

Технологии инженерного образования

УДК 378.146:51:681.3

СРАВНИТЕЛЬНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ОЦЕНКИ МАТЕМАТИЧЕСКИХ ЗНАНИЙ СТУДЕНТОВ ПЕРВОГО КУРСА

А.Ю. Трифонов, А.А. Михальчук

Томский политехнический университет
E-mail: trifonov@phtd.tpu.edu.ru

Обсуждены методы оценок знаний студентов вузов. На базе результатов текущего контроля математических знаний студентов первого курса Томского политехнического университета проведен сравнительный статистический анализ системы оценки знаний в форме тестирования с машинной проверкой, аналогичной ЕГЭ (в объеме групп заданий А и В, без группы С) или централизованному тестированию, и системы традиционного контроля знаний на основе аудиторных контрольных работ с проверкой их преподавателями. Сделан вывод о статистически значимых различиях в результатах оценки знаний этими методами. Обсуждены причины выявленных существенных различий в рассматриваемых системах оценки знаний и место этих систем в технологии контроля качества знаний студентов.

Введение

Стратегической целью ТПУ является его дальнейшая интеграция в международную научно-образовательную систему и превращение в одного из мировых лидеров высшей инженерной школы [1]. Одной из составляющих этого процесса является разработка и использование методик оценки качества образования, сопоставимых с мировыми аналогами, в частности, создание эффективной системы контроля качества знаний [2–4].

В Томском политехническом университете (ТПУ) проводится эксперимент по введению системы независимой оценки знаний студентов набора 2003 г. в тестовой форме. Тестированием в форме, аналогичной билетам ЕГЭ (в объеме групп заданий А и В, без группы С), с последующей машинной проверкой оценивались текущие знания по изучаемым в первом семестре по трем разделам курса высшей математики: ЛА – линейная алгебра (*тест₁*), ВА – векторная алгебра (*тест₂*) и Пр – пределы (введение в анализ) (*тест₃*). В отличие от результатов тестирования *тест_к* посредством *ТЕСТ_к* обозначены результаты тестирования студентов, обучающихся на кафедре высшей математики и математической физики (ВММФ) (первый курс физико-технического, электрофизического и химико-технологического факультетов). Конечной целью данного эксперимента является распространение его на все изучаемые разделы математики вплоть до проведения экзамена в тестовой форме. Как правило, выделяют следующие преимущества тестовой системы контроля знаний студентов: **объективность** оценки знаний; **снижение трудоёмкости**

в процессе оценки знаний со стороны преподавателей; популярность тестовой системы контроля знаний в **большинстве развитых стран мира**.

Следует отметить, что анализу тестовых испытаний посвящено значительное количество публикаций (см. например [5–10]). Однако, в подавляющем большинстве работ анализ проводится исходя из внутренней структуры теста и полученных результатов тестирования. В связи с этим представляет интерес сравнение результатов тестирования *ТЕСТ_к* с соответствующими результатами традиционного контроля знаний *КТ_к*, полученными преподавателями кафедры ВММФ в ходе текущего семестрового контроля оценки знаний студентов. Такое сравнение поможет определить роль и место тестовой формы контроля знаний студентов в курсах математических дисциплин.

Статистический метод

Сравнение результатов *ТЕСТ* и *КТ* проведено методами статистического анализа. Все числовые результаты контроля качества знаний студентов были приведены к единой пяти бальной шкале (делением результата на соответствующий максимальный результат и умножением на число пять). База данных, созданная таким образом в MS Excel, использовалась далее в пакете STATISTICA 6.0 для статистического анализа данных [11, 12]. Строились частотные группированные распределения выборок и вычислялись выборочные числовые характеристики (среднее *m*, стандартное отклонение *σ*, асимметрия *A* и эксцесс *E*, а также стандартные ошибки (Std. Err.) *A* и

Е). Использовался двумерный визуальный анализ данных (гистограммы, линейные графики). Применялся *t*-критерий Стьюдента (*t* Test) для сравнения средних двух независимых нормальных выборок (при этом нормальность распределения выборок проверялась с помощью оценок асимметрии *A* и эксцесса *E*, χ^2 -критерия Пирсона и критерия Колмогорова-Смирнова). В случае невыполнения условий применения *t*-критерия Стьюдента применялись его непараметрические аналоги: *U* – критерий Манна-Уитни (M-W U Test), двухвыборочный критерий Колмогорова-Смирнова (K-S Test).

Результаты и их анализ

Сначала было проведено сравнение результатов тестирования в рамках ТПУ и кафедры ВММФ в 2003 г. Результаты суммарного (ЛА + ВА + Пр) тестового контроля ($mest_{1-3}$)₂₀₀₃ и ($TEST_{1-3}$)₂₀₀₃ отображены на рис. 1. Над столбцами указаны значения относительных частот в %. Для визуальной оценки сходства наблюдаемых распределений (гистограммы) с теоретическим распределением по нормальному закону (соответствующие кривые) использована равномерная 5-бальная шкала.

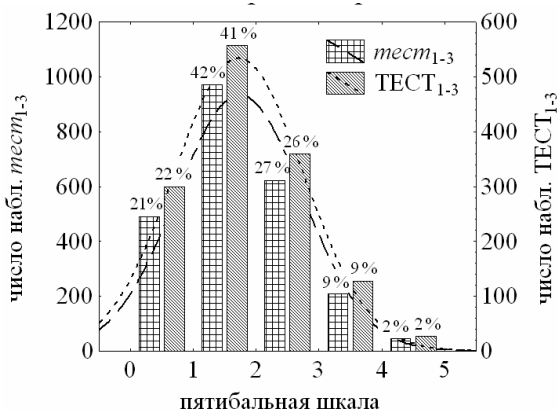


Рис. 1. Составная гистограмма результатов суммарного тестирования ($mest_{1-3}$)₂₀₀₃ и ($TEST_{1-3}$)₂₀₀₃

Числовые характеристики выборок ($mest_{1-3}$)₂₀₀₃ и ($TEST_{1-3}$)₂₀₀₃ приведены в табл. 1. Несмотря на существенную разницу в объемах *N* выборок результаты тестирования схожи по остальным числовым характеристикам. Проверка нормальности распределения выборок с помощью асимметрии *A* и эксцесса *E* (*A* и *E* по модулю не должны превышать утроенных соответствующих стандартных ошибок) показывает существенное отклонение от нормальности в случае ($mest_{1-3}$)₂₀₀₃ и ($TEST_{1-3}$)₂₀₀₃ по асимметрии (см. табл.1).

Таблица 1. Числовые характеристики выборок ($mest_{1-3}$)₂₀₀₃ и ($TEST_{1-3}$)₂₀₀₃

Выборка	<i>N</i>	<i>m</i>	σ	<i>A</i>	Std. Err. <i>A</i>	<i>E</i>	Std. Err. <i>E</i>
($mest_{1-3}$) ₂₀₀₃	2340	1,7299	0,9998	0,5334	0,0506	-0,0633	0,1012
($TEST_{1-3}$) ₂₀₀₃	1371	1,7162	1,0245	0,5041	0,0661	-0,1718	0,1321

Проверка нормальности распределения выборок с помощью χ^2 -критерия Пирсона и критерия Колмогорова-Смирнова дают уровни значимости α , приводящие к выводу о высоко значимом ($\alpha < 0,001$) отличии распределения выборок от нормального закона. В связи с нарушением условия нормальности распределения выборок далее применялись непараметрические критерии сравнения средних для двух независимых выборок (см. табл. 2), приводящие к выводу о не значимом различии ($\alpha > 0,1$) в средних баллах результатов тестирования в рамках ТПУ ($mest_{1-3}$)₂₀₀₃ и кафедры ВММФ ($TEST_{1-3}$)₂₀₀₃.

Таблица 2. Критерии сравнения средних в двух выборках ($mest_{1-3}$)₂₀₀₃ – ($TEST_{1-3}$)₂₀₀₃

Выборки	α -level <i>t</i> Test	α -level M-W U Test	α -level K-S Test
($mest_{1-3}$) ₂₀₀₃ – ($TEST_{1-3}$) ₂₀₀₃	0,6994	0,6835	>0,10

В табл. 2. приведены также результаты применения *t*-критерия Стьюдента, подтверждающие вывод об однородности выборок ($mest_{1-3}$)₂₀₀₃ и ($TEST_{1-3}$)₂₀₀₃, что свидетельствует об устойчивости *t*-критерия Стьюдента относительно нарушений предположения о нормальности распределений сравниваемых выборок [11]. Таким образом, далее можно ограничиться рассмотрением выборок $TEST_k$ и сравнением их с соответствующими выборками KT_k в рамках кафедры ВММФ.

Объемы выборок KT_k внутри каждого из наборов 2003 и 2004 г. практически одинаковы. Как видно из рис. 2, объемы выборок $TEST_k$ тем меньше, чем позже проводилось соответствующее тестирование. Такую потерю интереса студентов к тестированию можно объяснить недостаточной организованностью в проведении этого эксперимента и слабой привязкой тестирования к действующей рейтинговой системе оценки успеваемости студента.

К сожалению, тенденция незаинтересованности студентов в участии в тестировании усилилась в дальнейшем, что привело к катастрофическому снижению объемов выборок $TEST_k$ во втором семестре для набора 2003 г., а в третьем семестре тестирование проводилось лишь эпизодически в отдельных группах. С набором 2004 г. ситуация начинает повторяться. Все это ограничивает сравнительный анализ объемом первого семестра.

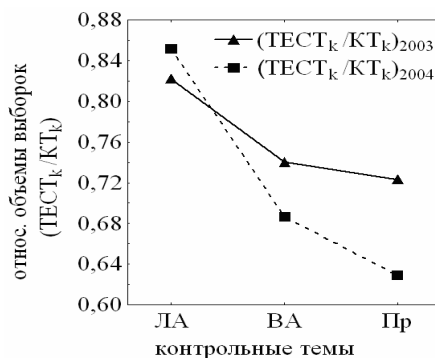


Рис. 2. Относительное количество студентов ВММФ, принявших участие в контрольных испытаниях в 2003 и 2004 гг.

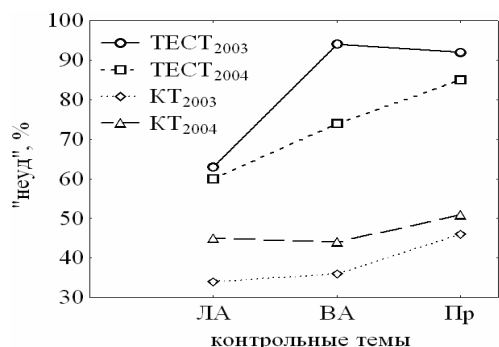


Рис. 3. Сравнение неудовлетворительных результатов КТ и ТЕСТ студентов ВМФ в 2003 и 2004 гг.

Различия КТ и ТЕСТ показаны на рис. 3 сравнением неудовлетворительных оценок ($\leq 2,5$ по 5-бальной шкале) в контрольных темах 2003 и 2004 г. Можно заметить, что по результатам суммарного контроля в 2003 г. (см. рис. 6) неудовлетворительных результатов в $(ТЕСТ_{1-3})_{2003}$ было более чем в 2 раза больше в сравнении с результатами $(КТ_{1-3})_{2003}$ (82 и 39% соответственно). В значительной степени такой контраст результатов объясняется организационными сдвигами в проведении тестирования ТЕСТ₂ и ТЕСТ₃ (тестирование неизученных до конца тем). Поэтому результаты КТ можно считать более адекватно отражающими реальную ситуацию с успеваемостью студентов. Различия результатов КТ₂₀₀₃ и КТ₂₀₀₄ на 5...10% можно объяснить, во-первых, более слабым набором 2004 г. [12] и, во-вторых, неучастием в данном контроле лучших студентов, отделенных в 2004 г. в группы элитного технического образования [13].

Сравним суммарные (ЛА+ВА+Пр) выборки ТЕСТ₁₋₃ и КТ₁₋₃ по 2003 г. (рис. 4) и 2004 г. (рис. 5). Визуально очевидны различия между ТЕСТ₁₋₃ и КТ₁₋₃ как по 2003 г., так и по 2004 г.

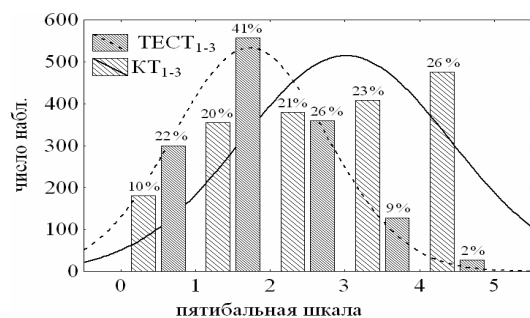


Рис. 4. Составная гистограмма $(КТ_{1-3})_{2003}$ и $(ТЕСТ_{1-3})_{2003}$ по равномерной 5-бальной шкале

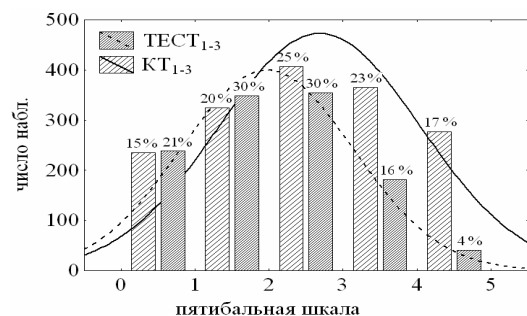


Рис. 5. Составная гистограмма $(КТ_{1-3})_{2004}$ и $(ТЕСТ_{1-3})_{2004}$ по равномерной 5-бальной шкале

На рис. 6 и 7 для простоты восприятия использована неравномерная 5-бальная шкала: [0; 2,5] – «неуд»; (2,5; 3,5] – «удовл»; (3,5; 4,5] – «хор» и (4,5; 5] – «отл», в которой отражено визуальное сравнение результатов КТ₁₋₃ и ТЕСТ₁₋₃. Здесь различия между ТЕСТ₁₋₃ и КТ₁₋₃ как по 2003 г., так и по 2004 г. еще более очевидны.

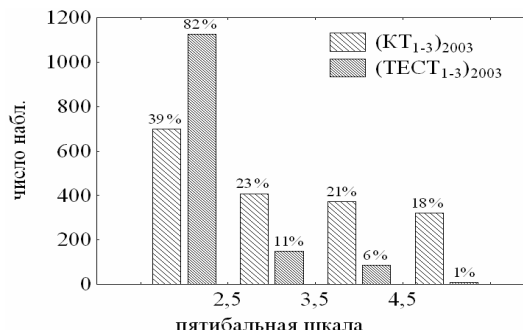


Рис. 6. Составная гистограмма $(КТ_{1-3})_{2003}$ и $(ТЕСТ_{1-3})_{2003}$ по неравномерной 5-бальной шкале

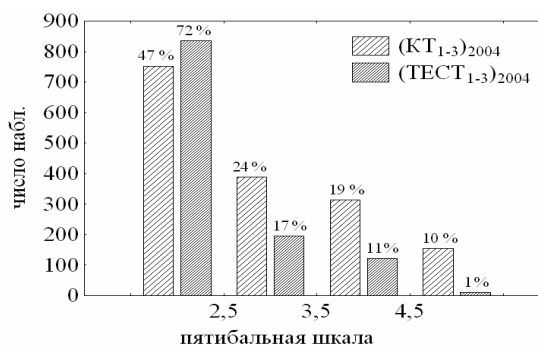


Рис. 7. Составная гистограмма $(КТ_{1-3})_{2004}$ и $(ТЕСТ_{1-3})_{2004}$ по неравномерной 5-бальной шкале

Уровень значимости различия средних результатов ТЕСТ₁₋₃ и КТ₁₋₃ определялся с помощью статистических критериев. Числовые характеристики выборок ТЕСТ₁₋₃ и КТ₁₋₃ приведены в табл. 3. Следует отметить существенные отличия от нормального закона распределения выборок $(КТ_{1-3})_{2003}$ и $(ТЕСТ_{1-3})_{2003}$ по асимметрии *A* и выборок $(КТ_{1-3})_{2003}$, $(ТЕСТ_{1-3})_{2004}$ и $(КТ_{1-3})_{2004}$ по эксцессу *E*, что визуально очевидно из рис. 4, 5. Это подтверждается также проверкой нормальности распределения выборок с помощью χ^2 -критерия Пирсона и критерия Колмогорова-Смирнова, дающих высоко значимые ($\alpha < 0,001$) отличия распределений выборок от нормального закона.

Таблица 3. Числовые характеристики выборок ТЕСТ₁₋₃ и КТ₁₋₃ 2003 и 2004 гг.

Выборка	<i>N</i>	<i>m</i>	σ	<i>A</i>	Std. Err. <i>A</i>	<i>E</i>	Std. Err. <i>E</i>
$(ТЕСТ_{1-3})_{2003}$	1371	1,7162	1,0245	0,5041	0,0661	-0,1718	0,1321
$(КТ_{1-3})_{2003}$	1799	2,9954	1,3944	-0,2384	0,0577	-0,9694	0,1153
$(ТЕСТ_{1-3})_{2004}$	1165	1,9599	1,1621	0,1164	0,0717	-0,7163	0,1432
$(КТ_{1-3})_{2004}$	1609	2,6682	1,3568	-0,1204	0,0610	-0,8480	0,1219

В связи с нарушением условий применимости *t*-критерия Стьюдента для сравнения средних двух независимых выборок использовались его непара-

метрические аналоги, давшие высоко значимые ($\alpha < 0,001$) различия в средних баллах результатов суммарного контроля в 2003 г. (1,7162 для (ТЕСТ₁₋₃)₂₀₀₃ и 2,9954 для (КТ₁₋₃)₂₀₀₃), а также в средних баллах результатов суммарного контроля в 2004 г. (1,9599 для (ТЕСТ₁₋₃)₂₀₀₄ и 2,6682 для (КТ₁₋₃)₂₀₀₄).

Как следует из выше изложенного, ТЕСТ определяет уровень знаний студентов ниже, чем КТ, в силу как организационных, так и методических причин. Но если организационный сбой в проведении ТЕСТ еще можно устранить, то с методическими проблемами сложнее. Предложенные тестовые механизмы оценки знаний являются более грубыми, чем существующие традиционные. По аналогии с билетами ЕГЭ билеты ТЕСТ содержат только задания на «угадывание» (группа А) и задания, правильность решения которых проверяются только по конечному числовому результату (группа В). Например, предлагается вычислить определитель четвертого порядка, что в принципе допускает множество способов решения. Проверка метода решения подобных задач не под силу ТЕСТ. В то же время КТ, опираясь на ручную проверку и собеседование со студентом, основной акцент делает на практические задания (группа В), проверяя методы их решения, и теоретические задания (группа С) например, доказательства теорем, которые развивают строгое логическое мышление. Причем контроль теоретического материала в КТ составляет треть общей суммы баллов по используемой рейтинговой системе. Таким образом, без дальнейшего совершенствования ТЕСТ своими результатами может свести на «нет» все свои «положительные моменты».

Хорошо известно, что при текущей системе контроля знаний студентов в итоговой оценке присутствует составляющая, связанная с субъективными пристрастиями преподавателя. Однако в среднем вклад этой составляющей нивелируется, и среднюю оценку в (допустимых пределах) можно считать объективной. Кроме того, существующая система предусматривает целый ряд механизмов, уменьшающих влияние этого фактора и исправления ситуации, когда такое влияние выходит за допустимые рамки (возможность апелляции, создание комиссий и т.д.). Предлагаемая система полностью исключает субъективную составляющую оценки, однако она вводит другую составляющую. Последняя носит чисто вероятностный характер, и связана с возможностью угадать ответ (достаточно заметить, что вероятность угадать правильно три задания из предложенных восьми с четырьмя вариантами ответов (группа А) по формуле Бернулли [14] равна 0,21), неверно переписать ответ на бланк, ошибиться клеточкой или строчкой (точность попадания имеет фатальный характер) и т.д. Возникает естественный вопрос – стала ли, при новой системе оценки знаний, итоговая оценка более объективной. Следует отметить организационные сбои в проведении тестирования ВА и Пр в 2003 г. (рис. 3), не адекватно отразившие реальную ситуацию с успеваемостью студентов. Полученные

результаты позволяют оценить вклад случайной вероятностной составляющей в итоговую студенческую оценку как существенный. При этом контрольные испытания ВА и Пр, проведенные по традиционной системе, ничем не выделяются из остальных. Вклад случайной вероятностной составляющей в итоговую студенческую оценку особенно существенен на краях интервала (рис. 4, 5). Случайный фактор не позволяет сильным студентам набрать высокий балл, а слабым низкий. В результате распределение ТЕСТ более похоже на нормальное с малой дисперсией по сравнению с распределением КТ.

При традиционном стандартном подходе экзамен является составной частью учебного процесса, и его роль не исчерпывается контролирующими функциями. Во время экзамена студент учится логически мыслить, связно говорить, рассуждать и многое другое. После экзамена студент знает больше, чем до него. Тестовое испытание несет только контролирующие функции, поэтому его проведение, естественно, менее трудоёмко, чем традиционное стандартное испытание. В условиях значительного сокращения аудиторной нагрузки экономия на экзаменах отрицательно скажется на качестве учебного процесса. Тестовые испытания показывают относительные знания студентов. Поэтому критерии тестовой оценки знаний должны привноситься извне (например, приведением в соответствие с традиционной рейтинговой системой оценки текущих знаний студентов).

Тестовая форма оценки знаний получила широкое распространение в мире, однако это не означает, что речь идет о тестах типа ЕГЭ [5] или тестам типа [6–8]. Анализ экзаменационных билетов значительного числа университетов Европы, Америки и Канады по математическим дисциплинам показывает, что они аналогичны билетам письменного экзамена Российской системы образования. В этом смысле предлагаемая система контроля знаний находится значительно дальше от мировых образцов, чем традиционная.

Выводы

1. Различия в результатах оценки качества знаний студентов первого курса по математике методами ТЕСТ и КТ являются статистически значимыми.
2. Приходится констатировать, что предложенные тестовые механизмы оценки знаний являются более грубыми, чем существующие традиционные, и отсутствуют какие-либо основания для утверждения, что предлагаемая система тестового контроля знаний дает более объективную оценку знаний студентов.
3. ТЕСТ проверяет знания студентов, отвечающие нижнему базовому уровню.
4. Тестовые испытания показывают относительные знания студентов. Поэтому критерии оценки знаний должны привноситься извне (например, приведением в соответствие с традиционной рейтинговой системой оценки текущих знаний студентов).

5. Система тестовых испытаний позволяет получить быстро, но грубо (по системе «зачтено – не зачтено») срез знаний. Поэтому система тестовых испытаний должна, прежде всего, применять-

ся для оценки остаточных знаний и контроля самостоятельной работы студентов (особенно по темам, вынесенным на самостоятельное изучение).

СПИСОК ЛИТЕРАТУРЫ

1. Похолков Ю.П. Проблемы и основные направления совершенствования инженерного образования // *Alma Mater. Вестник высшей школы*. – 2003. – № 10. – С. 3–8.
2. Гребнев Л., Попов В. Об организации высшего технического образования в США // *Высшее образование в России*. – 2004. – № 11. – С. 150–165.
3. Сенашенко В., Ткач Г. Болонский процесс и качество образования // *Alma Mater. Вестник высшей школы*. – 2003. – № 8. – С. 8–14.
4. Мотова Г. Система оценки качества образования в странах СНГ и Балтии // *Alma Mater. Вестник высшей школы*. – 2004. – № 1. – С. 37–40.
5. Свиридов А.П. Основы статистической теории обучения и контроля знаний. – М: Высшая школа, 1981. – 262 с.
6. Карпенко Д.С. и др. Автоматизированная система мониторинга эффективности усвоения знаний и качества тестовых заданий // *Инновации в образовании*. – 2001. – № 2. – С. 69–85.
7. Евтюхин Н.В. и др. Современное состояние методов тестирования знаний и умений за рубежом и в России // *Инновации в образовании*. – 2004. – № 1. – С. 27–38.
8. Берестнева О.Г., Марухина О.В. Компьютерные технологии в оценке качества обучения // *Известия Томского политехнического университета*. – 2003. – Т. 306. – № 6. – С. 106–112.
9. Марухина О.В., Берестнева О.Г. Анализ и обработка информации в задачах оценивания качества обучения студентов вуза // *Известия Томского политехнического университета*. – 2004. – Т. 307. – № 4. – С. 136–141.
10. Минин М.Г. Диагностика качества знаний и компьютерные технологии обучения. – Томск: Изд-во ТГПУ, 2000. – 216 с.
11. Боровиков В.П. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. – СПб.: Питер, 2003. – 688 с.
12. Арефьев В.П., Михальчук А.А., Кулебакина Н.Н. Компьютерный статистический анализ качества инженерного образования. Входной контроль математических знаний // *Известия Томского политехнического университета*. – 2005. – Т. 308. – № 2. – С. 201–205.
13. Похолков Ю.П., Вайсбурд Д.И., Чубик П.С. Элитное образование в традиционном техническом университете // *Элитное техническое образование: Труды Междунар. конф. в рамках симпозиума*. – Томск, 2003. – С. 6–8.
14. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 2002. – 473