

СТАТИСТИЧЕСКИЕ МЕТОДЫ В ДИДАКТИЧЕСКОМ ТЕСТИРОВАНИИ

Е.А. Михайлычев

По результатам тестовой диагностики составляется таблица частотного распределения набранных баллов (табл. 1), которая поможет разобраться с основными статистическими методами.

Таблица 1

Частотное распределение результатов тестирования

Тестовые баллы	Количество учащихся	Средний результат	Отклонение от среднего	Стандарт
0	20	3,60	-3,6	1,62
1	90		-2,6	
2	140		-1,6	
3	200		-0,6	
4	280		0,4	
5	170		1,4	
6	50		2,4	
7	40		3,4	
8	10		4,4	

Частота — количество случаев, в которых встречается данный результат тестового испытания. Например, **170** учащихся из выборки в **1000** человек справились с 5-ю тестовыми заданиями из 8-ми. Следова-

тельно, частота данного тестового результата **170**. Если тестовые результаты разделены на интервалы, то частота определяется числом испытуемых, набравших сумму тестовых баллов, относящихся к данному интервалу. Например, интервалу от 6 до 8-ми тестовых заданий соответствует **90** человек. Частота данного тестового результата (6–8) равна **90**.

Меры центральной тенденции

Средний результат — наиболее типичный (репрезентативный) результат, характеризующий выполнение теста всей группой. Например, при выполнении теста из 8-ми заданий индивидуальные результаты 1000 учащихся распределились так, как показано в таблице 1. Чтобы найти средний результат, необходимо сложить все результаты отдельных учащихся и разделить сумму на количество испытуемых (среднее арифметическое значение):

$$X_{cp} = (X_1 + X_2 + \dots + X_N)/N,$$

где X_i — индивидуальный тестовый результат i -го учащегося, N — количество учащихся в выборке. В нашем примере средний результат составляет **3,6**. Такого конкретного результата не может показать никто, но он характеризует усреднённое выполнение теста всей группой. Ближе всего в распределении располагается результат **280** учащихся, которые набрали **4** тестовых балла.

Для упрощения расчёта среднего результата можно индивидуальное число баллов умножить на количество набравших их учащихся. Например, **140** человек выполнили по **2** тестовых задания, результат перемножения этих чисел (**280**) необходимо сложить с подобными произведениями по каждой строке таблицы 1. Затем полученную сумму делят на количество учащихся в выборке (**1000**).

Мода — наиболее часто встречающийся результат, т.е. обладает максимальной частотой. Например, в таблице 1 мода распределения составляет **4** тестовых балла, так как их набрали **280** учащихся — это максимальная частота.

Медиана — тестовый результат, находящийся в середине последовательности показателей, если их расположить в порядке возрастания или убывания. Медиана делит статистическое распределение пополам. В нашем примере медиана соответствует 4 тестовым баллам и совпадает с модой распределения.

Нормальное распределение — строго симметричное статистическое распределение, при котором мода совпадает со средним результатом и медианой, а график зависимости «частота — количество тестовых бал-

лов» имеет колоколообразный вид. Нормальное распределение описывается функцией Гаусса

$$\varphi(X) = \exp(-X^2/2\sigma^2)/\sigma\sqrt{2\pi},$$

где σ — стандарт статистического распределения.

Меры разброса

Среднее отклонение — среднее арифметическое значение отклонений индивидуальных тестовых результатов от среднего, взятых по модулю

$$x_{cp} = (|x_1| + |x_2| + \dots + |x_N|)/N,$$

где x_i — отклонение тестового результата i -го учащегося от среднего, которое вычисляется с помощью вычитания среднего результата (3,6) из суммы тестовых баллов, набранных i -м учащимся. Например, учащийся №251 обладает индивидуальным тестовым результатом 3 и отклонением от среднего $x_{251}=0,6$ (табл. 1). Если не пренебрегать знаками индивидуальных отклонений от среднего результата (четвёртый столбец табл. 1), их сумма окажется равной нулю. В нашем примере необходимо перемножить конкретные значения отклонений на число учащихся, показавших соответствующий результат, а затем разделить полученную сумму на число учащихся в выборке. Среднее отклонение оказывается равным 1,3.

Стандарт — стандартное отклонение от среднестатистического результата. Квадрат этой величины называется **дисперсией** и определяется по формуле

$$\sigma^2 = \sum x_i^2 / (N1).$$

Индивидуальные отклонения от среднего результата, взятые из четвёртого столбца таблицы, необходимо возвести в квадрат и просуммировать. На это указывает значок « Σ » в формуле. Затем полученную сумму делят на число испытуемых за вычетом единицы. Приведённый пример статистического распределения (табл. 1) обладает дисперсией 2,62 и стандартом 1,62.

В случае нормального статистического распределения в интервале от $(X_{cp} - \sigma)$ до $(X_{cp} + \sigma)$ расположены 68,26% тестовых результатов, в интервале от $(X_{cp} - 2\sigma)$ до $(X_{cp} + 2\sigma)$ — 95,44%, а в интервале от $(X_{cp} - 3\sigma)$ до $(X_{cp} + 3\sigma)$ — 99,72%.

Коэффициент IQ — отношение умственного возраста индивида к его биологическому возрасту.

$$IQ = 100(MA/CA),$$

где MA — умственный возраст; CA — биологический возраст. IQ=100 означает нормальное или обычное выполнение теста. Например, 12-лет-

ний ребёнок, выполняющий тестовые задания, рассчитанные на 15-тилетний возраст, обладает $IQ = 125$. Такой же интеллектуальный коэффициент 10-летнего ребёнка указывает на его умственный возраст: $MA = CA \times IQ \times 100 = 12,5$ лет. Однако это не значит, что опережение интеллектуального развития в обоих случаях одинаково.

Только определяя положение тестового результата индивида относительно нормального статистического распределения можно сравнить интеллект разновозрастных учащихся. Например, опережение развития в 12-летнем возрасте на 3 года ($IQ = 125$) совпадает с опережением на 1,5 года интеллектуального развития ребёнка в возрасте 10 лет ($IQ = 115$), так как оба значения коэффициента IQ превышают средний результат соответствующих возрастных распределений на величину стандарта (1). Это позволяет заменить традиционный интеллектуальный коэффициент стандартным IQ , в котором рассматривается не отношение умственного и биологического возрастов, а положение тестового результата испытуемого относительно моды статистического распределения, принимаемой за $IQ = 100$. Зная величину стандарта данного теста интеллекта, можно определить стандартную величину индивидуального интеллектуального коэффициента. Например, в результате применения теста интеллекта, состоящего из 100 заданий, испытуемый правильно выполнил 65. Процедура стандартизации теста на репрезентативной выборке ровесников продемонстрировала моду статистического распределения $M = 50$ и стандарт $= 15$. Следовательно, индивидуальный тестовый результат превышает моду на величину стандарта ($M + 1\sigma$). Заменяя моду на нормальный интеллектуальный уровень (100), получаем $IQ = 115$. Такой способ вычислений стандартного IQ позволяет не акцентировать внимание на возрасте репрезентативной выборки и не использовать термин «умственный возраст».

Надёжность методом ретеста — измеряется при повторном применении того же самого теста на том же контингенте, в тех же условиях. Корреляция между данными двух сеансов тестирования будет отражать случайные колебания в выполнении теста, вызываемые неконтролируемыми факторами, например: изменением погоды, неожиданными отвлекающими шумами, изменением в состоянии здоровья испытуемых, эмоциональным напряжением и т.п. Чем выше ретестовая надёжность, тем менее чувствительны результаты тестирования к обычным изменениям состояния испытуемого. Указывая в руководстве к тесту ретестовую надёжность, необходимо указывать, в каком интервале времени она изменена. Расчёт ретестовой надёжности совпадает с определением надёжности взаимозаменяемых форм.

Надёжность взаимозаменяемых форм — предполагает на стадии разработки теста создание нескольких независимо построенных форм, от-

вечающих одинаковым требованиям. Они должны включать в себя одинаковое число заданий, представленных в одном и том же виде и с однотипным содержанием, с одинаковым диапазоном и уровнем трудности заданий. При этом можно изменять порядок следования дистракторов.

Надёжность взаимозаменяемых форм, как и ретестовая надёжность, определяется с помощью **коэффициента корреляции произведения моментов Пирсона**

$$r_{xy} = \Sigma(x_i \times y_i) / (N \times \sigma_x \times \sigma_y),$$

где x_i и y_i — отклонения индивидуальных тестовых результатов от средних значений в различных сериях испытаний; σ_x и σ_y — стандартны соответствующих статистических распределений; N — количество учащихся. Например, необходимо определить по результатам повторного тестирования десяти учащихся (табл. 2) ретестовую надёжность или надёжность взаимозаменяемых форм. В таблице результат первичного испытания указан в числителе, а вторичного — в знаменателе.

Таблица 2

Результаты повторного тестирования

№	Отметка о выполнении соответствующего задания								Баллы	
	1	2	3	4	5	6	7	8	I	II
1	1/1	0/0	1/1	0/0	1/1	1/1	0/1	0/0	4	5
2	1/1	0/0	0/0	0/0	0/0	1/1	0/0	1/0	3	2
3	0/0	0/0	0/1	0/0	0/0	1/1	0/0	0/0	1	2
4	1/1	0/0	0/0	1/1	1/1	1/1	1/0	0/1	5	5
5	1/1	1/1	0/0	0/0	1/1	1/1	0/1	1/1	5	6
6	1/1	0/0	0/0	1/1	0/0	1/0	0/0	0/1	3	3
7	0/0	1/1	0/0	0/0	1/1	1/0	1/1	1/0	5	3
8	1/1	0/0	1/1	0/0	0/1	0/0	0/0	1/1	3	4
9	1/1	0/1	0/0	0/0	0/0	1/1	0/0	0/0	2	3
10	0/0	0/0	0/0	0/0	1/0	0/0	0/1	0/0	1	1

На основе данных, которые приведены в таблице 2, можно составить таблицу частотного распределения результатов первой и второй серии испытаний (табл. 3), в которой указано количество учащихся (n_1, n_2), набравших соответствующее число тестовых баллов. Это позволяет вычислить стандартные распределения первичного и вторичного испытаний (σ_x, σ_y) также, как описано в пункте **стандарт**.

Чтобы воспользоваться формулой Пирсона, необходимо перемножить индивидуальные отклонения от среднего результата в каждом ис-

пытании. С помощью таблицы 2 можно определить количество тестовых заданий, которые правильно выполнил каждый ученик в обоих испытаниях. Ученик №1 набрал в первом испытании 4 балла ($x_1 = 0,8$), а во втором — 5 баллов ($y_1 = 1,6$). Следовательно, произведение $x_1 \times y_1 = 1,28$. Ученик №2 характеризуется произведением $x_2 \times y_2 = 0,28$, ученик №3 — $x_3 \times y_3 = 3,03$, ученик №4 — $x_4 \times y_4 = 2,88$, ученик №5 — $x_5 \times y_5 = 4,68$, ученик №6 — $x_6 \times y_6 = 0,08$, ученик №7 — $x_7 \times y_7 = 0,72$, ученик №8 — $x_8 \times y_8 = 0,12$, ученик №9 — $x_9 \times y_9 = 0,48$, ученик №10 — $x_{10} \times y_{10} = 5,28$. Два произведения (№7 и 8) оказались отрицательными, так как в различных испытаниях отклонения тестовых результатов этих учащихся отличаются знаками. Теперь необходимо сложить все полученные произведения с учётом знака и разделить на количество учеников в группе и на стандарты обоих испытаний. Получаем количественное выражение надёжности теста $r_{xy} = 0,70$.

Таблица 3

Частотное распределение результатов повторного тестирования

Баллы	n_1	n_2	Средние значения		x_i	y_i	σ_x	σ_y
0	0	0	3,2	3,4	-3,2	-3,4	1,55	1,58
1	2	1			-2,2	-2,4		
2	1	2			-1,2	-1,4		
3	3	3			-0,2	-0,4		
4	1	1			0,8	0,6		
5	3	2			1,8	1,6		
6	0	1			2,8	2,6		
7	0	0			3,8	3,6		
8	0	0			4,8	4,6		

Надёжность методом Кьюдера-Ричардсона — находится по данным одноразового применения единственной формы теста с использованием результатов выполнения испытуемыми каждого задания. Для расчёта надёжности в этом случае применяется формула Кьюдера-Ричардсона

$$r_{pq} = m \times (\sigma^2 - \sum p_j \cdot q_j) / ((m - 1) \times \sigma^2),$$

где m — число тестовых заданий, σ^2 — дисперсия суммарных показателей теста; p_j и q_j — доля испытуемых, которые справились или не справились с j -м тестовым заданием. Например, необходимо определить надёжность теста и по результатам первичного испытания, взятого из таб-

лицы 2 ($\sigma = 1,55$). Составим таблицу 4 распределения доли выполнения и невыполнения тестовых заданий.

Таблица 4

Доли выполнения и невыполнения тестовых заданий

Выполнение (p_i) или невыполнен. (q_i)	1	2	3	4	5	6	7	8
p_i	0,7	0,2	0,2	0,2	0,5	0,8	0,2	0,4
q_i	0,3	0,8	0,8	0,8	0,5	0,2	0,8	0,6
$p_i \times q_i$	0,21	0,16	0,16	0,16	0,25	0,16	0,16	0,24

Результаты перемножения долей выполнения и невыполнения каждого тестового задания необходимо сложить, а затем вычесть полученную сумму из дисперсии распределения ($\sigma^2 = 2,40$). Количество тестовых заданий $m = 8$, дисперсия и полученная разность подставляются в формулу Кьюдера-Ричардсона. Показатель надёжности теста $r_{pq} = 0,43$ оказался гораздо ниже, чем при повторном тестировании. Метод Кьюдера-Ричардсона позволяет определить не только минимальный уровень надёжности, но и согласованность тестовых заданий.

Надёжность методом расщепления — осуществляет путём разбиения единственной формы теста на две сопоставимые части и получения двух серий результатов испытаний. Коэффициент корреляции в этом случае будет характеризовать только половину теста. Надёжность методом расщепления определяют по формуле Рюлона

$$r_R = 1 - \sigma_R^2 / \sigma^2 \Sigma,$$

где σ_R^2 — дисперсия разностей между результатами каждого испытуемого по обеим половинам теста, а σ^2 — дисперсия суммарных результатов. Разделим тест, результаты первичного применения которого представлены в таблице 2, на две части, в одной из которых окажутся чётные задания, а в другой — нечётные. Суммарная дисперсия в этом случае $\sigma^2 = 2,40$, а для определения разностной дисперсии составим таблицу 5.

Полученный стандарт σ_R возводим в квадрат и подставляем в формулу Рюлона. Надёжность по методу расщепления $r_R \approx 0,91$.

Надёжность на скорость — определяется путём расщепления тестовых заданий на два субтеста, для которых устанавливаются отдельные лимиты времени. Вычисления производятся также как и в случае определения надёжности по методу расщепления. Если сравнить результаты выполнения теста без учёта времени и выполнения субтестов на скорость, можно вычислить «скоростные индексы». Например, тесты на

Определение разностной дисперсии расщепленных форм

№	Нечётная половина	Чётная половина	Разность результатов	Среднее значение	Отклонение от среднего	σ_R
1	3 1	2		1	1	0,47
2	1	2	1		0	
3	0	1	1		0	
4	3	2	1		0	
5	2	3	1		0	
6	1	2	1		0	
7	2	3	1		0	
8	2	1	1		0	
9	1	1	0		-1	
10	1	0	1		0	

рассуждение и пространственное восприятие в большой степени подвержены влиянию скорости. В то же время выполнение теста на понимание слов практически не зависит от временного фактора.

Коэффициент Спирмана-Брауна — оценивает влияние увеличения или же сокращения теста на его надёжность

$$r = n \times r' / (1 + (n - 1) \times r')$$

где r' — исходное значение коэффициента надёжности; r — его значение после расширения или сокращения теста; n — отношение нового числа заданий к первоначальному. Например, если был разработан дидактический тест итогового контроля, состоящий из **40** заданий, с надёжностью **0,90**, то после его сокращения до **10** заданий ($n = 0,25$) по формуле Спирмана-Брауна можно определить новое значение надёжности теста: $r = 0,69$. При сокращении количества тестовых заданий надёжность теста снижается.

Валидности индекс — оценивается корреляцией тестовых результатов с результатами независимых показателей (критериальная валидность) или согласованностью и диагностической направленностью тестовых заданий (конструктивная валидность). Для определения критериальной валидности применяется формула коэффициента корреляции произведения моментов **Пирсона**, приведённая при расчёте **надёжности взаимозаменяемых форм**. Чтобы количественно оценить конструктивную валидность, можно применить формулу **Кьюдера-Ричардсона**, приведённую при расчёте **надёжности методом Кьюдера-Ричардсона**.

Валидизация экспертная — производится путём опроса нескольких групп компетентных экспертов по концептуальным вопросам методологии, а также по содержательной значимости более мелких структурных элементов курса относительно более крупных. В случае применения дифференцированной формы представления результатов отдельных тестовых заданий для оценки их согласованности используется **коэффициент альфа Кронбаха**

$$\alpha = m \times (\sigma^2 - \sum \sigma_i^2) / ((m - 1) \times \sigma^2),$$

где σ_i^2 — дисперсия результатов отдельных заданий. Например, при разработке теста итогового контроля знаний по физике было необходимо установить, каким из методологических подходов к изложению курса можно руководствоваться при составлении тестовых заданий, если преподавание материала возможно в исторической последовательности (№1), по классической (№2) или квантово-релятивистской методологии (№3). Проведём статистическую обработку результатов экспертизы, определяя средний номер методологического подхода и стандарт распределения мнений экспертов (табл. 6).

Таблица 6

Экспертная оценка концептуальной валидности

Раздел	Учителя				Учёные			
	Ср. значение		Стандарт		Ср. значение		Стандарт	
	А	В	А	В	А	В	А	В
1	1,89	1,89	0,56	0,32	1,88	1,88	0,35	0,35
2	2,00	1,78	0,49	0,43	1,75	1,75	0,46	0,46
3	2,17	2,33	0,52	0,65	1,88	1,88	0,35	0,35
4	2,39	2,72	0,78	0,67	2,75	2,75	0,50	0,50

Показатели по наиболее удачным методологиям обозначены буквой «А», а по устойчиво применяемым — буквой «В». Зная стандарт суммарных показателей бланка экспертизы ($\sigma = 1,5$) и количество вопросов в нём ($m = 4$), можно подсчитать сумму дисперсий из таблицы 6 и подставить все данные в формулу коэффициента альфа Кронбаха. В результате вычислений получаем $\alpha = 0,91$ для экспертной группы учёных и $\alpha = 0,48$ для наиболее удачных методологических подходов, по мнению учителей. Устойчиво применяемые учителями подходы оцениваются коэффициентом альфа, равным **0,64**. Согласованность экспертных оценок учёных наибольшая, а абстрактные пожелания учителей отличаются минимальной согласованностью, так как не отличаются большой компетентностью.

Кси-квадрат — показатель статистической значимости коэффициента альфа, т. е. степени согласованности тестовых заданий в случае дифференцированной формы представления ответов. Для вычислений кси-квадрат применяется **формула Китса**

$$\chi^2 = m \times (N - 1) / (m \times (1 - \alpha) + \alpha),$$

где α — коэффициент альфа Кронбаха; m — количество тестовых заданий; N — количество испытуемых. Например, используя значения коэффициента альфа для концептуальной валидизации, проведённой 11-ю учёными (**0,91**) и 19-ю учителями (**0,48**), рассчитаем статистическую значимость соответствующих экспертиз, состоящих из четырёх вопросов, по формуле Китса. Значимость экспертной оценки учёных характеризует кси-квадрат, равный **31,5**, а учителей — **28,1**.

Таблицы Тейлора-Расселла — позволяют определить выигрыш в точности отбора от использования теста. В качестве примера рассмотрим таблицу 7 с базовым уровнем **0,60**, характеризующим отношение количества сотрудников, справившихся со своими обязанностями и набранных случайно (без использования теста).

В верхней части таблицы 7 приведены различные значения индекса отбора. В крайнем левом столбце указан коэффициент валидности, а в строках таблицы — относительное число успехов среди принятых на работу по результатам теста. Разность между такой величиной и **0,60** указывает на выигрыш от применения теста.

Когда индекс отбора равен единице и приёму подлежат все претенденты, ни один тест, как бы валиден он не был, не улучшит качества отбора. Например, конкурс абитуриентов составляет два человека на место. Соответствующий индекс отбора равен **0,50**. Используя тест с валидностью **0,80**, вы произведёте набор студентов, **88%** которых будут хорошо учиться (табл. 7). Выигрыш от применения теста составит **0,28**. Из таблицы Тейлора-Расселла легко понять, что эффективность тестовой диагностики теряется только при валидности конструкта ниже **0,10** и индексе отбора **0,90 ÷ 0,95**.

Бисериальная корреляция — мера валидности тестового задания, т. е. отношения между содержанием задания и критерием, который не зависит от его трудности. Чтобы вычислить бисериальную корреляцию, необходимо знать среднее значение и стандарт критериального показателя выполнивших и невыполнивших задание, а также процент справившихся с заданием по всей выборке. Однако способа точного расчёта уровней значимости бисериальной корреляции не существует. Поэтому используются специальные таблицы, позволяющие определить бисериальную корреляцию по проценту справившихся с заданием в группах, соответствующих верхним и нижним 27% распределения значений критерия.

Ожидаемая величина успехов для базового уровня 0,60

Валидность	Индекс отбора										
	0,05	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	0,95
0,00	0,69	0,60	0,60	0,60	0,60	0,60	0,60	0,60	0,60	0,60	0,60
0,05	0,64	0,63	0,63	0,62	0,62	0,62	0,61	0,61	0,61	0,60	0,60
0,10	0,68	0,67	0,65	0,64	0,64	0,63	0,62	0,61	0,61	0,61	0,60
0,15	0,71	0,70	0,68	0,67	0,66	0,65	0,64	0,63	0,62	0,61	0,61
0,20	0,75	0,73	0,71	0,69	0,67	0,66	0,65	0,64	0,63	0,62	0,61
0,25	0,78	0,76	0,73	0,71	0,69	0,68	0,66	0,65	0,63	0,62	0,61
0,30	0,82	0,79	0,76	0,73	0,71	0,69	0,68	0,66	0,64	0,62	0,61
0,35	0,85	0,82	0,78	0,75	0,73	0,71	0,69	0,67	0,65	0,63	0,62
0,40	0,88	0,85	0,81	0,78	0,75	0,73	0,70	0,68	0,66	0,63	0,62
0,45	0,90	0,87	0,83	0,80	0,77	0,74	0,72	0,69	0,66	0,64	0,62
0,50	0,93	0,90	0,86	0,82	0,79	0,76	0,73	0,70	0,67	0,64	0,62
0,55	0,95	0,92	0,88	0,84	0,81	0,78	0,75	0,71	0,68	0,64	0,62
0,60	0,96	0,94	0,90	0,87	0,83	0,80	0,76	0,73	0,69	0,65	0,63
0,65	0,98	0,96	0,92	0,89	0,85	0,82	0,78	0,74	0,70	0,65	0,63
0,70	0,99	0,97	0,94	0,91	0,87	0,84	0,80	0,75	0,71	0,66	0,63
0,75	0,99	0,99	0,96	0,93	0,90	0,86	0,81	0,77	0,71	0,66	0,63
0,80	1,00	0,99	0,98	0,95	0,92	0,88	0,83	0,78	0,72	0,66	0,63
0,85	1,00	1,00	0,99	0,97	0,95	0,91	0,86	0,80	0,73	0,66	0,63
0,90	1,00	1,00	1,00	0,99	0,97	0,94	0,88	0,82	0,74	0,67	0,63
0,95	1,00	1,00	1,00	1,00	0,99	0,97	0,92	0,84	0,75	0,67	0,63
1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,86	0,75	0,67	0,63

Можно оценить согласованность содержания тестового задания с выбранным критерием с помощью формулы Пирсона. Например, расположим данные о профессиональных успехах **1000** бывших курсантов мореходных высших учебных заведений в порядке их возрастания и проанализируем выполнение задания на закон Архимеда у тех из них, которые расположены с 1 по 270 и с 731 по 1000-е место. Выполнение задания курсантами, относящимися к верхней группе, или невыполнение его курсантами, относящимися к нижней группе, будем считать корреляционным слагаемым $x_i \times y_i$ равным +1. Если же успешные выпускники мореходных учебных заведений не знали закона Архимеда или неудачники правильно ответили на этот вопрос, то будем считать корреляционное

слагаемое равным — 1. Из **540** критериально значимых результатов **376** вносят положительный вклад в бисериальную корреляцию, а **124** — отрицательный.

Если **65%** всех курсантов справились с данным тестовым заданием (трудность $P=65$), то средний результат его выполнения равен **0,65**. В **650**-ти случаях из **1000** отклонение от среднего составит **+0,35**, а в **350** случаях — **0,65**. Стандарт статистического распределения вычисляется по формуле $\sigma^2 = \Sigma x_i^2 / (N - 1) = (0,35^2 \times 650 + 0,65^2 \times 350) / 999 \approx 0,23$. Извлекая квадратный корень, получаем $\sigma \approx 0,48$. Стандарт статистического распределения независимых критериальных показателей будем считать равным единице. Используя формулу Пирсона, можно рассчитать бисериальную корреляцию:

$$r_{\text{bis}} = \Sigma(x_i \times y_i) / (N \times \sigma) = (376 - 124) / 1000 \times 0,48 = 0,525.$$

Критерий эффективности обучения — используется для оценки качества образовательного процесса по проценту усвоения объёма изучаемого материала учащимися. Средний результат стандартизованного теста достижений с высокой содержательной валидностью, отнесённый к количеству тестовых заданий, позволяет определить эффективность обучения

$$\eta = X_{\text{ср}} / m.$$

Например, средний результат, показанный учащимися в процессе применения теста итогового контроля по физике, состоящего из **40** заданий, равен **21,5**. В результате расчёта получаем $\eta \approx 0,54$, что соответствует достаточно высокому качеству обучения ($\eta_{\text{эталон}} \approx 0,6$).