstract in Eng.)
11. Kondrateva, E. & Kondrateva O. (2018) Analysis of efficiency of forecasting prosperity on the basis of fuzzy logic. In: Dudov, S. (ed.) *Mathematical and computer modeling in economics, insurance and risk management : Proceedings of the International youth scientific and practical conference*, 14-17 November 2018, Saratov. pp. 79-82. (In Russ.)
12. Kotova, E. & Pisarev, A. (2019) Automated prediction of student learning outcomes. *Proceedings of Saint Petersburg Electrotechnical University*. 5, 31-39. (In Russ.)
13. Kuznetsov, V., Bayramov, R., Smirnov, E., Kosilova, E. & Kosilov K. (2019) The relationship of self-assessment of health status and morbidity with academic performance in senior students of medical specialties, taking into account the impact of socio-economic and demographic characteristics. *Medical Almanac*. 5-6 (61), 10-15. (In Russ.)
14. Martynov G. (2019) Statistical dependency analysis performance students from related factors. *Aktualnye voprosy obrazovaniya*. 3. 62-66. (In Russ.)
15. Nikulina, I. & Snezhkova, A. (2019) Study of the level of formation of motivation for learning activities of students. *Vestnik of Samara University. History, pedagogics, philology*. 25(3), 89-94. (In Russ.)
16. Sadovnichy, V. (2019). [Eurasian University Mission]. V. A. Sadovnichy et al. (Eds.). *Three University Missions : Education, Science, Society*. Moscow : Maks Press Publ., pp. 7-19. (In Russ.)
17. Stepanova, I., Ganzina, I., Atavina, O., Postnova, T. & Mugak V. (2018) Prognosis of training success in chemical disciplines according to results of the Unified state exam (USE) in medical university. *International journal of experimental education*. 11, 17-22. (In Russ.)
18. Tarasova, E. (2018) [Analysis of the influence of environmental factors on the annual dynamics of student performance at a northern university]. *Bulletin of the Northern State Medical University*. 1(40), 13-14. (In Russ.)
19. Temnyatkina, O. & Tokmeninova, D. (2019) Models of teacher performance evaluation applied in various countries. *Perspectives of science and education*. 3(39), 489-499. (In Russ.)
20. Shmarikhina, E. (2018) Investigation the factors of students performance. *Vestnik NSUEM*. 3, 130-143. (In Russ.)

Kanapukhin Pavel A., Dr. Sci. (Econ.), Assoc. Prof., Head of Economics, Marketing and Commerce Department, Voronezh State University, Voronezh, Russian Federation

E-mail: kanapukhin@econ.vsu.ru

ORCID ID: 0000-0002-2236-4871

Korotkikh Viacheslav V., Cand. Sci. (Econ.), Assoc. Prof., IT and Mathematical Methods in Economics Department, Voronezh State University, Voronezh, Russian Federation

E-mail: korotkikh@econ.vsu.ru

ORCID ID: 0000-0001-9029-7466

Shchekunskikh Svetlana S., Cand. Sci. (Phys.-Math.), Assoc. Prof., IT and Mathematical Methods in Economics Department, Voronezh State University, Voronezh, Russian Federation

E-mail: shchekunskikh@econ.vsu.ru

ORCID ID: 0000-0003-0830-4554

Received 15.03.2020

Accepted 12.04.2020