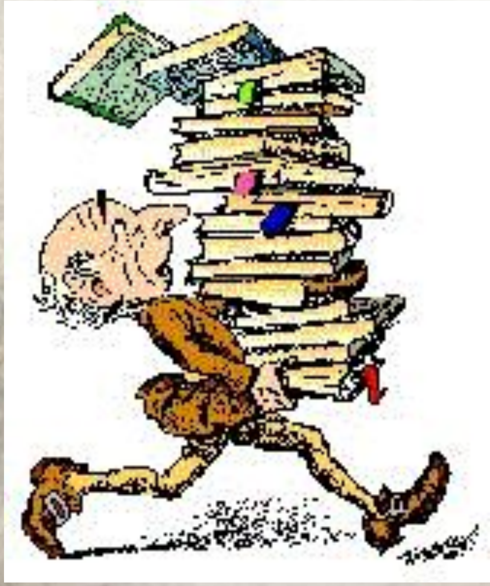
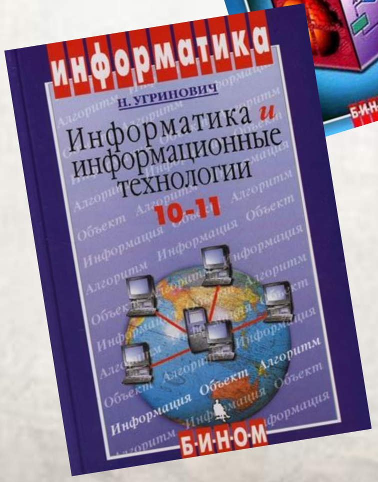
Элементы статистики, комбинаторики и теории вероятностей в изучении информатики

Все перспективные государственные образовательные документы последних лет содержат вероятностно-статистическую линию в курсе информатики 5-9 классов.





Информация – это …

«Информация есть информация, а не материя и не энергия».

Н. Винер, «Кибернетика, или Управление и связь в животном и машине»

Информация – одно из базовых понятий в науке (как *материя*, *энергия*), поэтому нет более четкого определения:

- невозможно выразить через более простые понятия
- объясняется только на примерах или в сравнении с другими понятиями

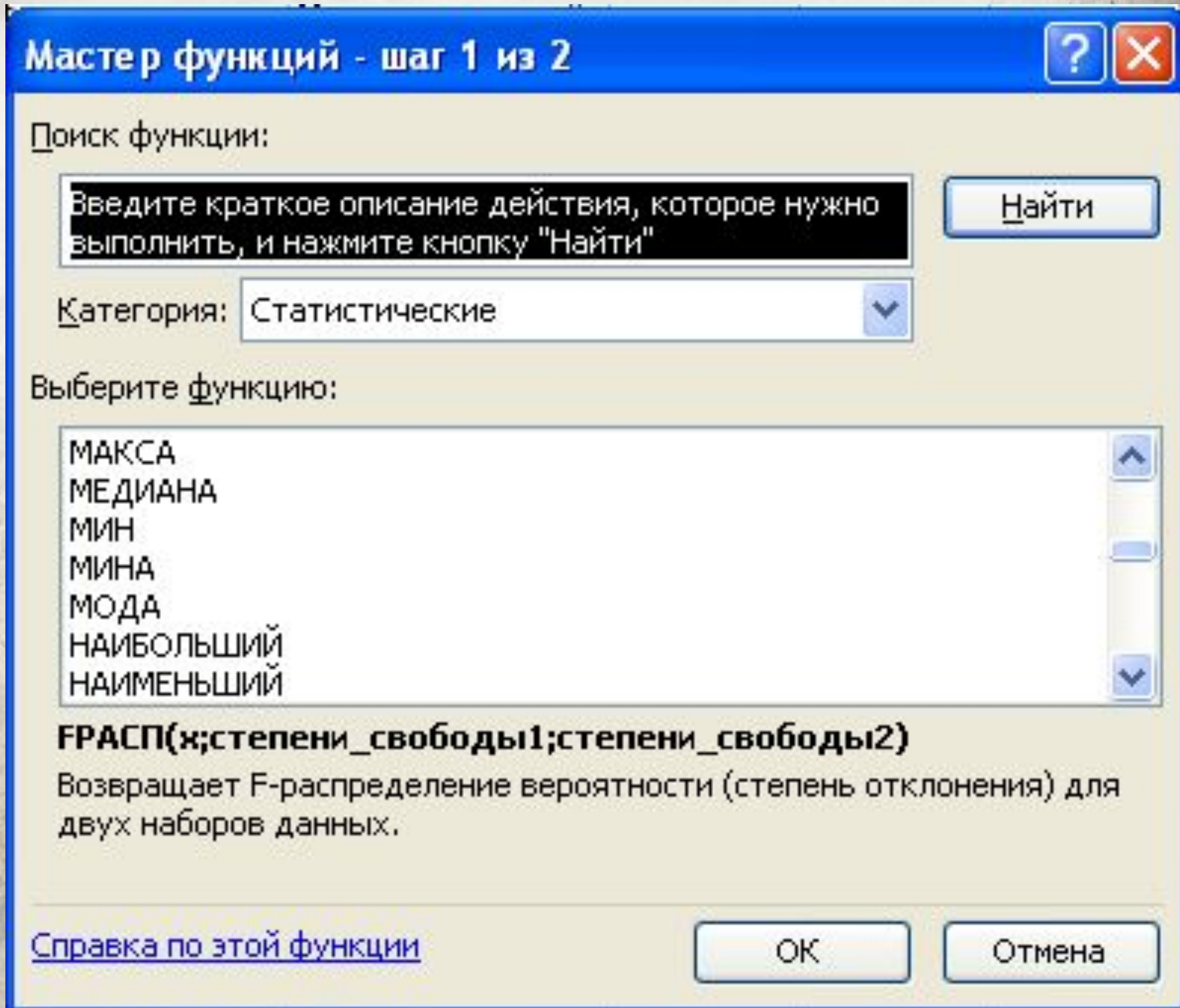
Статистика

Математика

Статистические данные	Представление данных в виде таблиц, диаграмм, графиков	**5 часов**	**Математика**	**6 класс**
	Средние результаты измерений Понятие о статистическом выводе на основе выборки Понятие и примеры случайных событий	**5 часов** **5 часов**	**Алгебра** **Алгебра**	**7 класс** **8 класс**

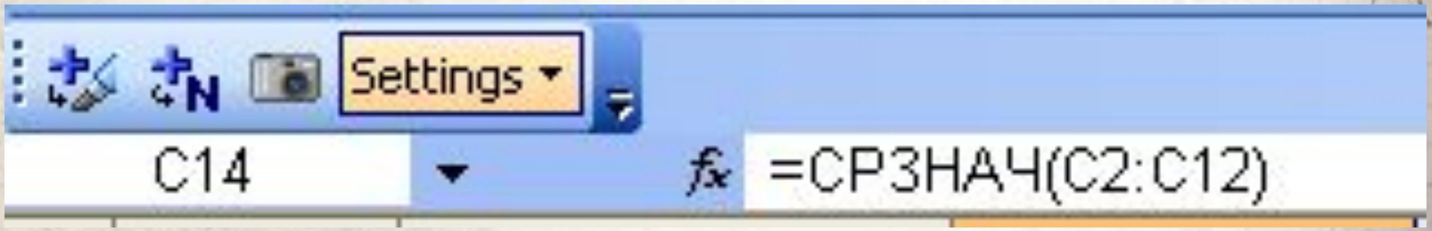
Информатика (7 класс)

Табличные вычисления на компьютере (18 час)	0	0
Электронные таблицы MS Excel. Назначение и возможности.	1	1
Операции с ячейками таблицы.	1	1
Проверка орфографии. Автоматизация ввода данных.	0,5	0,5
Работа с формулами. Абсолютная и относительная адресация.	1	1
Обработка данных с помощью математических функций.	0,5	1,5
Обработка данных с помощью статистических функций.	1	1
Создание и редактирование диаграмм.	1	1
Печать электронных таблиц.	0,5	0,5
Практикум: работа VI	0	4

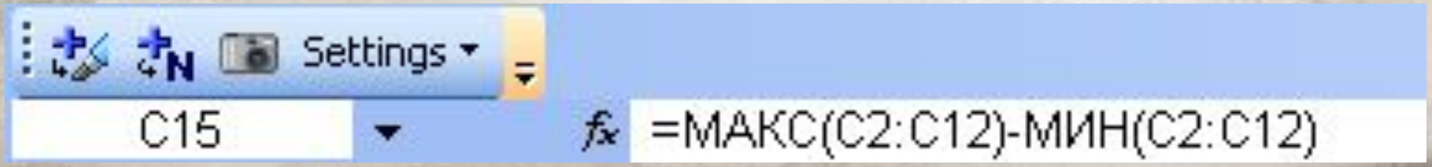

Мастер функций - шаг 1 из 2
Поиск функции:
Введите краткое описание действия, которое нужно выполнить, и нажмите кнопку "Найти"
Найти
Категория: Статистические
Выберите функцию:
МАКСА
МЕДИАНА
МИН
МИНА
МОДА
НАИБОЛЬШИЙ
НАИМЕНЬШИЙ
FРАСП(x;степени_свободы1;степени_свободы2)
Возвращает F-распределение вероятности (степень отклонения) для двух наборов данных.
Справка по этой функции
ОК
Отмена

Статистические характеристики

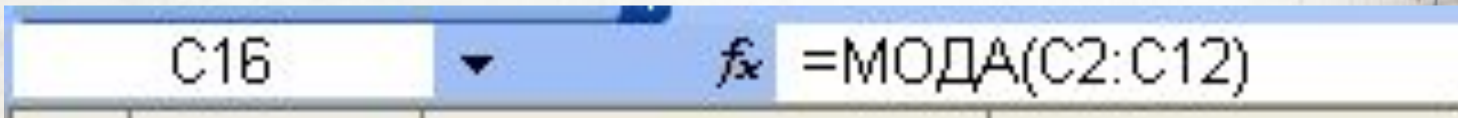
Средним арифметическим ряда чисел называется частное от деления суммы этих чисел на число слагаемых.



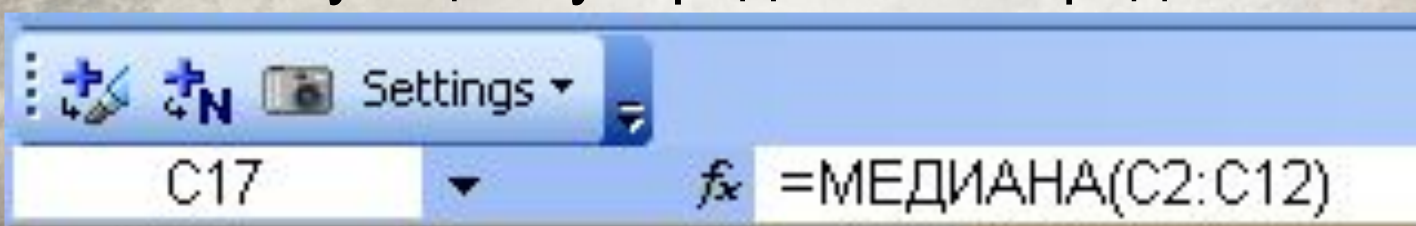
Размахом ряда чисел называется разность между наибольшим и наименьшим из этих чисел.



Модой ряда чисел называется число, наиболее часто встречающееся в данном ряду.

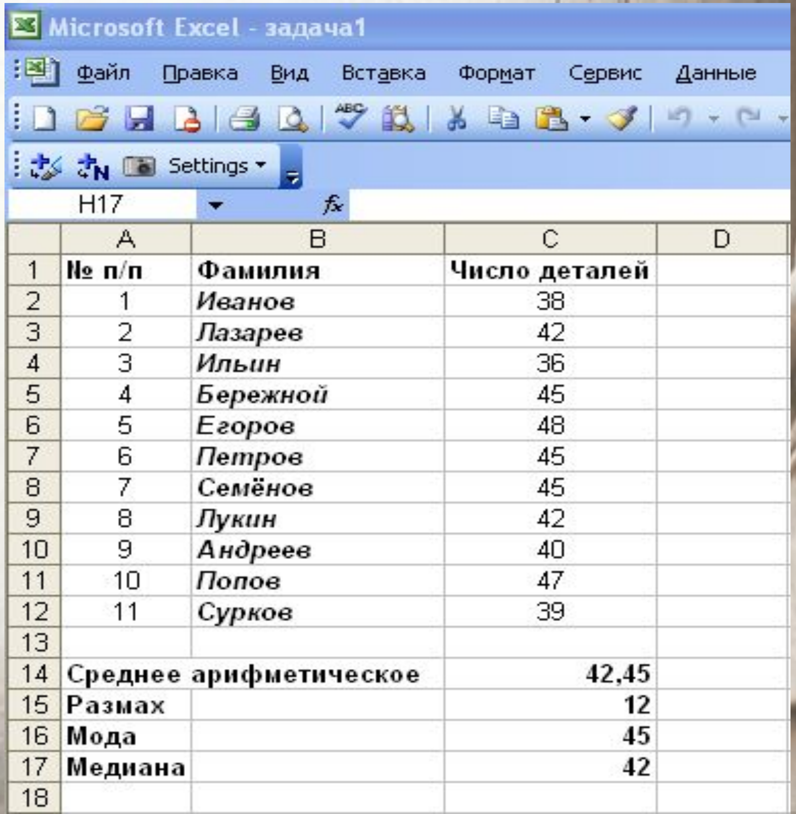


Медианой упорядоченного ряда чисел с нечётным числом членов называется число, записанное посередине, а медианой упорядоченного ряда чисел с чётным числом членов называется среднее арифметическое двух чисел, записанных посередине. **Медианой произвольного ряда чисел** называется медиана соответствующего упорядоченного ряда.

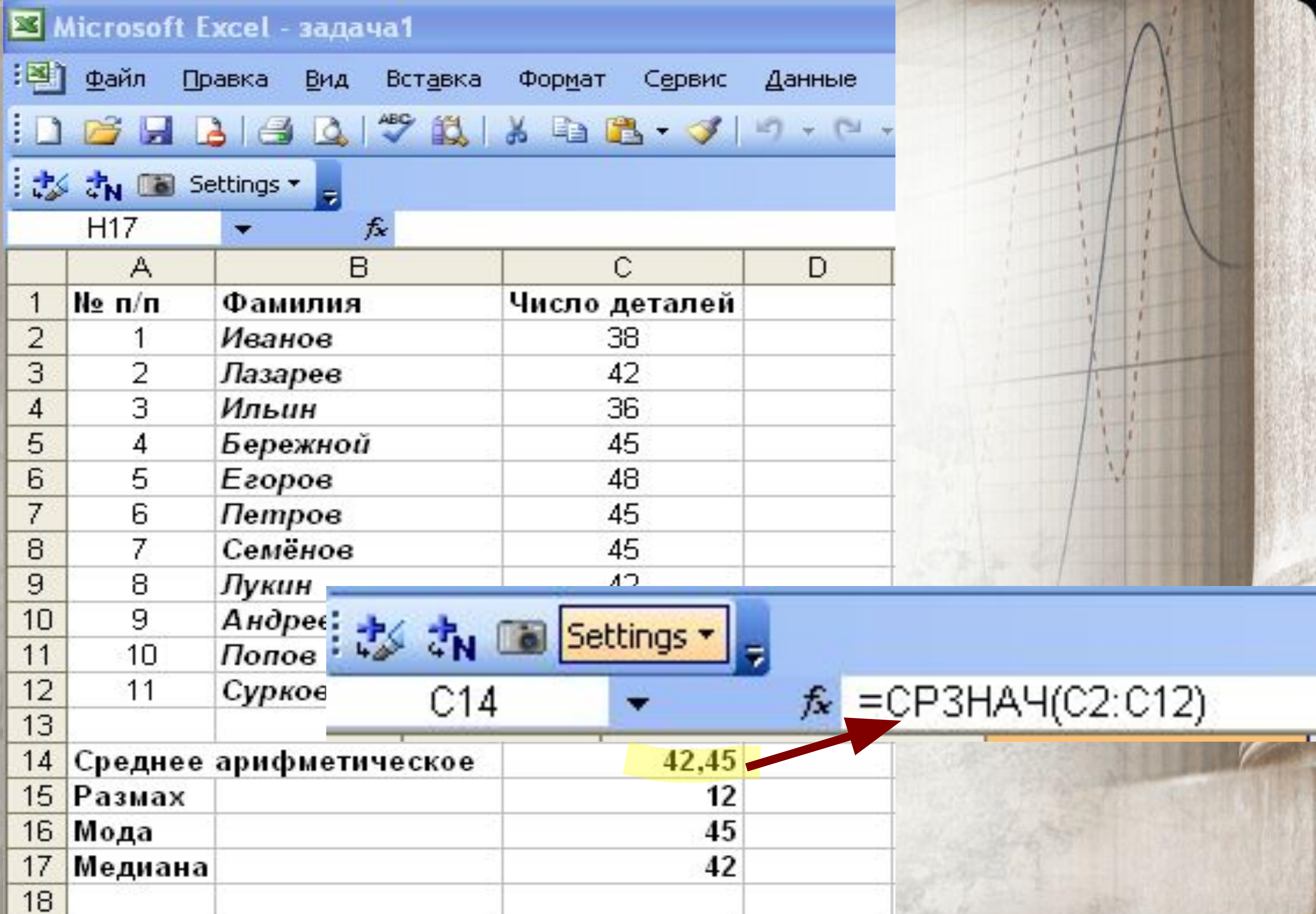


Задача

В таблице показано число деталей, изготовленных за смену рабочими одной бригады. Для представленного в таблице ряда чисел найти среднее арифметическое, размах и моду. Какой смысл каждого из этих показателей?



	A	B	C	D
1	№ п/п	Фамилия	Число деталей	
2	1	Иванов	38	
3	2	Лазарев	42	
4	3	Ильин	36	
5	4	Бережной	45	
6	5	Егоров	48	
7	6	Петров	45	
8	7	Семёнов	45	
9	8	Лукин	42	
10	9	Андреев	40	
11	10	Попов	47	
12	11	Сурков	39	
13				
14	Среднее арифметическое		42,45	
15	Размах		12	
16	Мода		45	
17	Медиана		42	
18				



	A	B	C	D
1	**№ п/п**	**Фамилия**	**Число деталей**	
2	1	***Иванов***	38	
3	2	***Лазарев***	42	
4	3	***Ильин***	36	
5	4	***Бережной***	45	
6	5	***Егоров***	48	
7	6	***Петров***	45	
8	7	***Семёнов***	45	
9	8	***Лукин***	[illegible]	
10	9	***Андрее***		
11	10	***Попов***		
12	11	***Сурко***		
13				
14	**Среднее арифметическое**		**42,45**	
15	**Размах**		**12**	
16	**Мода**		**45**	
17	**Медиана**		**42**	
18				

Microsoft Excel - задача1

Файл Правка Вид Вставка Формат Сервис Данные

Settings

H17 f_x

	A	B	C	D
1	**№ п/п**	**Фамилия**	**Число деталей**	
2	1	***Иванов***	38	
3	2	***Лазарев***	42	
4	3	***Ильин***	36	
5	4	***Бережной***	45	
6	5	***Егоров***	48	
7	6	***Петров***	45	
8	7	***Семёнов***	45	
9	8	***Лукин***	42	
10	9	***Андреев***	40	
11	10	***Попов***	47	
12	11	***Сурков***	39	
13				
14	**Среднее арифметическое**		**42,45**	
15	**Размах**		**12**	
16	**Мода**		**45**	
17	**Медиана**		**42**	
18				

f_x =МАКС(C2:C12)-МИН(C2:C12)

Microsoft Excel - задача1

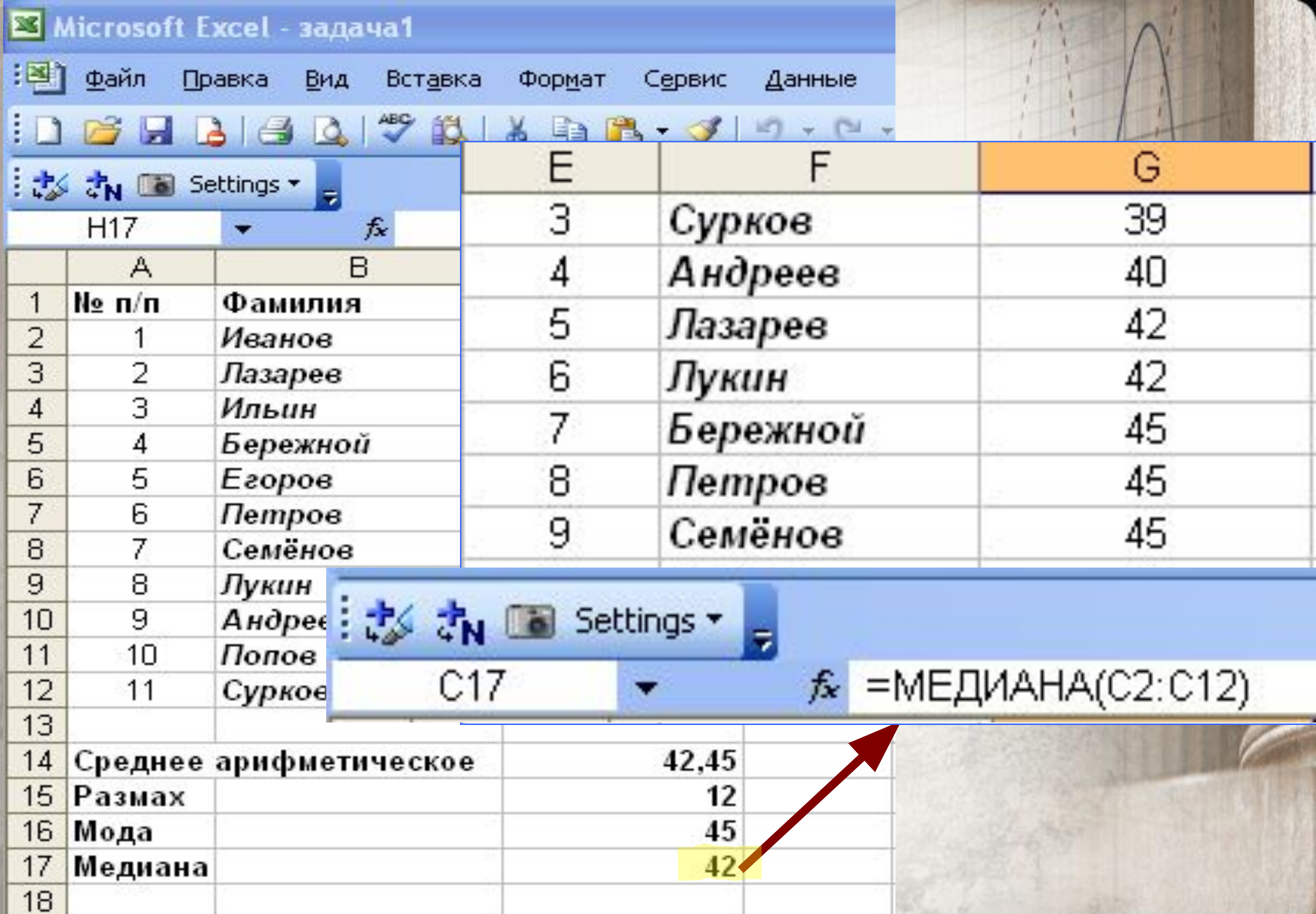
Файл Правка Вид Вставка Формат Сервис Данные

Settings

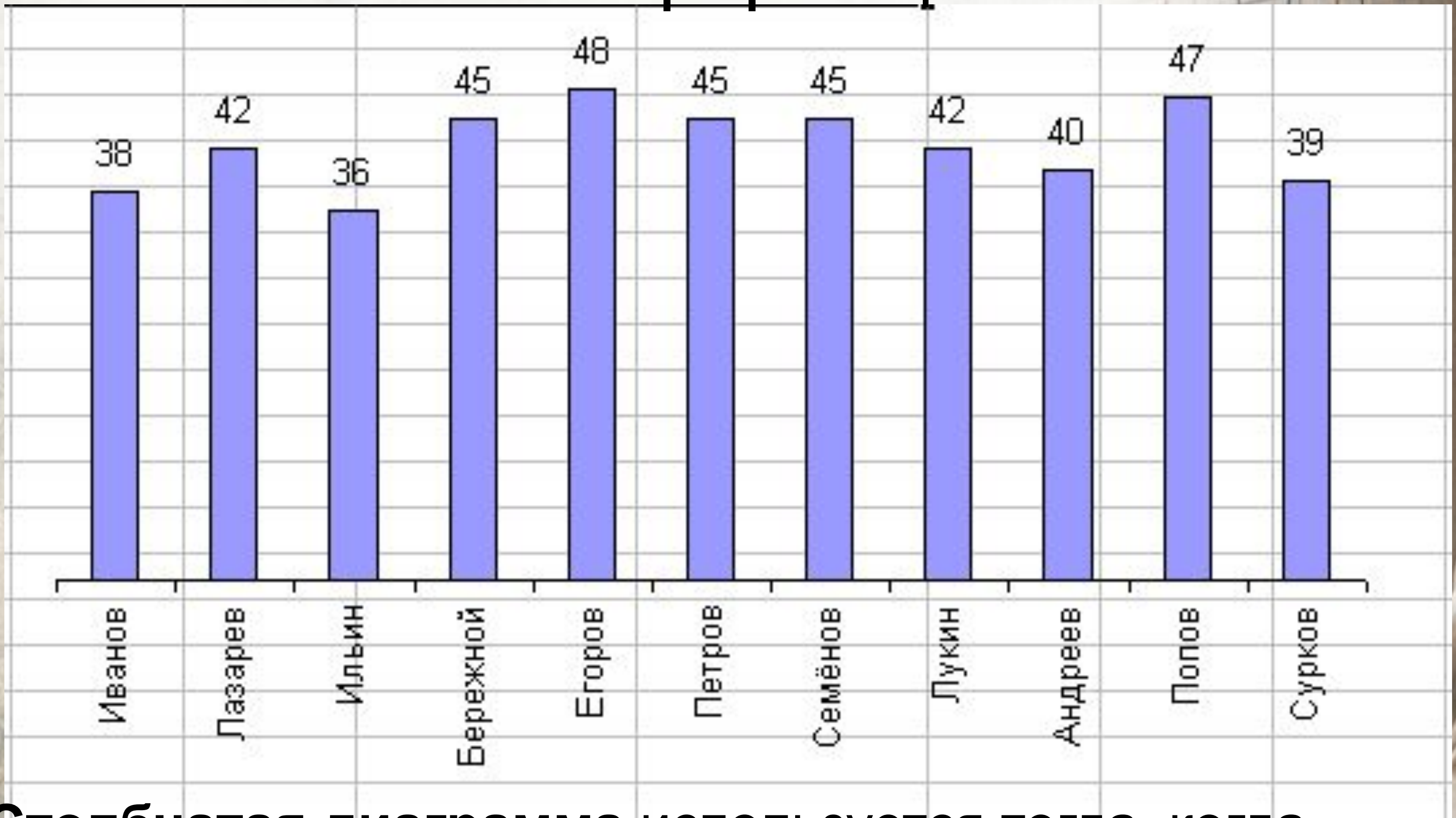
H17

	A	B	C	D
1	**№ п/п**	**Фамилия**	**Число деталей**	
2	1	***Иванов***	38	
3	2	***Лазарев***	42	
4	3	***Ильин***	36	
5	4	***Бережной***	45	
6	5	***Егоров***	48	
7	6	***Петров***	45	
8	7	***Семёнов***	45	
9	8	***Лукин***	42	
10	9	***Андреев***	40	
11	10	***Попов***	47	
12	11	***Сурк***	[illegible]	
13				
14	**Среднее арифметическое**		**42,45**	
15	**Размах**		**12**	
16	**Мода**		**45**	
17	**Медиана**		**42**	
18				

C16 f_x =МОДА(C2:C12)


Microsoft Excel - задача1
Файл Правка Вид Вставка Формат Сервис Данные
Settings
H17
A B
1 № п/п Фамилия
2 1 Иванов
3 2 Лазарев
4 3 Ильин
5 4 Бережной
6 5 Егоров
7 6 Петров
8 7 Семёнов
9 8 Лукин
10 9 Андрее
11 10 Попов
12 11 Сурков
13
14 Среднее арифметическое 42,45
15 Размах 12
16 Мода 45
17 Медиана 42
18
E F G
3 Сурков 39
4 Андреев 40
5 Лазарев 42
6 Лукин 42
7 Бережной 45
8 Петров 45
9 Семёнов 45
Settings
C17
=МЕДИАНА(C2:C12)

Наглядное представление статистической информации



Столбчатая диаграмма используется тогда, когда хотят проиллюстрировать распределение данных, полученных в результате статистических исследований.

Задача

В таблице записаны результаты ежедневного измерения на метеостанции в полдень температуры воздуха (в градусах Цельсия) в течение первой декады марта. Найти среднюю температуру в полдень в эту декаду. Составить таблицу отклонений от средней температуры воздуха в полдень в каждый день из дней декады.

	A	B	C	D	E	F	G	H	I	J	K
1	**Число месяца**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	**Температура, C^0**	-2	-1	-3	0	1	2	2	3	4	3
3	**Средняя температура**	**0,9**									
4	**Отклонение от ср. температуры**	**2,9**	**1,9**	**3,9**	**0,9**	**-0,1**	**-1,1**	**-1,1**	**-2,1**	**-3**	**-2,1**
5											
6											
7											
8											
9											

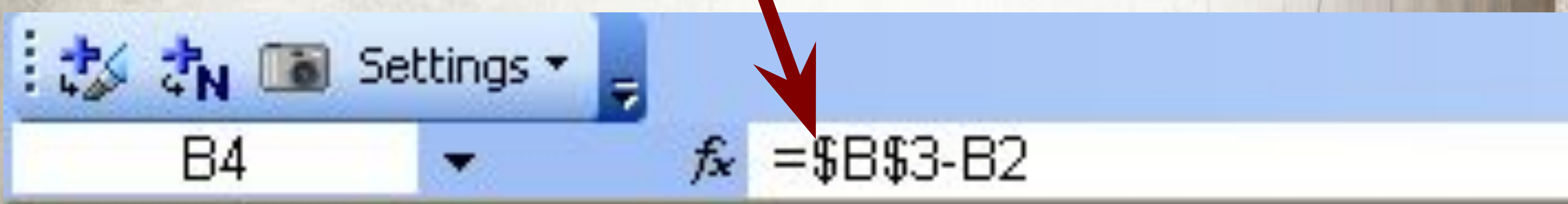
Решение

	A	B	C	D	E	F	G	H	I	J	K
1	**Число месяца**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	**Температура, C^0**	-2	-1	-3	0	1	2	2	3	4	3
3	**Средняя температура**	**0,9**									
4	**Отклонение от ср. температуры**	**2,9**	**1,9**	**3,9**	**0,9**	**-0,1**	**-1,1**	**-1,1**	**-2,1**	**-3**	**-2,1**

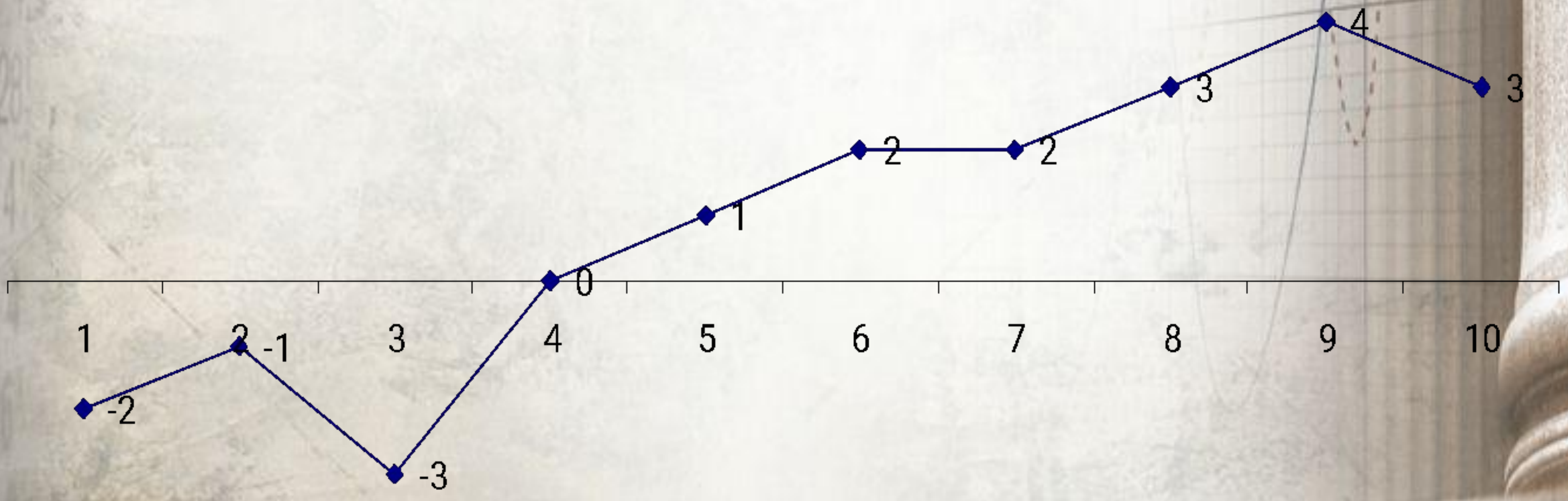
Settings

B3 f_x =СРЗНАЧ(B2:K2)

	A	B	C	D	E	F	G	H	I	J	K
1	**Число месяца**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	**Температура, C^0**	-2	-1	-3	0	1	2	2	3	4	3
3	**Средняя температура**	**0,9**									
4	**Отклонение от ср. температуры**	**2,9**	**1,9**	**3,9**	**0,9**	**-0,1**	**-1,1**	**-1,1**	**-2,1**	**-3**	**-2,1**



Температура, C^0

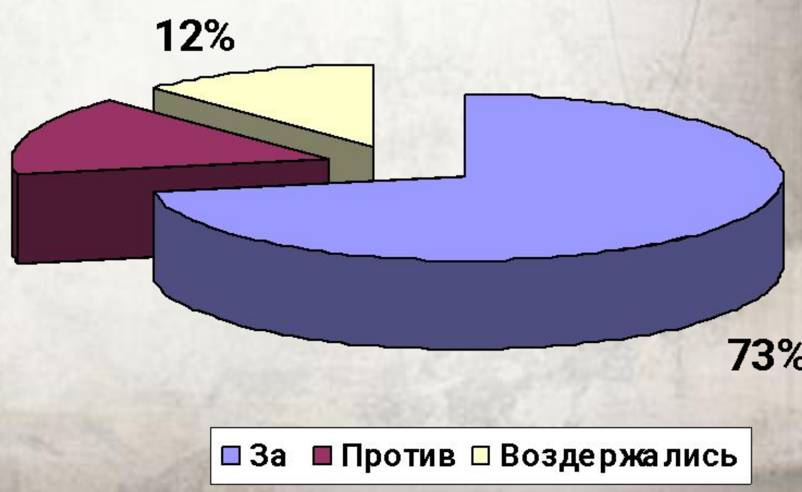


Динамику изменения статистических данных во времени иллюстрируют с помощью ***полигона (графика)***

Обработка результатов исследований (опросов)

Проект «Школьная форма – «ЗА» и «ПРОТИВ»

За	523
Против	112
Воздержались	88



Комбинаторика

Математика

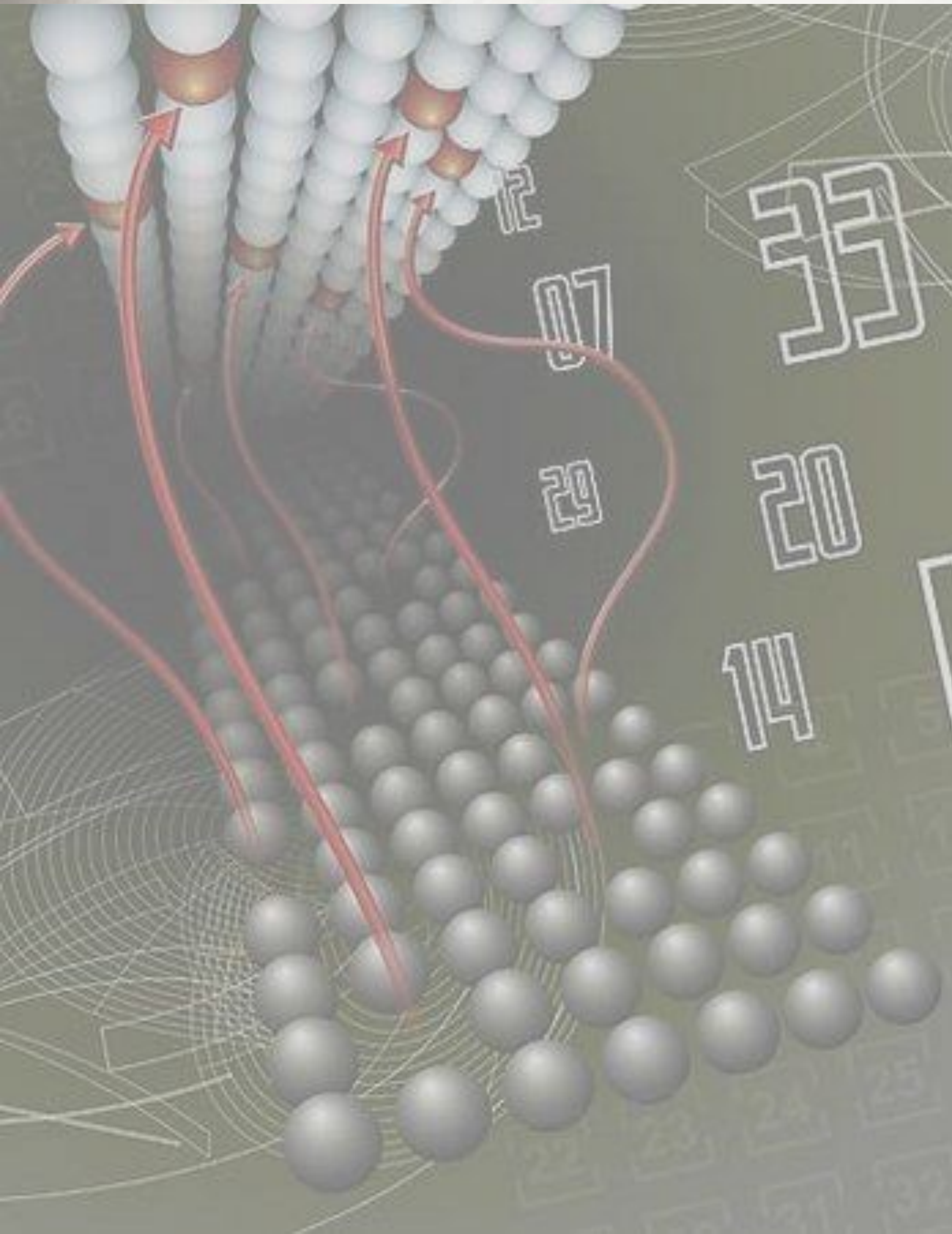
Множество и комбинаторика	Множество. Элемент множества Подмножество Объединение и пересечение множеств. Диаграммы Эйлера	5 часов	7 класс 8 класс
	Примеры решения комбинаторных задач: перебор вариантов, правило умножения.	10 часов	9 класс

Информатика

! **НЕТ в программе!!!**

! **ЕГЭ по информатике (А12) - есть**

КОМБИНАТОРИКА



– раздел математики, в котором изучаются вопросы о том, сколько различных комбинаций, подчиненных тем или иным условиям, можно составить из заданных объектов.

Что нужно знать!!!

- если на каждом шаге известно количество возможных вариантов выбора, то для вычисления общего количества вариантов нужно все эти числа **перемножить**; **например**, в двузначном числе мы можем выбрать первую цифру **9** способами (она не может быть нулем), а вторую – **10** способами, поэтому всего есть **9·10=90** двузначных чисел

Что нужно знать!!!

- если мы разбили все нужные нам комбинации на несколько групп (*не имеющих общих элементов*!) и подсчитали количество вариантов в каждой группе, то для вычисления общего количества вариантов нужно все эти числа **сложить**;
  например, есть **9·10=90** трехзначных чисел, оканчивающихся на **5**, и **9·10=90** трехзначных чисел, оканчивающихся на **2**, поэтому **90+90=180** трехзначных чисел оканчиваются на **2 или** на **5**

Что нужно знать!!!

- если в предыдущем случае группы имеют общие элементы, их количество нужно **вычесть** из полученной суммы;
  например, есть **9·10=90** трехзначных чисел, оканчивающихся на **5**, и **10·10=100** трехзначных чисел, начинающихся на **5**; в обе группы входят числа, которые начинаются и заканчиваются на **5**, их всего **10 штук**, поэтому количество чисел, которые начинаются ***или*** заканчиваются на **5**, равно **90+100-10=180.**

Что не мешает знать!!!

если есть **n** различных элементов, число их различных перестановок равно факториалу числа **n**, то есть произведению всех натуральных чисел от 1 до **n**:

n! = 1·2·3·…·(n-1)·n

например, три объекта **(А, Б и В)** можно переставить 6 способами (**3!=1·2·3=6**):

(А, Б, В), (А, В, Б), (Б, А, В), (Б, В, А), (В, А, Б) и (В, Б, А)

.

Что не мешает знать!!!

если нужно выбрать **m** элементов из **n** (где **n≥m**) и две комбинации, состоящие из одних и тех же элементов, расположенных в разном порядке, считаются различными, число таких комбинаций (они называются ***размещениями***) равно

$$A_n^m = \frac{n!}{(n-m)!} = (n-m+1)\cdot(n-m+2)\cdot\ldots\cdot(n-1)\cdot n$$

например, в соревновании **пяти** спортсменов призовые места **(первые три)** могут распределиться **60** способами, поскольку

.

$$A_5^3 = \frac{5!}{(5-3)!} = 3\cdot4\cdot5 = 60$$

Что не мешает знать!!!

если нужно выбрать **m** элементов из **n** (где **n≥m**) и порядок их расположения не играет роли, число таких комбинаций (они называются ***сочетаниями***) равно

$$C_n^m = \frac{n!}{m!(n-m)!}$$

например, выбрать двух дежурных из пяти человек можно 10 способами, поскольку

.

$$C_5^2 = \frac{5!}{2!(5-2)!} = \frac{120}{2 \cdot 6} = 10$$

Задача

Сколько существует различных четырехзначных чисел, в записи которых используются только четные цифры?

1) 125 2) 250 3) 500 4) 625

Решение:

1) первой цифрой может быть любая четная цифра, кроме нуля (иначе число не будет четырехзначным) – это **2, 4, 6 или 8**, всего **4** варианта

	x	?	?	?
Вариантов	4			

2) предположим, что первая цифра выбрана; независимо от нее на втором месте может стоять любая из четных цифр – **0, 2, 4, 6 или 8**, всего **5** вариантов:

	x	y	?	?
Вариантов	4	5		

3) аналогично находим, что последние две цифры также могут быть выбраны 5-ю способами каждая, независимо друг от друга и от других цифр (первой и второй):

	x	*y*	*z*	*w*
Вариантов	**4**	**5**	**5**	**5**

4) общее количество комбинаций равно произведению

4 · 5 · 5 · 5 = 500

5) таким образом, правильный ответ – **3.**

Ещё пример задания

Сколько существует различных четырехзначных чисел, в записи которых ровно две девятки, стоящие рядом?

1) 212 2) 225 3) 243 4) 280

Решение:

1) возможны три случая: 99··, ·99· и ··99, где жирная точка обозначает некоторую цифру, не равную 9

2) для каждого из этих случаев нужно подсчитать количество вариантов и эти числа сложить

3) в варианте 99·· две последних цифры могут быть любыми, кроме девятки (по 9 вариантов выбора):

	9	*9*	*x*	*y*
Вариантов	**1**	**1**	**9**	**9**

поэтому всего получаем 1·1·9·9 = 81 вариант

4) в варианте $\cdot 99 \cdot$ первая цифра не может быть нулем и девяткой (остается 8 вариантов), а последняя может быть любой, кроме девятки (9 вариантов):

	x	9	9	y
Вариантов	**8**	**1**	**1**	**9**

поэтому всего получаем 8·1·1·9 = 72 варианта

5) в варианте $\cdot\cdot 99$ первая цифра не может быть нулем и девяткой (остается 8 вариантов), а последняя может быть любой, кроме девятки (9 вариантов):

	x	x	9	9
Вариантов	**8**	**9**	**1**	**1**

поэтому всего получаем **8 · 9 · 1 · 1 = 72** варианта

6) общее количество вариантов равно сумме
81 + 72 + 72 = 225

7) таким образом, правильный ответ – **2**.

Еще пример задания:

Виктор хочет купить пять разных книг, но денег у него хватает только на три (любые) книги. Сколькими способами Виктор может выбрать три книги из пяти?

1) 10 2) 20 3) 30 4) 60

Решение (вариант 2, формулы комбинаторики):

1) нам нужно выбрать 3 объекта из 5, причем порядок выбора здесь не важен – нам нужны разные ***сочетания***

2) зная формулу для вычисления количества ***сочетаний***, сразу находим (при **m = 3** и **n = 5**)

$$C_5^3 = \frac{5!}{3! \cdot 2!} = \frac{120}{6 \cdot 2} = 10$$

3) таким образом, правильный ответ – 1.

Вероятность и информация

Математика

Вероятность	Частота события, вероятность. Равновозможные события и подсчет их вероятности. Представление о геометрической вероятности.	**5 часов**	**9 класс**

Информатика (9 класс)

Информация и информационные процессы (10 час)
Язык как способ представления информации: естественные и формальные языки
Дискретная форма представления информации. Компьютерное представление текстовой информации
Количество информации как мера уменьшения неопределенности знания. Определения количества информации. **Алфавитный и вероятностный подход к определению количества информации.**

Информатика (11 класс, профильный курс)

Информация. Системы счисления (30 часов)
Понятие «информация» в науках о неживой и живой природе, обществе и технике
Количество информации как мера уменьшения неопределенности знаний. Задание «Определение количества информации».
Практическое задание «Перевод единиц измерения количества информации».
Алфавитный и вероятностный подход к определению количества информации.
Формула Шеннона
Решение задач. Практическое задание «Определение количества информации».

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

– это раздел математики, изучающий закономерности, основанные на взаимодействии большого числа случайных явлений (статистические закономерности).

Вероятность - отношение числа случаев благоприятствующих событию А, к числу всех возможных случаев называют вероятностью события А.

Вероятностный подход

Вероятность события – число от **0 до 1,** показывающее, как часто случается это событие в большой серии одинаковых опытов.

- **p = 0** событие **никогда** не происходит (нет неопределенности)
- **p = 0,5** событие происходит в половине случаев (есть неопределенность)
- **p = 1** событие происходит **всегда** (нет неопределенности)

Полная система событий: одно из N событий обязательно произойдет (и только одно!).

p_i – вероятность выбора i-ого варианта ($i=1,\ldots,N$)

$$0 \le p_i \le 1, \quad p_1 + p_2 + \ldots + p_N = 1$$

Количество информации в сообщении о некотором событии зависит от вероятности этого события.

Сообщение о том, что произошло одно событие из двух равновероятных, несёт 1 бит информации.

Формула Хартли (1928)

$$I = \log_2 N$$

$$N = 2^I$$

I – количество информации в битах

N – количество равновероятных событий

Пример:

В аэропорту стоит 6 самолетов, из них один летит в Москву. Сколько информации в сообщении «В Москву летит второй самолет»?

$6 = 2^I \quad = 2{,}585$ бит

$$I = \log_2 6 = \frac{\ln 6}{\ln 2} = \frac{\lg 6}{\lg 2} = 2{,}585 \text{ бит}$$

Таблица 1.1. Количество информации в сообщении об одном из N равновероятных событий

N	i	N	i	N	i	N	i
1	0,00000	17	4,08746	33	5,04439	49	5,61471
2	1,00000	18	4,16993	34	5,08746	50	5,64386
3	1,58496	19	4,24793	35	5,12928	51	5,67243
4	2,00000	20	4,32193	36	5,16993	52	5,70044
5	2,32193	21	4,39232	37	5,20945	53	5,72792
6	2,58496	22	4,45943	38	5,24793	54	5,75489
7	2,80735	23	4,52356	39	5,28540	55	5,78136
8	3,00000	24	4,58496	40	5,32193	56	5,80735
9	3,16993	25	4,64386	41	5,35755	57	5,83289
10	3,32193	26	4,70044	42	5,39232	58	5,85798
11	3,45943	27	4,75489	43	5,42626	59	5,88264
12	3,58496	28	4,80735	44	5,45943	60	5,90689
13	3,70044	29	4,85798	45	5,49185	61	5,93074
14	3,80735	30	4,90689	46	5,52356	62	5,95420
15	3,90689	31	4,95420	47	5,55459	63	5,97728
16	4,00000	32	5,00000	48	5,58496	64	6,00000

Вероятностный подход

Вычисление вероятности

Задача. В пруду живут 100 рыб, из них 20 карасей, 30 пескарей, а остальные – окуни. Какова вероятность поймать карася (пескаря, окуня), если все рыбы одинаково голодны?

Формула: $$p_i = \frac{n_i}{N}$$

число «нужных» событий

общее число событий

Решение:

караси $p_1 = \frac{n_1}{N} = \frac{20}{100} = 0,2$

пескари $p_2 = \frac{n_2}{N} = \frac{30}{100} = 0,3$

окуни $p_3 = \frac{n_3}{N} = \frac{(100-20-30)}{100} = \frac{50}{100} = 0,5$

? **Как иначе посчитать p_3?**

$$p_3 = 1 - p_1 - p_2 = 0,5$$

Вероятностный подход

Как посчитать информацию, если варианты не равновероятны?

Клод Шеннон (1916 —2001) американский математик и электротехник, один из создателей математической теории информации и криптографии.

Идея: если случается менее вероятное событие, мы получаем больше информации.

$0 \le p_i \le 1$ – вероятность выбора i-ого варианта (i=1,…,N)

Если произошло событие i, мы получаем информацию

$$I_i = -\log_2 p_i = \log_2 \frac{1}{p_i}$$

Вероятностный подход

Задача 1. В пруду живут 100 рыб, из них 20 карасей, 30 пескарей, а остальные – окуни. *Сколько информации* несет сообщение о том, что рыбак поймал карася (пескаря, окуня), если все рыбы одинаково голодны?

Формула:

$$I_i = -\log_2 p_i = \log_2 \frac{1}{p_i}$$

Решение:

карась $p_1 = \frac{20}{100} = 0{,}2$ $I_1 = -\log_2 0{,}2 = \log_2 5 \approx 2{,}32$ бита

пескарь $p_2 = \frac{30}{100} = 0{,}3$ $I_2 = -\log_2 0{,}3 \approx \log_2 3{,}33 \approx 1{,}74$ бита

окунь $p_3 = \frac{50}{100} = 0{,}5$ $I_3 = -\log_2 0{,}5 = \log_2 2 = 1$ бит

Вероятностный подход

Задача 2. Посчитать, чему равна информация в сообщении «Сейчас идет снег» зимой и летом.

Что еще нужно для решения?

Событие 1 – *идет снег*, событие 2 – *снег не идет*.

летом

$$p_1 = 0{,}001; \quad p_2 = 0{,}999$$

зимой

$$p_1 = 0{,}5; \quad p_2 = 0{,}5$$

Решение:

летом

$I_1 = -\log_2 0{,}001 = 9{,}97$ бита

$I_2 = -\log_2 0{,}999 = 0{,}0014$ бита

зимой

$I_1 = I_2 = -\log_2 0{,}5 =$

$= \log_2 2 = 1$ бит

Формула Шеннона (1948)

Неопределенность (*энтропия* системы)

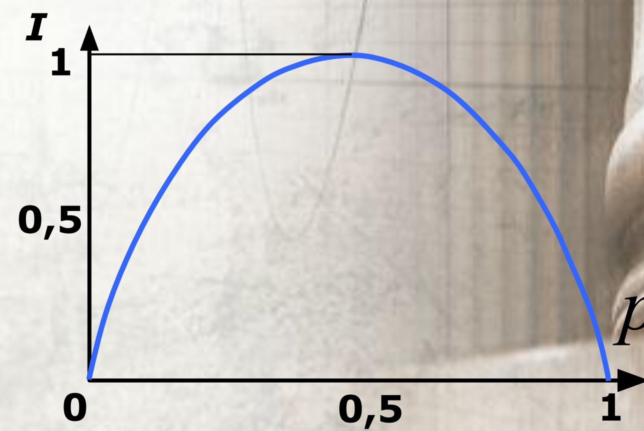
$$I = -\sum_{1}^{N} p_i \log_2 p_i = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - \ldots - p_N \log_2 p_N$$

Информация = снятая неопределенность!

? Когда неопределенность наибольшая?

Система двух событий: $p_2 = 1 - p_1$

Средняя информация (неопределенность) **максимальна**, когда все события **равновероятны**.



$$p_1 = p_2 = \ldots = p_N = \frac{1}{N}$$

$$I = -\sum_{i=1}^{N} \frac{1}{N} \log_2 \frac{1}{N} = \log_2 N$$

Литература

- Н. Угринович. Информатика и информационные технологии (10-11 кл.)
- Ю.Н. Макарычев. Алгебра: элементы статистики и теории вероятности (7-9кл.)
- И.Семакин. Базовый курс (7-9 кл.)
- http://kpolyakov.narod.ru