УДК 303.4-057.875 DOI: 10.12737/16791

ИССЛЕДОВАНИЕ ФАКТОРОВ АДАПТАЦИИ СТУДЕНЧЕСТВА 2000-х ГОДОВ К ОБУЧЕНИЮ В ВУЗЕ МЕТОДАМИ DATA MINING

Шуметов В.Г., Лясковская О.В.[1]

Статья содержит результаты исследования факторов адаптации студентов к условиям вузовского обучения с помощью одной из эффективных технологий Data Mining – метода дисперсионного анализа. Представлены результаты моделирования факторов адаптации к учебе студентов периферийных вузов новым методом Data Mining, реализуемого с помощью процедуры обобщенной линейной модели пакета анализа данных общественных наук SPSS Base. Эмпирической базой моделирования являлись результаты социологического опроса студентов вузов городов Орел, Брянск, Курск и Белгород. Показано, что в 2000-х гг. основным фактором, определяющим адаптацию студентов к обучению в вузе, являлся курс обучения и гендерный фактор, факультет обучения оказался менее значимым фактором.

Ключевые слова: факторы адаптации, высшее учебное заведение, студенчество, социологический опрос, технологии Data Mining, дисперсионный анализ, общая линейная модель.

STUDY OF ADAPTATION FACTORS OF STUDENTS OF 2000S FOR TRAINING IN HIGHER EDUCATIONAL INSTITUTION BY DATA MINING METHODS

SHUMETOV V.G. — Doctor of Economic Sciences, Professor, Russian Presidential Academy of National Economy and Public Administration (Russian Federation, Orel), e-mail: shumetov@list.ru.
LYASKOVSKAYA O.V. — Candidate of Sociological Sciences, Docent, Associate Professor at the Department of Applied Informatics and Mathematics, Russian Presidential Academy of National Economy and Public Administration (Russian Federation, Orel), e-mail: olga0873@list.ru.

The article contains the results of studying the factors of students' adaptation to the conditions of university training with one of the most effective technologies Data Mining - the method of analysis of variance. The results of modeling factors of adaptation of students of peripheral universities to the new method of Data Mining, implemented by the procedure of generalized linear model analysis of the data packet of Social Sciences SPSS Base are presented. The empirical base of modeling is the result of the poll of university students in Orel, Bryansk, Kursk and Belgorod. It is shown that in 2000s the main factor determining the adaptation of students to training in university, was the course of training and the gender factor, the faculty of training was less significant factor.

Keywords: factors of adaptation, university, students, opinion poll, technology Data Mining. analysis of variance, general linear model.

В статье[1] показана эффективность такого метода математического моделирования результатов прикладных социологических исследований, как дисперсионный анализ. В частности, одна из новых

Шуметов Вадим Георгиевич — доктор экономических наук, профессор, Российская академия народного хозяйства и государственной службы при Президенте РФ (Российская Федерация, г. Орел), e-mail: shumetov@list.ru.

Лясковская Ольга Вадимовна — кандидат социологических наук, доцент, доцент кафедры прикладной информатики и математики, Российская академия народного хозяйства и государственной службы при Президенте РФ, Орловский филиал (Российская Федерация, г. Орел), e-mail: olga0873@list.ru.

процедур Data Mining – обобщенная (общая) линейная модель (*General Linear Model*) – оказалась полезной в исследовании факторов, определяющих эффективность процесса адаптации бывшего абитуриента к обучению в вузе[2]. Основной проблемой реализации процедуры общей линейной модели является ограничение, накладываемое недостаточным объемом выборки, – ее необходимо разбивать на несколько подвыборок, отвечающих сочетаниям градаций исследуемых факторов[3]. Чем больше факторов включается в анализ и чем больше их градаций, тем менее надежными становятся оценки параметров моделей, а следовательно, и их социологическая интерпретация. Это вынуждает ограничиваться небольшим – двумя-тремя факторами, каждый из которых варьируется на двух – четырех уровнях.

Следует отметить, что в условиях низкого уровня шкал измерения социологических данных (в шкалах наименований и, в лучшем случае, в ранговых шкалах) процедура общей линейной модели, реализуемая программными средствами пакета анализа данных общественных наук SPSS Base[4], позволяет выявить степень и характер влияния одной или нескольких независимых переменных, которые, как правило, принимают дискретные значения, на одну или несколько зависимых переменных, измеренных в шкале интервалов, отношений или абсолютной шкале. В отличие от традиционного «классического» метода дисперсионного анализа по Фишеру, когда совокупная дисперсия всех наблюдаемых значений раскладывается на дисперсию внутри отдельных групп и дисперсию между группами, в основе обобщенной линейной модели лежит корреляционный или регрессионный анализ[5].

Процедура общей линейной модели была использована нами при исследовании зависимости индекса удовлетворенности студентов своим выбором от их пола и степени адаптации к обучению в вузе. В этой работе из-за ограниченного объема анализируемой выборки (было опрошено 858 респондентов) мы ограничились ее разбиением на шесть групп, отвечающих сочетанию двух категорий фактора «пол» с тремя градациями фактора «степень адаптации», определяемыми вариантами ответов на вопрос анкеты «Считаете ли Вы вуз своим вторым домом?»: «да», «нет», «затрудняюсь ответить». Предполагалось, что респондентов, положительно ответивших на этот вопрос, можно считать адаптировавшимися к обучению в вузе, отрицательно – неадаптировавшимися, затруднившихся ответить – не определившимися. Количественной результативной переменной служил «индекс удовлетворенности выбором вуза», рассчитываемый по формуле:

$$И_{удовл} = w_{да} - w_{нет}, \quad (1)$$

где $w_{да}$ и $w_{нет}$ – частости положительных и отрицательных ответов на вопрос «В какой степени Вы удовлетворены сделанным выбором?».

Эмпирической базой исследования в цитируемой работе послужили результаты социологического опроса студентов московских и орловских вузов, выполненного нами в 2010 – 2012 гг. по методике, аналогичной использованной ранее (в 2000-х гг.) при опросе студентов вузов ряда городов Центральной России – в Орле, Брянске, Курске и Белгороде. Результаты этих более ранних исследований

[1]Шуметов В.Г., Лясковская О.В., Гудова Т.Г. Data Mining в эмпирической социологии: методология количественного моделирования: монография / Под общей ред. В.Г. Шуметова. – Орел: Изд-во ОРАГС, 2011.

[2] Редько Л.Л., Лобейко Ю.А. Психолого-педагогическая поддержка адаптации студента-первокурсника в вузе. – М.: Илекса, 2008.

[3]См. подробнее: Шуметов В.Г., Крюкова О.А. Методология и практика анализа данных в управлении: методы одномерного и двумерного анализа. Монография. – Орел: Изд-во ОФ РАНХиГС, 2013. – 178 с.

[4] SPSS Base 8.0 для Windows. Руководство по применению. – М.: СПСС Русь, 1998.

[5] Бююль А., Цёфель П. SPSS: Искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. – СПб.: ООО «ДиаСофтЮП», 2002. – С. 323.

опубликованы в монографии[1] и оказались полезными при разработке мероприятий по психологической работе со студентами-первокурсниками, направленной на содействие их адаптации к обучению в вузе. Представляет несомненный интерес сопоставить полученные нами спустя десятилетие результаты с результатами выполненных ранее социологических исследований. В соответствии с поставленной таким образом задачей на первом этапе выполнено моделирование связи результирующего показателя – индекса удовлетворенности выбором вуза – с двумя факторами – гендерный фактор (пол студентов) и факультет обучения (гуманитарный, технический)(таб. 1).

Таблица 1 — Индекс удовлетворенности студентами выбором вуза в зависимости от пола и факультета обучения (проценты)

Table 1 — The index of satisfaction of students with the choice of university depending on the gender and the faculty of education (percentage)

Пол	Факультет	Статистика	Удовлетворенность выбором вуза		Индекс удовлетворенности, %
			да	нет	
Мужской	Гуманитарный	частота	48	32	20,0
		частость	60,0%	40,0%	
	Технический	частота	139	108	12,6
		частость	56,3%	43,7%	
	Итого	частота	187	140	14,4
		частость	57,2%	42,8%	
Женский	Гуманитарный	частота	154	129	8,8
		частость	54,4%	45,6%	
	Технический	частота	54	64	-8,4
		частость	45,8%	54,2%	
	Итого	частота	208	193	3,8
		частость	51,9%	48,1%	

Эмпирическая база моделирования представлена в таблице 1, где указаны не только частости положительных и отрицательных ответов на ключевой вопрос «В какой степени Вы удовлетворены сделанным выбором?», но и их частоты, что позволяет оценить надежность исходных данных (статистика по нейтральным ответам респондентов – «затрудняюсь ответить» – в таблице не приводится). Из итоговых строк таблицы 1 следует, что студенты-юноши в целом заметно опережают студентов-девушек по уровню удовлетворенности выбором вуза, – индекс удовлетворенности у первых равен 14,4% против 3,8% у вторых. Гендерные различия проявляются и в зависимости уровня удовлетворенности выбором вуза от факультета обучения: студенты-«технари» в меньшей степени удовлетворены выбором вуза – 12,6% против 20,0% для юношей и -8,4% против 8,8% для девушек, т.е. разность в значениях индекса составляет 7,4% у первых и 17,2% у вторых.

Несмотря на приведенные цифры, это лишь предварительная оценка, поскольку при этом не учтены ошибки, обусловленные ограниченным объемом сравниваемых подвыборок опрошенных респондентов.

[1] Социально-психологическая и профессиональная адаптация студентов вузов Центрального региона России (по материалам социологического исследования) / В.И. Уварова, В.Г. Шуметов, Т.Н. Афонина, Т.А. Иваненко. // Под ред. канд. филос. наук В.И. Уваровой. – Орел: Изд-во ОрелГАУ, 2001.

Учет этих ошибок может быть проведен при статистическом подходе к анализу результатов опроса, и здесь полезным оказывается такой новый метод дисперсионного анализа, как процедура общей линейной модели. Задача при этом формулируется следующим образом: оценить параметры полной факторной линейной модели

$$Y_{ij} = \mu_0 + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij}, \quad (2)$$

где Y_{ij} – значение индекса удовлетворенности выбором вуза $И_{удовл}$ для *i*-го пола и *j*-го факультета; μ_0 – оценка свободного коэффициента модели; α_i – оценка эффекта *i*-го пола; β_j – оценка эффекта *j*-го факультета обучения; γ_{ij} – оценка взаимодействия *i*-го пола и *j*-го факультета; ε_{ij} – случайная ошибка.

Однако из-за отсутствия повторных наблюдений хотя бы для некоторых групп респондентов мы ограничились факторной линейной моделью без учета взаимодействия. Это позволяет случайную ошибку приравнивать к эффекту взаимодействия, но модель становится более простой:

$$Y_{ij} = \mu_0 + \alpha_i + \beta_j + \varepsilon_{ij}. \quad (3)$$

Однако из приведенных ниже результатов дисперсионного анализа, выполненного по схеме общей линейной модели, следует, что оба фактора статистически не значимы: хотя эмпирические значения критерия Фишера достаточно большие – 10,8 для гендерного фактора и 6,3 для фактора «факультет», их уровни значимости 0,188 и 0,241 превышают нормативную величину *p*=0,10, принимаемую обычно в анализе данных социологических исследований.

Таблица 2 — Дисперсионный анализ двухфакторной модели индекса удовлетворенности студентов выбором вуза (факторы «пол» и «факультет» без учета их взаимодействия)

Table 2 — Analysis of variance of two-factor model of the index of satisfaction of students with the choice of university (factors "gender" and "faculty" are without regard to their interaction)

Источник изменчивости	Сумма квадратов	Степени свободы	Средний квадрат	Критерий Фишера *F*	Уровень значимости
Пол	259,210	1	259,210	10,796	0,188
Факультет	151,290	1	151,290	6,301	0,241
Ошибка	24,010	1	24,010		

Несмотря на, казалось бы, отрицательные результаты выполненного моделирования (модель неадекватна), можно утверждать, что влияние гендерного фактора на уровень удовлетворенности студентов выбором вуза сильнее по сравнению с влиянием факультета обучения. Этот вывод имеет методическое значение: в дальнейшем фактор «факультет» можно исключить из анализа, а данные по подвыборкам для разных факультетов использовать как «повторные наблюдения». Эффективность этого вывода проявилась на следующем этапе исследований, при моделировании влияния на уровень удовлетворенности студентов выбором вуза гендерного фактора и курса обучения. Здесь мы уже формировали две эмпирические базы моделирования – для студентов, обучающихся на гуманитарных факультетах, с одной стороны, и технических – с другой. Эти базы приведены ниже в таблицах 3 и 4; в них, как и в предыдущем случае, помимо частостей положительных и отрицательных ответов на ключевой вопрос анкеты «В какой степени Вы удовлетворены сделанным выбором?» приведены их частоты.

Таблица 3 — Индекс удовлетворенности студентов-гуманитариев выбором вуза в зависимости от пола и курса обучения (проценты)

Table 3 — The index of satisfaction of students in humanities with the choice of university depending on the gender and the course of education (percentage)

Пол	Курс	Статистика	Удовлетворенность выбором вуза		Индекс удовлетворенности, %
			да	нет	
мужской	1	частота	12	13	-4,0
		частость	48,0%	52,0%	
	2	частота	17	7	41,6
		частость	70,8%	29,2%	
	3	частота	8	3	45,4
		частость	72,7%	27,3%	
	4	частота	11	9	10,0
		частость	55,0%	45,0%	
	Итого	частота	48	32	20,0
		частость	60,0%	40,0%	
женский	1	частота	67	39	26,4
		частость	63,2%	36,8%	
	2	частота	32	24	14,2
		частость	57,1%	42,9%	
	3	частота	37	26	17,4
		частость	58,7%	41,3%	
	4	частота	18	40	-38,0
		частость	31,0%	69,0%	
	Итого	частота	154	129	8,8
		частость	54,4%	45,6%	

Понятно, что при «дроблении» исходной выборки на столь большое число подвыборок (16 подвыборок для двух факультетов, двух полов и четырех курсов) прямое сравнение средних по подвыборкам не имеет смысла и единственным способом сделать обоснованные выводы является статистическое моделирование. Так, в таблице 3 фигурируют подвыборки, содержащие данные по 25-ти, 14-ти и даже всего по 11-ти респондентам (студенты-юноши 1-го, 2-го и 3-го курсов, обучающиеся на гуманитарных факультетах), тогда как расчетный объем выборки, обеспечивающий ошибку около 5% при доверительной вероятности 95%, составляет 400 респондентов.

Задача этого этапа моделирования формулируется следующим образом: оценить параметры полной факторной линейной модели (2), где Y_{ij} – наблюдаемое значение индекса удовлетворенности выбором вуза $И_{удовл}$ для *i*-го пола и *j*-го курса; μ_0 – оценка свободного коэффициента модели; α_i – оценка эффекта *i*-го пола; β_j – оценка эффекта *j*-го курса обучения; γ_{ij} – оценка взаимодействия *i*-го пола и *j*-го курса; ε_{ij} – случайная ошибка.

Реализация дисперсионного анализа по схеме общей линейной модели, однако, показала, что статистически значимым на достаточно высоком уровне (p=0,014) является лишь взаимодействие факторов пол и курс обучения – (таб. 5).

Таблица 4 — Индекс удовлетворенности студентов-«технарей» выбором вуза в зависимости от пола и курса обучения (проценты)

Table 4 — The index of satisfaction of students - “techies”in with the choice of university depending on the gender and the course of education (percentage)

Пол	Курс	Статистика	Удовлетворенность выбором вуза		Индекс удовлетворенности, %
			да	нет	
мужской	1	частота	18	26	-18,2
		частость	40,9%	59,1%	
	2	частота	32	23	16,4
		частость	58,2%	41,8%	
	3	частота	38	20	31,0
		частость	65,5%	34,5%	
	4	частота	51	39	13,4
		частость	56,7%	43,3%	
	Итого	частота	139	108	12,6
		частость	56,3%	43,7%	
женский	1	частота	11	3	57,2
		частость	78,6%	21,4%	
	2	частота	6	8	-14,2
		частость	42,9%	57,1%	
	3	частота	26	38	-18,8
		частость	40,6%	59,4%	
	4	частота	11	15	-15,4
		частость	42,3%	57,7%	
	Итого	частота	54	64	-8,4
		частость	45,8%	54,2%	

Таблица 5 — Дисперсионный анализ двухфакторной модели индекса удовлетворенности студентов выбором вуза (факторы «пол студента», «курс обучения» и их взаимодействие)

Table 5 — Analysis of variance of two-factor model of the index of satisfaction of students with the choice of university (factors «gender», «course of training» and their interaction)

Источник изменчивости	Сумма квадратов	Степени свободы	Средний квадрат	Критерий Фишера *F*	Уровень значимости
Пол студентов	712,890	1	712,890	2,462	0,155
Курс обучения	1725,530	3	575,177	1,987	0,195
Взаимодействие Пол*Курс	5914,290	3	1971,430	6,810	0,014
Ошибка	2316,000	8	289,500		

С формальной точки зрения полученная двухфакторная модель объясняет достаточно большую часть общей дисперсии – 78,3%, что само по себе является положительным фактом, но она не может считаться адекватной эмпирическим данным, поскольку значения критерия Фишера для обоих факторов – пола студентов и курса обучения – статистически не значимы.

Так, можно заметить, что наибольшее различие в характере зависимости индекса удовлетворенности выбором вуза от курса обучения студентов-юношей и студентов-девушек наблюдается для первокурсников: если у студентов-юношей на первом курсе преобладают отрицательные ответы на ключевой вопрос об их удовлетворенности выбором вуза, то у студентов-девушек, напротив, преобладают положительные ответы. По мере роста курса обучения у респондентов-юношей, обучающихся на 2–4 курсах, преобладают положительные ответы, у респондентов-девушек – отрицательные.

Выявленный нами факт полярности суждений юношей и девушек, обучающихся на первом курсе, показателен – он свидетельствует в пользу гипотезы о значительной роли психологических факторов в процессе адаптации студентов к условиям вуза: если у девушек-первокурсниц преобладает эмоциональное восприятие новой обстановки, ожидание хорошего, то у юношей-первокурсников, напротив, наблюдается настороженное отношение к новым для них условиям. Показательным является и наблюдаемый факт снижения индекса удовлетворенности выбором вуза у юношей и в особенности у девушек, для которых характерно преобладание отрицательных ответов на ключевой вопрос об их удовлетворенности выбором вуза.

Отметим также важный методический аспект двухфакторной модели с парным взаимодействием - она в точности соответствует исходным данным. В отличие от этого, факторная модель, не включающая взаимодействия, «сглаживает» исходные данные. Правомочность использования результатов моделирования в анализе выявленных закономерностей в определенной мере подтверждается анализом характеристик качества полной факторной модели. Коэффициент детерминации составляет величину R^2=0,783, т.е. полная модель объясняет 78,3% общей дисперсии, а оставшиеся 21,7% объясняются другими, не учтенными моделью факторами, а также погрешностью социологических измерений.

В данной статье не ставится задача сравнения моделей, построенных по данным социологических исследований факторов адаптации студентов 2000-х и 2010-х гг. Этот вопрос, который представляет несомненный теоретический и практический интерес, будет предметом обсуждения в наших последующих публикациях.

Библиография / References:

1. Анисимова Т.Г., Лясковская О.В. Исследование факторов адаптации современного студенчества к обучению в вузе методами дисперсионного анализа // Среднерусский вестник общественных наук. – 2014. – №5.

2. Анисимова, Т.Г., Шуметов, В.Г. Адаптация современного студенчества к обучению в вузе: количественный анализ результатов социологического исследования // Управление в условиях динамично развивающегося общества. Материалы IV Орловских социологических чтений 7 декабря 2012 г. – Орел: Изд-во ОФ РАНХиГС, 2013. – С. 176 - 180.

3. Бююль, А., Цёфель, П. SPSS: Искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. – СПб.: ООО «ДиаСофтЮП», 2002. – 608 с.

4. Редько, Л.Л., Лобейко, Ю.А. Психолого-педагогическая поддержка адаптации студента-первокурсника в вузе. – М.: Илекса, 2008. – 296 с.

5. Социально-психологическая и профессиональная адаптация студентов вузов Центрального региона России (по материалам социологического исследования) / В.И. Уварова, В.Г. Шуметов, Т.Н. Афонина, Т.А. Иваненко. Под ред. канд. филос. наук В.И. Уваровой. – Орел: Изд-во ОрелГАУ, 2001. – 197 с.

6. Шуметов В.Г., Лясковская О.В., Гудова Т.Г. Data Mining в эмпирической социологии: методология количественного моделирования: монография / Под общей ред. В.Г. Шуметова. – Орел: Изд-во ОРАГС, 2011. – 275 с.

7.Шуметов В.Г., Крюкова О.А. Методология и практика анализа данных в управлении: методы одномерного и двумерного анализа. Монография. – Орел: Изд-во ОФ РАНХиГС, 2013. – 178 с.

8. SPSS Base 8.0 для Windows. Руководство по применению. – М.: СПСС Русь, 1998. – 397 с.

1. Anisimova, T.G., Ljaskovskaja, O.V. (2014) Issledovanija faktorov adaptacii sovremennogo studenchestva k obucheniju v vuze metodami dispersionnogo analiza [Study on adaptation of modern students to training in high school by variance analysis] // Srednerusskij vestnik obschestvennyh nauk [Central Russian Journal of Social Sciences]. - №.5. (In Russ.)

2. Anisimova, T.G., Shumetov, V.G. (2013) Adaptacija sovremennogo studenchestva k obucheniju v vuze: kolichestvennyj analiz rezul'tatov sociologicheskogo issledovanija [Adaptation of modern students to learn in high school: a quantitative analysis of the results of sociological research] // Upravlenie v uslovijah dinamichno razvivajuschegosja obschestva. Materialy IV Orlovskih sociologicheskih chtenij 7 dekabrja 2012 g. Orjol: Izd-vo OF RANHiGS – P. 176 – 180. (In Russ.)

3. Bjujul', A., Cjefel', P. (2002) SPSS: Iskusstvo obrabotki informacii. Analiz statisticheskih dannyh i vosstanovlenije skrytyh zakonomernostej [Art information processing. Analysis of statistical data and restore hidden patterns]. - SPb.: OOO "DiaSoftJuP" – 608 p. (In Russ.)

4. Red'ko, L.L., Lobejko, Ju.A. (2008) Psihologo-pedagogicheskaja podderzhka adaptacii studenta-pervokursnika v vuze [Psychology and pedagogical support of adaptation of the first-year student in higher education institution]. - M.: Ileksa – 296 p. (In Russ.)

5. Social'no-psihologicheskaja i professional'naja adaptacija studentov vuzov Central'nogo regiona Rossii [Social and psychological and professional adaptation of students of higher education institutions of the Central region of Russia]) / V.I. Uvarova, V.G. Shumetov, T.N. Afonina, T.A. Ivanenko, pod. red. kand. filos. nauk V.I. Uvarovoj, (2001) Orjol: Izd-vo OrjolGAU – 197 p. (In Russ.)

6. Shumetov, V.G., Ljaskovskaja, O.V., Gudova, T.G. (2011) Data Mining v empiricheskoj sociologii: metodologija kolichestvennogo modelirovanija: monografija [Data Mining in empirical Sociology: Methodology of quantitative Modeling[/ Pod obschej red. V.G. Shumetova. - Orjol: Izd-vo ORAGS – 275 p. (In Russ.)

7. Shumetov, V.G., Kriukova, O.A. (2013) Metodologiia i praktika analiza dannykh v upravlenii: metody odnomernogo i dvumernogo analiza [The methodology and practice of data analysis in the administration: the methods of one-dimensional and two-dimensional analysis]. Monografiia. – Orel: Izd-vo OF RANKhiGS. – 178 p. (In Russ.)

8. SPSS Base 8.0 dlja Windows. Rukovodstvo po primeneniju (Application Guide) (1998) — M.: SPSS Rus'— 397 p. (In Russ.)