

## 10 класс

### Проект к главе 5 «Обработка статистических данных»

#### **Теоретический минимум**

Статистические методы обработки данных распространены очень широко. Среди всего множества их применений можно указать как минимум одно, которое непосредственно касается лично вас - это обработка результатов контрольных тестов. На статистике построена вся специальная дисциплина, посвященная созданию и применению тестов - тестология. Именно с помощью этого инструментария готовятся и обрабатываются материалы ЕГЭ, так что вам будет полезно знать как именно это делается.

В предлагаемом проекте вам предстоит проанализировать качество контрольных тестов по результатам их выполнения группой учащихся. При подготовке контрольного теста проверку на качество проходят как отдельные задания так и тест в целом.

Мы не будем готовить тест, мы проанализируем уже результаты его применения.

Первое, что нам нужно будет сделать - оценить качество заданий, из которых тест составляется. Основные показатели качества тестовых заданий вводятся в классической теории тестов. К ним относят такие статистические характеристики как трудность и дискриминативность (дифференцирующая способность).

*Трудность* определяется как доля учащихся, справившихся с заданием. Более точное определение такое: характеристика тестового задания, выраженная процентом от количества испытуемых репрезентативной выборки, верно выполнивших задание. Поэтому для оценки качества самих заданий необходимо опробовать составленные тестовые задания на репрезентативной выборке учащихся. В обычных школьных условиях обеспечить репрезентативность невозможно, это задача специалистов. Но для обычных тематических тестов в этом нет необходимости. Достаточно обеспечить необходимый объем выборки. Следует помнить, что чем больше выборка, тем достовернее данные. Заслуживают внимание статистические данные по заданию, которое выполняли не менее 250 тестируемых.

Необходимо правильно истолковывать понятие трудности. На элементарном примере рассмотрим её значение. Из 100 учащихся выполнили первое задание 30 учащихся, а второе – 60. Это означает, что второе задание менее трудное и его надо поставить вначале теста (если это тест для текущей проверки усвоения материала, т.е. обучающий). Некоторые специалисты пользуются обратной величиной (доля тех, кто с заданием не справился), называемой индекс трудности. В ЕГЭ используют просто процент выполнения от всех приступивших к выполнению. Значение трудности – величина условная, поскольку зависит от выборки. Для

сильных и слабых групп это значение будет меняться. Значения трудности меньше 20 и больше 80 считают критическими, и задания с такими значениями трудности в тест стараются не включать.

Вторая характеристика качества заданий называется *дискриминативность* или дифференцирующая способность. Эта характеристика определяется как способность отделять испытуемых с высоким общим баллом по тесту от тех, кто получил низкий балл, т.е. насколько точно задание дает возможность провести различие по определенному измеряемому признаку между экзаменуемыми с хорошей и не очень хорошей подготовкой. Задание, на которое одинаково хорошо могут ответить все экзаменуемые не обладает хорошей дифференцирующей способностью, поскольку не дает никакой информации об относительных уровнях результатов. Самый простой и наглядный способ вычисления дискриминативности – это применение метода крайних групп (метод 27), когда при расчете учитываются результаты учащихся наиболее и наименее успешно справившихся с тестом.

Определяют  $N_{\text{сильные}}$  – общее количество испытуемых в сильной группе,

$N_{1\text{сильные}}$  – количество учащихся в сильной группе верно выполнивших задание,

$N_{\text{слабые}}$  – общее количество испытуемых в слабой группе,

$N_{1\text{слабые}}$  – количество учащихся в слабой группе верно выполнивших данное задание.

Дискриминативность вычисляют как разность долей испытуемых из сильной (27%) и слабой (27%) групп, правильно выполнивших задание.

$$D = \frac{N_{1\text{сильн}}}{N_{\text{сильн}}} - \frac{N_{1\text{слаб}}}{N_{\text{слаб}}}$$

Значение дискриминативности может изменяться от -1 до +1. Задание со значением близким к 1 правильно разделяет учащихся, говорит о том, что большинство сильных учащихся справились с заданием, а слабым это не удалось. Нулевое значение говорит о том, что доли справившихся с заданием в сильной и слабой группах будут равны, задание нуждается в корректировке.

Причин низкой дискриминативности задания может быть много, например::

- а) Излишняя сложность или запутанность формулировки;
- б) Неоднозначно понимаемое условие;
- в) Подсказка в условии;
- г) Опора на память, а не на мыслительные навыки при выполнении задания;
- д) Наличие двух или более правильных ответов;
- е) Наличие «терминологической или логической ловушки» в условии или ответах.

Параметр не может вам указать на конкретную ошибку - но фиксирует сам факт ее наличия.

На основании значения дискриминативности можно решить что делать с каждым конкретным заданием: задание с отрицательным значением чаще всего требуется удалить, либо существенно переработать. В тест должны попасть задания со значением дискриминативности выше 0,2.

Существуют и другие способы вычисления дискриминативности, которые используются в практике измерений. Рассмотренный нами самый простой и вполне надежный. После первичной апробации теста (на малой выборке испытуемых) его разработчик должен организовать выполнение этого теста большой представительной (репрезентативной) выборкой. Это делается для того, чтобы определить, как часто встречается тот или иной тестовый балл.

Накопив эти данные, мы можем оценивать уже не отдельные задания, а тест в целом и результаты его применения. Рассмотрим некоторые понятия и расчет основных статистических характеристик по тесту в целом (меры центральной тенденции, мода, дисперсия, стандартное отклонение, коэффициенты корреляции).

На основе анализ матрицы результатов тестирования поэтапно получим основные статистические характеристики по тесту в целом и интерпретируем их. Затем эту же матрицу используем для расчета дискриминативности каждого задания. Для упрощения работы в качестве наглядности будем использовать небольшую выборку 20 тестируемых и дихотомический способ оценивания выполнения каждого задания (1 – правильно выполнено, 0 – неправильно выполнено). Вначале сформируем таблицу, так называемую матрицу результатов тестирования. По вертикали, которой располагаются профили (линейная матрица из 0 и 1) ответов тестируемых на каждое задание теста, по горизонтали результаты выполнения каждым тестируемым заданий теста.

Введем обозначения:

$i$  – номер тестируемого,  $j$  - номер задания,

$x_{ij}$  – результат выполнения  $i$ -м тестируемым  $j$ -го задания

$$x_{ij} = \begin{cases} 1, & \text{при правильном ответе } i\text{-го тестируемого на } j\text{-е задание;} \\ 0, & \text{при неправильном ответе } i\text{-го тестируемого на } j\text{-е задание;} \end{cases}$$

Таблица 1. Матрица результатов тестирования

Номер тестируемого $i$	Номер задания $j$										Индивидуальный балл $X_i$
	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
3	0	0	0	0	0	0	0	1	0	0	1
4	1	1	0	1	1	1	1	1	1	1	9
5	1	0	1	1	1	1	0	0	0	0	5
6	1	1	1	0	0	0	0	1	0	0	4
7	1	1	1	1	0	1	0	0	0	0	5
8	1	1	1	1	0	0	0	0	0	0	4
9	1	1	1	1	1	1	1	1	1	0	9
10	1	1	1	1	1	0	1	0	0	0	6
11	0	0	0	0	0	0	0	0	0	0	0(удалить)
12	1	1	1	1	1	1	1	1	1	1	10(удалить)
13	0	1	1	0	0	0	0	0	1	1	4
14	1	0	1	1	0	1	1	1	1	0	7
15	0	1	1	0	1	1	1	0	0	0	5
16	1	0	1	1	1	1	0	0	0	0	5
17	1	1	1	1	1	1	1	0	1	0	8
18	1	1	1	1	1	0	1	0	0	0	6
19	1	0	0	0	0	0	0	0	0	0	1
20	1	1	1	0	1	0	0	1	0	0	5
Число правильных ответов $R_j$ для 20 заданий	16	14	15	12	11	10	8	7	6	3	

Первое, что мы сделаем - очистим данные. Из таблицы удаляются строки 11 и 12 поскольку они не несут необходимой информации. По профилям этих учеников получается, что данный тест непригоден для истинной оценки уровня подготовленности этих тестируемых. Для 11 этот тест слишком сложный, а для 12 – легкий. В реальной

педагогической практике такая ситуация тоже может возникнуть. Безусловно, из обработки такие результаты исключаются.

Хороший тест для итогового контроля знаний и умений учащихся обладает приблизительно 70% точностью измерения результатов, находящихся в центре распределения, результаты 3-5% учащихся, приходящихся по краям распределения отбрасываются, так как отражают уровень подготовленности учащихся с очень большой ошибкой измерения. Удаляемые в таблице строки – это и есть крайние значения.

Далее считаются индивидуальные баллы тестируемых  $X_i$  и количество правильных ответов  $R_j$  простым суммированием.

Для графической интерпретации результатов тестирования необходимо упорядочить матрицу, располагая числа  $R_j$  в порядке убывания (по возможности), а значения  $X_i$  в порядке возрастания значений сверху вниз.

Таблица 2. Упорядоченная матрица тестовых результатов.

Номер тестируемого $i$	Номер задания $j$										Индивидуальный балл $X_i$
	1	2	3	4	5	6	7	8	9	10	
19	1	0	0	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	1	0	0	1
2	1	1	0	0	0	0	0	0	0	0	2
8	1	1	1	1	0	0	0	0	0	0	4
6	1	1	1	0	0	0	0	1	0	0	4
13	0	1	1	0	0	0	0	0	1	1	4
7	1	1	1	1	0	1	0	0	0	0	5
20	1	1	1	0	1	0	0	1	0	0	5
5	1	0	1	1	1	1	0	0	0	0	5
16	1	0	1	1	1	1	0	0	0	0	5
15	0	1	1	0	1	1	1	0	0	0	5
1	1	1	1	1	1	1	0	0	0	0	6
10	1	1	1	1	1	0	1	0	0	0	6
18	1	1	1	1	1	0	1	0	0	0	6
14	1	0	1	1	0	1	1	1	1	0	7
17	1	1	1	1	1	1	1	0	1	0	8
9	1	1	1	1	1	1	1	1	1	0	9
4	1	1	0	1	1	1	1	1	1	1	9
Число	15	13	14	11	10	9	7	6	5	2	

правильных ответов $R_j$ для 20 заданий																		
---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Для графического представления построим частотное распределение в виде гистограммы. Но вначале представим результаты в виде ранжированного ряда.

Таблица 3. Ранжированный ряд.

Ранг	1	1	2	3	3	3	4	4	4	4	4	5	5	5	6	7	8	8
Номер	19	3	2	8	6	13	7	20	5	16	15	1	10	18	14	17	9	4
балл	1	1	2	4	4	4	5	5	5	5	5	6	6	6	7	8	9	9

Таблица 4. Частотное распределение

балл	1	2	4	5	6	7	8	9
частота	2	1	3	5	3	1	1	2

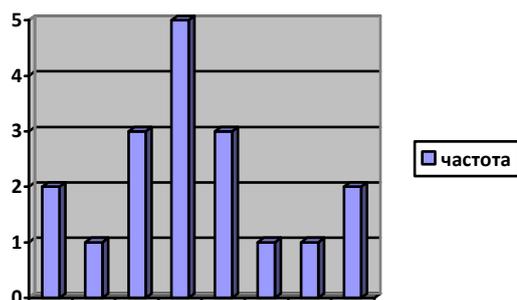


рис. 1

Сумма всех частот в группе должна быть равна числу тестируемых учеников.

Получилась гистограмма - последовательность столбцов, каждый из которых опирается на единичный интервал, а высота каждого пропорциональна частоте наблюдаемых баллов.

Гистограмма будет в большей мере похожа на нормальное распределение, если её строить для сгруппированных данных (обычное количество таких групп 12-15). В нашем примере выборка очень маленькая, поэтому ограничимся тремя группами.

Таблица 5. Сгруппированное частотное распределение.

Интервал баллов	Частота
1-3	3
4-6	12
7-9	4

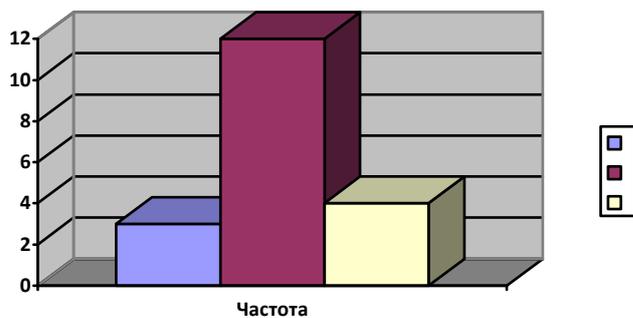


рис. 2

По этой гистограмме вырисовывается картина нормального распределения, которое в идеальном варианте должна выглядеть примерно так:

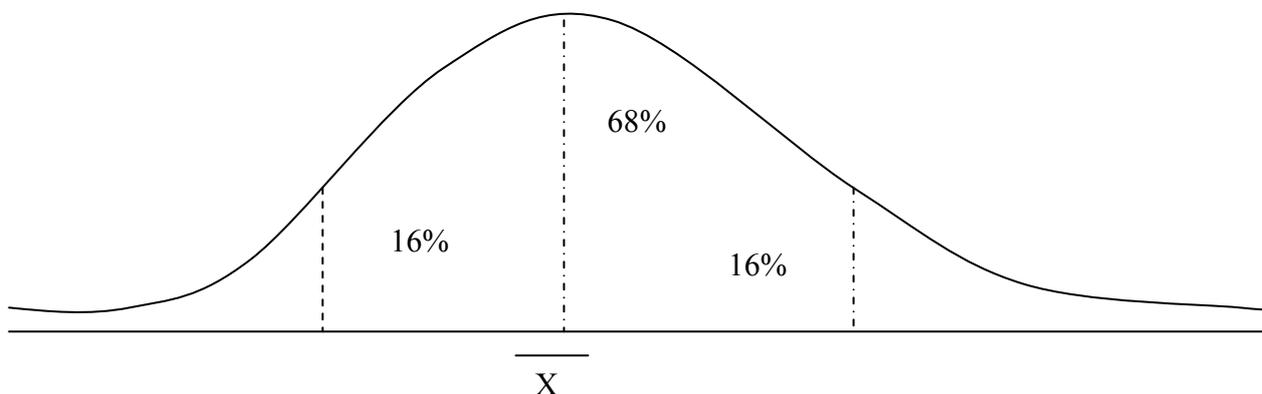


рис. 3

Визуально, по полученной кривой распределения можно оценить асимметрию (обычно она рассчитывается и на этом основании делается вывод о нормальности распределения). Нулевая асимметрия говорит о том, что тест сбалансирован по трудности заданий (рис. 3). Асимметрия распределения положительна, когда в тесте слишком много легких заданий (рис. 4)

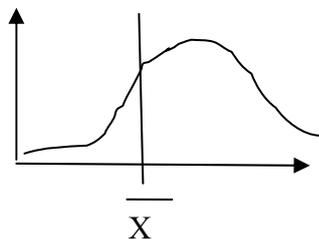


рис. 4

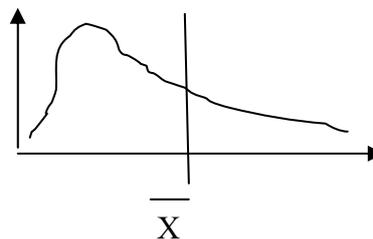


рис.5

Асимметрия распределения отрицательна при избытке в тесте трудных заданий (рис.5).

Определим некоторые другие характеристики. *Мода* – наиболее часто встречающееся значение среди результатов выполнения теста. Для нашего случая модой является 5, поскольку встретилось чаще, чем другие значения (5 раз). Если встречается два значения

моды, то распределение называют бимодальным. Нормальное распределение результатов должно быть унимодальным и симметричным. Бимодальное распределение говорит о неудачно построенном тесте, требующим внимательного анализа других характеристик для выявления причин неудачного построения. На рисунке представлено нормальное распределение результатов для абстрактных данных, это идеальный вид кривой, достаточно редко встречающийся на практике. Но если распределение приближается к такой картине, то с определенной допустимой ошибкой измерения говорят о распределении по нормальному закону.

Хороший тест обеспечивает нормальное распределение индивидуальных баллов репрезентативной выборки тестируемых, если среднее значение баллов ( $\bar{X}$ ) находится в центре распределения, остальные концентрируются вокруг, примерно 70% в центре, остальные сходят до минимума по краям. *Смещение среднего значения влево говорит о слишком трудной подборке заданий теста и наоборот, смещение вправо – о слишком легкой подборке заданий (см. выше асимметрию).* Среднее выборочное значение ( $\bar{X}$ ) вычисляется просто, поскольку это и есть среднее арифметическое индивидуальных баллов тестируемого.

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$$

Среднее значение индивидуального балла важно не само по себе, а для анализа других описательных характеристик, позволяющих сравнивать различные распределения по тестам. Можно выявить различия в качестве тестов, сравнивая несколько распределений с одинаковым средним значением. Необходимо оценить, как разбросаны эмпирические данные вокруг среднего значения, сгруппированы тесно или наоборот сильно удалены. Используют такие характеристики как дисперсия и стандартное отклонение.

*Дисперсия* отражает меру неоднородности результатов по тесту и вычисляется по формуле:

$$S_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{N - 1}$$

Можно подсчитать значение дисперсии вручную. Но на практике среднее значение чаще всего число дробное, подсчет вручную делать утомительно, лучше написать небольшой фрагмент программы, которая будет проводить автоматический подсчет, если данная характеристика будет необходима для дальнейшей обработки результатов.

```
program Dispersion;
```

```
const N = 18;
```

```

var
  balls: array[1..N] of word;
  m,d : real;
  i : Word;
  s : longint;
begin
  s:=0;
  for i:=1 to N do
  begin
    write('Введите балл ',i,'-го ученика:');
    readln(balls[i]);
    s := s + balls[i];
  end;
  m := s/n;
  d:=0;
  for i:=1 to N do
    d := d+(balls[i]-m)*(balls[i]-m);
  d := d/(N-1);
  Writeln('Дисперсия:',d);
end.

```

Для наших данных результат работы:

*Дисперсия: 5.39869281045752*

В наших условиях его можно округлить до двух знаков: 5.40

Если дисперсия не единственная статистическая характеристика, которую вы вычисляете, может оказаться удобнее спланировать расчет в электронной таблице. Для Microsoft Excel достаточно ввести данные единым диапазоном и рассчитать параметры, вставив формулы.

Для нашего случая:

1 1 2 4 4 4 5 5 5 5 5 6 6 6 7 8 9 9 =СРЗНАЧ(A1:R1) =ДИСП(A1:R1)  
 Результат вычисления, конечно, будет точно таким же.

Использование электронной таблицы может оказаться менее удобным - если объем данных будет большим, или вычисляемые характеристики не предусмотрены в стандартном пакете функций, или ресурсы ограничены.

Низкая дисперсия говорит о слабом разделении тестируемых по уровню подготовки, а излишне высокое значение дисперсии говорит об искаженной картине распределения, а значит проблемах в тесте.

Для анализа чаще используют стандартное отклонение, которое находят как корень квадратный из дисперсии, и это значение увеличенное в 3 раза сравнивают со средним выборочным значением. Считается, что они должны быть приблизительно равны.

$$S_x = \sqrt{S_x^2}; \quad \bar{X} \approx 3 * S_x$$

Существует еще несколько характеристик для детального анализа результатов тестирования. Наибольшее внимание обращают на корреляцию (связь между исследуемыми объектами). Необходимо не только установить наличие связи, но и выбрать вид и форму показателя для оценки этой связи. Это одна из характеристик обеспечивающих валидность теста. *Валидность* теста - соответствие теста целям контроля.

Для определения связи между различными наборами данных используют коэффициент корреляции Пирсона. В случае необходимости определения связи между заданиями в одном тесте используют преобразованный коэффициент Пирсона называемый «коэффициент  $\varphi$ ».

$$\varphi_{jl} = \frac{p_{jl} - p_j p_l}{\sqrt{p_j q_j * p_l q_l}}; \text{ где } p_{jl} - \text{доля тестируемых верно выполнивших оба задания, т.е.}$$

получивших по 1 баллу за оба задания;  $p_j$  – доля тестируемых, правильно выполнивших  $j$ -е задание;  $q_j=1-p_j$ ;  $p_l$  – доля тестируемых, правильно выполнивших  $l$ -е задание;  $q_l=1-p_l$ .

Результаты подсчета коэффициентов корреляции между результатами по отдельным заданиям теста сводят в таблицу для удобства интерпретации. Коэффициенты корреляции для итоговых тестов должны быть в пределах (0; 0,3). Поскольку итоговый предметный тест гомогенный, то корреляция должна быть невысокой положительной. Высокое значение говорит о зависимости заданий друг от друга, это недопустимо в итоговом тесте. Отрицательные значения говорят об отсутствии предметной чистоты содержания теста, такие задания, как правило, удаляются. Для тематических тестов корреляция будет достаточно высокой, так как и предполагается, результаты выполнения заданий тематического теста слабо варьируются, поскольку отражают исходное содержание.

Такой коэффициент корреляции используется в случае, когда в тесте задания одного типа, значит и распределение задается в одной шкале. На практике в итоговом тесте используют задания разных типов. В случае, когда один набор значений распределения задается в дихотомической шкале, а другой в интервальной, используют коэффициент бисериальной корреляции. В практической работе используют коэффициент точечно-бисериальной корреляции, поскольку он проще в расчетах и обладает существенным преимуществом: его расчетное значение не выходит за границы интервала  $[-1; +1]$ , в отличие от значений коэффициента бисериальной корреляции.

$$(r_{pbis})_j = \frac{(\overline{X_1})_j - (\overline{X_0})_j}{S_x} \sqrt{\frac{(N_1)_j * (N_0)_j}{N(N-1)}};$$

Где  $(\overline{X_1})_j$  - среднее значение индивидуальных баллов тестируемых, выполнивших верно  $j$ -е задание теста;  $(\overline{X_0})_j$  - среднее значение индивидуальных баллов тестируемых выполнивших  $j$ -е задание теста неверно;  $S_x$  – стандартное отклонение по множеству значений

индивидуальных баллов;  $N$  – общее число тестируемых;  $(N_1)_j$  – число тестируемых, верно выполнивших  $j$ -е задание теста;  $(N_0)_j$  – число тестируемых, неверно выполнивших  $j$ -е задание теста;

*Например*, для нашей таблицы такой расчет в Excel будет выглядеть примерно так:

	Задание	1	2	3	4	5	6	7	8	9	10	Балл
1												
2	Ученик											
3	1	1	0	0	0	0	0	0	0	0	0	1
4	2	0	0	0	0	0	0	0	1	0	0	1
5	3	1	1	0	0	0	0	0	0	0	0	2
6	4	1	1	1	1	0	0	0	0	0	0	4
7	5	1	1	1	0	0	0	0	1	0	0	4
8	6	0	1	1	0	0	0	0	0	1	1	4
9	7	1	1	1	1	0	1	0	0	0	0	5
10	8	1	1	1	0	1	0	0	1	0	0	5
11	9	1	0	1	1	1	1	0	0	0	0	5
12	10	1	0	1	1	1	1	0	0	0	0	5
13	11	0	1	1	0	1	1	1	0	0	0	5
14	12	1	1	1	1	1	1	0	0	0	0	6
15	13	1	1	1	1	1	0	1	0	0	0	6
16	14	1	1	1	1	1	0	1	0	0	0	6
17	15	1	0	1	1	0	1	1	1	1	0	7
18	16	1	1	1	1	1	1	1	0	1	0	8
19	17	1	1	1	1	1	1	1	1	1	0	9
20	18	1	1	0	1	1	1	1	1	1	1	9
21	Средний балл - выполнивших	5,47	5,62	5,64	6,36	6,4	6,56	7,14	5,83	7,4	6,5	
22	Средний балл - невыполнивших	3,33	3,8	3,25	3,14	3,5	3,67	3,82	4,75	4,23	4,94	2,323509
23												
24	К-т бисериальной корреляции:	0,35	0,36	0,44	0,7	0,64	0,64	0,72	0,23	0,63	0,22	
25												

Рис.6.

Расчет выполняется в три этапа:

1. Рассчитывается суммарный балл по каждому ученику (например: =СУММ(B3:K3) - для первого ученика, и т.д.) и стандартное отклонение (=СТАНДОТКЛОН(L3:L20) - ячейка L22) - с помощью встроенных функций.

2. Рассчитывается средний балл выполнивших и не выполнивших каждое задание. Формула для подсчета среднего балла выполнивших задание выглядит в этом примере для первого задания так: =СУММЕСЛИ(B3:B20;1;L\$3:L\$20)/СЧЁТЕСЛИ(B3:B20;1)

То есть мы суммируем с условием и подсчитываем количество с условием. Обратите внимание, один из диапазонов зафиксирован.

3. Рассчитываем коэффициент корреляции. Для первого столбца по такой формуле: =(B21-B22)\*КОРЕНЬ(СЧЁТЕСЛИ(B3:B20;1)\*СЧЁТЕСЛИ(B3:B20;0)/(18\*17))/L\$22

Анализ значений точечно–бисериальной корреляции позволяет сделать вывод о валидности заданий, т.е. насколько задание пригодно для измерения в соответствии с общей целью построения теста. Если эта цель итогового тестирования – дифференцировать по уровню подготовки учащихся, то валидные задания должны четко разделять сильно и слабо подготовленных в тестируемой группе. Значение коэффициента должно приближаться к 0,5. Рекомендуется оценить разность  $(\bar{X}_1)_j - (\bar{X}_0)_j$ . Чем выше значение этой разности, тем лучше работает задание на дифференциацию учащихся.

### **Задание 1.**

Рассмотрим расчет дискриминативности (дифференцирующей способности) для 10 заданий на примере имеющихся результатов выполнения теста 18 учащимися. Напоминаем, что это слишком малый объем выборки для корректных выводов.

К сильной группе  $N_{\text{сильные}}$  (таблица 3) можно отнести четырех учащихся, получивших 7, 8 и 9 баллов (27% от начала рейтинга), к слабой группе  $N_{\text{слабые}}$  относятся учащиеся с 1, 2 и 4 баллами, таких - шесть (27% от конца рейтинга).

Воспользуемся приведенной ранее формулой: 
$$D = \frac{N_{1\text{сильн}}}{N_{\text{сильн}}} - \frac{N_{1\text{слаб}}}{N_{\text{слаб}}}$$

$$D_1 = 4/4 - 4/6 = 1 - 0,66 = 0,34$$

$$D_2 = 3/4 - 4/6 = 0,75 - 0,66 = 0,09$$

$$D_3 = 3/4 - 3/6 = 0,75 - 0,5 = 0,25$$

$$D_4 = 1 - 1/6 = 0,83$$

$$D_5 = 3/4 - 0 = 0,75$$

$$D_6, D_7 = 1 - 0 = 1$$

$$D_8 = 3/4 - 2/6 = 0,75 - 0,33 = 0,42$$

$$D_9=1-1/6=0,84$$

$$D_{10}=1/4-1/6=0,25-0,16=0,09$$

В итоговый тест могут быть включены задания с дискриминативностью более 0,2. Задания с нулевым или отрицательным значением дискриминативности следует удалить из теста, поскольку их формулировки некорректны. Чем ближе значение к единице, тем лучше задание разделяет тестируемых по уровню подготовленности.

В нашем тесте задания 2 и 10 следует существенно переработать.

Все перечисленные нами характеристики позволяют оценить качество отдельных заданий, весь тест - но не влияют на оценку тестируемого. Хотелось бы взять в качестве оценки тестируемого набранные баллы - но сделать этого нельзя, потому что в этом случае данные разных тестов нельзя сравнивать. Сравнивать же их необходимо - иначе как можно будет, например, провести конкурс аттестатов? Вариантов итогового теста должно быть очень много - как собрать полностью эквивалентные варианты? Это тоже непростая проблема в настоящее время не разрешенная.

Поэтому нам нужен механизм, который позволяет превратить "сырой балл" в "итоговый". Таким инструментом является специальная шкала. То есть набранный "сырой" балл должен быть пересчитан в специальные абсолютные единицы. Конечно, в текущих тестах - обучающих или проверочных никакого смысла в построении единой шкалы нет - все равно результаты сравниваться не будут, а вот в итоговых тестах, например, в ЕГЭ такой механизм необходим. Эта процедура перевода первичных или "сырых" баллов в тестовый балл называется шкалирование.

### ***Задание 2.***

Воспользуйтесь результатами выполнения теста (таблица 6) и рассчитайте основные характеристики заданий и теста в целом (трудность, дискриминативность, мода, дисперсия, стандартное отклонение, коэффициенты корреляции между результатами выполнения заданий и результатами отдельных учеников).

Постройте распределение и сделайте обоснованный вывод о качестве тестового материала. Для удобства проведения расчетов воспользуйтесь электронной таблицей.

### ***Задание 3.***

Проанализируйте предложенное распределение участников экзамена по информатике по полученным первичным баллам в ходе ЕГЭ 2012 года и ответьте на следующие вопросы:

1. Можно ли считать данное распределение нормальным?

2. Какие выводы можно сделать по качеству тестового материала?

3. Можно ли вычленить основную проблему, которую необходимо разрешить?

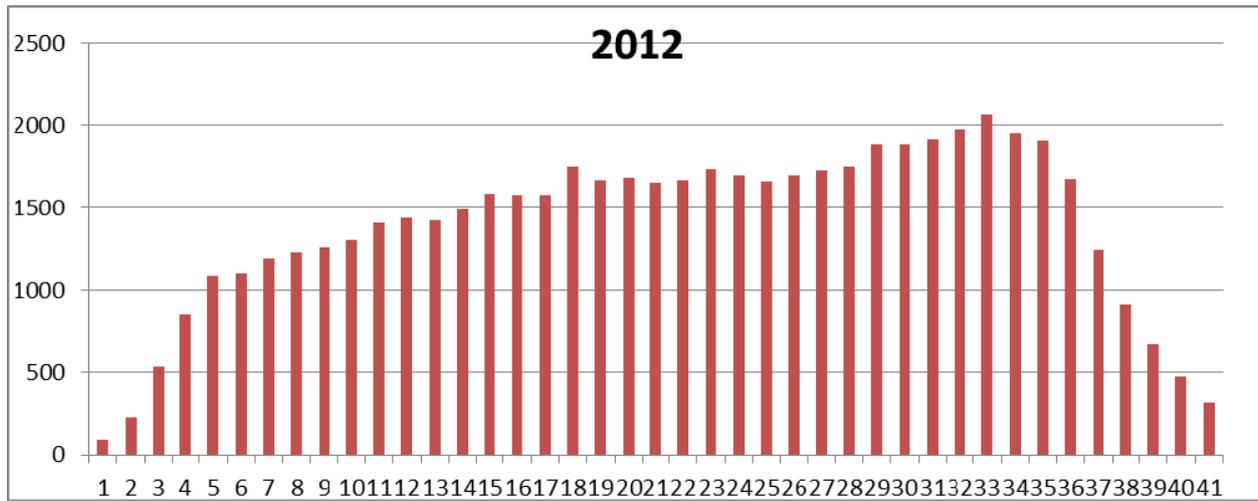


Рисунок 7. Распределение участников экзамена по полученным первичным баллам

