Представление результатов педагогических измерений

1. Матрица результатов тестирования
2. Подготовка матрицы ответов к математико-статистическому анализу

Матрица результатов тестирования

№	ФИО тестирующегося	Номер задания									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
		Правильный ответ									
		C	***D***	***A***	***C***	***C***	***B***	***D***	***C***	***D***	***B***
1	Иванов В.А.	C	D	A	C	C	B	A	A	C	D
2	Петров П.П.	C	D	C	D	A	D	C	D	C	C
3	Дроздов П.У.	A	C	B	A	A	A	A	A	A	D
4	Вахненко М.К.	A	C	D	A	D	A	C	C	A	A
5	Елизарова М.М.	C	D	D	C	C	B	D	C	D	B
6	Лысенко Т.О.	C	B	A	C	C	B	D	C	D	B
7	Голуб А.В.	C	D	A	D	D	D	A	C	C	C
8	Файзулин М.Р.	C	D	A	C	A	B	A	D	C	A
9	Игнатович Р.Т.	C	D	A	C	C	B	D	C	D	D
10	Роднина М.В.	C	D	A	C	C	D	D	A	C	C
11	Правдина М.Л.	C	D	A	C	C	B	D	C	D	B

Матрица результатов тестирования – матрица ответов

Номер испытуемого *i*	Номер задания *j*									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0
5	1	1	0	1	1	1	1	1	1	1
6	1	0	1	1	1	1	1	1	1	1
7	1	1	1	0	0	0	0	1	0	0
8	1	1	1	1	0	1	0	0	0	0
9	1	1	1	1	1	1	1	1	1	0
10	1	1	1	1	1	0	1	0	0	0
11	1	1	1	1	1	1	1	1	1	1

n – общее количество испытуемых (объектов исследования);
p – общее количество заданий теста;
“*j*”-е задание; “*i*”-й испытуемый;
xij – результат выполнения “*j*”-го задания “*i*”-м испытуемым.

Подготовка матрицы ответов к статистическому анализу

Номер испытуемого *i*	Номер задания *j*									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0
5	1	1	0	1	1	1	1	1	1	1
6	1	0	1	1	1	1	1	1	1	1
7	1	1	1	0	0	0	0	1	0	0
8	1	1	1	1	0	1	0	0	0	0
9	1	1	1	1	1	1	1	1	1	0
10	1	1	1	1	1	0	1	0	0	0
11	1	1	1	1	1	1	1	1	1	1

- Правильный и неправильный профиль ответов испытуемого

Подготовка матрицы ответов к статистическому анализу

Номер испытуемого *i*	Номер задания *j*									
	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0
5	1	1	0	1	1	1	1	1	1	1
6	1	0	1	1	1	1	1	1	1	1
7	1	1	1	0	0	0	0	1	0	0
8	1	1	1	1	0	1	0	0	0	0
9	1	1	1	1	1	1	1	1	1	0
10	1	1	1	1	1	0	1	0	0	0
11	1	1	1	1	1	1	1	1	1	1

- ***Удаление строк и столбцов, состоящих <u>только</u> из нулей и единиц.***

Подготовка матрицы ответов к статистическому анализу

Подсчет индивидуальных баллов испытуемых и количества правильных ответов испытуемых на каждое задание теста.

Номер испытуемого i	Номер задания j										Y_i
	1	2	3	4	5	6	7	8	9	10	
1	1	1	1	1	1	1	0	0	0	0	**6**
2	1	1	0	0	0	0	0	0	0	0	**2**
4	0	0	0	0	0	0	0	1	0	0	**1**
5	1	1	0	1	1	1	1	1	1	1	**9**
6	1	0	1	1	1	1	1	1	1	1	**9**
7	1	1	1	0	0	0	0	1	0	0	**4**
8	1	1	1	1	0	1	0	0	0	0	**5**
9	1	1	1	1	1	1	1	1	1	0	**9**
10	1	1	1	1	1	0	1	0	0	0	**6**
R_j	**8**	**7**	**6**	**6**	**5**	**5**	**4**	**5**	**3**	**2**	**51**

Подготовка матрицы ответов к статистическому анализу

Упорядочение матрицы результатов тестирования.

R_j *в порядке убывания*

Y_i *в порядке возрастания*

Номер испытуемого *i*	Номер задания *j*										***Yi***
	1	2	3	4	5	6	8	7	9	10	
5	1	1	0	1	1	1	1	1	1	1	**9**
6	1	0	1	1	1	1	1	1	1	1	**9**
9	1	1	1	1	1	1	1	1	1	0	**9**
1	1	1	1	1	1	1	0	0	0	0	**6**
10	1	1	1	1	1	0	0	1	0	0	**6**
8	1	1	1	1	0	1	0	0	0	0	**5**
7	1	1	1	0	0	0	1	0	0	0	**4**
2	1	1	0	0	0	0	0	0	0	0	**2**
4	0	0	0	0	0	0	1	0	0	0	**1**
Rj	**8**	**7**	**6**	**6**	**5**	**5**	**5**	**4**	**3**	**2**	**51**

Основные статистические характеристики

1. Тестовые задания
2. Тест
3. Распределение результатов тестирования

Статистические характеристики тестовых заданий

- Известная мера трудности является обязательным требованием к тестовым заданиям. Если неизвестна эмпирическая мера трудности задания, то это задание **не тестовое.**
- ***Трудность задания*** определяется после апробации теста. В качестве показателя трудности используется статистика – доля неправильных ответов q_j.

$$p_j = \frac{R_j}{n} \qquad q_j = 1 - p_j$$

- Вариация баллов является вторым обязательным требованием к тестовым заданиям. *Если на какое-то задание отвечают все тестируемые, или наоборот, не отвечают, то никакой вариации нет, а, следовательно, задание не дифференцирует знающих испытуемых от незнающих.*
- **Дисперсия (вариация ответов)** определяется по формуле:

$$s_j^2 = p_j q_j,$$

где p_j и q_j – доля правильных и неправильных ответов в каждом задании.

Статистические характеристики тестовых заданий

Номер испытуемого i	Номер задания j										Y_i
	1	2	3	4	5	6	8	7	9	10	
4	0	0	0	0	0	0	1	0	0	0	**1**
2	1	1	0	0	0	0	0	0	0	0	**2**
7	1	1	1	0	0	0	1	0	0	0	**4**
8	1	1	1	1	0	1	0	0	0	0	**5**
1	1	1	1	1	1	1	0	0	0	0	**6**
10	1	1	1	1	1	0	0	1	0	0	**6**
5	1	1	0	1	1	1	1	1	1	1	**9**
6	1	0	1	1	1	1	1	1	1	1	**9**
9	1	1	1	1	1	1	1	1	1	0	**9**
R_j	**8**	**7**	**6**	**6**	**5**	**5**	**5**	**4**	**3**	**2**	**51**
W_j	**1**	**2**	**3**	**3**	**4**	**4**	**4**	**5**	**6**	**7**	
p_j	**0.89**	**0.78**	**0.67**	**0.67**	**0.56**	**0.56**	**0.56**	**0.44**	**0.33**	**0.22**	**5.67**
q_j	**0.11**	**0.22**	**0.33**	**0.33**	**0.44**	**0.44**	**0.44**	**0.56**	**0.67**	**0.78**	
$p_j q_j$	**0.10**	**0.17**	**0.22**	**0.22**	**0.25**	**0.25**	**0.25**	**0.25**	**0.22**	**0.17**	

График трудности тестовых заданий (неупорядоченный)

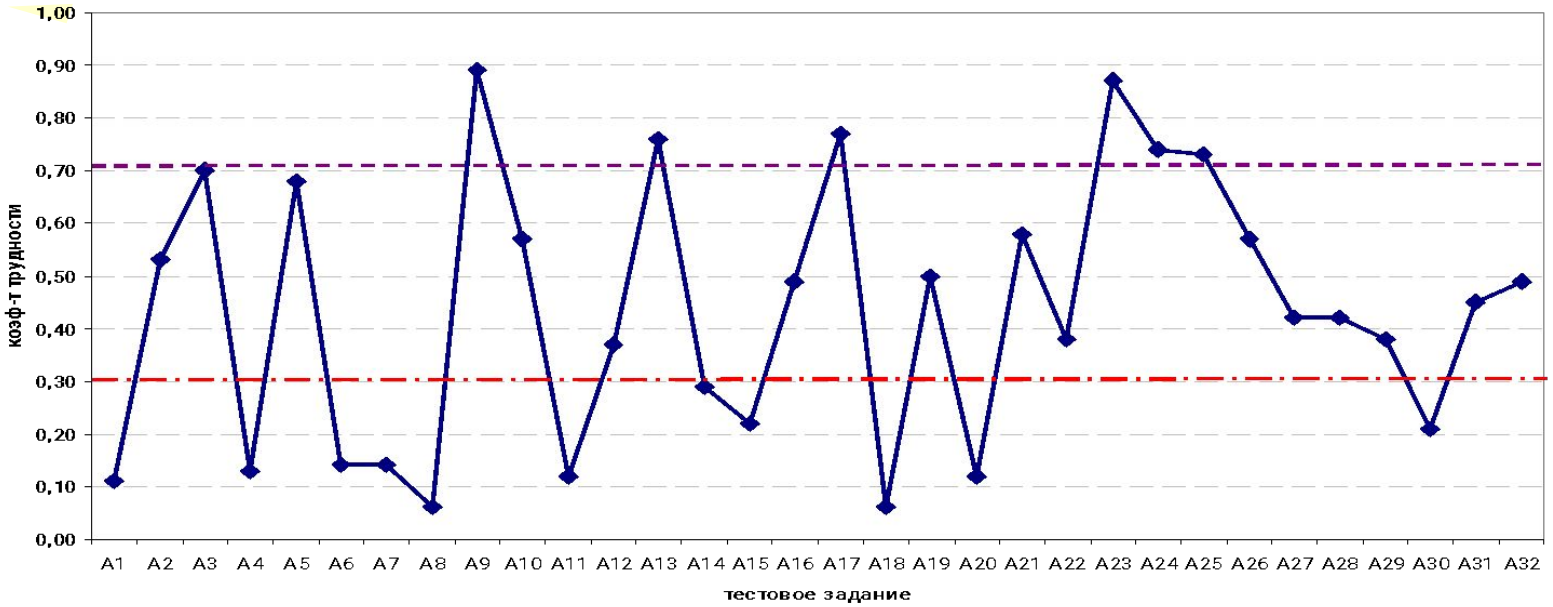
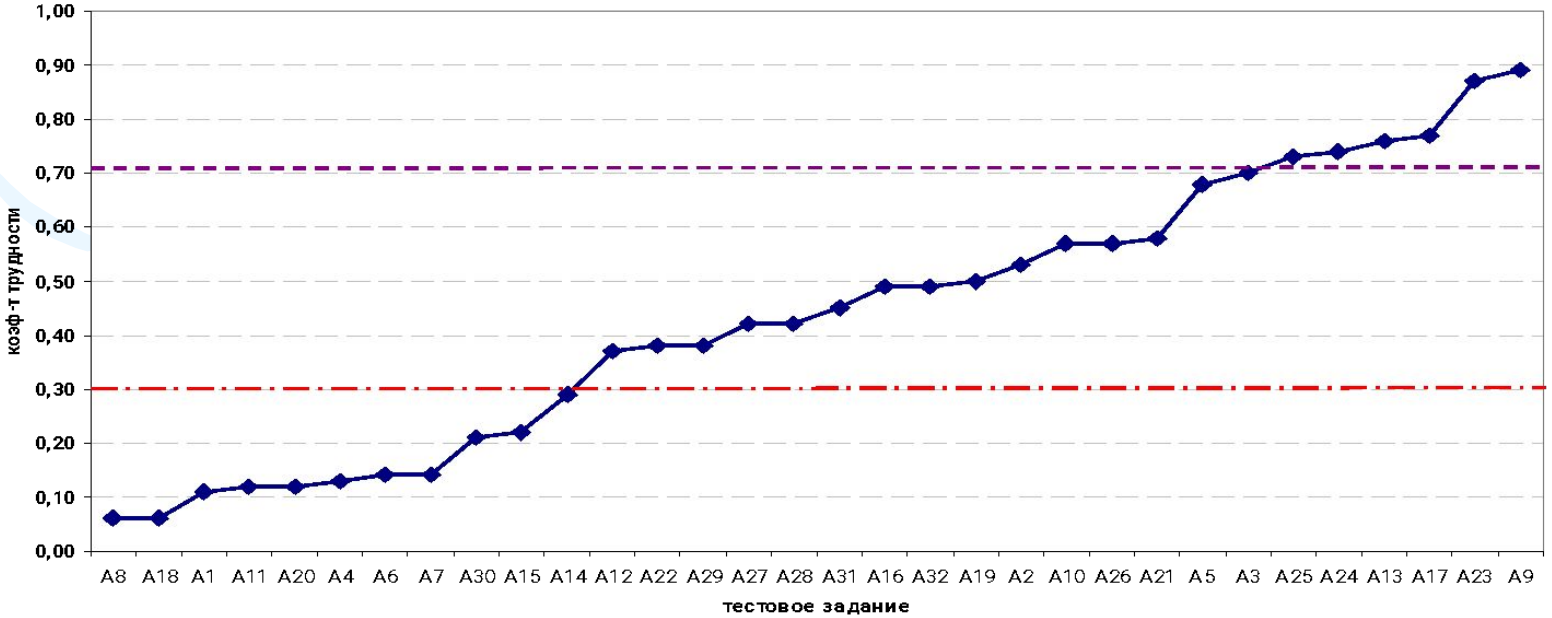


График трудности тестовых заданий (упорядоченный)



Статистические характеристики тестовых заданий

Дифференцирующая способность (discriminatory power) задания, способность задания разделять обучающихся на «знающих» и «незнающих».

$$D_j = p_j^{хор} - p_j^{сл}$$

$p_j^{хор}$ – доля правильных ответов на задание в лучшей группе испытуемых (27-30% испытуемых, имеющих высокие тестовые баллы),

$p_j^{сл}$ – доля правильных ответов на то же задание в худшей группе (27-30% испытуемых, имеющих низкие тестовые баллы).

№	D_j	Оценка / Рекомендации
1	$0{,}4 < D_j \leq 1$	Отличное задание
2	$0{,}3 < D_j \leq 0{,}4$	Хорошее задание
3	$0{,}2 < D_j \leq 0{,}3$	Желательна дальнейшая доработка. Если трудность задания выше 80%, то задание можно оставить без доработки
4	$0{,}1 < D_j \leq 0{,}2$	Критическое значение. Необходим тщательный анализ содержания.
5	$-1 \leq D_j \leq 0{,}1$	Задание нельзя использовать. Отрицательное значение ДСЗ указывает на ошибки в задании, например, указан неверный ответ.

Достижение дифференцирующего эффекта – главная цель создания <u>нормативно-ориентированных тестов</u>.

Статистические характеристики тестовых заданий

Номер испытуемого i	Номер задания j										Y_i
	1	2	3	4	5	6	8	7	9	10	
4	0	0	0	0	0	0	1	0	0	0	**1**
2	1	1	0	0	0	0	0	0	0	0	**2**
7	1	1	1	0	0	0	1	0	0	0	**4**
8	1	1	1	1	0	1	0	0	0	0	**5**
1	1	1	1	1	1	1	0	0	0	0	**6**
10	1	1	1	1	1	0	0	1	0	0	**6**
5	1	1	0	1	1	1	1	1	1	1	**9**
6	1	0	1	1	1	1	1	1	1	1	**9**
9	1	1	1	1	1	1	1	1	1	0	**9**
R_j	**8**	**7**	**6**	**6**	**5**	**5**	**5**	**4**	**3**	**2**	**51**
W_j	**1**	**2**	**3**	**3**	**4**	**4**	**4**	**5**	**6**	**7**	
p_j	**0.89**	**0.78**	**0.67**	**0.67**	**0.56**	**0.56**	**0.56**	**0.44**	**0.33**	**0.22**	**5.67**
q_j	**0.11**	**0.22**	**0.33**	**0.33**	**0.44**	**0.44**	**0.44**	**0.56**	**0.67**	**0.78**	
$p_j q_j$	**0.10**	**0.17**	**0.22**	**0.22**	**0.25**	**0.25**	**0.25**	**0.25**	**0.22**	**0.17**	
$p_{j_слаб}$	**0,67**	**0,67**	**0,33**	**0,00**	**0,00**	**0,00**	**0,67**	**0,00**	**0,00**	**0,00**	
$p_{j_хор}$	**1,00**	**0,67**	**0,67**	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	**1,00**	**0,67**	
ДСЗ	**0,33**	**0,00**	**0,33**	**1,00**	**1,00**	**1,00**	**0,33**	**1,00**	**1,00**	**0,67**	

График дифференцирующей способности заданий

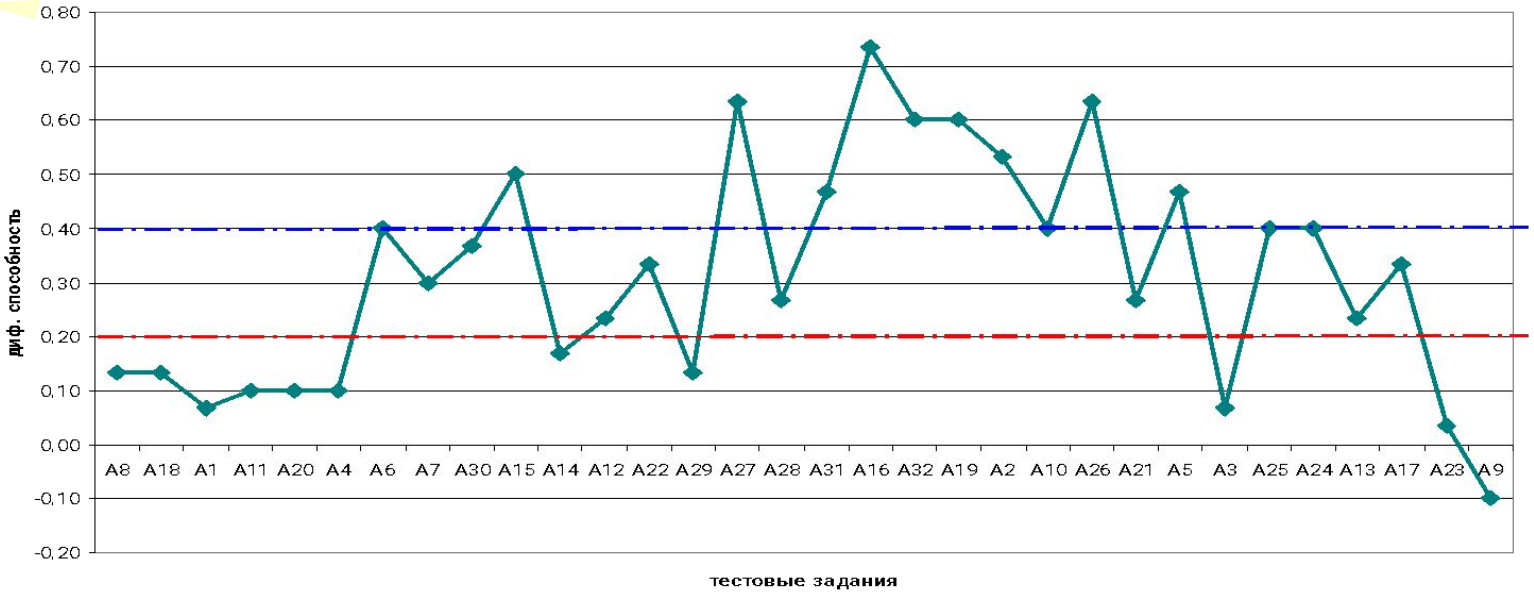
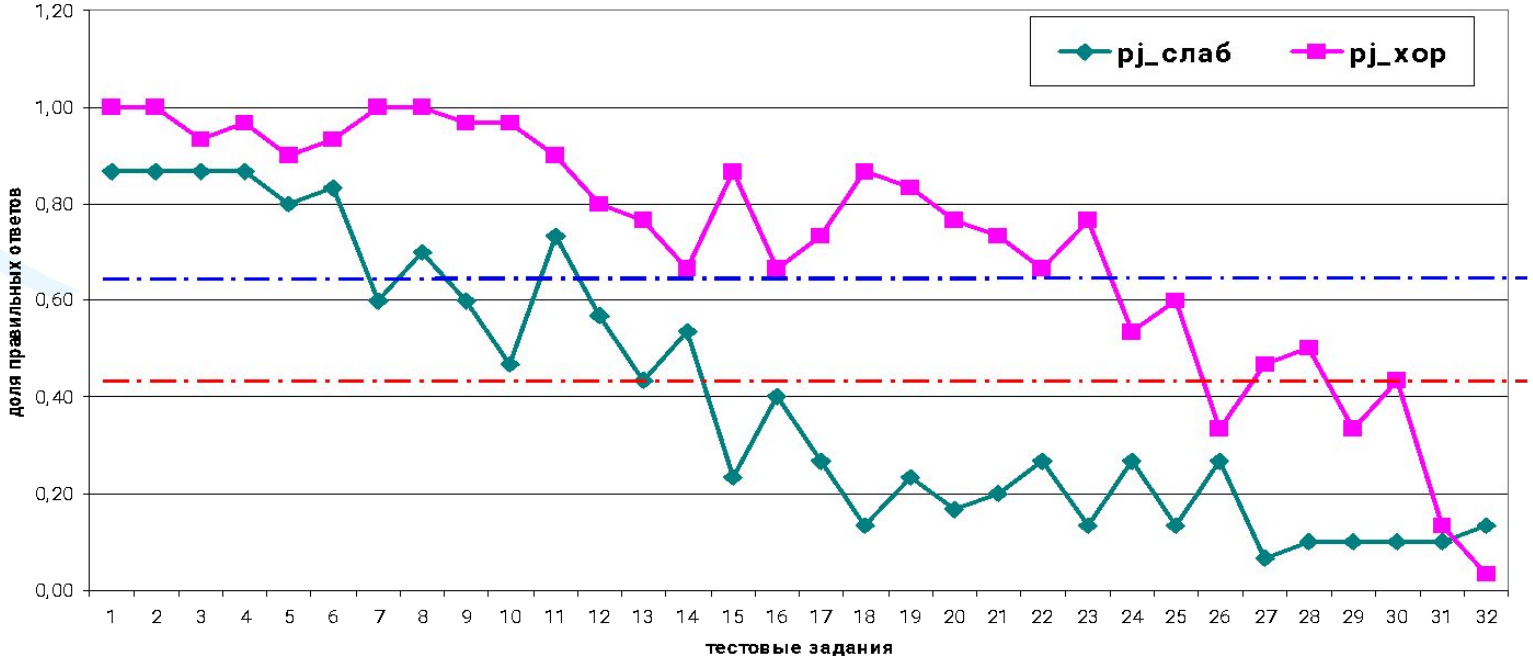


График дифференцирующей способности заданий



Оценка системности (валидности) тестового задания

- Расчет корреляционной матрицы позволяет оценить тестовые свойства задания.
- В корреляционной матрице представлены корреляции между двумя заданиями, а также корреляция тестового задания с суммой баллов.
- Желательно, чтобы между заданиями были низкие корреляции (значения меньше 0,3).
- Чем выше значение корреляции задания с суммой баллов, тем выше шансы задания называться тестовым и попасть в тест (говорят о системности (валидности) задания).
- **Квадрат корреляции** задания с суммой баллов, представленный в процентах (коэффициент детерминации), указывает на вклад задания в общую дисперсию тестовых баллов.
- Задания с значением корреляции ниже 0,2 из теста удаляются, как не выдержавшие эмпирической проверки.

Корреляционная матрица

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*Yi*
1	1										
2	0,56	1									
3	0,36	0,65	1								
4	**0,30**	0,68	0,88	1							
5	**0,24**	0,66	0,85	0,97	1						
6	**0,22**	0,67	0,86	0,97	0,96	1					
7	**0,04**	0,59	0,81	0,88	0,93	0,95	1				
8	**0,18**	0,65	0,84	0,95	0,98	0,96	0,97	1			
9	**0,15**	0,65	0,83	0,93	0,96	0,98	0,99	0,99	1		
10	**0,14**	0,64	0,82	0,93	0,96	0,97	0,98	0,98	0,99	1	
Yi	**0,58**	**0,13**	**0,42**	**0,83**	**0,84**	**0,76**	**0,29**	**0,82**	**0,83**	**0,63**	**1**
	34%	**2%**	**17%**	**69%**	**71%**	**58%**	**8%**	**67%**	**69%**	**40%**	

График коэффицентов корреляции тестовых заданий с суммой баллов

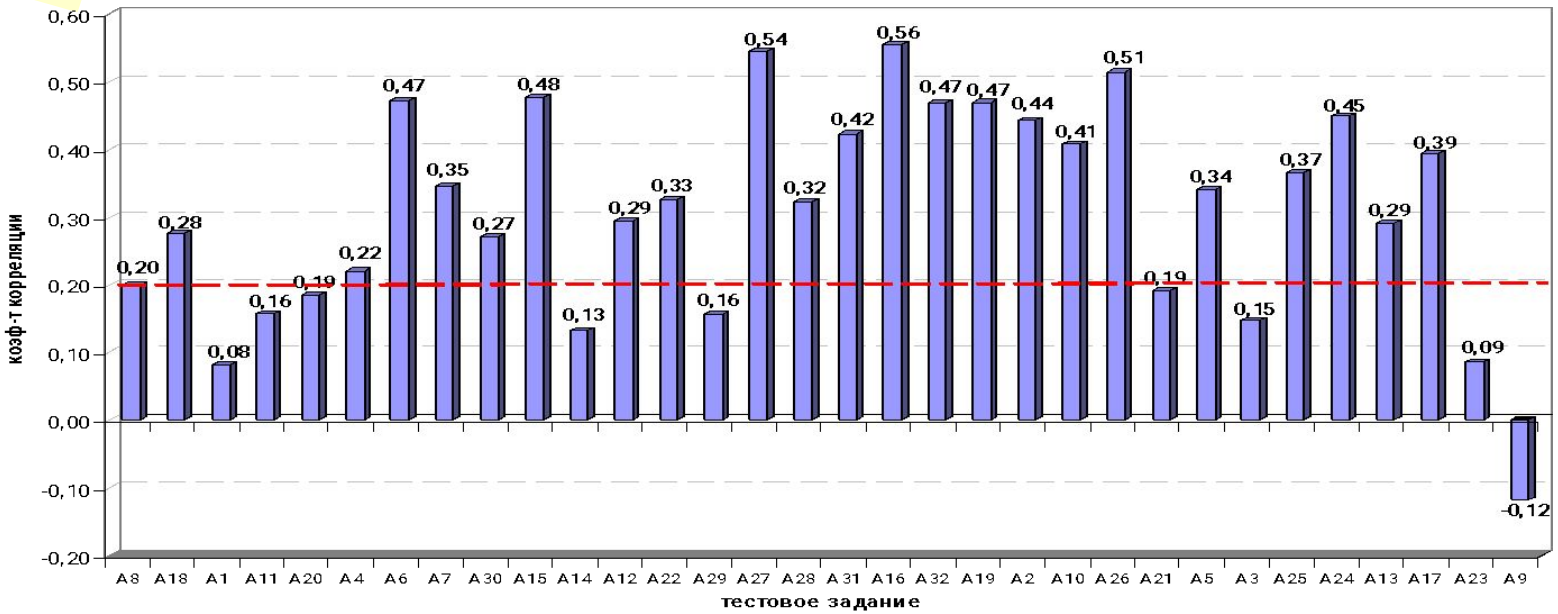
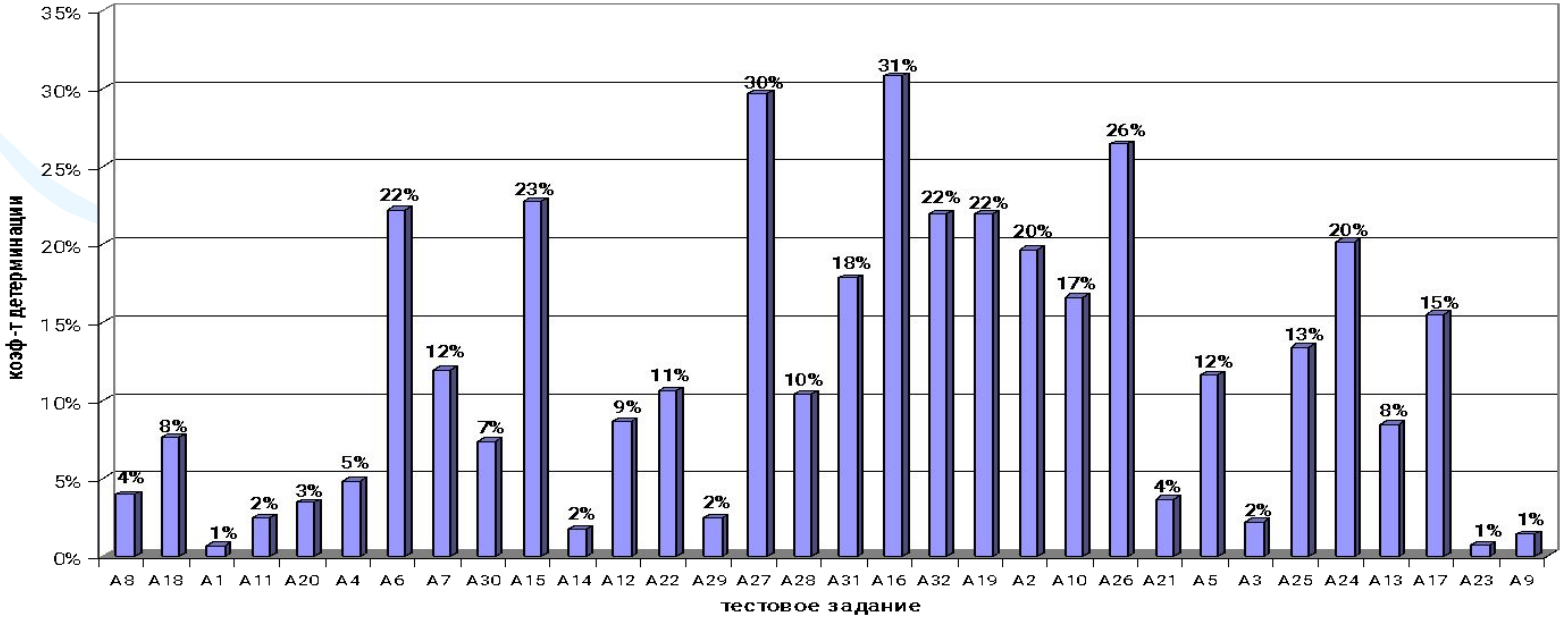


График коэффицентов детерминации тестовых заданий



Коэффициент надежности теста

- Коэффициент надежности теста показывает насколько можно доверять полученным результатам.
- Говорят «данные получены с высокой надежностью», если коэффициент выше 0,8. Чем больше данное значение, тем уже доверительный интервал для истинного балла.
- Коэффициент надежности позволяет судить о качестве теста в целом.

$$(r_н)_{KR-20} = \frac{n}{n-1} \cdot \left(1 - \frac{\sum_{j=1}^{p} p_j q_j}{s_Y^2} \right)$$

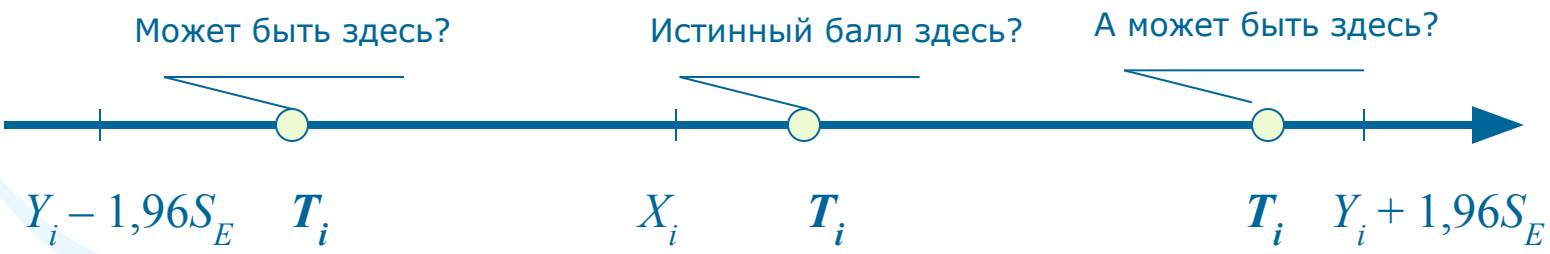
где *pj* и *qj* – доля правильных и неправильных ответов в каждом задании;
n – количество тестируемых; р – количество заданий;
s_Y^2 – дисперсия тестового балла.

Ошибка измерения

- Один из аспектов применения коэффициента надежности связан с определением стандартной ошибки измерения (standard error of measurement).

$$S_E = s_Y \sqrt{(1 - r_н)}$$

- Стандартная ошибка измерения является стандартной погрешностью оценки истинных баллов на основании наблюдаемых результатов тестовых измерений.



- При использовании в своей работе «готовых» тестов, стремитесь узнать (определить) коэффициент надежности теста, с целью вычисления ошибки измерения.

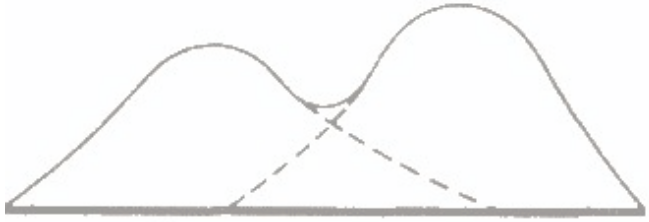
Описательные характеристики распределения тестовых результатов

1. **Меры среднего положения** (меры центральной тенденции).
   - Мода, медиана, среднее
2. **Меры вариации** (рассеяния, изменчивости данных).
   - Размах, дисперсия, стандартное отклонение, коэффициент вариации
3. **Меры формы** (меры симметрии и островершинности кривой распределения).
   - Коэффициенты асимметрии, эксцесса
4. Квантили.
   - Квартили, процентили

Меры среднего положения

- Это обобщающие показатели, характеризующие типичное значение, присущее большинству единиц совокупности, позволяющие выявлять закономерности.

- ***Мода, М –*** это наиболее часто встречающееся значение признака в исследуемой совокупности.
  - **Унимодальное распределение,**
  - ***Бимодальное распределение,***
  - ***Мультимодальное распределение.***



Бимодальное распределение

В данной совокупности имеются две относительно самостоятельные группы.

- ***Медиана, Md*** – это значение, которое делит упорядоченную совокупность данных пополам, так что одна половина значений больше медианы, а другая – меньше.

Меры среднего положения

- ***Среднее,*** $\bar{x}$, – это значение, которое отражает типичное значение для исследуемой совокупности в данных условиях:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n},$$

где n – объем совокупности, x_i – i-е значение совокупности.

– Чтобы средняя величина была действительно обобщающей характеристикой, улавливающей закономерность, она должна применяться к достаточно однородной совокупности.

– Средняя величина рассчитывается <u>только для количественных</u> признаков.

Соотношение мер среднего положения

- У симметричных унимодальных распределений:

$$\bar{x} = M = Md$$

- У унимодальных распределений с правой асимметрией:

$$M < Md < \bar{x}$$

- У унимодальных распределений с левой асимметрией:

$$M > Md > \bar{x}$$

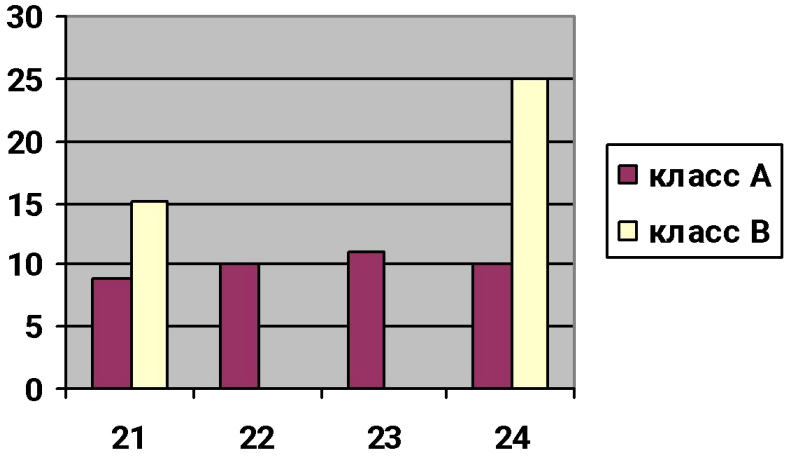
Выбор меры центральной тенденции в зависимости от типа измерительной шкалы

Тип шкалы	Меры центральной тенденции
Номинальная	Мода
Ранговая	Мода, медиана
Интервальная	Мода, медиана, среднее
Отношений	Мода, медиана, среднее

Меры изменчивости

- ***Меры изменчивости*** – показатели, измеряющие вариацию (разброс) значений совокупности. К мерам изменчивости относятся: размах, дисперсия, стандартное отклонение, коэффициент вариации, и др.

- ***Размах вариации*** (***R***), отражает пределы изменчивости значений совокупности.
- Представляет собой разность между максимальным (x_{max}) и минимальным значением (x_{min}) совокупности:



$$R = x_{max} - x_{min}.$$

- Размах не учитывает всех значений в выборке и определяется только двумя значениями.

Меры изменчивости

- ***Дисперсия,*** δ^2 **или** s^2, – это значение, которое отражает внутреннюю изменчивость значений исследуемой совокупности:

$$\delta^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1},$$

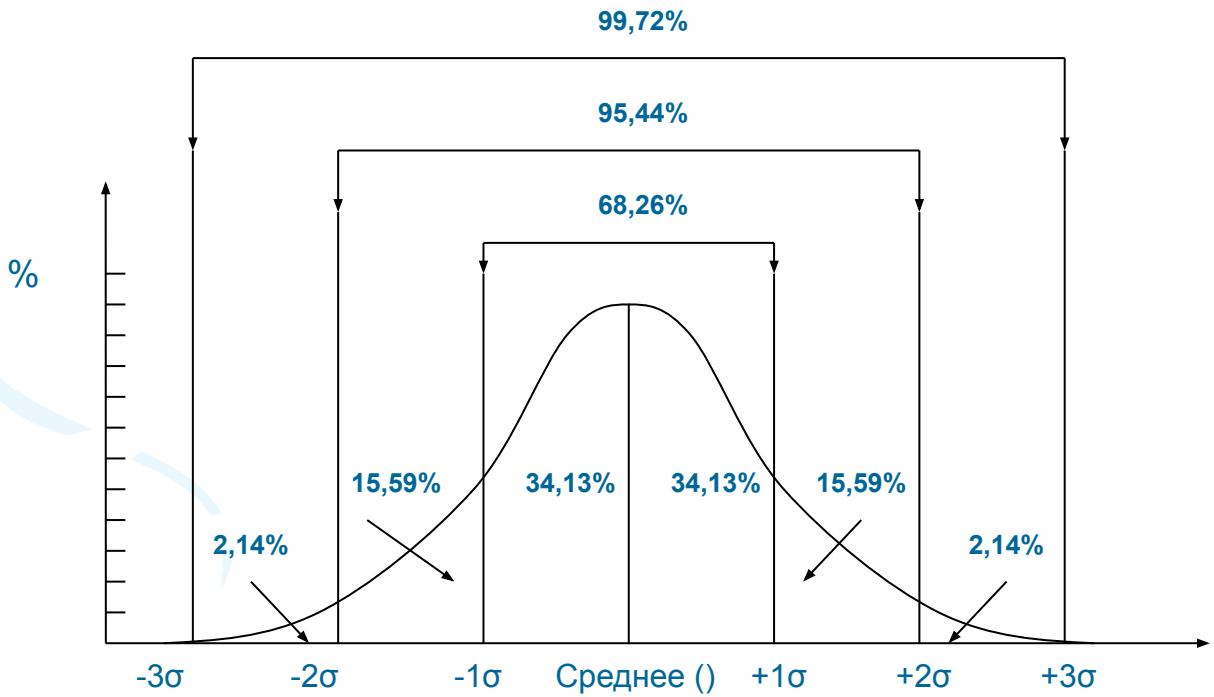
- где n – объем совокупности,
  x_i – i-е значение, $\bar{x}$ – среднее значение.

- ***Стандартное отклонение,*** δ **или** s, показывает насколько в среднем отклоняется каждое значение (x_i) от среднего $\bar{x}$:

$$\delta = \sqrt{\delta^2},$$

Меры изменчивости

- **<u>В случае нормальности</u>** распределения исследуемой совокупности значений справедливо следующее (правило «Трех сигм»):
- 70% значений лежит между $\bar{x} - \sigma$ и $\bar{x} + \sigma$,
- 95% значений лежит между $\bar{x} - 2\sigma$ и $\bar{x} + 2\sigma$,
- 99% значений лежит между $\bar{x} - 3\sigma$ и $\bar{x} + 3\sigma$.



Процентное распределение наблюдений под нормальной кривой

Меры изменчивости

- Для сопоставления изменчивости признаков используют ***коэффициент вариации***

$$Cv = \frac{\sigma}{\bar{x}} \cdot 100\% .$$

- При симметричных распределениях коэффициент вариации Cv не превышает 50%. При сильно асимметричных рядах распределения коэффициент вариации Cv может принимать значение 100% и даже выше.

Варьирование считается

- слабым, если $Cv < 10\%$,
- средним, если $10\% < Cv \leq 33\%$,
- значительным, при $Cv > 33\ \%$.

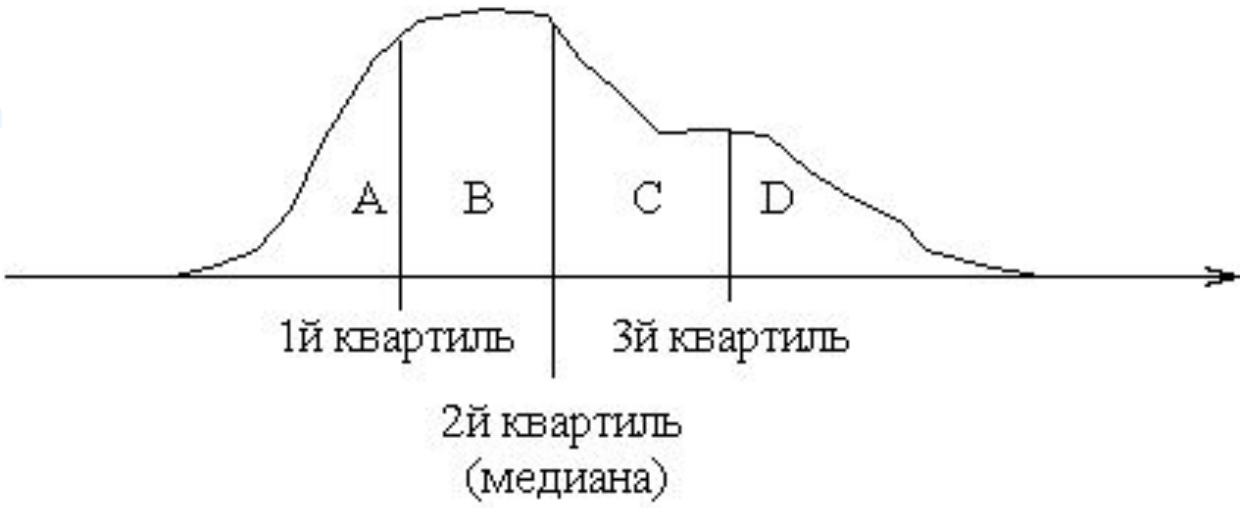
- Если значение коэффициента вариации Cv не превышает 33%, то изучаемая совокупность считается однородной.

Квантили

Квантиль – делит совокупность на заданное число равных частей по количеству значений.

Число частей может быть различным, отсюда и разные квантили – **квартили**, **децили**, **перцентили**.

Квартиль делит совокупность на четыре равные части, по 25% значений а каждой части.



Квантили

- ***Дециль*** делит совокупность на 10 равных частей, по 10% значений в каждой части.

- ***Перцентиль*** делит совокупность на 100 равных частей.

- Например,

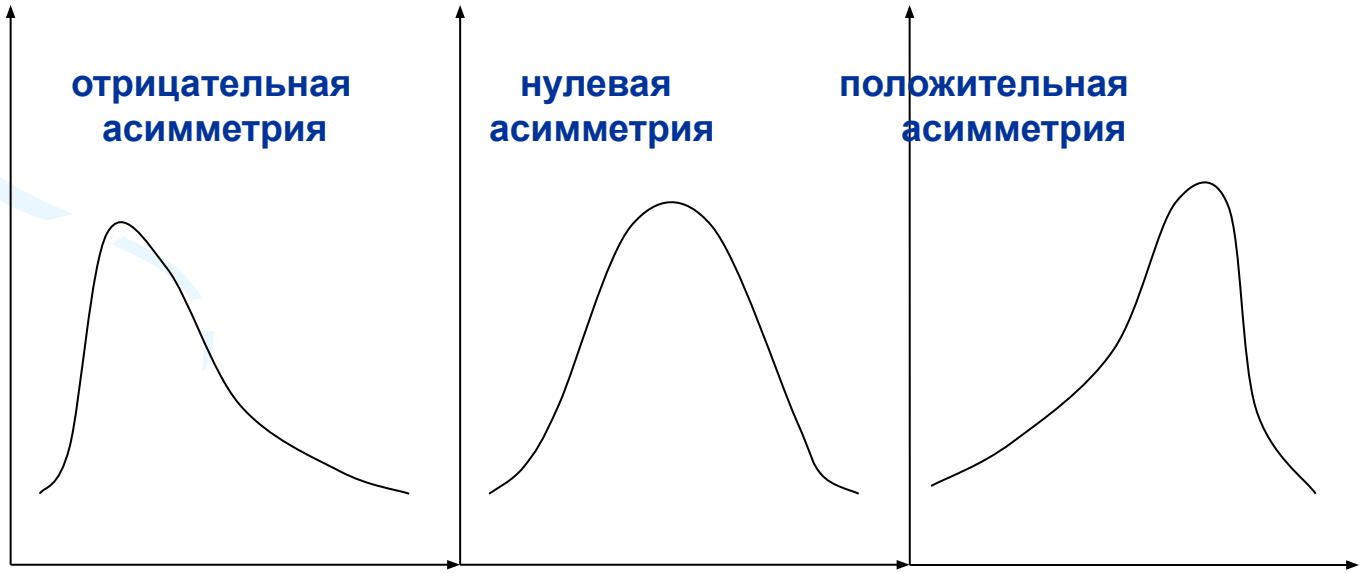
$$D_1 = C_{10},\ Q_1 = C_{25},\ Md = C_{50},\ Q_3 = C_{75} \text{ и т.д.}$$

- **<u>Необходимость расчета квантилей</u>** вызвана теми же причинами, что и расчет медианы: низкая чувствительность к случайным резким отклонениям значений признака.

Меры формы

Асимметрия (As) – степень отклонения распределения от симметричного распределения, характерного для нормальной кривой.

- Асимметрия *As* принимает значения в диапазоне от –3 до +3.
- As = 0, распределение симметрично;
- As < 0, левосторонняя асимметрия,
- As > 0, правосторонняя асимметрия.



Меры формы

Эксцесс (Ex) – степень островершинности кривой распределения.

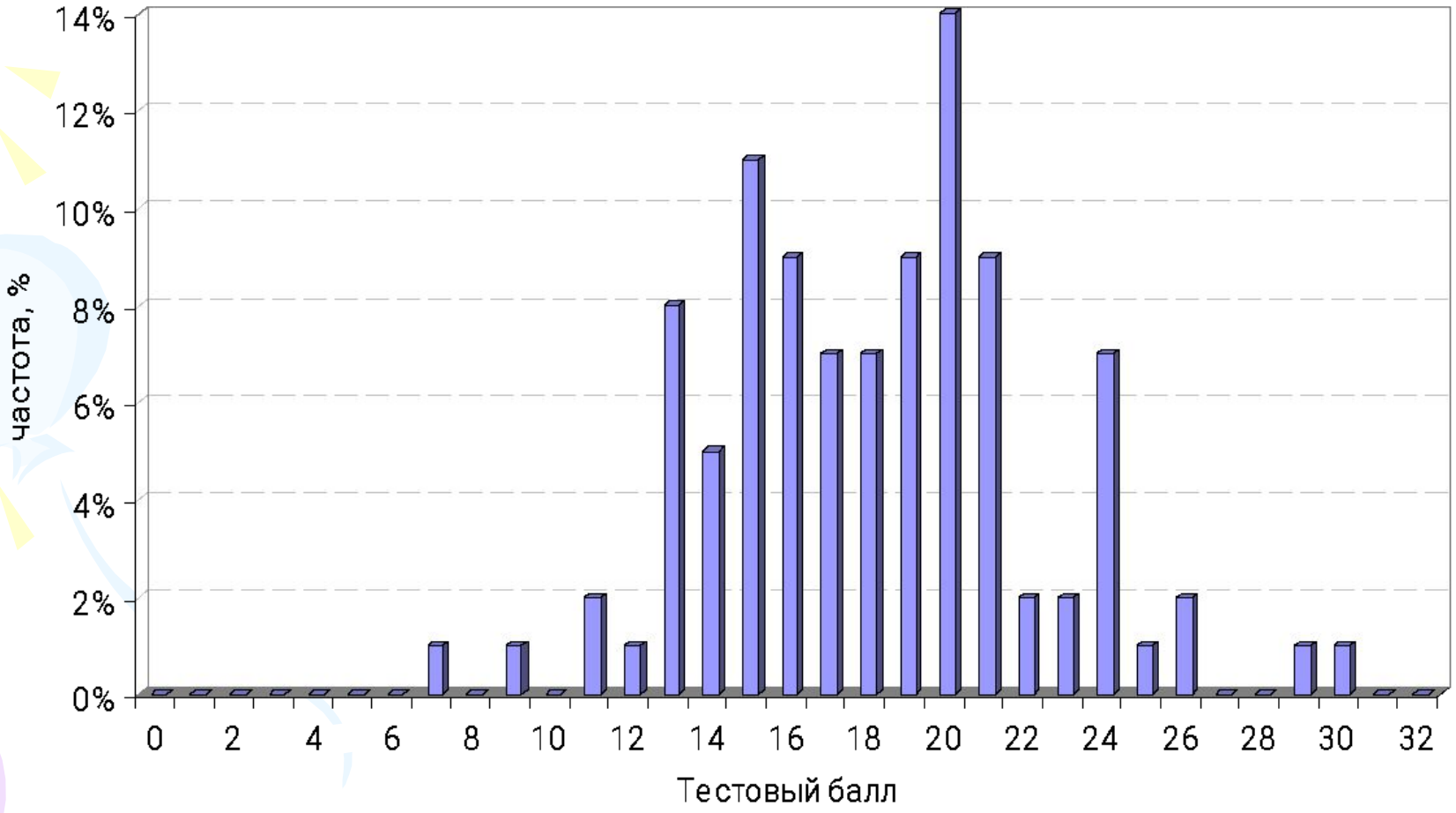
Эксцесс (Ex) принимает значения в диапазоне от –3 до +3.

- ***Ex*** = 0, распределение средневершинно;
- ***Ex*** < 0, плосковершинная кривая,
- ***Ex*** > 0, островершинная кривая.

Понятие «эксцесс» применимо лишь к унимодальным распределениям. Если две моды, то говорят об эксцессе кривой в окрестности каждой моды.

Меры асимметрии и эксцесса можно использовать для сравнения различных распределений.

Гистограмма распределения результатов тестирования



План апробации тестовых заданий

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____

Спасибо за внимание!

Ваши вопросы
и пожелания