

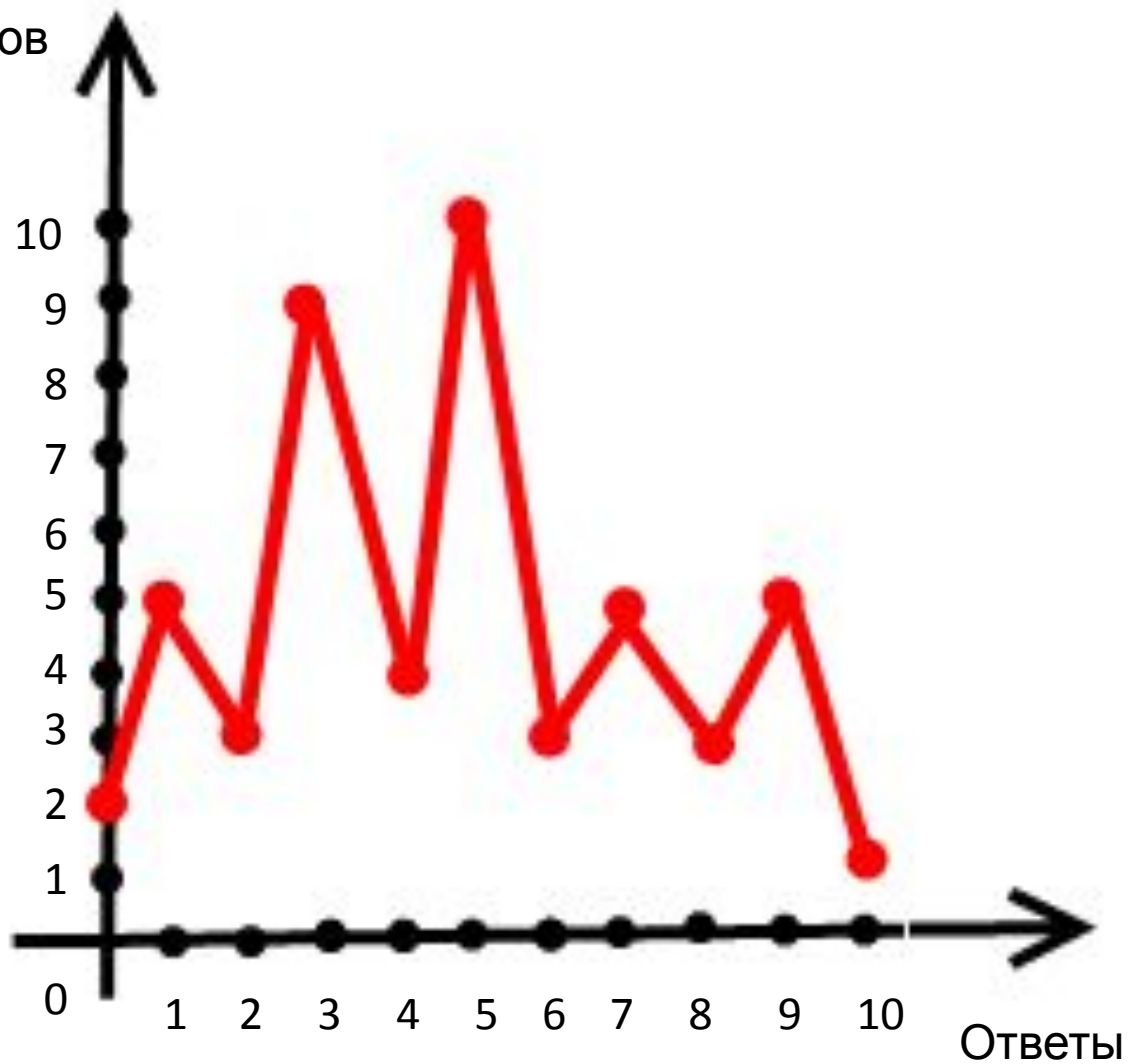
Статистическая обработка данных

На праздничном вечере среди учеников 10 «А» и 10 «Б» классов провели лотерею. Каждый из 50 учащихся произвольно задумал одну цифру от 0 до 9 и записал ее на левой и правой половине своего лотерейного билета. Правые половинки билетов остались у учеников, а левые половинки положили на стол перед организатором лотереи. Итак, на столе 50 листочков, как в них разобраться?

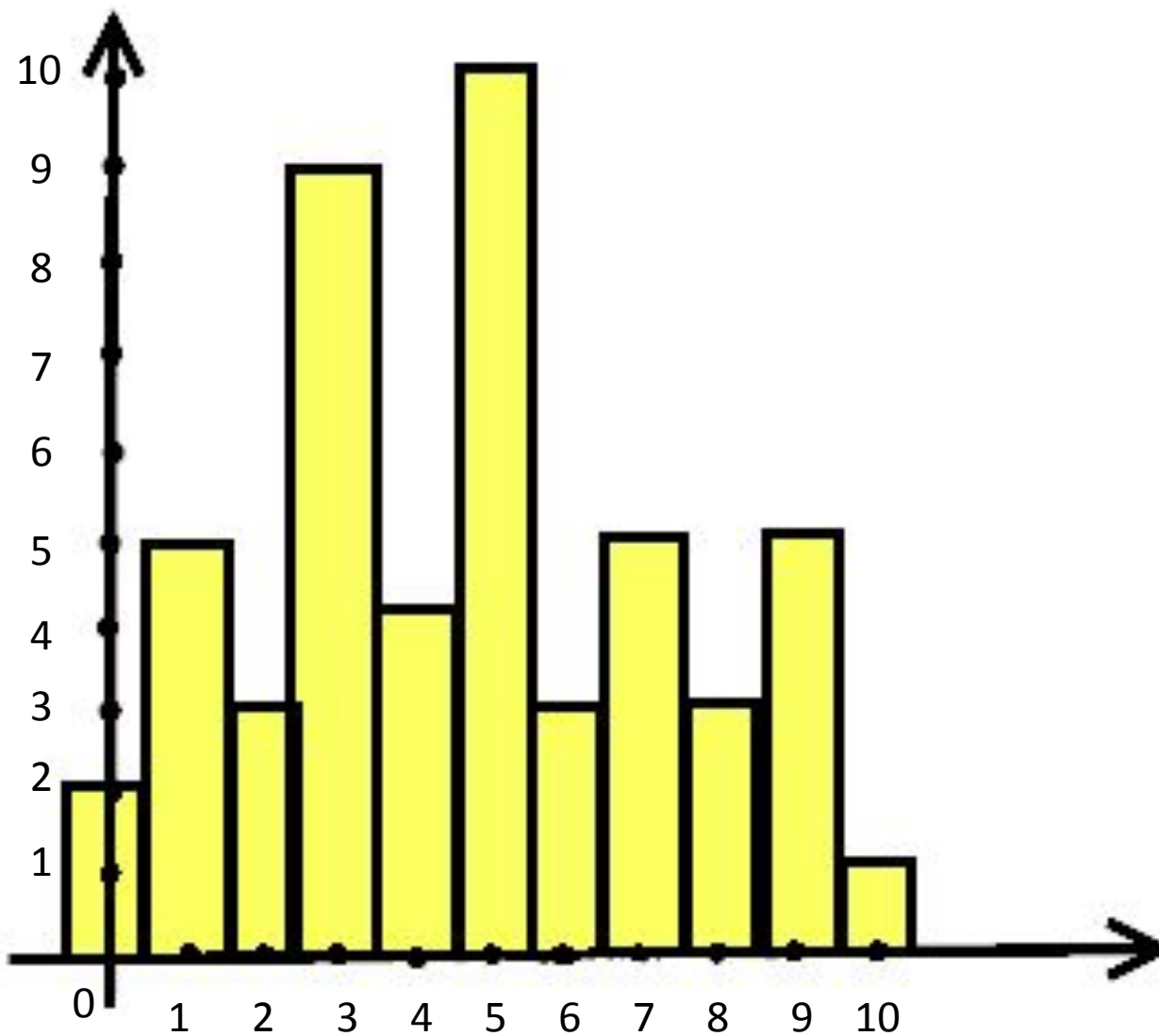
Ответ	0	1	2	3	4	5	6	7	8	9	10
Количество ответов	2	5	3	9	4	10	3	5	3	5	1

«Самый популярный» (ответ 5), «Почти самый популярный» (ответ 3), «Оригинально, но не верно» (ответ 10), «Сладкая парочка» (ответ 0), «Три богатыря» (ответ 2, 6, 8), «Отличники» (ответы 1,7,9), «Хорошисты» (ответ 4).

Количество ответов



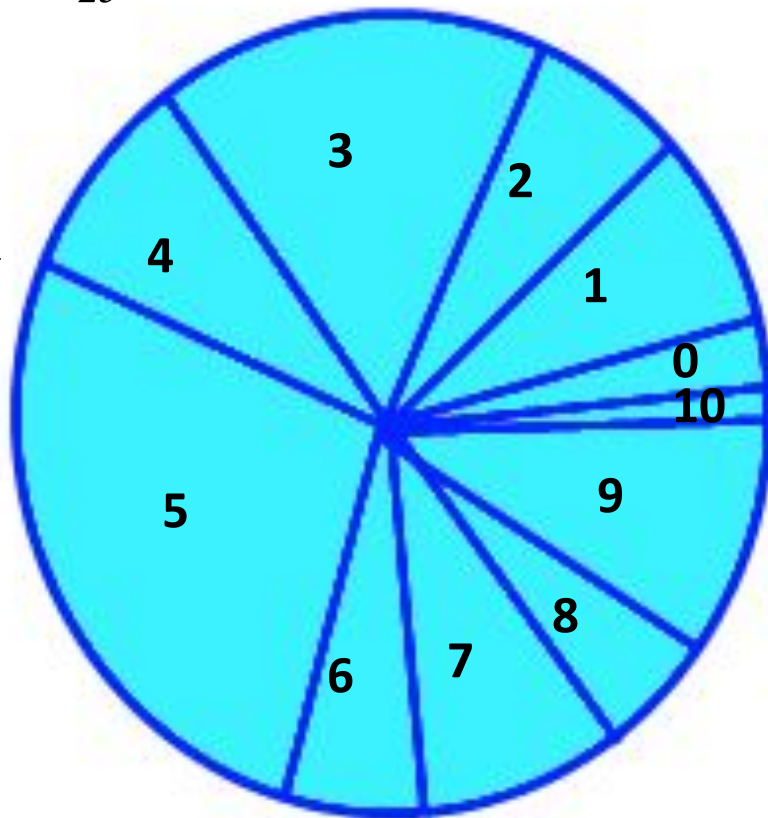
***Многоугольник
распределения***



**Столбчатая диаграмма – гистограмма
распределения**

Сектор «0» занимает $\frac{1}{25}$ часть круга

Два полученных
ответа «0»
составляют $\frac{2}{50} = \frac{1}{25}$
часть от общего
круга. Значит
центральный
угол сектора «0»
равен $\frac{360^\circ}{25} = 14,4^\circ$



Круговая диаграмма

Основные этапы статистической обработки данных:

- Сначала данные измерений **упорядочивают** и **группируют**;
- Затем составляют **таблицы распределения данных**;
- Таблицы распределения позволяют построить **графики распределения данных** в виде многоугольника распределения, гистограммы распределения или круговой диаграммы;

К этим трем этапам позже, как правило, добавляют еще один:

- Получение **паспорта данных** измерения, который состоит из небольшого количества основных **числовых характеристик** полученной информации.

Числовые характеристики:

1. **Объем измерения.**

В данном случае он равен 50, т.к. обрабатывались ответы 50 участников.

2. **Размах измерения.**

В данном случае он равен 10, т.е.

Разности между наибольшим (10) и наименьшим (0) результатами измерения (10-0=10)

3. **Мода измерения.**

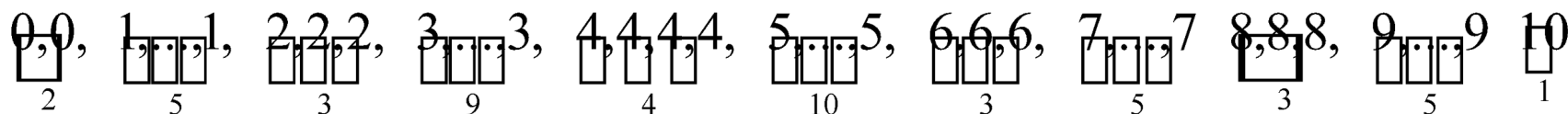
В данном случае она равна 5, т.к. ответ «5» – самый «модный», самый популярный, он встречается чаще других.

4. **Среднее (или среднее арифметическое).**

Это частное от деления суммы всех результатов измерения на объем измерения,

$$\frac{0*2 + 1*5 + 2*3 + 3*9 + 4*4 + 5*10 + 6*3 + 7*5 + 8*3 + 9*5 + 10*1}{50} =$$
$$= \frac{0 + 5 + 6 + 27 + 16 + 50 + 18 + 35 + 24 + 45 + 10}{50} = \frac{236}{50} = 4,72$$

Чаще всего результатами измерения являются числа. Каждое число, встретившееся в конкретном измерении, называют **вариантой измерения**. Варианты измерения могут быть никак не упорядочены, но если их записать по порядку (например по времени поступления), то получится **ряд данных** измерения. Если же записать по возрастанию, то получится **сгруппированный ряд данных**.



Среднюю варианту ряда называют **медианой измерения**. Так в рассмотренном примере 50 вариантов, среди них – две, это варианты №25 и №26. Обе они равны 5, значит медиана равна 5.

Ответ «0» встречается два раза. В статистике в этом случае говорят, что **абсолютная частота** варианты «0» равна 2. **Абсолютная частота** варианты «7» равна 5.

Мы заменим термин **абсолютная частота** более кратким.

ОПРЕДЕЛЕНИЕ. Если среди всех данных конкретного измерения одна из вариантов встретилась k раз, что число k называют **кратностью** этой варианты.

Ответ	0	1	2	3	4	5	6	7	8	9	10
Количество ответов	2	5	3	9	4	10	3	5	3	5	1

	Варианта											Сумма
	0	1	2	3	4	5	6	7	8	9	10	
Кратность	2	5	3	9	4	10	3	5	3	5	1	50

Так получается ***таблица распределения данных*** измерения. Клетку «Сумма» добавляют для контроля: число в графе «Сумма» всегда должно равняться объему измерения.

Пример 1.

На уроке физкультуры 14 школьников прыгали в высоту, а учитель записывал их результаты. Получился такой ряд данных (в сантиметрах): 125, 110, 130, 125, 120, 130, 140, 125, 110, 130, 120, 125, 120, 125. Требуется сгруппировать данные, составить таблицу их распределения, найти размах, моду и медиану измерения.

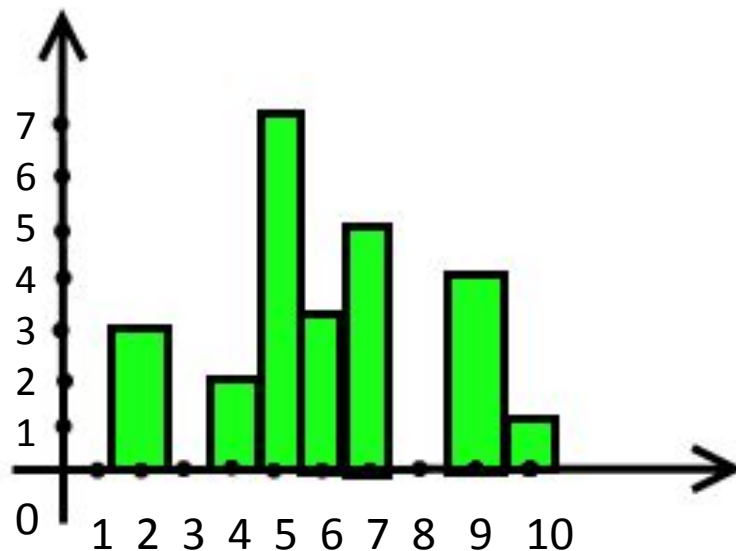
Пример 2.

В таблице распределения данных часть информации была утеряна. Восстановите ее, если известно, что объем измерения равен 20, размах равен 6, а мода равна 2.

	Варианта					Сумма
		-1	0		3	
Кратность	5	1		7	3	

Пример 3.

По приведенной гистограмме распределенных данных найти: количество вариантов измерения, объем, размах, моду измерения; наиболее удаленную от моды варианту и ее кратность. Составить таблицу распределения данных.



Если кратность данной варианты разделить на объем измерения, то получится **частота варианты**.

$$\text{Частота варианты} = \frac{\text{кратность варианты}}{\text{объем измерения}}$$

Частоту варианту можно измерять в процентах:

$$\text{Частота варианты} = \frac{\text{кратность варианты}}{\text{объем измерения}} * 100\%$$

Пример 4.

Зная таблицу распределения данных составьте распределение частот данных.

	Варианта					Сумма
	-3	-1	0	2	3	
Кратность	5	1	4	7	3	20
Частота						
Частота , %						

Пример 5.

На испытательном стенде оружейного завода пристреливают готовые ружья, т.е. уточняют и корректируют их прицел. В таблице приведены измерения горизонтальных отклонений (в сантиметрах) от цели при стрельбе из трех ружей. Вычислите среднее значения результатов испытания.

	Выстрел									
	№1	№2	№3	№4	№5	№6	№7	№8	№9	№10
Ружье А	+1,0	+1,0	+2,0	+1,5	+2,0	+2,0	+1,5	+1,5	+0,5	+1,0
Ружье Б	+1,0	0	-1,5	+1,5	-0,5	-1,5	+2,0	+1,0	-1,0	+2,0
Ружье В	-0,5	-1,0	0	-105	-1,0	+1,0	+1,0	+1,5	+1,0	+3,0

Числовую характеристику данных измерения, отвечающую за разброс (рассеивание) данных вокруг их среднего значения, называют дисперсией (от лат. Disperses – рассыпанный, разогнанный, рассеянный) и обозначают буквой D ; число \sqrt{D} называют **средним квадратическим отклонением**.

Пример 6.

Найти дисперсию результатов измерений для ружей А и Б из примера 5.