

# ИСПОЛЬЗОВАНИЕ КАТЕГОРИЗОВАННЫХ ПЕРЕМЕННЫХ ПРИ ВЫЯВЛЕНИИ ЗАВИСИМОСТИ МЕЖДУ ХАРАКТЕРИСТИКАМИ АБИТУРИЕНТА И УСПЕШНОСТЬЮ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ БАКАЛАВРА

## THE USE OF CATEGORIZED VARIABLES IN IDENTIFYING THE DEPENDENCE BETWEEN THE CHARACTERISTICS OF THE ENTRANT AND THE SUCCESS OF THE BACHELOR'S EDUCATIONAL PROGRAM

**T. Novgorodtseva**  
**A. Burdukovskaya**  
**E. Ivanova**  
**N. Kuzmina**  
**I. Sagalueva**

*Summary:* This article is devoted to the issue of applying the method of analysis of categorized variables to identify the relationship between the characteristics of the applicant and the success of mastering the educational program of the bachelor. Objectives of the study: on the basis of real data on applicants who entered the university, to highlight those characteristics of the applicant that later influenced the result of the successful mastering of the bachelor's educational program. Students of the Faculty of Geography of Irkutsk State University were selected as a base for the study. In the course of the study, the document analysis method and the mathematical method for the analysis of categorized variables were used. Using the method of analysis of categorized variables, the characteristics of applicants that affect the success of their studies at a university were identified, their statistical significance was proved based on the coefficients of quadratic contingency and Cramer. The results obtained made it possible to identify the most significant characteristics of the applicant in terms of the success of higher education. The scientific and practical significance of the work lies in the fact that one of the possible approaches is proposed to identify the main factors - the characteristics of the applicant, which determine the success of obtaining higher education. A similar approach can be used in any other educational institution.

*Keywords:* Coefficient of quadratic contingency, Cramer's coefficient, method of categorized variables,  $\chi^2$ -distribution, characteristics of the applicant.

**Новгородцева Татьяна Юрьевна**

к.т.н., доцент, ФГБОУ ВО «Иркутский государственный университет»  
 nfyz-31@mail.ru

**Бурдуковская Анна Валерьевна**

к.ф.-м.н., доцент, ФГБОУ ВО «Иркутский государственный университет»  
 buran\_baikal@mail.ru

**Иванова Елена Николаевна**

к.п.н, ФГБОУ ВО «Иркутский государственный университет»  
 iitmoi@mail.ru

**Кузьмина Надежда Дмитриевна**

к.ф.-м.н., доцент, ФГБОУ ВО «Иркутский государственный университет»  
 nd-kuzmina@mail.ru

**Сагалуева Инга Николаевна**

ФГБОУ ВО «Иркутский государственный университет»  
 sgh@mail.ru

*Аннотация:* Настоящая статья посвящена вопросу применения метода анализа категоризованных переменных для выявления зависимости между характеристиками абитуриента и успешности освоения образовательной программы бакалавра.

Цели исследования: на основе реальных данных об абитуриентах, поступивших в ВУЗ, выделить те характеристики абитуриента, которые повлияли в дальнейшем на результат успешного освоения образовательной программы бакалавра. В качестве базы для проведения исследования выбраны студенты Географического факультета Иркутского государственного университета. В процессе проведения исследования использовались метод анализа документов и математический метод анализа категоризованных переменных. С помощью метода анализа категоризованных переменных выделены характеристика абитуриентов, влияющие на успешность обучения в ВУЗе, доказана их статистическая значимость на основе коэффициентов квадратической сопряженности и Крамера. Полученные результаты позволили выявить наиболее значимые характеристики абитуриента с точки зрения успешности получения высшего образования.

Научно-практическая значимость работы заключается в том, что предложен один из возможных подходов к выявлению основных факторов – характеристик абитуриента, обуславливающих успешность получения высшего образования. Подобный подход может быть использован в любых других учебных заведениях.

*Ключевые слова:* коэффициент квадратической сопряженности, коэффициент Крамера, метод категоризованных переменных,  $\chi^2$ -распределение, характеристики абитуриента.

Проблема успешного освоения образовательной программы высшей школы рассматривается с различных сторон – это и анализ абитуриентов при выборе ВУЗа, и вопросы адаптации студентов-первокурсников и социализации студенчества в современной высшей школе, и анализ факторов, влияющих на выбор профессии и др. [2-6].

Состав методов, используемых в данном исследовании, определен целью исследования.

На этапе формирования исходных данных использован метод анализа документов. Цель применения метода – получить информацию о характеристиках абитуриента, имеющихся в официальных документах.

Метод категоризованных переменных применен для отбора факторов, статистически значимо влияющих на успешность получения высшего образования и получения диплома бакалавра по выбранному направлению при поступлении. Суть метода заключается в статистическом анализе таблицы сопряженности двух категоризованных переменных [1].

В общем виде при статистическом анализе двух категоризованных переменных исходные данные представ-

ляют в виде таблицы сопряженности или таблицы перекрестных частот.

В общем виде двухходовая таблица сопряженности имеет следующую структуру:

	1	2	...	j	...	m <sub>2</sub>	n <sub>i</sub>
1	n <sub>11</sub>	n <sub>12</sub>	...	n <sub>1j</sub>	...	n <sub>1m<sub>2</sub></sub>	∑ <sub>j=1</sub> <sup>m<sub>2</sub></sup> n <sub>1j</sub>
2	n <sub>21</sub>	n <sub>22</sub>	...	n <sub>2j</sub>	...	n <sub>2m<sub>2</sub></sub>	∑ <sub>j=1</sub> <sup>m<sub>2</sub></sup> n <sub>2j</sub>
...	...	...	...	...	...	...	...
i	n <sub>i1</sub>	n <sub>i2</sub>	...	n <sub>ij</sub>	...	n <sub>im<sub>2</sub></sub>	∑ <sub>j=1</sub> <sup>m<sub>2</sub></sup> n <sub>ij</sub>
...	...	...	...	...	...	...	...
m <sub>1</sub>	n <sub>m<sub>1</sub>1</sub>	n <sub>m<sub>1</sub>2</sub>	...	n <sub>m<sub>1</sub>j</sub>	...	n <sub>m<sub>1</sub>m<sub>2</sub></sub>	∑ <sub>j=1</sub> <sup>m<sub>2</sub></sup> n <sub>m<sub>1</sub>j</sub>
n <sub>j</sub>	∑ <sub>i=1</sub> <sup>m<sub>1</sub></sup> n <sub>i1</sub>	∑ <sub>i=1</sub> <sup>m<sub>1</sub></sup> n <sub>i2</sub>	...	∑ <sub>i=1</sub> <sup>m<sub>1</sub></sup> n <sub>ij</sub>	...	∑ <sub>i=1</sub> <sup>m<sub>1</sub></sup> n <sub>im<sub>2</sub></sub>	n

где n – общее число исследуемых объектов; x<sub>1</sub> и x<sub>2</sub> – измеряемые переменные (характеристики); m<sub>1</sub> – количество градаций по первой переменной;

m<sub>2</sub> – количество градаций по второй переменной; n<sub>ij</sub> – число объектов, у которых значение характеристики x<sub>1</sub> зафиксировано на i-ой градации, i = 1, m<sub>1</sub>, значение характеристики x<sub>2</sub> на уровне j-ой градации, j = 1, m<sub>2</sub>.

В качестве показателей тесноты парной статистической связи между переменными использованы коэффициент квадратической сопряженности и коэффициент Крамера. Статистическая значимость коэффициента ква-

Таблица 2.

Характеристики (параметры абитуриента)

Измеряемая характеристика	Условное обозначение (x <sub>j</sub> )	Число градаций (m <sub>j</sub> )	Комментарий
Пол	x <sub>1</sub>	m <sub>1</sub> = 2	1 – женский; 0 – мужской.
Возраст на момент зачисления	x <sub>2</sub>	m <sub>2</sub> = 3	1 – до 20 лет; 2 – от 20 до 30 лет; 3 – старше 30 лет.
Место постоянного проживания	x <sub>3</sub>	m <sub>3</sub> = 3	1 – Иркутск; 2 – Иркутская область; 3 – другие регионы.
Вид образовательного учреждения до поступления в ВУЗ	x <sub>4</sub>	m <sub>4</sub> = 6	1 – школа; 2 – лицей; 3 – гимназия; 4 – училище; 5 – техникум; 6 – ВУЗ.
Сумма баллов по ЕГЭ	x <sub>5</sub>	m <sub>5</sub> = 3	1 – до 100 баллов; 2 – от 100 до 200 баллов; 3 – выше 200 баллов.
Средний балл за период обучения	x <sub>6</sub>	m <sub>6</sub> = 4	1 – меньше 3 баллов; 2 – от 3 до 4 баллов; 3 – от 4 до 5 баллов; 4 – 5 баллов.
Результат обучения	x <sub>7</sub>	m <sub>7</sub> = 2	0 – не закончил (отчислен по различным причинам); 1 – закончил обучение и получил диплом бакалавра по выбранному направлению при поступлении.

Таблица сопряженности для переменных  $x_7$  и  $x_1$

	<i>1</i>	<i>2</i>	<i>сумма</i>
<i>0</i>	124	62	186
<i>1</i>	82	115	197
<i>сумма</i>	206	177	383

Таблица сопряженности для переменных  $x_7$  и  $x_2$

	<i>1</i>	<i>2</i>	<i>3</i>	<i>сумма</i>
<i>0</i>	150	49	7	206
<i>1</i>	145	24	8	177
<i>сумма</i>	295	73	15	383

Таблица сопряженности для переменных  $x_7$  и  $x_3$

	<i>1</i>	<i>2</i>	<i>3</i>	<i>сумма</i>
<i>0</i>	60	99	47	206
<i>1</i>	45	100	32	177
<i>сумма</i>	105	199	79	383

Таблица сопряженности для переменных  $x_7$  и  $x_4$

	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>сумма</i>
<i>0</i>	147	11	15	5	2	26	206
<i>1</i>	131	9	13	4	2	18	177
<i>сумма</i>	278	20	28	9	4	44	383

Таблица сопряженности для переменных  $x_7$  и  $x_5$

	<i>1</i>	<i>2</i>	<i>3</i>	<i>сумма</i>
<i>0</i>	5	178	23	206
<i>1</i>	1	144	32	177
<i>сумма</i>	6	322	55	383

Таблица сопряженности для переменных  $x_7$  и  $x_6$

	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>сумма</i>
<i>0</i>	95	113	13	0	221
<i>1</i>	0	60	100	2	162
<i>сумма</i>	95	173	113	2	383

Оценка тесноты связи между категоризованными переменными

Коэффициенты	Результаты анализа таблиц сопряженности					
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>
Коэффициент квадратической сопряженности	24,14	6,55	2,82	0,64	5,56	175,29
$\chi^2$ - квадрат (табличное)	3,84	5,99	5,99	11,07	5,99	7,81
Коэффициент Крамера	0,25	0,13	0,09	0,04	0,12	0,68

Рис.1. Результаты проведения анализа категоризованных переменных

дратической сопряженности проверяется с помощью  $\chi^2$ -распределения; коэффициент Крамера принимает значения от 0 до 1. Нулевое значение свидетельствует о строгой статистической независимости анализируемых факторов, а значение, близкое к единице – о зависимости между характеристиками.

Анализ документов абитуриента позволил сформировать набор характеристик, представленных в таблице 2: пол, возраст на момент зачисления, место постоянного проживания, вид образовательного учреждения до поступления в ВУЗ, сумма баллов по ЕГЭ. Причем, все рассматриваемые характеристики носят качественный характер. В качестве показателя, описывающего процесс освоения образовательной программы бакалавриата, использован средний балл за период обучения. Эндogenous переменной (зависимой, объясняемой) выбран результат обучения (не закончил обучение – отчислен по различным причинам, закончил обучение).

Общая численность абитуриентов, включенных в исследование, составила 383 человека. Требовалось выяснить:

- существует ли связь между результатом обучения и характеристиками абитуриента;
- если «да», то какова степень тесноты этой связи.

В рамках исследования построены и проанализированы 6 таблиц сопряженности, представленных на рис. 1, в каждой из которых одной из категоризованной переменной является переменная «Результат обучения» (переменная  $x_7$ ).

Анализ коэффициентов квадратической сопряженности и расчетных значений коэффициентов Крамера позволил сделать выводы.

Статистически значимыми характеристиками (при уровне значимости  $\alpha = 0,05$ ) оказались следующие: «Пол», «Возраст на момент зачисления», «Средний балл

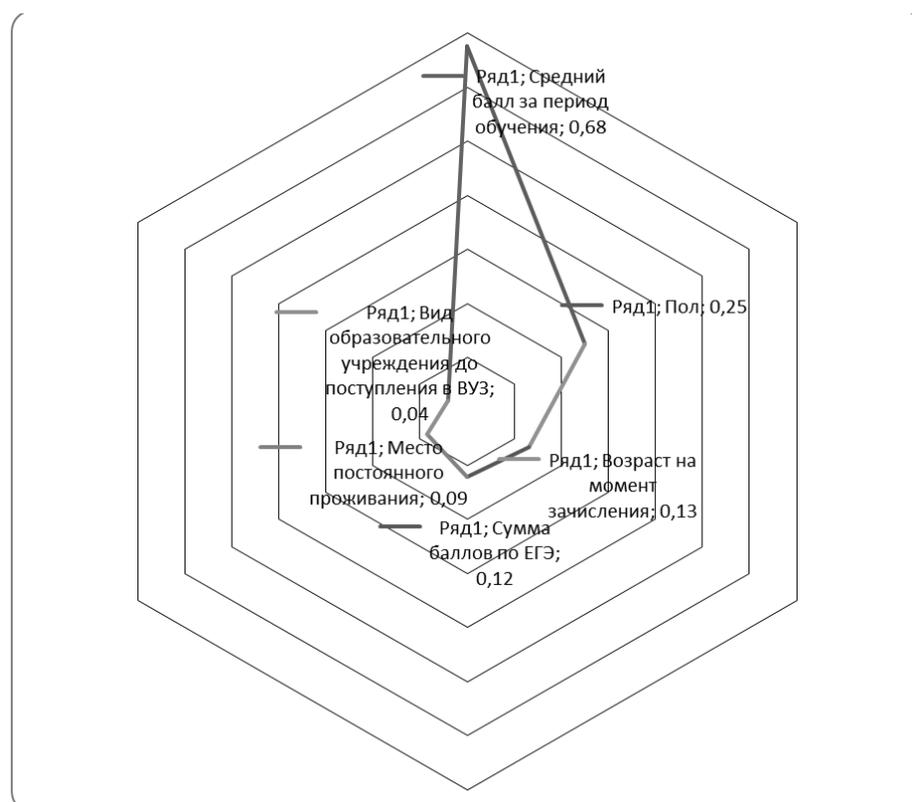


Рис. 2. Анализ характеристик абитуриента

за период обучения». Связь между «Результатом обучения» и вышеперечисленными характеристиками абитуриента есть, характеризуется достаточно высокой степенью тесноты.

На рисунке 2 представлены результаты расчета коэффициента Крамера. Из шести анализируемых переменных наличие достаточно тесной связи с переменной «Результат обучения» выявлено у переменных «Пол», «Возраст на момент зачисления» и «Средний балл за пе-

риод обучения».

Ранжировка (по убыванию) анализируемых характеристик с точки зрения их влияния на переменную «Результат обучения» выглядит следующим образом: «Средний балл за период обучения», «Пол», «Возраст на момент зачисления», «Сумма баллов по ЕГЭ», «Место постоянного проживания», «Вид образовательного учреждения до поступления в ВУЗ».

#### ЛИТЕРАТУРА

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. 2-е изд., испр. — Т. 1: Теория вероятностей и прикладная статистика. — М: ЮНИТИ-ДАНА, 2001. — 656 с.
2. Бережной Л.Н. К вопросу повышения качества учебного процесса в системе «Школа – ВУЗ» // Современное образование: содержание, технологии, качество. Материалы XXV международной научно-методической конференции. СПб.: Изд-во СПбГЭТИ «ЛЭТИ», 2019. С. 612-615.
3. Збанацкая А.Б. Личностные характеристики студентов вуза как основа индивидуального стиля учебной деятельности: дис. канд. псих. наук: 19.00.07 / А.Б. Збанацкая. Московский госуд. Университет. М., 2017. – 208с.
4. Рябоконт М.В. Модели поведения абитуриентов при выборе ВУЗа // Вестник Нижегородского университета им. Н.И. Лобачевского. 2016. №4. С.177-185.
5. Фомина Т.П., Карлова М.Ю. Исследование факторов, влияющих на выбор профессии и успешность обучения студентов по направлению подготовки «Педагогическое образование», средствами статистического анализа. // Педагогика. Вопросы теории и практики. Тамбов: Грамота. 2018. №2. С 59-63.
6. Хавенсон Т.Е., Соловьева А.А. Связь результатов единого государственного экзамена и успеваемости в ВУЗе // Вопросы образования. 2014. № 1. С. 176-199.

© Новгородцева Татьяна Юрьевна (nfyz-31@mail.ru), Бурдуковская Анна Валерьевна (buran\_baikal@mail.ru), Иванова Елена Николаевна (iimoi@mail.ru), Кузьмина Надежда Дмитриевна (nd-kuzmina@mail.ru), Сагалуева Инга Николаевна (sgh@mail.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»