

УДК 658.562

ВЕРИФИКАЦИЯ КОНТРОЛЬНО-ИЗМЕРИТЕЛЬНЫХ МАТЕРИАЛОВ И ПРОВЕРКА ЗНАНИЙ СТУДЕНТОВ

В. В. Алексеев, П. Г. Королёв, А. В. Утушкина

*Кафедра информационно-измерительных систем и технологий,
ФГБОУ ВПО «Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)»;
PGKorolev@gmail.com*

Ключевые слова и фразы: дисперсия; контрольно-измерительные материалы; малая выборка; математическое ожидание; статистическая гипотеза.

Аннотация: Рассмотрены вопросы корректности применения контрольно-измерительных материалов для дисциплин профиля бакалаврской подготовки и дисциплин магистерских программ. Описаны статистики для анализа данных при малых объемах выборок. Проведен анализ результатов тестирования ряда групп. Разработан алгоритм применения контрольно-измерительных материалов для небольших групп студентов.

Обозначения

H_0 – гипотеза о принадлежности двух выборок одной и той же генеральной совокупности;	U – значение критерия Манна и Уитни;
n_1 – объем выборки 1;	μ – среднее значение;
n_2 – объем выборки 2;	\tilde{y} – медиана.
R – сумма рангов;	
T – большая сумма рангов;	

Индексы

эмп – эмпирический;
кр – табличное значение критерия.

Объективность оценки качества образования необходима для осуществления управления как на уровне вуза, так и на уровне министерства. Если для дисциплин федеральной компоненты (унифицированной части учебных планов) возможна разработка универсальных тестов для всей страны, то для дисциплин, составляющих специфику кафедры, контрольно-измерительные материалы (**КИМ**) должны разрабатываться непосредственно авторами дисциплины и верифицироваться на уровне вуза. Для формирования объективных показателей, характеризующих деятельность вуза и качество подготовки выпускников, применяются методы тестирования [1] и статистической обработки [2]. Данная статья посвящена разработке алгоритма верификации КИМ и оцениванию знаний студентов малочисленных групп, что характерно для дисциплин профиля бакалаврской подготовки и дисциплин магистерских программ.

С точки зрения технических измерений, термин «измерительные материалы» нельзя считать применительно к КИМ абсолютно корректным, именно из-за отсутствия меры, по которой можно их «откалибровать». Тем не менее, в литературе встречаются термины «психофизические измерения» и «психологические измерения». Из-за отсутствия меры при обработке данных психологических исследований используется либо сравнительный анализ по нескольким группам, либо

результаты долговременных наблюдений за одной группой. Одной из проблем является нерепрезентативность выборки, на которой КИМ по дисциплинам специальности будут применяться. Численность учебных групп колеблется от 12 до 22 человек и даже если тестирование проводится в 2–3 группах, для набора статистики этого недостаточно.

В ходе тестирования необходимо оценить:

– адекватность КИМ, разрабатываемых в вузе, содержанию и требуемому уровню освоения материала;

– качество подготовки студентов;

– качество работы преподавателя.

Калибровка КИМ может быть: внутривузовской, перекрестной.

Внутривузовские мероприятия по оцениванию КИМ включают:

– экспертизу на уровне кафедр;

– тестирование в группе преподавателей;

– тестирование в группе студентов.

Целями внутривузовских мероприятий по оцениванию КИМ являются: подтверждение воспринимаемости вопросов; выявление некорректно сформулированных вопросов.

По каждому вопросу и по тесту в целом рассчитываются математическое ожидание и среднеквадратическое отклонение. Главная проблема – малые выборки.

Перекрестное тестирование КИМ может быть организовано в целях увеличения репрезентативности выборки и получения более высокой достоверности оценки, однако, вследствие того что вузы в настоящее время конкурируют друг с другом за абитуриентов и финансирование, в результатах внешнего оценивания не всегда можно быть абсолютно уверенным.

Контрольно-измерительные материалы по специализированным дисциплинам состоят, как правило, из двух видов вопросов:

1) общетеоретические сведения по изучаемой дисциплине;

2) специфические сведения (переменная часть, актуализируемая, зависящая от уровня развития техники, технологий и элементной базы).

Возможностей по проверке первой части вопросов значительно больше. Особенность контроля знаний по дисциплинам профиля бакалаврской подготовки и дисциплинам магистерских программ состоит в том, что такие дисциплины обеспечиваются одним – двумя преподавателями, что приводит к некоторой субъективности оценивания. Проверка гипотез о математических ожиданиях и дисперсиях призвана дать объективный инструмент для оценивания уровня подготовки студентов, качества преподавания и обучения студентов.

Проведем классификацию существующих критериев проверки гипотез о математических ожиданиях. Выберем признаки классификации:

– равенство дисперсий двух и более выборок;

– принадлежность выборки нормальному распределению.

Анализ тестового метода контроля знаний с применением математической статистики для оценивания надежности тестов проведен в работе [3]. Приводимые соотношения опираются на ряд гипотез, одна из которых гласит, что высокий уровень способностей встречается редко, причем способности у людей распределяются в соответствии с законом Гаусса. Большинство критериев можно применять только при нормальном распределении выборки.

Рассмотрим результаты тестирования нескольких групп по дисциплине бакалаврской подготовки «Микропроцессорные устройства в приборостроении». Максимально возможный результат – 32 балла. Полученные результаты приведены в табл. 1. Исходя из гипотезы о принадлежности результатов к одной генеральной совокупности, построена гистограмма (рис. 1).

Предположение о нормальности распределения спорно, скорее речь может идти о равномерном или двумодальном распределении. Для проверки гипотезы

Таблица 1

Результаты тестирования групп 2011 – 2013 гг.

Группа / тест	Объем выборки	Минимальное значение	Максимальное значение	Математическое ожидание	Среднеквадратичное отклонение
2011 г.:					
1/1	17	0	30	13,8	9,9
1/2	17	0	32	18,9	11,2
2/1	9	1	28	16,2	7,4
2/2	12	1	24	13,8	7,9
2012 г.:					
3/1	20	0	18	6,1	5,3
3/2	21	0	24	12,0	8,3
2013 г.:					
4/1	18	3	22	10,4	6,6
4/2	18	7	27	18,9	5,3

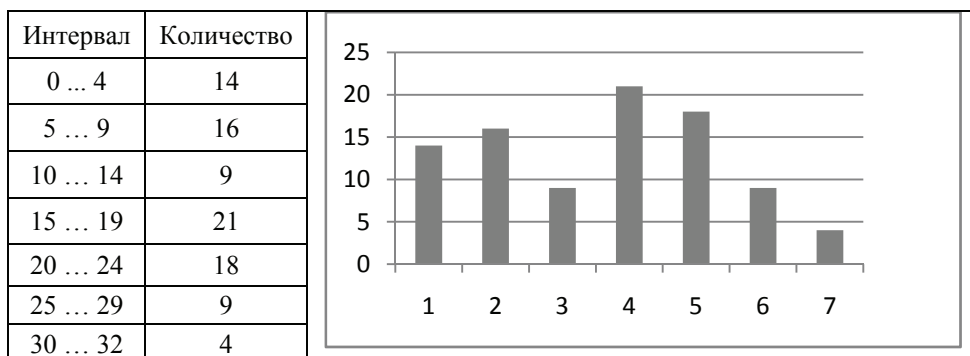


Рис. 1. Распределение результатов тестирования

о равенстве математических ожиданий при распределении, отличном от нормального, и неравных объемах выборок $n_1 \neq n_2$, можно использовать критерий Манна и Уитни [4, 5].

Ранговый критерий Манна и Уитни основан на критерии Уилкоксона для независимых выборок. Он является непараметрическим аналогом t -критерия для сравнения двух средних значений непрерывных распределений.

Гипотезу о принадлежности двух выборок одной и той же генеральной совокупности $H_0: F_1(x) = F_2(x)$ проверяет U -критерий Уилкоксона, Манна и Уитни. Эта гипотеза включает также равенство значений медиан $\tilde{\mu}_1 = \tilde{\mu}_2$ и равенство средних значений $\mu_1 = \mu_2$.

Для вычисления статистики упорядочивается объем $(n_1 + n_2)$ значений объединенной выборки по возрастанию и все элементы ранжируются. В случае совпадения нескольких значений, ранг такого значения следует считать как среднее арифметическое тех позиций, которые занимают данные значения при их расположении в порядке возрастания. Далее определяется сумма рангов R_1 , соответствующая элементам первой выборки, и сумма рангов второй выборки R_2 определяется большая из ранговых сумм T_x . Вычисляется эмпирическое значение $U_{\text{эмп}}$

$$U_{\text{эмп}} = n_1 n_2 + \frac{n_x(n_x + 1)}{2} - T_x,$$

где n_x – количество значений в выборке с большей суммой рангов.

При заданном уровне значимости находится табличное значение $U_{\text{кр}}$, и проверяется неравенство $U_{\text{эмп}} > U_{\text{кр}}$, если оно истинно, то гипотеза H_0 принимается, в противном случае – отвергается, и различие математических ожиданий определяется как существенное.

Результаты проверок по двум параллельным группам и сравнения групп двух лет приведены в табл. 2. На критерии Уилкоксона основан непараметрический критерий для проверки расхождения дисперсий. Проверяемая гипотеза H_0 заключается в том, что две независимые выборки принадлежат к общей генеральной совокупности с одинаковыми характеристиками рассеяния [3].

В работе [5] говорится, что в случае возрастания различия между средними значениями двух выборок растет вероятность ошибки второго рода для данного критерия. В то же время критерий чувствителен к разнице дисперсий при равных параметрах положения.

При использовании критерия объединенная выборка объемом $(n_1 + n_2)$ при $n_1 \leq n_2$ упорядочивается (строится вариационный ряд). Ранги элементам такой выборки присваиваются следующим образом: наименьшее значение получает ранг 1, два наибольших значения получают ранги 2 и 3, ранги 4 и 5 получают следующие наименьшие значения, 6 и 7 – следующие наибольшие значения и т.д. Если число наблюдений нечетно, то среднее наблюдение не получает никакого ранга, если четное – оно получает наивысший ранг.

Для каждой выборки определяют сумму рангов R_1 и R_2 . При $n_1 = n_2$ проверяемой гипотезе H_0 соответствует соотношение $R_1 \approx R_2$. Чем больше отличаются R_1 и R_2 , тем больше выборки отличаются по своим дисперсиям. Для оценки разности при малых выборках ($n_1 \leq n_2 \leq 20$) известны точные критические значения [6]. Для выборок ($n_1 > 9$ и $n_2 > 9$) или ($n_1 > 2$ и $n_2 > 20$) используют статистику

Таблица 2

Результаты проверок групп (2011–2012 гг.)

Группа/тест	Объем выборки	Сравнение математических ожиданий по критерию Манна и Уитни
1(2011 г.)/1	17	Расхождение незначимо
2(2011 г.)/1	9	
1(2011 г.)/2	17	
2(2011 г.)/2	12	
1(2011 г.)/1	17	Неопределенность
3(2012 г.)/1	20	
2(2011 г.)/1	9	Расхождение значимо
3(2012 г.)/1	20	
1(2011 г.)/2	17	
3(2012 г.)/2	21	
2(2011 г.)/2	12	Расхождение незначимо
3(2012 г.)/2	21	

$$z = \frac{2R_1 - n_1(n_1 + n_2 + 1) + 1}{\sqrt{n_1 n_2 (n_1 + n_2 + 1) / 3}},$$

где R_1 – сумма рангов меньшей выборки, которая приближенно подчиняется нормальному закону. Если $2R_1 > n_1(n_1 + n_2 + 1)$, то в выражении для статистики $+1$ заменяют на -1 [6].

Для оценки адекватности КИМ и качества обучения реализуется алгоритм, приведенный на рис. 2.

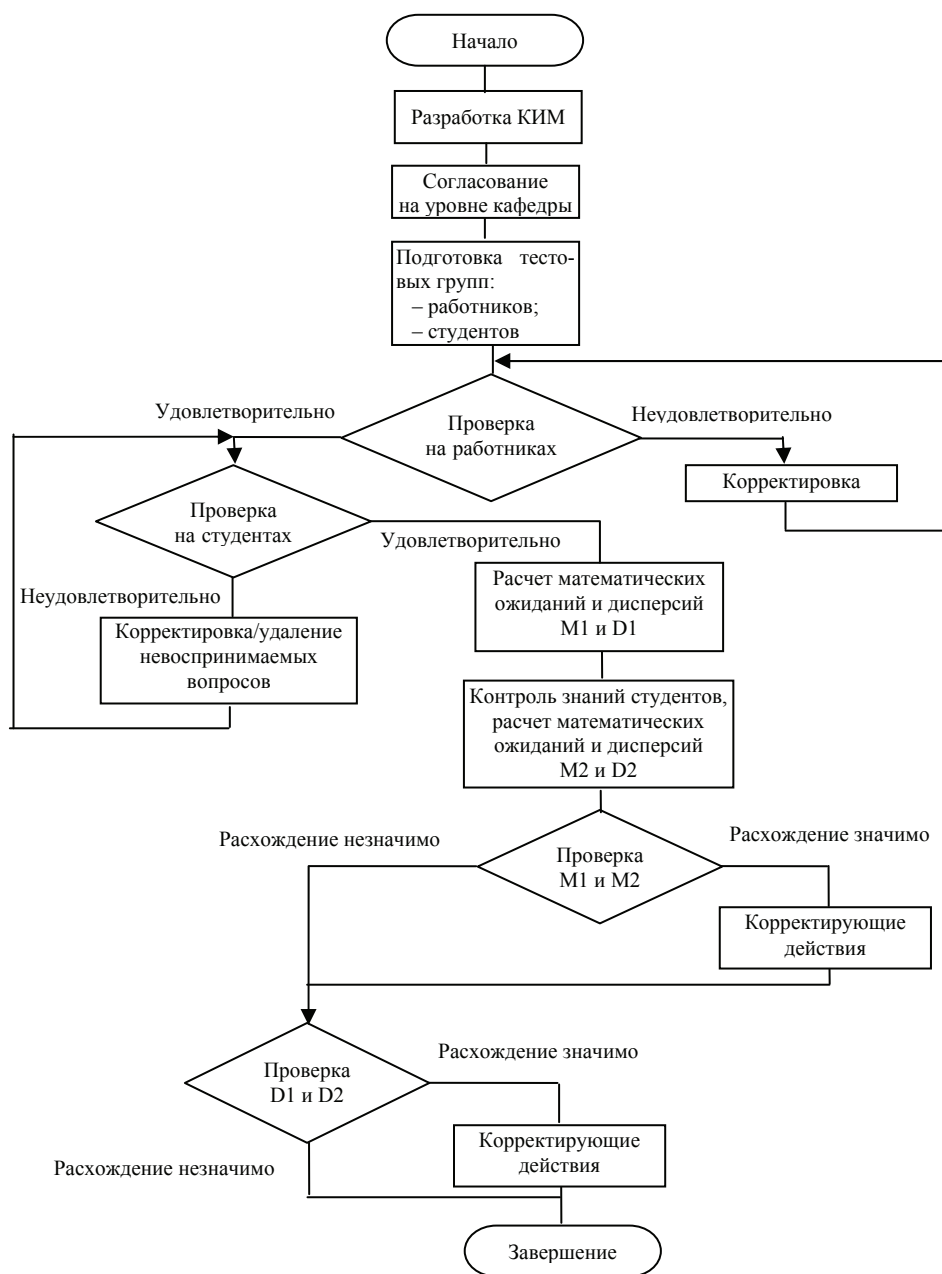


Рис. 2. Алгоритм применения КИМ для дисциплин специализации

Подтвержденная или опровергнутая гипотеза о равенстве математических ожиданий и дисперсий для оценивания результатов учебного процесса показывает, что:

1) незначимые расхождения математических ожиданий и дисперсий результатов учебных групп одного потока (при приемлемом среднем значении) позволяют сделать предположение о соответствии качества преподавания требованиям;

2) различие математических ожиданий результатов учебных групп одного потока в случае, если лабораторные или практические занятия велись разными преподавателями, и при известной априорной информации о равенстве предшествующих результатов обучения, говорит о необходимости корректирующих действий в этих видах занятий. Если группы и ранее различались по результативности, корректирующих действий не требуется;

3) различие математических ожиданий результатов в группах разных лет (см. табл. 2, значимые различия между гр. 1 и 3 и гр. 2 и 3) позволяет сделать вывод об ухудшении результатов.

4) различие дисперсий результатов групп одного потока предполагает наличие слабо подготовленных студентов, что позволяет предпринять корректирующие действия на стадии отбора обучающихся и формирования учебных групп.

Таким образом, показана возможность оценивания качества подготовки студентов с применением КИМ и методов математической статистики, ориентированных на применение для малых выборок.

Список литературы

1. Алексеев, В. В. Электронно-лабораторный практикум для учебных дисциплин направления «Приборостроение» / В. В. Алексеев, П. Г. Королев, В. В. Поливанов // Вестн. метролог. акад. – 2009. – Спец. вып. – С. 82 – 89.

2. Попов, Г. В. Квалиметрические подходы к оценке качества подготовки специалиста / Г. В. Попов, Т. В. Забегалина, Л. И. Назина // Вестн. Тамб. гос. техн. ун-та. – 2009. – Т. 15, № 2. – С. 422 – 428.

3. Карпенко, А. П. Тестовый метод контроля качества обучения и критерии качества образовательных тестов. Обзор [Электронный ресурс] / А. П. Карпенко, А. С. Домников, В. В. Белоус // Наука и образование : электрон. науч.-техн. издание. – Режим доступа : <http://technomag.edu.ru/doc/184741.html>. – Загл. с экрана.

4. Лемешко, Б. Ю. Проверка гипотез о математических ожиданиях и дисперсиях в задачах метрологии и контроля качества при вероятностных законах, отличающихся от нормального / Б. Ю. Лемешко, С. С. Помадин // Метрология. – 2004. – № 3. – С. 3 – 15.

5. Закс, Л. Статистическое оценивание / Л. Закс ; пер. с нем. В. Н. Варыгина ; под ред. Ю. П. Адлера, В. Г. Горского. – Москва : Статистика, 1976. – 598 с.

6. Siegel, S. A Nonparametric Sum of Ranks Procedure for Relative Spread in Unpaired Samples / S. Siegel, J. W. Tukey // J. Amer. Statist. Assoc. – 1960. – Vol. 55. – P. 429 – 455.

Test Materials Verification and Students' Examination

V. V. Alekseyev, P. G. Korolev, A. V. Utushkina

*Department "Information-Measuring Systems and Technologies",
St. Petersburg State Electrotechnical University "LETI"
named after V. I. Ulyanov (Lenin); PGKorolev@gmail.com*

Key words and phrases: expectation; monitoring and measuring equipment; small sample; statistical hypothesis; variance.

Abstract: The article is devoted to the correctness of the application of test materials for specialization disciplines. The statistics to analyze the data for small sample sizes is described. The test results of a number of groups have been analyzed. An algorithm for the use of test materials for small groups of students is developed.

Verifizierung der Kontrollmessmaterialien und Prüfung des Wissens der Studenten

Zusammenfassung: Es sind die Fragen der Korrektheit der Anwendung der Kontrollmessmaterialien für die Disziplinen der Spezialisierung betrachtet. Es sind die Statistiken für die Analyse der Daten bei den kleinen Umfängen des Abrufes beschrieben. Es ist die Analyse der Ergebnisse der Prüfung der Reihe der Gruppen durchgeführt. Es ist der Algorithmus der Anwendung der Kontrollmessmaterialien für die kleinen Gruppen der Studenten ausgearbeitet.

Vérification des matériaux de test et contrôle des connaissances des étudiants

Résumé: Sont examinés les problèmes de la correction de l'application des matériaux de test pour les disciplines de spécialisation. Sont décrites les statistiques pour l'analyse des données avec de petits volumes de choix. Est réalisée une analyse des résultats de test de certains groupes. Est élaboré l'algorithme de l'emploi des matériaux de test pour de petits groupes d'étudiants.

Авторы: *Алексеев Владимир Васильевич* – доктор технических наук, профессор, заведующий кафедрой информационно-измерительных систем и технологий; *Королёв Павел Геннадьевич* – кандидат технических наук, доцент кафедры информационно-измерительных систем и технологий; *Утушкина Анна Вячеславовна* – аспирант кафедры информационно-измерительных систем и технологий, СПбГЭТУ «ЛЭТИ», г. Санкт-Петербург.

Рецензент: *Юлдашев Зафар Мухамедович* – доктор технических наук, профессор, заведующий кафедрой биотехнических систем, СПбГЭТУ «ЛЭТИ», г. Санкт-Петербург.
