

МИНОБРНАУКИ РОССИИ



Орский гуманитарно-технологический институт (филиал)  
федерального государственного бюджетного образовательного  
учреждения  
высшего профессионального образования  
«Оренбургский государственный университет»

Факультет физико-математического и естественно-научного образования  
Кафедра математики, информатики, теории и методики обучения  
информатике

**Учебно-методическая разработка:**

**КОМПЬЮТЕРНАЯ ПРОГРАММА РАСЧЕТА  
ОСНОВНЫХ ХАРАКТЕРИСТИК ТЕСТА**

Руководитель к. п. н., доцент ка-  
федры математики, информати-  
ки, теории и методики обучения  
информатике

Абдулвелеева Рауза Рашитовна

Ведущий программист кафедры  
математики, информатики, тео-  
рии и методики обучения ин-  
форматике

Абдулвелеев Равиль Исмаилович

Студентка 2 курса факультета  
физико-математического и есте-  
ственно-научного образования,  
направление подготовки: педа-  
гогическое образование 050100;  
профиль информатика.

Воронина Наталья Викторовна

Орск 2015

## КОМПЬЮТЕРНАЯ ПРОГРАММА РАСЧЕТА ОСНОВНЫХ ХАРАКТЕРИСТИК ТЕСТА

Эффективное управление качеством обучения невозможно без оперативной информации об уровне знаний учащихся. Методическая готовность учителя включает в себя умение осуществлять компьютерную диагностику знаний, которая позволяет оптимизировать процесс оперативного определения результатов обучения, осуществлять оперативную обратную связь, корректировать имеющиеся недостатки в процессе подготовки [1].

Процедура диагностики достижений учащихся включает в себя процесс разработки диагностических материалов или теста, организацию процесса тестирования, обработку и интерпретацию результатов тестирования, анализ полученных результатов и качества диагностических заданий.

Не всегда у учителя для контроля знаний учащихся имеется нужный тест или комплекс диагностических заданий. Зачастую педагогу необходимо самостоятельно разработать тестовые задания, а значит необходимо обосновать их качество, воспользовавшись современной теорией педагогических измерений.

Основными характеристиками, которые измеряются при апробации и обработке результатов тестирования являются: величина трудности заданий теста, дискриминативности, корреляции заданий, дисперсии, надежности, валидности, корреляции теста [5]. Расчет этих характеристик состоит из большого количества громоздких формул и множества промежуточных вычислений. В случае коррекции тестовых заданий и повторной апробации теста все расчеты нужно повторить. Если расчеты делать вручную с помощью калькулятора, то это весьма затруднительно и требует много времени. Поэтому возникает потребность автоматизации расчета основных статистических характеристик теста.

Цель: разработать компьютерную программу в среде Excel с использованием языка программирования Visual Basic Applications (VBA), которая позволит значительно упростить процесс оценки качества теста.

Задачи:

1. Разработать инструментарий для выявления качества диагностических материалов и автоматизировать процесс его оценки.
2. Обеспечить представление педагога о теории педагогических измерений и ее применении в оценке качества разработанного теста.
3. Повысить профессиональную культуру и компетентность педагога в области современных средств оценивания результатов обучения.
4. Развить умение педагога применять информационные технологии в профессиональной деятельности.

В педагогике под понятием "тестирование" понимают методы исследования, с помощью которых конечный итог учебного процесса, может быть измерен, обработан и интерпретирован. Основная цель тестирования - использование результатов измерений в педагогической практике.

Проводить тестирование можно как над одним или несколькими испытуемыми, так и с несколькими группами людей. В педагогической практике чаще всего применяют групповые тесты, так как они являются более удобными для школьной практики в системе классно-урочной формы обучения. Использование индивидуальных тестов происходит тогда, когда исследуемое нельзя измерить в условиях проверки группы людей. Существовавшая гипотеза о том, что при индивидуальном тестировании можно достичь наиболее высоких результатов не подтвердилась. В наши дни групповое тестирование проводится чаще, чем индивидуальное, однако уровень развития компьютерной техники, программного обеспечения, сетевых и интернет технологий позволяют эффективно организовать индивидуальное тестирование.

Процесс информатизации процесса обучения позволил повысить возможности тестирования. Современные тестовые компьютерные программы и тестирующие системы имеют расширенные возможности в определении режима тестирования, режима работы программы. Они позволяют задать жесткий режим тестирования: запретить возвращаться к предыдущим заданиям, чтобы исправить ответ, ограничить время тестирования, ограничить количество попыток тестирования, осуществить случайный выбор определенного количества заданий по теме или разделу из обширной базы заданий, запретить выход из программы до окончания тестирования, отключить доступ к рабочему столу и панели задач.

Качество разработанного теста определяется значениями существенных показателей. Основными показателями качества теста являются коэффициенты надежности и валидности тестовых заданий.

Надежность теста является одним из критериев качества теста и показывает, насколько точно измеряет данный тест изучаемое явление, его «помехоустойчивость».

Точность теста как измерительного инструмента, прочность его к действию помех (состояния испытуемых, их отношения к процедуре тестирования, случайным ошибкам и т.п.) определяется коэффициентом надежности.

Тест будет надежным, если он является внутренне согласованным (гомогенным, консистентным). На практике применяются три ведущих метода оценивания надежности тестов [4]:

1. повторное тестирование (ре-тестирование) определение помехоустойчивости;
2. расщепление группы определение помехоустойчивости;
3. расщепление теста определение внутренней согласованности.

В большинстве случаев надежность определяется как коэффициент. Источники неудовлетворительной надежности тестов могут быть распре-

делены на три группы связанные с качеством теста, с процедурой его проведения и оценивания и связанные с испытуемыми.

Качество тестовых заданий может значительно повысить надежность тестов. Этому способствует увеличение длины теста, проверка объективности критериев оценки, снижение вероятности угадывания правильных ответов, ужесточение инструкций для учащихся, повышение качества инструктажа и подготовки ведущих. При определении надежности теста следует обратить внимание на критериально-ориентированные тесты. Такие тесты могут целиком состоять из заданий, с которыми справляются все учащиеся и это позволяет педагогу констатировать тот факт, что тему или раздел усвоили все учащиеся, или не справляется никто из учащихся. Тогда значение показателя надежности будет иметь неудовлетворительные значения.

Еще одной важнейшей характеристикой теста является его валидность. Валидность особенно важна для критериально-ориентированных тестов, так как определение надежности для этих тестов затруднено. Валидность и надежность — связанные понятия.

Валидность определяет, насколько тест отражает то, что он должен оценивать. Отметим несколько особенностей определения валидности [4]:

1. Валидность получается из экспертных оценок (не измеряется статистическими методами).
2. Валидность выражается степенью (высокая, средняя, низкая).
3. Валидность специфична для каждого конкретного использования.
4. Существует много способов определения валидности.

Часть показателей валидности может быть определена статистическими методами — корреляционным и факторным анализом. Однако судить только по ним о валидности теста некорректно. Для оценки валидности должен быть рассмотрен комплекс возможных показателей.

На современном этапе развития тестологии классификация валидности весьма условна. Этот факт определяется сложностью феномена валидности. Для различных видов валидности могут быть использованы одни и те же методы определения и, наоборот, одни и те же данные могут быть интерпретированы с точки зрения разных типов валидности.

Опишем процесс апробации теста и выявления его качества с помощью компьютерной программы [3]. Разработанный педагогом тест предлагается испытуемым. Форма предъявления заданий испытуемым может быть различной: в бумажном варианте или в виде тестовой компьютерной программы. Результаты выполнения тестовых заданий заносятся в матрицу результатов теста созданную в Excel.

Существует большое количество различных тестовых программ и оболочек для создания тестов. Для предъявления заданий испытуемым нами использовалась тестовая оболочка AD Tester. В программу в режиме конструктора вносятся разработанные тестовые задания, а затем в режиме тестирования испытуемые выполняют задания теста. В этой программе есть хорошая возможность – «экспорт результатов выполнения заданий теста в Excel». Таким образом, упрощается процесс переноса результатов выполнения тестов испытуемыми в матрицу результатов. Это производится простым копированием ячеек. Далее внесенные значения используются для обработки и расчета различных параметров теста. Эти параметры и промежуточные характеристики рассчитываются методом ввода формул в соответствующие ячейки таблицы. Для расчета различных параметров теста использовались формулы содержащиеся в методике разработанной и предложенной М.Б. Челышковой [5]. Полученные промежуточные результаты позволяют рассчитать валидность и надежность теста, которые дают возможность выявить необходимость внесения корректив в содержание тестовых заданий (заменить, усложнить или облегчить).

Тестовые задания, удовлетворяющие "внешним" формальным требованиям (правильность формулировки, логическая непротиворечивость,

адекватная форма и др.) не всегда являются качественными, позволяющими рассматривать такие задания априори или апостериори как эффективное средство оценки уровня подготовки обучаемых. Тестовые задания (тесты) измеряют свойства тех или иных выборок потенциальных испытуемых. Экспертиза качества заданий и всего теста – это оценка не только каждого его элемента, но и всего теста, его структуры, взаимоотношений с другими заданиями.

Главное окно программы автоматизации расчета характеристик теста показано на рисунке 1.

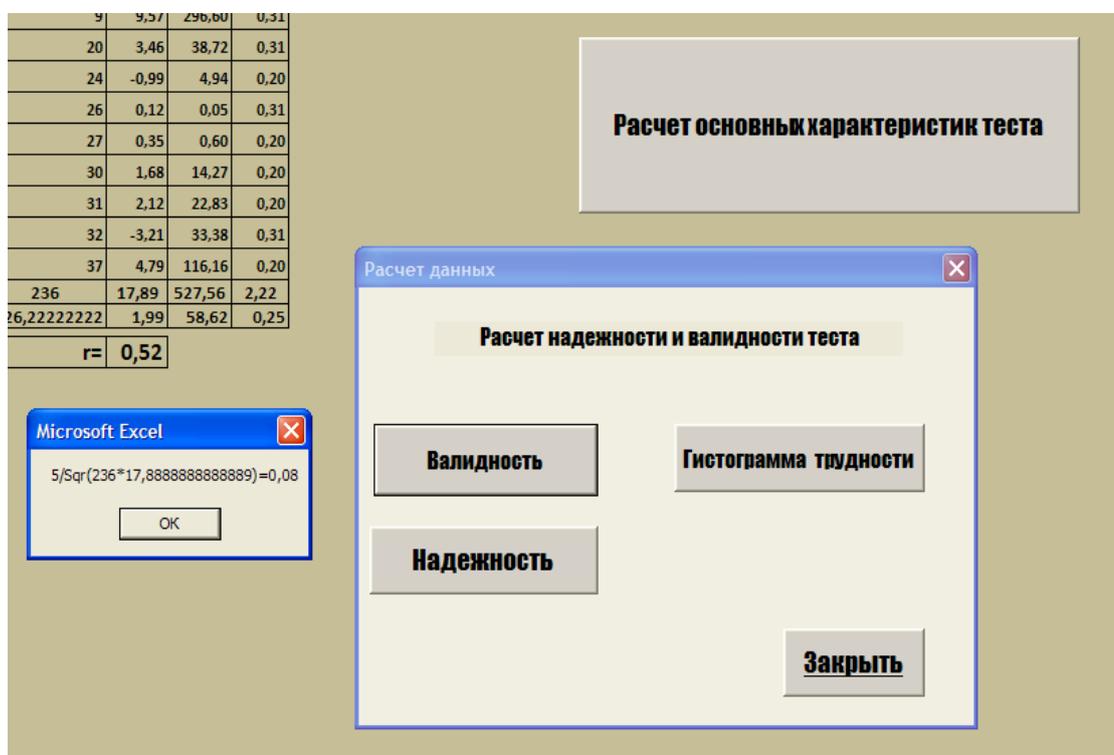


Рисунок 1 - Главное окно программы

Окно программы содержит кнопки расчета основных характеристик теста. Результаты расчета валидности, надежности теста, гистограмма трудности заданий теста выводятся на экран нажатием соответствующих кнопок.

Опишем методику расчета основных характеристик теста.

Одним из этапов апробации разработанного теста является расчет валидности и надежности теста. Сначала формируют результирующую матрицу ответов испытуемых в табличном процессоре Excel (рис.2), в ко-

торой строки соответствуют испытуемым, столбцы – заданиям теста, а в ячейки таблицы экспортируются ответы испытуемых, задаваемые в некоторой шкале. В разработанной компьютерной программе используется дихотомическая шкала (0 – не выполнено, 1 – выполнено).

Повторная обработка результатов тестирования в экспериментальной группе																												
#	ФИО	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	Студент 1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	1	0	1
2	Студент 2	1	1	1	0	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	0	1	0	1	1	0
3	Студент 3	1	1	1	0	1	1	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0
4	Студент 4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
5	Студент 5	1	0	1	0	0	1	0	1	1	1	1	1	1	0	0	0	1	1	0	1	0	1	0	1	1	1	0
6	Студент 6	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	0	1	0	1	0	0	1	0	0	0
7	Студент 7	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
8	Студент 8	1	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
9	Студент 9	1	0	0	0	1	1	0	1	0	1	1	0	1	1	0	1	1	1	0	0	1	0	1	1	0	1	0
число правильных ответов		9	6	8	3	7	8	3	7	6	8	7	7	8	7	5	6	7	8	5	6	4	5	4	4	6	5	4

Упорядочная матрица данных тестирования																												
#	ФИО	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
17	Студент 4	1	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
18	Студент 1	0	0	0	0	1	1	0	1	0	1	1	0	1	1	0	1	1	1	0	1	1	0	0	1	1	0	1
19	Студент 2	0	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	0	1	0	1	0	0	1	0	0	0
20	Студент 7	1	0	1	0	0	1	0	1	1	1	1	1	1	0	0	0	1	1	0	1	0	1	0	1	1	1	1
21	Студент 5	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0
22	Студент 3	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1
23	Студент 6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
24	Студент 9	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0

Рисунок 2 – Матрица результатов

В матрицу результатов вносят данные: ФИО испытуемых, количество заданий, результат ответа каждого испытуемого на каждое задание. Количество правильных ответов по каждому вопросу рассчитывают суммированием единиц по столбцам (рис. 3), количество верных ответов испытуемым по всем вопросам рассчитывают суммированием единиц по строкам (рис. 4).

#	ФИО	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Студент 1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1
2	Студент 2	1	1	1	0	1	1	0	1	0	1	1	1	1	1	0	1
3	Студент 3	1	1	1	0	1	1	0	1	0	1	0	1	1	1	1	1
4	Студент 4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	Студент 5	1	0	1	0	0	1	0	1	1	1	1	1	1	0	0	0
6	Студент 6	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0
7	Студент 7	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
8	Студент 8	1	1	1	0	0	0	0	1	0	0	0	0	0	1	0	0
9	Студент 9	1	0	0	0	1	1	0	1	0	1	1	0	1	1	0	1
	число правильных ответов	9	6	8	3	7	8	3	7	6	8	7	7	8	7	5	6

Рисунок 3 – Расчет числа правильных ответов испытуемых

Для подсчета общего количества правильных ответов вначале суммируют количество верных ответов каждого испытуемого (сумма единиц по строке) как показано на рисунке 4, и затем суммируют количество верных ответов всех испытуемых. В последнем столбце, получаем результат суммы.

#	ФИО	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	Итого
1	Студент 1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	1	1	0	1	0	1	0	1	0	1	1	1	1	1	1	1	30
2	Студент 2	1	1	1	0	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1	28	
3	Студент 3	1	1	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	1	0	1	1	1	1	0	0	1	0	1	25		
4	Студент 4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	35			
5	Студент 5	1	0	1	0	0	1	0	1	1	1	1	1	0	0	0	1	1	0	1	0	1	0	1	0	1	1	0	0	1	0	1	1	1	1	0	0	1	24		
6	Студент 6	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	0	1	0	1	0	0	1	0	0	0	0	0	0	1	1	0	1	0	0	1	21			
7	Студент 7	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	1	1	0	1	1	0	0	28			
8	Студент 8	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6			
9	Студент 9	1	0	0	1	1	0	1	0	1	1	0	1	1	1	0	1	1	0	1	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	21		
	число правильных ответов	9	6	8	3	7	8	3	7	6	8	7	7	8	7	5	6	7	8	5	6	4	5	4	4	6	5	4	2	4	5	3	7	6	5	6	2	4	2	6	215

Рисунок 4 – Расчет количества верных ответов

Далее вносят данные во вторую матрицу результатов испытуемых – упорядоченную по строкам результатов правильных ответов от общего количества заданий, от наименьшего к наибольшему значению.

Результаты каждого испытуемого упорядочим по возрастанию, то есть распределим от наименьшего результата к наибольшему (рис. 5).

	33	34	35	36	37	38	39	всего
0	0	0	0	0	0	0	0	6
0	1	0	1	1	0	0	0	18
1	0	1	1	0	0	1	1	23
1	1	1	1	1	0	1	1	29
1	1	0	0	0	1	1	1	29
0	1	1	1	1	1	1	1	34
1	0	1	1	1	1	0	0	34
1	1	1	1	1	1	1	1	35
1	1	1	0	1	1	1	1	38
6	6	6	6	5	6	6	6	246
0,78	0,67	0,67	0,67	0,56	0,67	0,67	0,67	
0,22	0,33	0,33	0,33	0,44	0,33	0,33	0,33	
0,82	0,54	0,11	0,70	0,37	0,32	0,19	0,57	
0,78	0,67	0,67	0,67	0,56	0,67	0,67	0,67	27,33
коэффициент надежности теста							pj*qj	39

Рисунок 5 – Упорядоченная по результатам матрица

Произведем расчет математико-статистических характеристик значений, таких как валидность, среднее арифметическое значение верных и неверных ответов для каждого задания [5].

Например, расчет среднего арифметического значения правильных ответов для первого задания показано на рисунке 6.

Рассчитывается среднее арифметическое значения правильных ответов следующим образом: количество правильных ответов делим на общее количество вопросов теста.

#	ФИО	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
17	Студент 4	1	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0
18	Студент 1	0	0	0	0	1	1	0	1	0	1	0	1	1	0	1	1	1	0	0	0	1
19	Студент 2	0	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1	0	1	0	0
20	Студент 7	1	0	1	0	0	1	0	1	1	1	1	1	1	0	0	0	1	1	0	1	0
21	Студент 5	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1
22	Студент 3	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1	1	1
23	Студент 6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
24	Студент 9	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
25	Студент 8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
26	число правильных ответов	7	6	8	5	7	8	5	7	6	7	7	7	8	7	5	6	7	8	6	6	6
27	(p)	0,78	0,67	0,89	0,56	0,78	0,89	0,56	0,78	0,67	0,78	0,78	0,78	0,89	0,78	0,56	0,67	0,78	0,89	0,67	0,67	0,67
28	(q)	0,22	0,33	0,11	0,44	0,22	0,11	0,44	0,22	0,33	0,22	0,22	0,11	0,22	0,22	0,44	0,33	0,22	0,11	0,33	0,33	0,33
29	валидность r <sub>т</sub>	0,25	0,02	0,29	0,52	0,61	0,80	0,44	0,68	0,05	0,80	0,57	0,82	0,80	0,09	0,58	0,61	0,40	0,10	0,52	0,79	0,38
30	ср знач (необх)	0,78	0,67	0,89	0,56	0,78	0,89	0,56	0,78	0,67	0,78	0,78	0,78	0,89	0,78	0,56	0,67	0,78	0,89	0,67	0,67	0,67

Рисунок 6 – Среднее арифметическое правильных ответов

Среднее арифметическое неправильных ответов для каждого задания находим по формуле:  $q_i = (1 - p_i)$ . Исходя из исходных данных, например, для ячейки E28 это значение получено так:  $1 - 0,89 = 0,11$  (рис.6).

#	ФИО	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
17	Студент 4	1	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0
18	Студент 1	0	0	0	0	1	1	0	1	0	1	1	0	1	1	0	1	1	1	0	0	1
19	Студент 2	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	0	1	0	0	0
20	Студент 7	1	0	1	0	0	1	0	1	1	1	1	1	0	0	0	1	1	0	1	0	0
21	Студент 5	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
22	Студент 3	1	1	1	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1
23	Студент 6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
24	Студент 9	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
25	Студент 8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
26	число правильных ответов	7	6	8	5	7	8	5	7	6	7	7	8	7	5	6	7	8	6	6	6	6
27	(p)	0,78	0,67	0,89	0,56	0,78	0,89	0,56	0,78	0,67	0,78	0,78	0,89	0,78	0,56	0,67	0,78	0,89	0,67	0,67	0,67	
28	(q)	0,22	0,33	0,11	0,44	0,22	0,11	0,44	0,22	0,33	0,22	0,22	0,11	0,22	0,44	0,33	0,22	0,11	0,33	0,33	0,33	
29	валидность r	0,29	0,02	0,29	0,52	0,61	0,80	0,44	0,68	0,05	0,80	0,57	0,82	0,80	0,09	0,58	0,61	0,40	0,10	0,52	0,79	0,38
30	ср знач (необх)	0,78	0,67	0,89	0,56	0,78	0,89	0,56	0,78	0,67	0,78	0,78	0,89	0,78	0,56	0,67	0,78	0,89	0,67	0,67	0,67	

Рисунок 7 – Среднее арифметическое неправильных ответов

Продемонстрируем, как производится расчет основных математико-статистических значений: расчет индивидуальных баллов теста, дисперсии, среднее выборочное значение, надежность теста, валидность.

1) Размах индивидуальных баллов:

Для расчета размаха индивидуальных баллов вычитаем из наибольшего значения правильных ответов наименьшее, т.е. испытуемый 4 хуже всех справился с тестом а испытуемый 8 лучше всех, поэтому из наибольшего результата вычитаем наименьший ( $38 - 6 = 32$ ), получаем размах индивидуальных баллов.

32	Размах индивидуальных баллов	32	рассчитывается по формуле: $X_{\max} - X_{\min}$
33	Среднее выборочное	27,3	рассчитывается по формуле: $X_{\text{ср}} = (X_1 + X_2 + \dots + X_k) / K$
34	дисперсия	104	рассчитывается по формуле: $S_x^2 = \frac{K \sum_{i=1}^K X_i^2 - \left( \sum_{i=1}^K X_i \right)^2}{K(K-1)}$
35	Надежности теста	0,7	0,7-0,79 оценка надежности удовлетворительная. $KR_{20} = \frac{N}{N-1} \left( 1 - \frac{\sum_{i=1}^N p_i q_i}{S_x^2} \right)$
36	Валидность каждого вопроса (представлены в таблице выше)		рассчитывается по формуле: $r = \frac{\sum_{i=1}^K X_i Y_i - \frac{1}{K} \sum_{i=1}^K X_i \sum_{i=1}^K Y_i}{\sqrt{\sum_{i=1}^K X_i^2 - \frac{1}{K} \left( \sum_{i=1}^K X_i \right)^2} \cdot \sqrt{\sum_{i=1}^K Y_i^2 - \frac{1}{K} \left( \sum_{i=1}^K Y_i \right)^2}}$

Рисунок 8 - Размах индивидуальных баллов

2) Среднее выборочное значение.

Среднее выборочное (среднее арифметическое) значение определяется суммированием всех значений совокупности и последующим делением на их число.

$$\bar{x} = \frac{(x_1 + x_2 + \dots + x_N)}{N}$$

Пример:

$$\bar{x} = \frac{(6 + 2 + 1 + 9 + 4 + 4 + 5 + 9 + 6)}{10} = 5$$

Процесс вычисления значительно упрощается, если отдельные значения в совокупности повторяются. Сумма всех результатов определяется умножением каждого значения балла на его частоту и затем производится суммирование полученных произведений. Среднее значение будет равно сумме всех правильных ответов деленных на количество испытуемых, например, все вместе испытуемые ответили правильно на 246 вопросов, всего испытуемых 9, поэтому делим 246 на 9.  $\frac{246}{9} = 27,3$  получаем среднее выборочное (среднее арифметическое) значение.

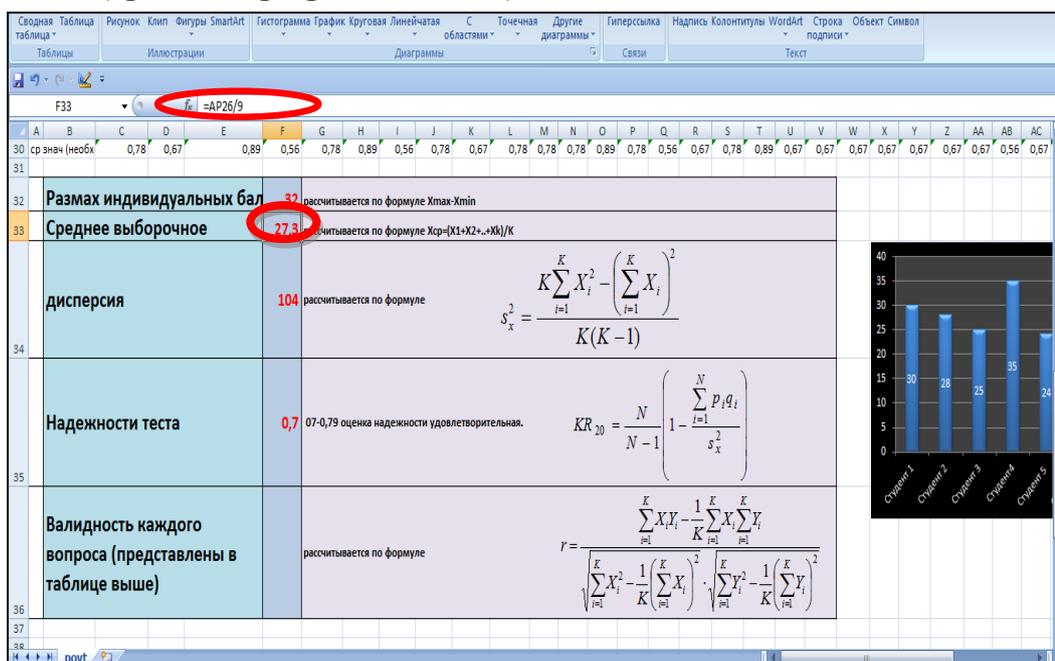


Рисунок 9 – Среднее выборочное значение

3) Дисперсия.

Подсчет значения дисперсии основан на вычислении отклонений каждого значения показателя от среднего арифметического в распределении. Для индивидуальных баллов значения отклонений  $x_i - \bar{x} (i = 1, 2, \dots, N)$  позволяют судить о вариации совокупности значений баллов  $N$  испытуемых, т.е. отражают меру неоднородности результатов по тесту. Совокупность с большей неоднородностью будет иметь большие по модулю отклонения, наоборот, для однородных распределений отклонения должны быть близки к нулю. Знак отклонения указывает место результата ученика по отношению к среднему арифметическому по тесту. Для испытуемого с индивидуальным баллом выше среднего значение разности  $x_i - \bar{x}$  будет положительно, а для тех, у кого результат ниже  $\bar{x}$ , отклонение  $x_i - \bar{x}$  меньше нуля.

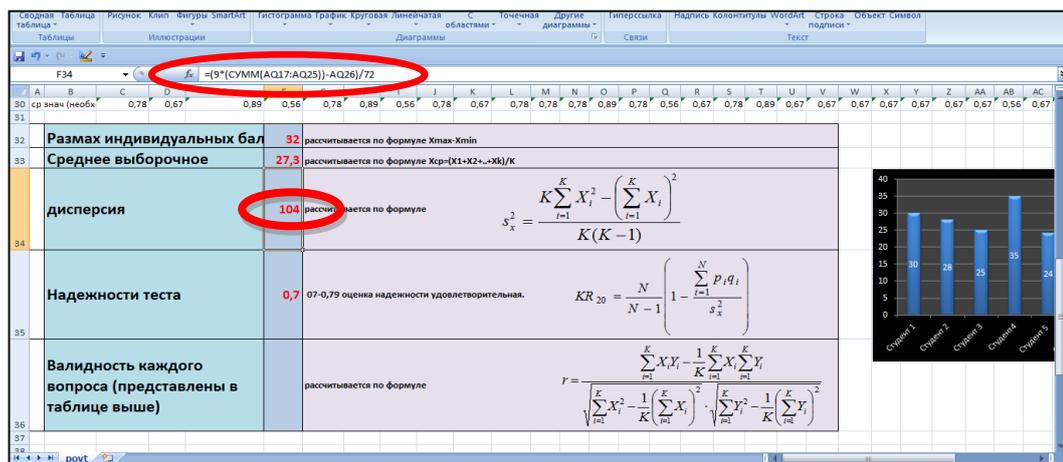


Рисунок 10 – Расчет дисперсии

### 3) Надежность теста:

Для расчета надежности теста необходимо:

Вычислить произведение среднего арифметического правильных и неправильных ответов. Умножаем строку  $p_j$  на  $q_j$  получаем 39.

Далее вычисляем по формуле:  $KR = \frac{9(1-(p_j * q_j))}{S_x}$ , где  $p_j$  - среднее арифметическое правильных ответов,  $q_j$  – среднее арифметическое неправильных ответов,  $S_x$  – дисперсия.

Отсюда следует, что  $KR = \frac{9(1-(AO31))}{F34} = \frac{9(1-39)}{104} = 0,7$

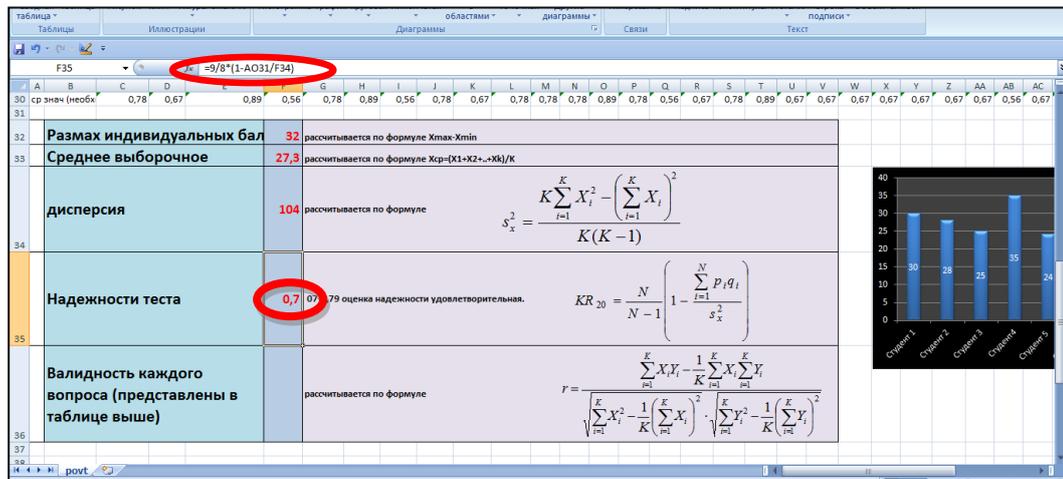


Рисунок 11 – Надежность теста

4) По результатам полученных данных строим диаграмму успеваемости студентов по предложенному им тесту.

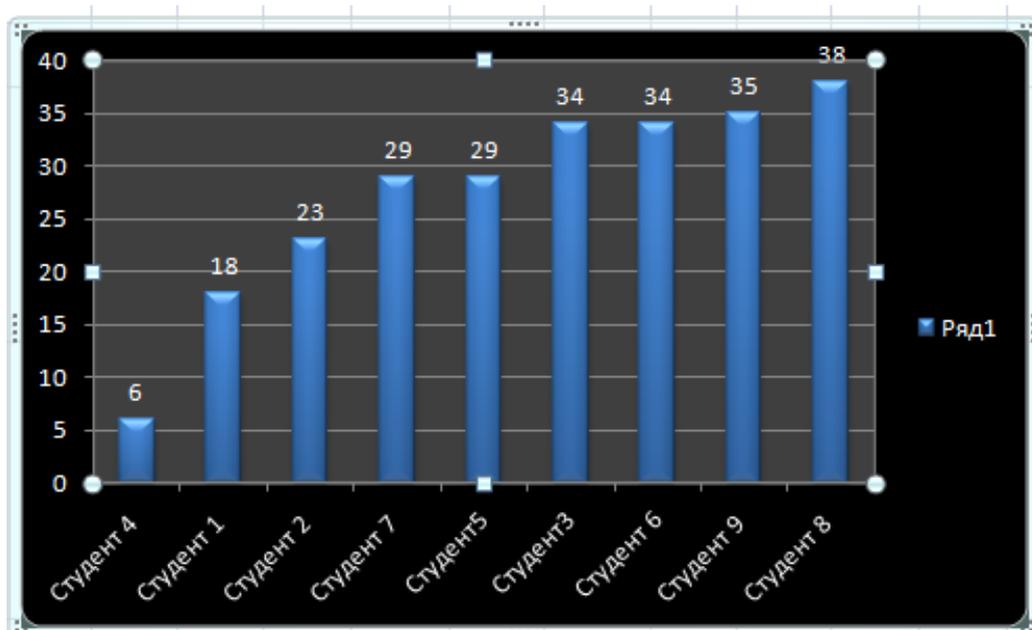


Рисунок 12 – Диаграмма успеваемости

5) Расчет валидности каждого задания:

Валидность позволяет определить способность тестовых заданий выявить то, ради чего проводилось тестирование.

Валидность каждого задания представлена в упорядоченной матрице, для нахождения валидности всего теста, строим отдельную таблицу. В данной таблице подсчитываем:

- 1) количество правильных ответов испытуемых по всему тесту.
- 2) Произведение  $(X_i - X_{sr})(Y_i - Y_{sr})$ , где  $X_i$  и  $Y_i$  – индивидуальные баллы  $i$ -го испытуемого в первом и во втором тестированиях;  $X_{sr}$  и  $Y_{sr}$  – среднее значение испытуемого в первом и во втором тестированиях.
- 3) Возводим в квадрат  $(X_i - X_{sr})$ .
- 4) Возводим в квадрат  $(Y_i - Y_{sr})$ .
- 5) Находим общую сумму и среднее значение.
- 6) Значение валидности вычисляем по формуле:

$$r_B = \frac{\sum_{i=1}^N (x_i - \bar{x})(x_{m_i} - \bar{x}_3)}{N_m \sqrt{S_x^2 \cdot S_{m_x}^2}}$$

- где  $x_i - \bar{x}$  - отклонение тестового балла  $i$ -го испытуемого от среднего балла по тесту;  $x_{m_i} - \bar{x}_3$  - отклонение балла  $i$ -го испытуемого у экспертов от  $x_3$  - среднего арифметического экспертных оценок;  $S_x^2$  – дисперсия баллов испытуемого по тесту;  $S_{m_x}^2$  - дисперсия баллов экспертов;  $m$  – число экспертов.

#	ФИО	1	всего	интервал	(х-хsr)*2	(у-usr)*2
8	Студент4	0	9	9,57	296,60	0,31
9	Студент1	0	20	3,46	38,72	0,31
6	Студент2	1	24	-0,99	4,94	0,20
5	Студент7	0	26	0,12	0,05	0,31
3	Студент5	1	27	0,35	0,60	0,20
2	Студент3	1	30	1,68	14,27	0,20
7	Студент6	1	31	2,12	22,83	0,20
1	Студент9	0	32	-3,21	33,38	0,31
4	Студент8	1	37	1,78	116,16	0,20
сумма	5	236	17,89	7,56	2,22	
среднее	0,556	26,22222222	3,578	58,62	0,25	
			г=	0,08		

Рисунок 13 – Расчет валидности теста

Для удобства пользователя в использовании программы при получении результатов расчета валидности и надежности теста используем возможности программы VBA. Создадим кнопочную форму.

Основная кнопка формы «Расчет основных характеристик теста» показана на рисунке 14.

	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
40	1	всего	количество	$(M-KSI)^2$	$(M-PI)^2$																			
41	0	9	9,57	296,60	0,31																			
42	0	20	3,46	38,72	0,31																			
43	1	24	-0,99	4,94	0,20																			
44	0	26	0,12	0,05	0,31																			
45	1	27	0,35	0,60	0,20																			
46	1	30	1,68	14,27	0,20																			
47	1	31	2,12	22,83	0,20																			
48	0	32	-3,21	33,38	0,31																			
49	1	37	4,79	116,16	0,20																			
50	5	236	17,89	527,56	2,22																			
51	0,556	26,22222222	1,99	58,62	0,25																			
52																								
53			r =	0,08																				
54																								
55																								

Рисунок 14 – Кнопка расчета характеристик теста

Вид формы в среде VBA показан на рисунке 15. В этом окне создаем форму, которая содержит кнопки «валидность», «надежность» «выход».

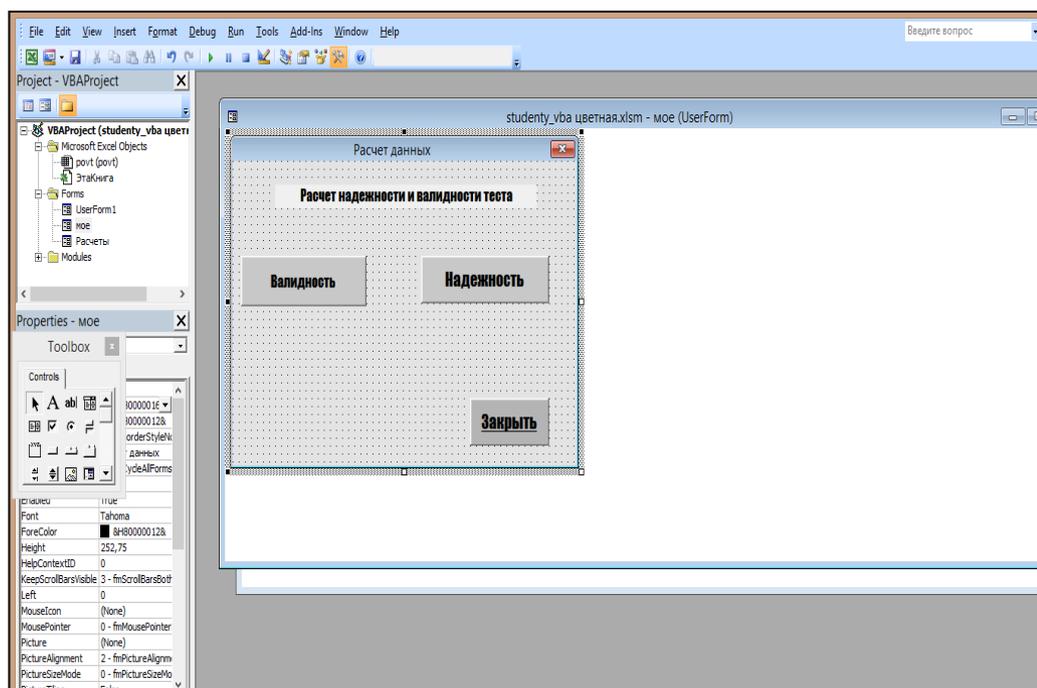


Рисунок 15 - Форма окна расчета характеристик теста

Команды для кнопок «валидность», «надежность» и «выход» представлены на рисунках 16, 17 (в виде кодов).

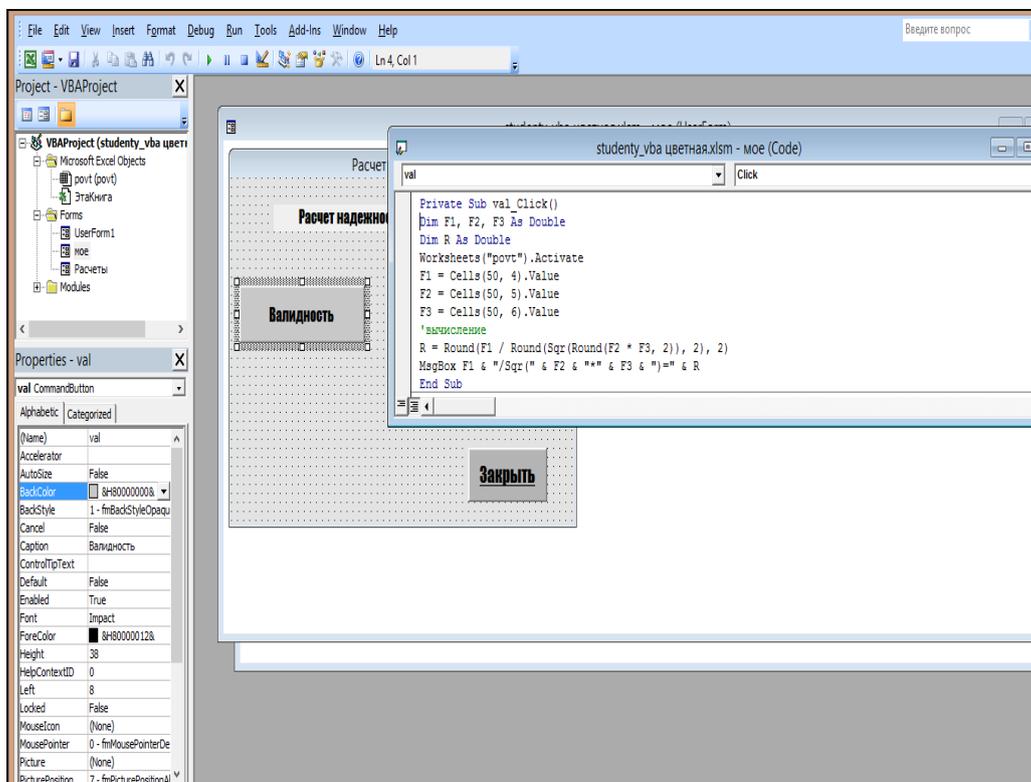


Рисунок 16 – Программа кнопки «Валидность»

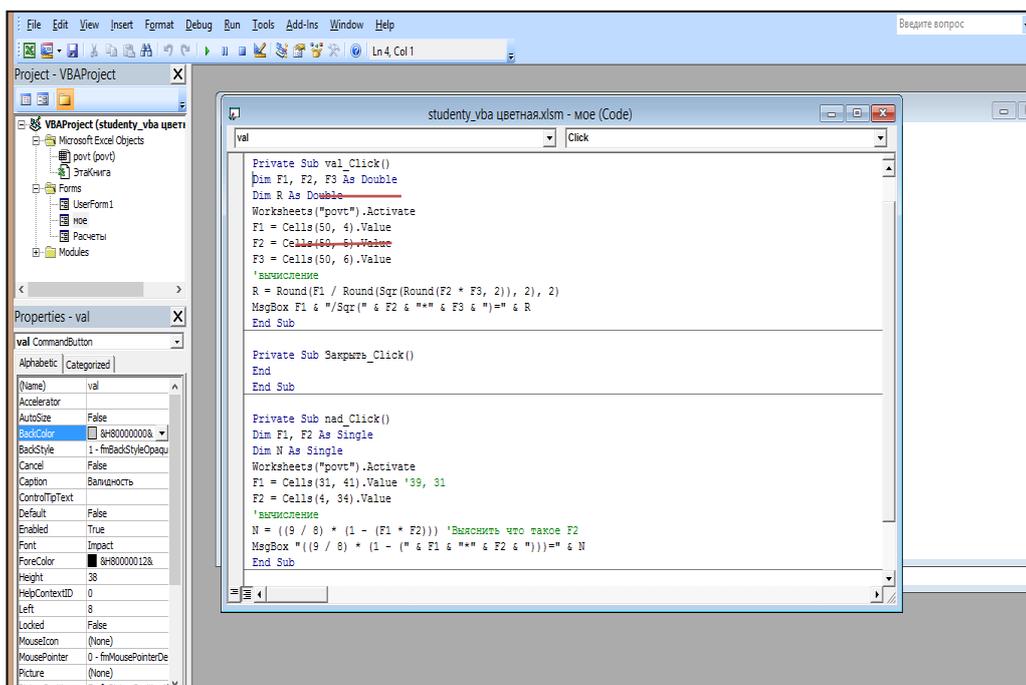


Рисунок 17 – Коды кнопок «Надежность» и «Выход»

При нажатии на кнопку «Расчет характеристик теста» Появляется окно формы «Расчет надежности и валидности теста» (рис.18).

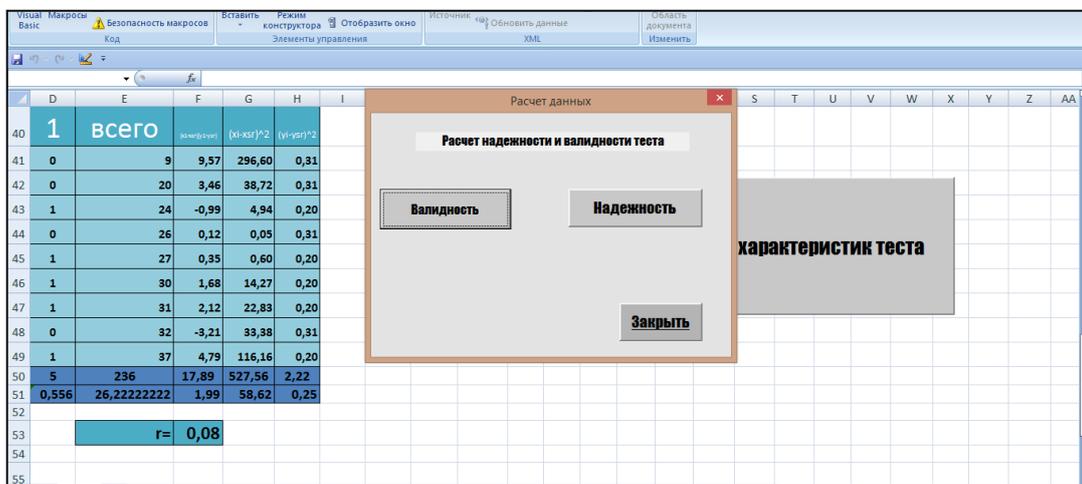


Рисунок 18 - Окно формы «Расчет надежности и валидности теста»

Результат нажатия на кнопку «валидность», показан на рисунке 19.

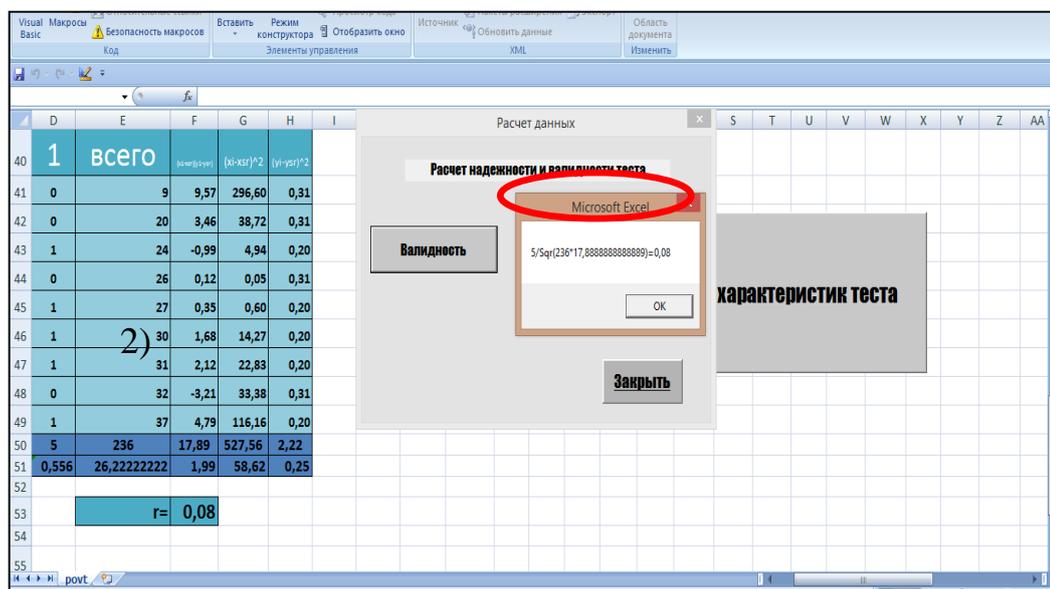


Рисунок 19 – Результат расчета валидности.

Нажатие на кнопку «надежность», выдает значение надежности теста. На рисунке 20 показан результат, и он имеет отрицательное значение. Исходя из заданных параметров (0,7-0,79), это означает, что значение надежности теста является неудовлетворительным.

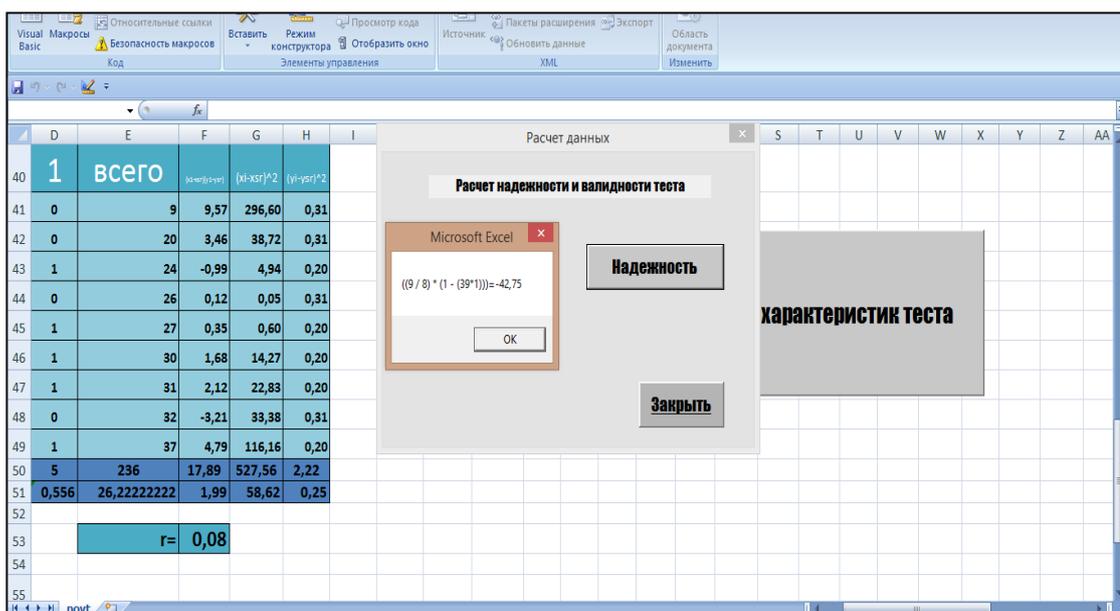


Рисунок 20 – Результат расчета надежности

Разработанная компьютерная программа, предназначена для расчета основных характеристик теста, вычисления валидности и надежности теста. В случае неудовлетворительных результатов, разработчик теста производит коррекцию содержания тестовых заданий, количества заданий и осуществляет повторную апробацию теста.

Создание таких программ для апробации результатов тестов позволит автоматизировать и повысить качество экспертизы, а также стимулировать будущих педагогов к изучению возможностей информационных технологий с целью применения в исследовательской деятельности, рефлексии своей профессиональной деятельности [2].

Таким образом, компьютерная программа позволяет, во-первых, повысить эффективность процесса апробации разработанного теста, во-вторых, обеспечить достоверное определение результатов тестирования и, в-третьих, свести к минимуму недостатки тестового контроля знаний.

Литература:

1. Абдулвелеева Р.Р. Технология моделирования компьютерной диагностики методической готовности студентов педагогического вуза // Информатизация образования и науки. – М.: ФГАУ ГНИИ ИТТ «Информика», 2014. - № 3(23). – С.42-54

2. Абдулвелеева Р.Р. Диагностика уровня развития рефлексии учебно-профессиональной деятельности студента // Сборник научных трудов SWorld. – № 1. Т.19. – Одесса, 2013. – С. 68-72.

3. Компьютерная программа оценки качества теста / Р. Р. Абдулвелеева, Воронина Н.В. // Международная научно-практическая молодежная Интернет-конференция «Инновационные взгляды научной молодежи '2015», 21-30 апреля 2015 г, Секция ПЕДАГОГИКА, ПСИХОЛОГИЯ И СОЦИОЛОГИЯ, Подсекция «Интерактивные технологии обучения и инновации в области образования» URL: <http://www.sworld.education/index.php/ru/pedagogy-psychology-and-sociology-m115/interactive-learning-technologies-and-innovations-in-education-m115/25171-m115-082>

4. Воронин К.Ю., Петров В.Ю., Правила создания и способы оценивания при конструировании тестов. Интернет-источник. URL: <http://www.scienceforum.ru/2013/pdf/8069.pdf>

5. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. – М.: Логос, 2002. – 432 с.