

Е.И. Кочергина, А.А. Юсуфов, М.В. Карпова  
ФГБОУ ВО Тверской государственный медицинский университет Минздрава РФ  
Кафедра лучевой диагностики  
**АНАЛИЗ ТЕСТА ПО КЛИНИЧЕСКОЙ РЕНТГЕНОСТОМАТОЛОГИИ В СДО  
MOODLE**

**Аннотация:** рассматривается пример использования возможности статистического анализа системы дистанционного обучения MOODLE для получения характеристик теста и заданий из банка вопросов.

**Ключевые слова:** тест, задание, оценка теста, оценка вопроса, статистические характеристики, MOODLE.

В систему дистанционного обучения (СДО) MOODLE встроена процедура статистического анализа результатов тестирования студентов [1]. По результатам ее работы можно провести статистический анализ качества отдельных вопросов и теста в целом.

Для анализа качества тестовых заданий был взят «Итоговый тест по клинической рентгеностоматологии». Банк вопросов для этого теста содержит 184 вопроса. Каждый вопрос максимально оценивается в 1 балл. Тестирование проходило для студентов 5 курса стоматологического факультета. Тест содержал 50 вопросов, по категориям не разделен.

Для получения статистических характеристик выбирается тест, справа значок «действие», затем пункт меню «статистика».

Перед выполнением статистического анализа можно указать/уточнить выборку для исследования: для всех попыток, последних, первых попыток. Результаты анализа можно передать в MS EXCEL. В процессе выполнения расчета характеристик генерируется достаточно содержательный отчет, который по содержанию делится на 2 части: первая относится к тесту в целом, вторая характеризует позицию вопроса в тесте и каждый вопрос банка вопросов, попадающих в тест.

Полученные результаты отчета представлены на рисунке 1.

Название теста	Тест – Клиническая рентгеностоматология
Название курса	Клиническая рентгеностоматология
Начало тестирования	Tuesday, 7 December 2021, 09:05
Окончание тестирования	Tuesday, 7 December 2021, 09:35
Открыть на	30 мин.
Количество полностью оцененных первых попыток	122
Общее количество полностью оцененных попыток	123
Средняя оценка первых попыток	87,95%
Средняя оценка по всем попыткам	87,94%
Средняя оценка последних попыток	88,13%
средняя оценка из лучших оцененных попыток	88,13%
Медиана оценок (для лучшая из оцененных попыток)	91,50%
Стандартное отклонение (для лучшая из оцененных попыток)	13,13%
Оценка асимметрии распределения (для лучшая из оцененных попыток)	-4,1566
Оценка распределения эксцесса (для лучшая из оцененных попыток)	21,9790
Коэффициент внутренней согласованности (для лучшая из оцененных попыток)	90,65%
Соотношение ошибок (для лучшая из оцененных попыток)	30,59%
Стандартная ошибка ( для лучшая из оцененных попыток)	4,02%

Рисунок 1. Отчет за тест.

Первая часть отчета содержит информацию о названии теста, курса, количестве попыток выполнения теста и нижеследующие статистические характеристики (в скобках указаны значения по данному тесту). Средняя оценка первых попыток (87,95%), средняя оценка по всем попыткам (87,94%), средняя оценка последних попыток (88,13%), средняя

оценка из лучших оцененных попыток (88,13%) – средняя оценка испытуемых, представляющая собой среднее арифметическое по оценкам всех выполнивших этот тест.

**Медиана оценок (91,50%)** – срединное значение оценок испытуемых.

**Стандартное отклонение (13.13%)** – мера вариации оценок, полученных при прохождении теста. Характеризует дифференцирующую способность теста. Величина дисперсии тестовых баллов позволяет судить о качестве теста, о его дифференцирующей способности [2]. Малая величина дисперсии говорит о том, что тест плохо различает испытуемых по уровню знаний, не позволяет с приемлемой точностью ранжировать их. Слишком большая дисперсия указывает на сильную неоднородность группы испытуемых, на возможные нарушения процедуры тестирования, на недостаточно ясные формулировки заданий и т.п. В случае оптимальной величины дисперсии, распределение тестовых баллов близко к нормальному.

**Оценка асимметрии распределения (-4.15)** показывает смещение максимума влево (отрицательная асимметрия) или вправо (положительная) от положения максимума на нормальной кривой (кривой нормального распределения). Положительное значение асимметрии указывает на то, что тест возможно легкий, а отрицательное значение – возможно излишне сложный.

**Оценка распределения эксцесса (20.97)** – характеризует отличие формы распределения полученных оценок за тест от нормального распределения. При положительном значении эксцесса кривая имеет более острую вершину, чем кривая нормального распределения, а при отрицательном наоборот.

На рисунке 2 показано распределение оценок исследуемого теста.

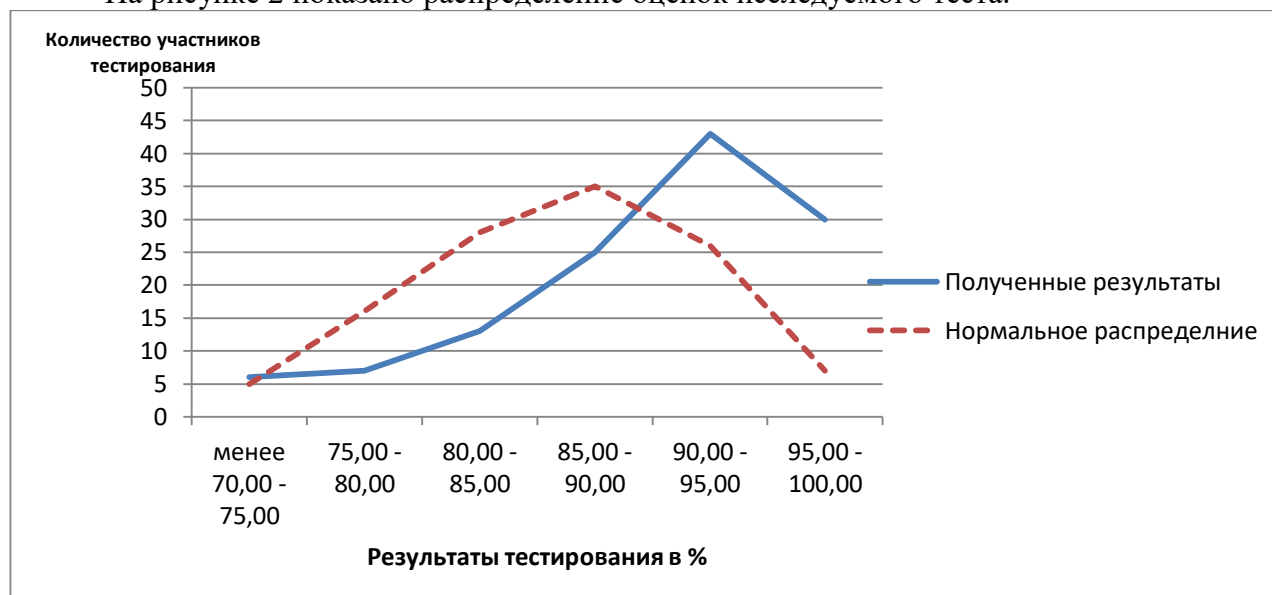


Рисунок 2. Распределение оценок за тест.

Кривая распределения оценок близка к нормальному распределению. Мода одна. Пик кривой острый, смещен влево. Считается, что в более хорошем тесте средний арифметический балл испытуемых равен медианному значению оценок используемых заданий, коэффициенты асимметрии и эксцесса не отклоняются от значений для стандартной кривой нормального распределения результатов [3]. Также хорошо, если значения среднего арифметического, моды и медианы совпадают.

**Коэффициент внутренней согласованности (90,65%)** – одна из разновидностей коэффициента надежности теста [2,4]. Для надежного теста с полностью согласованными заданиями равен 100%, для полностью несогласованного теста показатель надежности равен нулю. Считается, что внутренняя согласованность теста не должна быть ниже 80%, а для профессионально разработанных тестов должна быть на уровне не менее 90%.

**Соотношение ошибок (30,59%)** вычисляется на основании предыдущего показателя и характеризует степень внутренней согласованности теста. Чем это значение меньше, тем вопросы теста считаются более согласованными.

**Стандартная ошибка (4,02%)** указывает границы погрешности для оценки студента за тест.

**Вторая часть отчета** позволяет провести анализ и оценивание конкретных тестовых вопросов и позиций вопроса в тесте. Выводит гистограмму «Статистика по позиции вопроса». Анализ структуры теста в начале (рис.3) показывает характеристики вопроса в позиции теста, в конце таблицы показывает конкретные вопросы из базы тестов, решенные на 100% и нерешенные (0% правильных ответов). Для удобного изучения статистики теста результаты по всем вопросам банка вопросов можно скачать в MS Excel.

Анализ структуры теста

Скачать табличные данные как: Значения, разделяемые запятыми (.csv) Скачать

№	Название вопроса	Попытки	Индекс легкости	Стандартное отклонение	Балл случайного угадывания	Намеченный вес	Эффективный вес	Индекс дискриминации	Эффективность дискриминации
1	Случайный (Тесты Итоговые R-Стом)	122	89.69%	29.11%		2,00%	1,44%	19,04%	24,56%
2	Случайный (Тесты Итоговые R-Стом)	122	88.18%	29.90%		2,00%	1,51%	20,29%	26,02%
3	Случайный (Тесты Итоговые R-Стом)	122	86.07%	34.02%		2,00%	2,15%	39,88%	55,47%
4	Случайный (Тесты Итоговые R-Стом)	122	89.34%	28.81%		2,00%	1,84%	34,15%	43,88%

Рисунок 3 Анализ структуры теста.

Статистика теста содержит информационные поля (Тип вопроса, Название вопроса, Попытки – количество ответов на вопрос) и рассчитанные указанные ниже статистические показатели.

**Индекс легкости** – процент правильных ответов на этот вопрос. Вопросы, имеющие индекс легкости близкий к 0, считаются сложными, близкий к 100 – легкими. Такие вопросы следует убрать или изменить.

**Стандартное отклонение** характеризует разброс значений оценок студентов при ответе на конкретный вопрос теста. Низкое значение этой величины говорит о том, что студенты отвечают одинаково на этот вопрос, он имеет низкую дифференцирующую способность. Такие вопросы следует убрать или изменить.

**Балл случайного угадывания** – оценка, которую можно получить при случайном угадывании ответа.

**Намеченный вес** – вес, назначенный данному вопросу преподавателем.

**Эффективный вес** – характеристика того, какова фактическая доля данного вопроса в итоговой оценке.

**Индекс дискриминации** – соотношение связи между оценками за конкретный вопрос и за весь тест в целом. Этот коэффициент показывает, насколько взаимосвязаны правильность ответа на данный вопрос и остальные вопросы теста, характеризует способность тестового вопроса отличить сильных обучающихся от слабых. Предполагается, что для хорошего тестового вопроса студенты с высокими оценками за него также будут иметь более высокие оценки и за тест в целом [4]. На это укажет положительное значение коэффициента. Данный коэффициент может приобретать значение в диапазоне от -100% до 100%, является важной статистической характеристикой дифференцирующей способности заданий в тестовой форме.

**Эффективность дискриминации** – нормированный коэффициент дифференциации, этот коэффициент тоже демонстрирует соотношение ответов сильных и слабых обучающихся.

В хорошо составленных тестах отрицательных коэффициентов дискриминации быть не должно.

**Выводы.** 1. Представленный для анализа тест имеет неплохой коэффициент внутренней согласованности, распределение результатов тестирования близко к нормальному распределению, что свидетельствует о надежности теста.

2. По результатам анализа вопросов необходимо исправить задания с низким или высоким уровнем легкости и низким показателем эффективности дискриминации.

3. Встроенную процедуру статистики системы Moodle достаточно удобно использовать для анализа теста в целом и вопросов из банка данных.

Библиографический список:

1. [https://docs.moodle.org/dev/Quiz\\_statistics\\_calculations](https://docs.moodle.org/dev/Quiz_statistics_calculations)
2. Ким В. С. Тестирование учебных достижений. – Уссурийск: Издательство УГПИ, 2007. – 214 с.
3. Аванесов В.С. Метрическая система Георга Раша // Педагогические Измерения №3, 2011, Ред.2019. – 42 с.
4. Муханова А.А. Технология проектирования и оценки тестов в сдо moodle // <https://sworld.com.ua/konfer32/660.pdf>