

# Обзор метода статистической обработки результатов тестирования знаний

презентация по дисциплине

«Дополнительные главы математики»

Студент: Пиликов Д.А.

Группа: АММ-06

Преподаватели: к.т.н, доцент Казанская О.В

д.т.н, профессор Губарев В.В.

Тема магистерской диссертации:

Разработка и исследование автоматизированной системы мониторинга  
переподготовки кадров на предприятии

Руководитель:

д.т.н, профессор Губарев В.В.

Новосибирск, 2007

# Список литературы

- 1 Чельшкова М. Б. Теория и практика конструирования педагогических тестов: Учеб. пособие. – Москва.: Логос, 2002. – 432 с.
- 2 Педагогический тест: этапы и особенности конструирования и использования: учеб. пособие / С. В. Клишина, Н. А. Гулюкина. – 2-е изд., испр. и доп. – Новосибирск: Изд-во НГТУ, 2006. – 148 с.
- 3 Статистическая обработка результатов тестирования [Электронный ресурс]: сборник статей. –Тула., 2003. –Режим доступа: [www.tiei.ru/ppage/pages/57/HTML/Podsevalov.htm](http://www.tiei.ru/ppage/pages/57/HTML/Podsevalov.htm) - Свободный.

# Введение

Тест - специально подготовленный и прошедший экспериментальную проверку набор заданий, позволяющий объективно и надежно оценить исследуемые качества и свойства на основе использования экспертных оценок и статистических методов /1/.

Для анализа качества теста и его коррекции используется математико-статистическая обработка результатов.

# Постановка задачи обзора

## Исходные данные:

Результаты тестирования группы проверяемых на одних и тех же тестовых заданиях, либо на тестах схожих по структуре и сложности.

## Необходимо:

Использовать метод математико-статистической обработки для анализа качества теста и его коррекции исходя из полученных расчётов по заданным критериям качества.

## Этапы обработки результатов теста

1. Сбор эмпирических данных тестирования (проведение теста и составление матрицы результатов);
2. Статистическая обработка результатов тестирования
3. Анализ качества теста.

# Математическая модель зависимости результатов от сложности заданий

1958 г. Г. Раш: двухпараметрическая модель зависимости тестового результата  $i$ -го испытуемого в  $j$ -м задании, где

а) вероятности правильных ответов:

$$P_{ij} \{x_{ij} = 1 | \theta_i, \beta_j\} = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}}$$

б) вероятности неправильных ответов

$$q_{ij} \{x_{ij} = 0 | \theta_i, \beta_j\} = 1 - \frac{e^{(\theta_i - \beta_j)}}{1 + e^{(\theta_i - \beta_j)}}$$

где  $x_{ij} = 1$ , если ответ  $i$ -го испытуемого на  $j$ -е задание правильный;

$x_{ij} = 0$ , если ответ  $i$ -го испытуемого на  $j$ -е задание неправильный;

$\theta_i$ , уровень знаний  $i$ -го испытуемого;

$\beta_j$ , уровень трудности  $j$ -го задания;

$p_{ij}$  – вероятность правильного ответа  $i$ -го испытуемого на  $j$ -е задание;

$q_{ij}$  – вероятность неправильного ответа  $i$ -го испытуемого на  $j$ -е задание,

равная, как принято в теории вероятностей,  $q = 1 - p$ .

# Матрица результатов тестирования

Результаты тестирования представляются в виде матрицы размерами  $M \times N$ , где  $M$  – число заданий в тесте,  $N$  – число испытуемых

№ п/п	Тестируемые	Задания теста							$R$	$P_i$	$q$	$R_i^2$
		1	2	3	4	5	...	$M$				
1		1	1	0	1	1		1	$R_1$	$P_1$	$q_1$	$R_1^2$
2		0	1	1	1	0		0	$R_2$	$P_2$	$q_2$	$R_2^2$
.												
.												
.												
$N$		1	0	0	0	0		0	$R_n$	$P_n$	$q_n$	$R_n^2$
$R_j$		$R_1$	$R_2$					$R_m$	*			**
$W_j$												
$P_j$		$p_1$	$P_2$					$p_m$				
$q_j$		$q_1$	$q_2$					$q_m$				
$d_j$		$d_1$	$d_2$					$d_m$				
$\sigma_i$		$\sigma_1$	$\sigma_2$					$\sigma_m$				

$$R_i = \sum_{j=1}^M x_{ij}, (1 \leq i \leq N)$$

$$p_i = \frac{R_i}{M}$$

$$q_i = 1 - p_i = \frac{1 - R_i}{M}$$

$$R_j = \sum_{i=1}^N x_{ij} (1 \leq j \leq M)$$

$$p_j = \frac{R_j}{N} = \frac{\sum_{i=1}^N x_{ij}}{N}$$

$$d_j = p_j q_j$$

$$W_j = N - R_j (1 \leq j \leq M);$$

$$q_j = 1 - p_j = \frac{W_j}{N}$$

$$\sigma_i = \sqrt{p_j q_j}$$

# Матрица результатов тестирования

- $R_i$  – тестовый балл  $i$ -го испытуемого (кол-во единиц в соответствующей строке)  $R_i = \sum_{j=1}^M x_{ij}, (1 \leq i \leq N)$
  - $p_i$  – доля правильных ответов испытуемого по всем заданиям теста  $p_i = \frac{R_i}{M}$
  - $q_i$  – доля неправильных ответов испытуемого по всем заданиям теста:  $q_i = 1 - p_i = \frac{1 - R_i}{M}$
- 
- $R_j$  – количество правильных ответов по каждому заданию  $R_j = \sum_{i=1}^N x_{ij} (1 \leq j \leq M)$
  - $W_j$  – количество неправильных ответов по каждому заданию  $W_j = N - R_j (1 \leq j \leq M);$
  - $p_j$  – доля правильных ответов по каждому заданию всеми испытуемыми  $p_j = \frac{R_j}{N} = \frac{\sum_{i=1}^N x_{ij}}{N}$
  - $q_j$  – доля неправильных ответов по каждому заданию всеми испытуемыми  $q_j = 1 - p_j = \frac{W_j}{N}$
  - $d_j$  – дисперсия каждого задания  $d_j = p_j q_j$
  - $\sigma_i$  – среднее квадратическое отклонение по каждому заданию  $\sigma_i = \sqrt{p_j q_j}$

# Матрица результатов тестирования и предварительные выводы о тесте

Характеристики по всему тесту:

– суммарный тестовый балл всех испытуемых:

$$X = \sum_{i=1}^N R_i$$

– среднее арифметическое индивидуальных тестовых баллов испытуемых:

$$\tilde{X} = \frac{\sum_{i=1}^N R_i}{N}$$

*Оценка качества теста* проводится по результатам апробации теста на репрезентативной выборке тестируемых.

Где  $R_i$  – тестовый балл  $i$ -го испытуемого,  $N$  – количество испытуемых

# Матрица результатов тестирования и предварительные выводы о тесте

Для оценки вариации тестовых баллов по всему тесту используются три показателя.

1. Сумма квадратов отклонений от среднего:

$$SS_x = \sum_{i=1}^N (R_i - \bar{X})^2 = \sum_{i=1}^N R_i^2 - \frac{\left( \sum_{i=1}^N R_i \right)^2}{N}$$

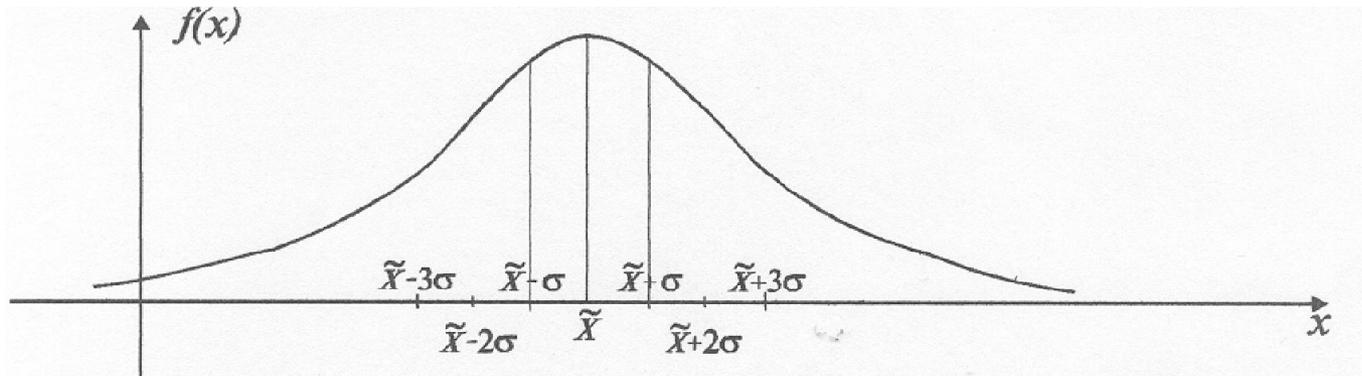
2. Дисперсия тестовых баллов испытуемых:

$$S_x^2 = \frac{SS_x}{N-1}$$

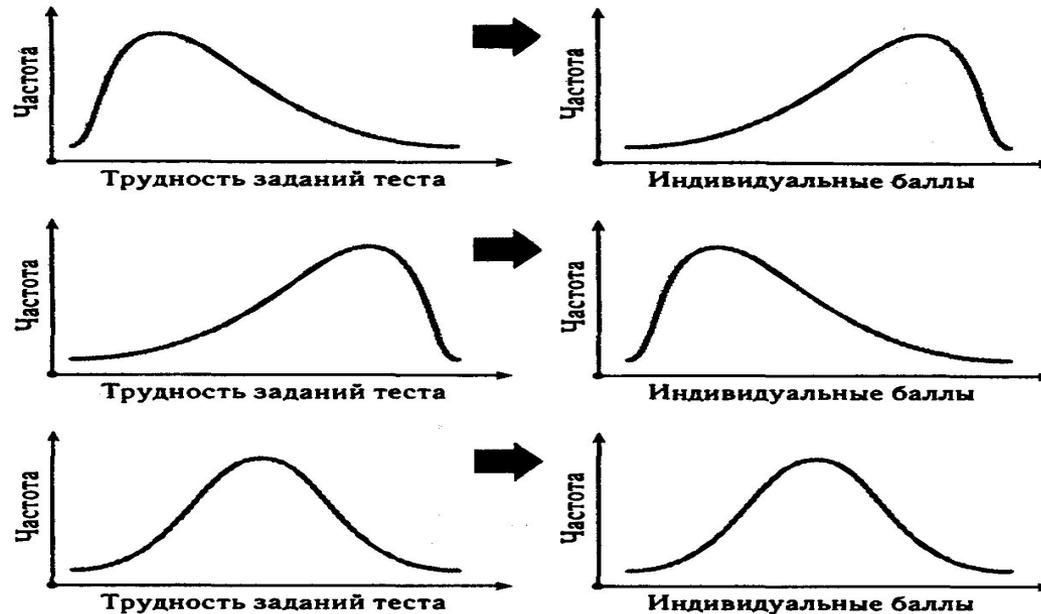
3. Стандартное (среднее квадратическое) отклонение:

$$S_x = \sqrt{S_x^2}$$

# Предварительные выводы о тесте



Кривая Гаусса (распределение индивидуальных баллов испытуемых)  
 $f(x)$  – плотность распределения случайной величины  $x$  – результатов тестирования



Связь распределения индивидуальных баллов и трудности заданий теста

# Предварительные выводы о тесте

Из-за сложности проверки нормальности распределения, на практике применяют величину среднего квадратического отклонения.

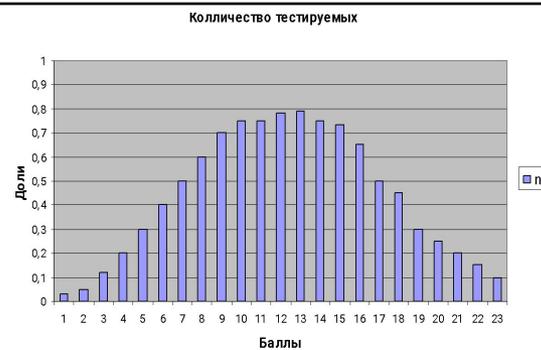
Если  $3Sx \approx \tilde{X}$ , то распределение близко к нормальному (гауссовскому) /2/

# Графическое представление результатов тестирования



*Гистограмма долей правильных ответов по всем заданиям ( $p_j$ ) дает представление о качестве теста.*

*Гистограмма упорядочена по уменьшению долей правильных ответов. Номера заданий по горизонтальной оси должны идти последовательно от первого до последнего*



*Гистограмма распределения испытуемых по результатам тестирования позволяет сделать вывод о распределении индивидуальных баллов и составе группы испытуемых.*

# Корреляционный анализ результатов

## **Бисериальный коэффициент корреляции (для дихотомической шкалы)**

$M_B$  – среднее арифметическое по всему тесту у студентов, успешно ответивших на  $j$ -е задание;

$M_H$  – среднее арифметическое по всему тесту у студентов, ответивших неверно на  $j$ -е задание;

$\sigma_j$  – стандартное отклонение по  $j$ -му заданию (3.9);

$S_x$  – стандартное отклонение по всему тесту, по всей группе испытуемых

$$r_{pbisx_j R_i} = \frac{M_B - M_H}{S_x} \sqrt{p_j q_j}$$

## **Коэффициент корреляции Пирсона**

$x, y$  – исследуемые выборки

$$SP_{x,y} = \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}$$

– сумма попарных произведений по исследуемым выборкам;

$$SS_x = \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}$$

– сумма квадратов отклонений от среднего элементов выборки  $x$ ;

$$SS_y = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}$$

– сумма квадратов отклонений от среднего элементов выборки  $y$ .

$$r_{x,y} = \frac{SP_{x,y}}{\sqrt{SS_x SS_y}}$$

# Корреляционный анализ результатов

На основании корреляционного анализа можно сделать следующие выводы /2/:

Если  $r_{xij} \geq 0,3$ , следовательно скорее всего присутствует предметная связь между заданиями

Если  $r_{xij}, Ri < 0,3$  (коэффициент корреляции каждого задания с суммарным тестовым балом испытуемых), следовательно такие задания требуют доработки

Если  $r_{xij} < 0$ , некорректность содержания заданий

# Оценка качества теста

## Надёжность

Коэффициентов надёжности теста –  $r_{HT}$

<p><b>С помощью коэффициента корреляции <math>r_k</math></b></p> <p>где <math>r_k</math> – это коэффициент корреляции, вычисляемый между суммой баллов по чётным и нечётным заданиям, либо между несколькими тестами, схожими по структуре и сложности</p>	$r_{HT} = \frac{2r_k}{1 + r_k}$
<p><b>Формула KR-20</b></p> <p>где <math>M</math> – число заданий в тесте</p> <p><math>\sum_{j=1}^M p_j g_j</math> – сумма произведений дисперсий всех заданий теста</p> <p><math>S_x^2</math> – дисперсия тестовых баллов испытуемых</p>	$r_{HT}(KR) = \frac{M}{M-1} \left( 1 - \frac{\sum_{j=1}^M p_j g_j}{S_x^2} \right)$
<p><b>С помощью среднего коэффициента корреляции по всему тесту <math>R_{cp}</math></b></p> <p>где <math>M</math> – число заданий в тесте</p> $R_{cp} = \frac{\sum_{i=1}^{M-1} r_i}{M}$ <p>, где <math>\sum_{i=1}^{M-1} r_i</math> – сумма средних коэффициент корреляции каждого задания со всеми остальными</p>	$r_{HT} = \frac{M R_{cp}}{1 + (M-1)R_{cp}}$

# Оценка качества теста

## Надёжность

Характеристики тестов, используемые как критерий качества по надёжности приведены в таблице /2/.

Коэффициент надёжности	Надёжность
0,90...0,99	Отличная
0,85...0,89	Очень хорошая
0,80...0,84	Удовлетворительная
0,75...0,79	Малоудовлетворительная
0,70...0,74	Сомнительная
0,60...0,69	Неудовлетворительная
0,50...0,59	Совсем неудовлетворительная
0,40...0,49	-
0,30...0,39	-
0,20...0,29	-
0,10...0,19	-
0,00...0,09	-

# Выводы

В ходе данной работы был рассмотрен метод статистической обработки результатов тестов, подходящий для дихотомических шкал (0,1; «да», «нет»; «правильно», «не правильно»); описан математический аппарат, необходимый для корректировки тестовых заданий в тесте.

На основании пробного тестирования, проведённого среди 30 испытуемых, были собраны данные по результатам, построены гистограммы долей правильных ответов и распределения количества испытуемых, вычислены числовые характеристики (среднего статистического, дисперсии и среднего квадратического отклонения), вычислена несколькими методами и проанализирована надёжность тестовых заданий. По результатам вычисления сделаны выводы о качестве теста из которых следовало то, что тест требует доработки.

## Планируется:

- рассмотреть методы анализа качества тестов для других шкал;
- рассмотреть различные виды тестирования (адаптивное);
- рассмотреть возможности применения описанных методов обработки для других видов тестирования.