

## СТАТИСТИЧЕСКИЙ АНАЛИЗ ПРОФИЛЬНОГО ВЫСШЕГО ОБРАЗОВАНИЯ НА ОСНОВЕ ВСТУПИТЕЛЬНЫХ ИСПЫТАНИЙ

Арефьев В.П., Михальчук А.А.

*ФГБОУ ВПО «Национальный исследовательский Томский политехнический университет», Томск, Россия  
(634050, г. Томск, пр. Ленина, 30), e-mail: [yva@ido.tpu.ru](mailto:yva@ido.tpu.ru)*

---

Проведен многомерный статистический анализ качества приема в восемь профильных групп российских государственных вузов на основе результатов вступительных испытаний 2012 г. Выявлена высокосignificant отрицательная корреляция между долевыми количествами абитуриентов, принятых по конкурсу баллов ЕГЭ и по целевому набору. В трехмерной системе показателей вступительных испытаний 2012 г. (средний балл ЕГЭ 2012 г., долевыми количества абитуриентов, принятых по конкурсу баллов ЕГЭ и по олимпиадам) сделан вывод о высокосignificant различии восьми профильных групп вузов по совокупности показателей вступительных испытаний, оценена их однородность по каждому показателю в рамках однофакторного дисперсионного анализа и построена статистически значимая пятикластерная качественная модель восьмипрофильного потенциального российского высшего образования. Такая модель позволила выделить из восьми профильных групп вузов пять однородных (по совокупности показателей) кластеров; 1 составной (архитектурно-строительные, классические, педагогические и технические вузы) и 4 монокластера, состоящие соответственно из гуманитарных, аграрных, социально-экономических и медицинских вузов. Для каждого показателя оценен тренд динамики профильно-групповых средних и его значимость. Результаты подобной кластеризации вузов могут быть учтены в рамках проходящей реформы высшего образования, например при принятии решений об оптимизации сети вузов и их финансировании.

---

Ключевые слова: многомерный статистический анализ, корреляционный анализ, кластерный анализ, дисперсионный анализ, рейтинг, вступительные испытания.

## STATISTICAL ANALYSIS OF PROFILE HIGHER EDUCATION ON THE BASIS OF INTRODUCTORY TRIALS

Arefyev V.P., Mihalchuk A.A.

*National research Tomsk polytechnic university, Tomsk, Russia (634050, Tomsk, Lenin's avenue, 30), e-mail:  
[yva@ido.tpu.ru](mailto:yva@ido.tpu.ru)*

---

The many-dimensional statistical analysis of quality of reception is spent to eight profile groups of the Russian state high schools on the basis of outcomes of introductory trials of 2012. Highly significant inverse correlation between share amounts of the entrants accepted on competition of points of Unified State Examination and on a target gang is revealed. In three-dimensional system of indicators of introductory trials of 2012 (the mean score of Unified State Examination of 2012, share amounts of the entrants accepted on competition of points of Unified State Examination and on olympiad) is drawn a conclusion on highly significant distinction of eight profile groups of high schools on a population of indicators of introductory trials, their homogeneity on each indicator within the limits of an one-factor analysis of variance is estimated and the significant five cluster qualitative model of eight profile potential Russian higher educations is constructed statistically. Such model has allowed to select from eight profile groups of high schools five homogeneous (on a population of indicators) clusters; 1 composite (is architectural-building, classical, pedagogical and technical colleges) and 4 mono a cluster, consisting accordingly from humanitarian, agrarian, social and economic and medical high schools. For each indicator the trend of dynamics of profilno-group averages and its importance is estimated. Outcomes similar clustering high schools can be considered within the limits of passing reform of higher education, for example, at a decision making about optimisation of a web of high schools and their financing.

---

Keywords: Many-dimensional statistical analysis, correlation analysis, component analysis, the cluster analysis, analysis of variance, rating, introductory trials.

В рамках реформирования современного российского высшего образования [1; 4; 7; 8] единый государственный экзамен (ЕГЭ) предлагается в качестве одного из критериев, по которым будет оцениваться деятельность высших учебных заведений [6]. Результаты ЕГЭ

рассматриваются как критерий оценки качества работы средней школы и качества набора абитуриентов в высшую школу [3; 5; 9]. При реорганизации сети вузов РФ предполагается сокращение их числа и изменение механизма перераспределения средств в государственном секторе образования таким образом, что финансирование вуза будет зависеть от качества сформированного им контингента студентов, то есть от рейтинга вуза по среднему баллу ЕГЭ ( $m_E$ ). Такой рейтинг [2], показывающий уровень знаний абитуриентов 2012 г., поступивших на бюджетные места в 491 государственный вуз страны, возглавляют 10 вузов, среди которых 5 социально-экономических (СЭВ), 2 классических (КВ), по одному техническому (ТВ), гуманитарному (ГВ), медицинскому (МВ) и ни одного педагогического (ПВ), архитектурно-строительного (АСВ), аграрного (АВ).

В данной работе проведен многомерный статистический (дисперсионный, корреляционный и кластерный) анализ профильного высшего образования на основе не только рейтинга вузов по среднему баллу ЕГЭ, но и целой системы показателей вступительных испытаний (ПВИ) 2012 г. Статистический анализ проведен в системе Statistica [10].

На основании ПВИ 2012 г. [2] были рассчитаны средние для профильных групп вузов по каждому ПВИ (табл. 1), среди которых, кроме  $m_E$ , приведено также долевое количество абитуриентов (в % от общего количества бюджетных мест), принятых по конкурсу баллов ЕГЭ ( $N_E$ ), по олимпиадам ( $N_O$ ), по льготам ( $N_L$ ) и по целевому набору ( $N_C$ ). Заметим, что подсистема долевых показателей является избыточной, так как  $N_E + N_C + N_O + N_L = 100\%$ , что позволяет исключить из рассмотрения одну долевую ПВИ, например  $N_L$ .

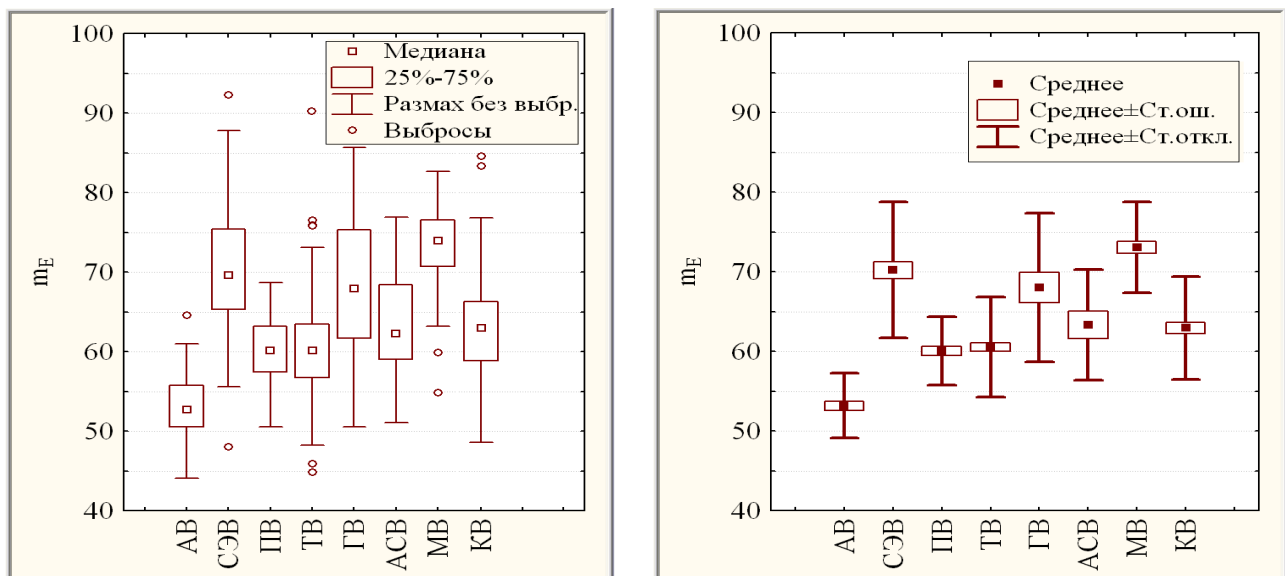
**Таблица 1 – Профильно-групповые средние ПВИ 2012 г.**

Профильные группы	100-процентная шкала					Стандартизированная шкала					Число вузов
	$m_E$	$N_E$	$N_C$	$N_O$	$N_L$	$m_E$	$N_E$	$N_C$	$N_O$	$N_L$	
МВ	73,06	55,35	36,54	0,50	7,60	1,166	-1,493	1,546	-0,159	0,475	51
СЭВ	70,19	78,33	8,84	3,48	9,35	0,827	-0,077	-0,327	0,471	0,869	65
ГВ	68,01	87,89	5,78	0,89	5,43	0,570	0,513	-0,533	-0,076	-0,015	24
АСВ	63,33	89,55	7,41	0,19	2,84	0,018	0,615	-0,423	-0,226	-0,599	16
КВ	62,95	85,39	7,01	1,01	6,60	-0,027	0,359	-0,451	-0,052	0,248	86
ТВ	60,52	81,36	13,77	1,76	3,12	-0,314	0,110	0,006	0,107	-0,537	139
ПВ	60,05	85,35	9,23	0,10	5,32	-0,369	0,356	-0,300	-0,244	-0,041	60
АВ	53,17	76,81	18,97	0,04	4,18	-1,181	-0,171	0,358	-0,257	-0,298	50
Всего	63,18	79,57	13,67	1,25	5,50	0,000	0,000	0,000	0,000	0,000	491

В результате проверки ПВИ на корреляционную зависимость выделена наиболее высокозначимая отрицательная корреляция между  $N_E$  и  $N_{Ц}$  (коэффициент парной корреляции Пирсона  $r \approx -0,921$ , а ранговый Спирмена  $R \approx -0,849$ ). Наличие сильной корреляционной связи между  $N_E$  и  $N_{Ц}$  позволяет уменьшить количество долевых ПВИ еще на одну, например,  $N_{Ц}$ , связанной с  $N_E$  уравнением линейной регрессии  $N_{Ц} = 80,525 - 0,84 * N_E$ , и провести дальнейший анализ в трехмерной системе ПВИ:  $\{m_E, N_E, N_O\}$ .

Согласно  $\lambda$ -критерию Уилкса восемь профильных групп вузов высокозначимо (на уровне значимости  $p < 0,0005$ ) различаются по совокупности показателей  $\{m_E, N_E, N_O\}$ . В рамках однофакторного дисперсионного анализа можно оценить их однородность по каждому показателю. Заметим, что в рассматриваемом случае параметрический  $F$ -критерий подтверждает выводы рангового критерия Краскела–Уоллиса и медианного теста о высокозначимом (на уровне  $p < 0,0005$ ) различии между профильными группами для каждого показателя, а использование параметрического апостериорного критерия наименьших значений разности приводит к большему количеству малочисленных объединений профильных однородных групп. При этом можно выделить для каждого показателя однородные (различающиеся незначимо, то есть на уровне значимости  $p > 0,10$ ) объединения профильных групп, расположенные в порядке убывания групповых средних:

→  $m_E$  (рис. 1):  $\{MB\}$ ,  $\{CЭВ, ГВ\}$ ,  $\{АСВ, KB\}$ ,  $\{АСВ, ТВ, ПВ\}$ ,  $\{AB\}$ . Имеется одна пара пересекающихся объединений профильных групп.



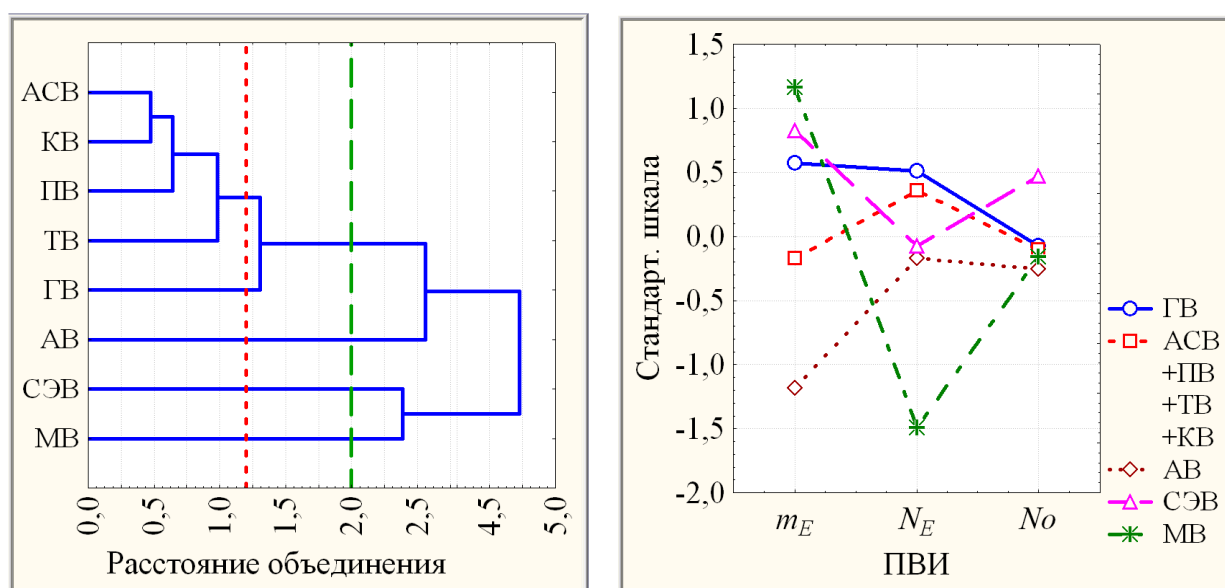
**Рис. 1. Непараметрическая (слева) и параметрическая (справа) диаграммы размаха профильных групп вузов по 100-балльной шкале  $m_E$ .**

→  $N_E$ :  $\{АСВ, ГВ, KB, ПВ\}$ ,  $\{KB, ПВ, ТВ\}$ ,  $\{ТВ, CЭВ, AB\}$ ,  $\{MB\}$ . Имеются две последовательные пары пересекающихся объединений профильных групп.

→  $N_0$ : {СЭВ, ТВ}, {ТВ, КВ, ГВ}, {ГВ, МВ, АСВ, ПВ, АВ}. Имеются две последовательные пары пересекающихся объединений профильных групп.

В связи с тем что неоднородность групп вузов оценивалась как с помощью параметрического  $F$ -критерия, так и с помощью рангового критерия Краскела–Уоллиса, наряду с параметрическими (рис. 1, справа) числовыми характеристиками (среднее и стандартное отклонение) приводятся и непараметрические (рис. 1, слева) их аналоги (медиана и квартили) для профильных групп вузов по  $m_E$ .

Наличие однородных объединений профильных групп вузов по каждому показателю  $m_E, N_E, N_0$  допускает проведение их кластеризации в трехмерной системе ПВИ:  $\{m_E, N_E, N_0\}$ . В качестве меры близости выбрано расстояние городских кварталов, а в качестве правила объединения двух кластеров использован метод Уорда. Методом иерархической кластеризации построена дендрограмма (рис. 2, слева). В зависимости от выбора расстояния объединения можно получить соответствующее число кластеров. Так, например, расстоянию объединения, равному 1,2 (красная пунктирная вертикальная прямая), соответствуют 5 кластеров: 1 составной (АСВ+КВ+ПВ+ТВ) и 4 монокластера (ГВ), (АВ), (СЭВ), (МВ), а расстоянию объединения, равного 2, соответствуют 4 кластера за счет объединения (АСВ+КВ+ПВ+ТВ) с (ГВ). Наряду с методом древовидной кластеризации применен также метод  $K$ -средних, проводящий классификацию объектов (профильных групп вузов) по заданному количеству кластеров. Предлагается 5-кластерная модель профильных групп вузов, согласно  $\lambda$ -критерию Уилкса высокозначимо (на уровне значимости  $p < 0,000005$ ) различающая 5 кластеров профильных групп вузов по совокупности  $\{m_E, N_E, N_0\}$ . Кластерные графики средних профильных групп вузов изображены на рис. 2 справа.



**Рис. 2.** Дендрограмма (слева) и кластерные графики (справа) средних профильных групп вузов по стандартизированной шкале  $\{m_E, N_E, N_0\}$ .

Алгоритм метода  $K$ -средних перемещает объекты в разные кластеры с целью минимизации изменчивости внутри кластеров и максимизации изменчивости между кластерами, оценивая качество кластеризации профильных групп вузов по каждому ПВИ посредством параметрического дисперсионного анализа (табл. 2) как статистически значимое (на уровне значимости  $0,01 < p < 0,05$ ) для  $m_E$  и  $N_E$ .

**Таблица 2 – Результаты дисперсионного анализа кластеризации профильных групп вузов по каждому ПВИ**

перемен.	Дисперсионный анализ					
	Между SS	сс	Внутри SS	сс	F	значим. р
$m_E$	3,824	4	0,116	3	24,733	<b>0,012</b>
$N_E$	3,040	4	0,128	3	17,855	<b>0,020</b>
$N_O$	0,339	4	0,082	3	3,097	<b>0,190</b>

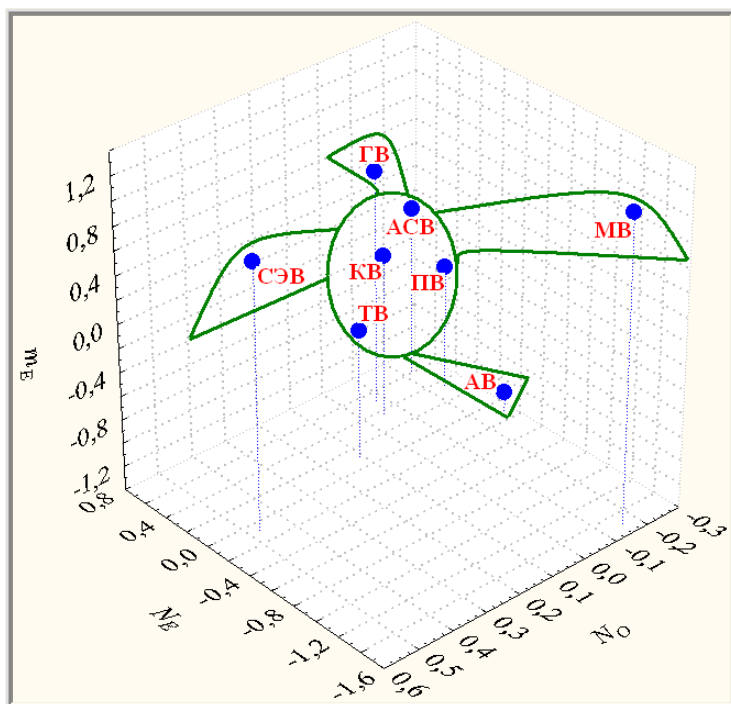
На основании рис. 2 в трехмерной системе ПВИ:  $\{m_E, N_E, N_O\}$  можно провести качественную классификацию профильных групп вузов в номинальной шкале измерений (табл. 3), полагая в качестве уровня «Средний» – стандартизированный интервал  $(-0,3; +0,3)$  для показателей. Аномально высокие значения  $(> +1,0)$  определяют уровень «**Лидер**», а аномально низкие значения  $(< -1,0)$  определяют уровень «**Аутсайдер**». Промежуточные значения между средними и аномальными определяют уровень «Выше среднего» и «Ниже среднего» соответственно. Таким образом, выделяются 5 объединений профильных групп.

**Таблица 3 – Качественная классификация профильных групп вузов**

Кластер	Уровень кластера на фоне среднего по ПВИ		
	$m_E$	$N_E$	$N_O$
МВ	<b>Лидер</b>	<b>Аутсайдер (Лидер по <math>N_{Ц}</math>)</b>	Средний
СЭВ	Выше среднего	Средний	Выше среднего
ГВ	Выше среднего	Выше среднего	Средний
АСВ+ПВ+ТВ+КВ	Средний	Средний – Выше среднего	Средний
АВ	<b>Аутсайдер</b>	Средний	Средний

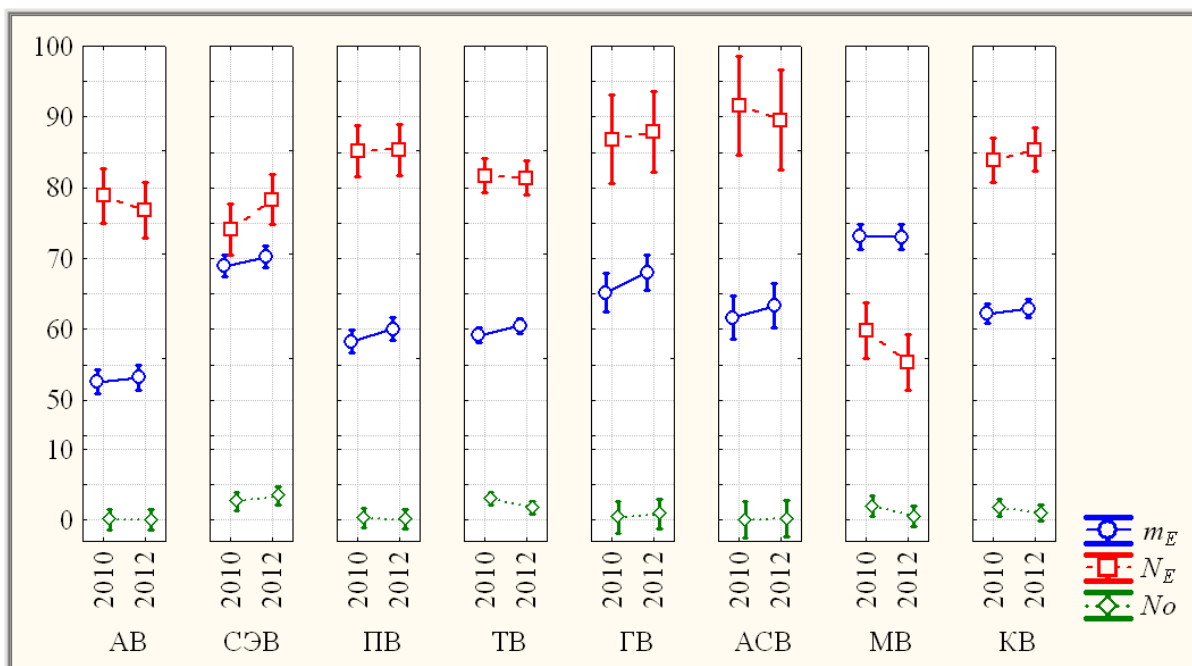
Результаты качественной классификации профильных групп вузов представлены на рис. 3 в виде образной формы «птицы потенциального профильного высшего образования», обладающей «гуманитарной» головкой с повышенным уровнем  $m_E$  и  $N_E$ , правым «медико-целевым» крылом, лидирующим по  $m_E$  и  $N_{Ц}$ , левым «социально-экономическим» крылом с

повышенным уровнем  $m_E$  и  $N_O$ , разнопрофильным туловищем (АСВ+ПВ+ТВ+КВ) среднего уровня по всем показателям и «аграрным» хвостом с сильно пониженным уровнем  $m_E$ .



**Рис. 3. Диаграмма рассеяния профильно-групповых средних в системе  $\{m_E, N_E, N_O\}$  ПВИ 2012 г.**

На основании ПВИ [2] в трехмерной системе  $\{m_E, N_E, N_O\}$  можно оценить тренд динамики профильно-групповых средних и его значимость (рис. 4).



**Рис. 4. Диаграмма рассеяния профильно-групповых средних 2010 и 2012гг. с 95% доверительными интервалами в трехмерной системе ПВИ  $\{m_E, N_E, N_O\}$ .**

К сожалению, ограниченность базы данных 2011 г. (включены вузы, принявшие не менее 200 человек по результатам ЕГЭ) уменьшает общее количество охваченных исследованием вузов более чем на 100 по сравнению с 2010 и 2012 гг., а по некоторым профильным группам вузов (СЭВ и МВ) уменьшает их число в 2 раза, что делает их сравнение некорректным. Поэтому далее профильно-групповые средние в трехмерной системе ПВИ  $\{m_E, N_E, N_O\}$  сравниваются по 2010 и 2012 гг.

Для каждого показателя оценим тренд динамики профильно-групповых средних и его значимость на основании параметрического  $t$ -критерия (уровень значимости  $p_t$ ) и непараметрического  $U$ -критерия Манна-Уитни (уровень значимости  $p_{M-V}$ ) для двух независимых групп:

- ⊗  $m_E$ : положительная динамика у всех профильных групп, кроме МВ. Однако значимой в разной степени оценивается положительная динамика почти на 2 балла у ПВ (уровень статистической значимости  $0,01 < p_t \approx 0,03 < 0,05$  и  $0,01 < p_{M-V} \approx 0,04 < 0,05$ ) и почти на 1,5 балла у ТВ (уровень слабой значимости  $0,01 < p_t \approx 0,07 < 0,05$  и уровень статистической значимости  $0,01 < p_{M-V} \approx 0,03 < 0,05$ ). Остальные случаи оцениваются как не значимые ( $p_t > 0,1$ ;  $p_{M-V} > 0,1$ ), в том числе наибольшая положительная динамика у ГВ (увеличение группового среднего почти на 3 балла).
- ⊗  $N_E$ : положительная не значимая ( $p_t > 0,1$ ;  $p_{M-V} > 0,1$ ) динамика у СЭВ (наибольшая почти на 4%), ПВ, ГВ, КВ и отрицательная не значимая ( $p_t > 0,1$ ;  $p_{M-V} > 0,1$ ) динамика у МВ (наибольшая почти на 4,5%), АСВ, ТВ, АВ.
- ⊗  $N_O$ : положительная не значимая ( $p_t > 0,1$ ;  $p_{M-V} > 0,1$ ) динамика у СЭВ (наибольшая почти на 1%), ГВ, АСВ и отрицательная не значимая ( $p_t > 0,1$ ;  $p_{M-V} > 0,1$ ) динамика у остальных, в том числе у МВ и ТВ ( $> 1\%$ ).

### Выводы

1. Выявлена высокозначимая отрицательная корреляция между долевыми количествами абитуриентов, принятых по конкурсу баллов ЕГЭ и по целевому набору.
2. В трехмерной системе ПВИ 2012 г. (средний балл ЕГЭ, долевыми количества абитуриентов, принятых по конкурсу баллов ЕГЭ и по олимпиадам) сделан вывод о высокозначимом различии восьми профильных групп вузов по совокупности показателей, оценена их однородность по каждому показателю в рамках дисперсионного анализа.

3. Дисперсионный анализ восьми профильных групп вузов по совокупности трех показателей позволил выделить пять однородных объединений и представить результаты в виде образной формы «птицы потенциального профильного высшего образования», обладающей «гуманитарной» головкой, правым (целевым) «медицинским» крылом, левым (олимпиадным) «социально-экономическим» крылом, «аграрным» хвостом и разнопрофильным туловищем из оставшихся четырех профильных групп вузов.
4. Для каждого показателя оценен тренд динамики профильно-групповых средних и его значимость. Статистически значимой оценена только положительная динамика по среднему баллу ЕГЭ почти на 2 балла у ПВ и почти на 1,5 балла у ТВ.
5. Результаты подобной кластеризации вузов могут быть учтены в рамках проходящей реформы высшего образования, например при принятии решений об оптимизации сети вузов и их финансировании.

#### **Список литературы**

1. Загвязинский В.И. Современная образовательная ситуация и задачи модернизации российского образования // Народное образование. – 2012. – № 5. – С. 11–16.
2. Качество приема в вузы (2010–2012 гг.). – URL: <http://www.hse.ru/ege> (дата обращения: 17.10.2012).
3. Лапотникова И.Н. Методы математической статистики для оценки результатов ЕГЭ // Ярославский педагогический вестник. – 2008. – № 1. – С. 17–23.
4. Ливанов: наука и образование в России нуждаются в радикальной реформе. – URL: [http://ria.ru/edu\\_news/20120621/678237385.html](http://ria.ru/edu_news/20120621/678237385.html) (дата обращения: 11.10.2012).
5. Малинецкий Г.Г., Подлазов А.В. ЕГЭ: Научно-педагогический диагноз // Профессиональное образование. Столица. – 2012. – № 7. – С. 29-34.
6. Минобрнауки разработало критерии оценки деятельности вузов. – URL: [http://ria.ru/edu\\_higher/20120730/712748694.html](http://ria.ru/edu_higher/20120730/712748694.html) (дата обращения: 11.10.2012).
7. Миронов В.В. О реформе российского образования (часть первая) // Социология образования. – 2012. – № 7. – С. 4-27.
8. Мисник М.В. Проблемы высшего профессионального образования в России // Качество. Инновации. Образование. – 2012. – № 2. – С. 11–17.
9. Талынев В.Е., Якушкин В.В. Отбор в систему современного высшего образования и отношение к ЕГЭ // Социология образования. – 2012. – № 4. – С. 66–72.
10. Халафян А.А. Statistica 6. Статистический анализ данных. – М. : Бином-Пресс, 2008. – 512 с.



*Работа выполнена в рамках государственного задания «Наука» № 1.604.2011 и поддержана ФЦП «Научные и научно-педагогические кадры инновационной России» по контрактам П691.*

#### **Рецензенты**

Трифонов Андрей Юрьевич, д.ф.-м.н., профессор кафедры высшей математики и математической физики, ФГБОУ ВПО «Национальный исследовательский Томский политехнический университет», г. Томск.

Арефьев Константин Петрович, д.ф.-м.н., профессор кафедры высшей математики, ФГБОУ ВПО «Национальный исследовательский Томский политехнический университет», г. Томск.