Элементы статистики.

Статистические характеристики

1. Среднее арифметическое
2. Размах
3. Мода
4. Медиана

Среднее арифметическое

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

Размах

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

Мода

Модой ряда чисел называется число, наиболее часто встречающееся в данном ряду. Ряд чисел может иметь более одной моды или не иметь моды совсем. Моду ряда данных обычно находят тогда, когда хотят выявить некоторые типичные показатели.

Медиана

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.

Статистические исследования.

Сбор и группировка статистических данных

Для исследования различных общественных и социально-экономических явлений, а так же некоторых процессов, происходящих в природе, проводятся специальные статистические исследования. Всякое статистическое исследование начинается с целенаправленного сбора информации об изучаемом явлении или процессе. Этот этап называется этапом статистического наблюдения.

Рассмотрим такой пример. Администрация школы решила проверить математическую подготовку 8 классников. С этой целью был составлен тест, содержащий 9 заданий. Работу выполняли 40 учащихся школы. При проверке каждой работы учитель отмечал число верно выполненных заданий. В результате был составлен такой ряд чисел:

6, 5, 4, 0, 4, 5, 7, 9, 1, 6, 8, 7, 9, 5, 8, 6, 7, 2, 5, 7, 6, 3, 4, 4, 5, 6, 8, 6, 7, 7, 4, 3, 5, 9, 6, 7, 8, 6, 9, 8.

Для того, чтобы удобно было анализировать полученные данные, упорядочим этот ряд:

0, 1, 2, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5,

6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7,

8, 8, 8, 8, 8, 9, 9, 9, 9.

Представим полученные данные в виде таблицы, в которой для каждого числа, верно выполненных заданий, записанного в верхней строке, укажем в нижней строке количество появлений этого числа в ряду, т.е. частоту:

Число верно выполненных заданий.	0	1	2	3	4	5	6	7	8	9
Абсолютная частота или кратность.	1	1	1	2	5	6	8	7	5	4

Такую таблицу называют таблицей частот.

В рассмотренном примере сумма частот равна общему числу проверяемых работ, т.е. 40.

Чаще всего результатами измерения являются числа. Каждое число, встретившееся в конкретном измерении, называют *вариантой измерения*[1].

[1] Использование женского рода слегка непривычно, но именно такой термин принят в статистике. Кстати, это помогает избежать путаницы с привычным использованием слова «вариант» в мужском роде (например, вариант контрольной работы).

Определение. Если среди всех данных конкретного измерения одна из вариант встретилась k раз, то число k называют **кратностью этой варианты.**

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

$$\boxed{\textbf{Частота варианты} = \frac{\textit{кратность варианты}}{\textit{объем измерения}}.}$$

$$\boxed{\textbf{Частота варианты (в процентах)} = \frac{\textit{кратность варианты}}{\textit{объем измерения}} \cdot 100\,\%.}$$

Относительная частота

Иногда составляют таблицу, в которой для каждого данного указывается не частота, а отношение частоты к общему числу данных в ряду. Это отношение, выраженной в процентах, называют относительной частотой, саму таблицу, таблицей относительных частот.

В нашем примере общая численность совокупности- это число учащихся, писавших работы, т.е. 40. Таблица относительных частот выглядит следующим образом:

Число верно выполненных заданий.	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Кратность	1	1	1	2	5	6	8	7	5	4
Относит. частота %	**2,5**	**2,5**	**2,5**	**5**	**12,5**	**15**	**20**	**17,5**	**12,5**	**10**

Наглядное представление статистической информации

1. Столбчатая диаграмма
2. Круговая диаграмма
3. Полигон
4. Гистограмма

В таблице показаны данные об успеваемости и качестве выполнения вычислительных навыков на входе и на выходе за три года. По этим данным была построена столбчатая диаграмма.

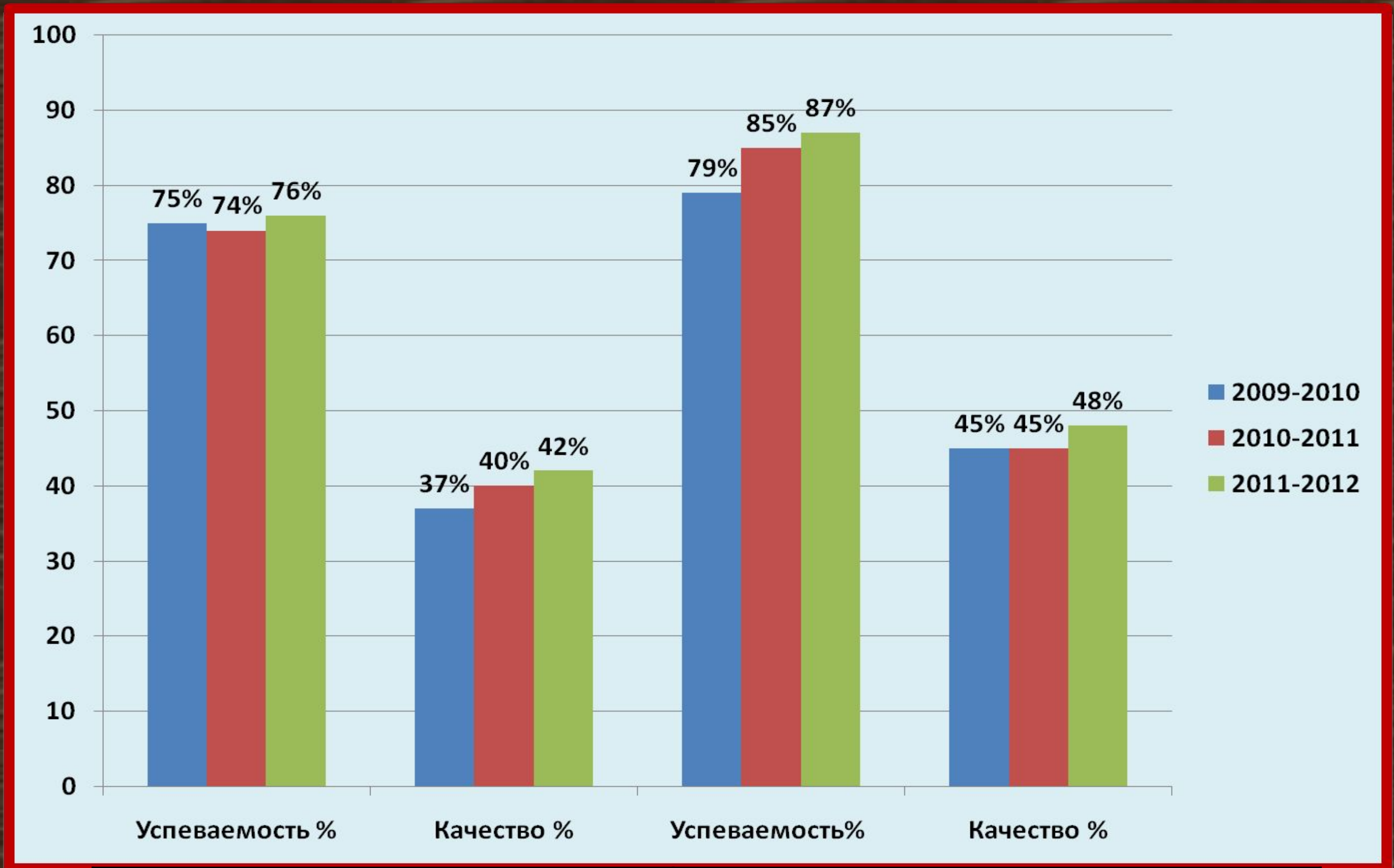
	На входе		На выходе	
Год	Успеваемость %	Качество %	Успеваемость %	Качество %
2009-2010	75	37	79	45
2010-2011	74	40	85	45
2011-2012	76	42	87	48

Столбчатые диаграммы

Для наглядного представления данных, полученных в результате статистического исследования, широко используются различные способы их изображения.

Одним из хорошо известных вам способов наглядного представления ряда данных является построение столбчатой диаграммы. Столбчатые диаграммы используют тогда, когда хотят проиллюстрировать динамику изменения данных во времени или распределения данных, полученных в результате статистического исследования.

Мониторинг вычислительных навыков за три года. МКОУ СОШ№236.Математика.



На входе	На выходе

Данная столбчатая диаграмма наглядно показывает, что успеваемость и качество в течении трех лет неуклонно растет.

Круговая диаграмма

Для наглядного изображения соотношения между частями исследуемой совокупности удобно использовать круговые диаграммы. Если результат статистического исследования представлен в виде таблицы относительных частот, то для построения круговой диаграммы круг разбивается на секторы, центральные углы которых пропорциональны относительным частотам, определённым для каждой группы данных. В тех случаях, когда результат статистического исследования представлен в виде таблицы частот, удобно для построения круговой диаграммы предварительно заменить её таблицей относительных частот.

Круговую диаграмму можно так же построить на основе таблицы в следующем примере, взятом из энциклопедии:

В таблице представлено количество вымирающих видов, взятых из энциклопедии.

	Растения	Птиц	Млекопитающие
Количество выми-рающих видов	5714	1192	1137
Амфибии	Рептилии	Рыбы	Беспозвоночные
157	293	742	1932

Заметим, что круговая диаграмма сохраняет свою наглядность и выразительность лишь при небольшом числе частей совокупности. В противном случае её применение малоэффективно.

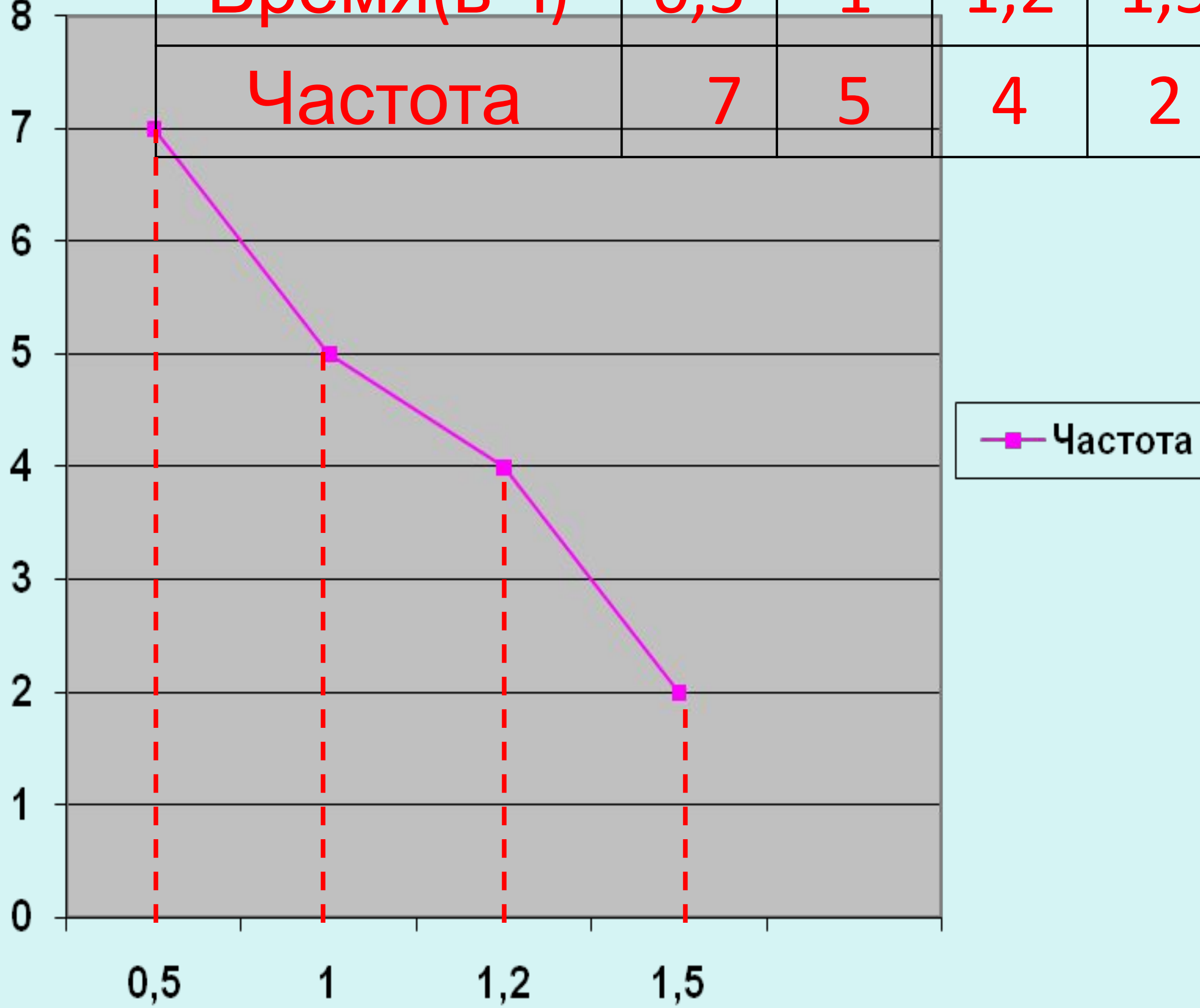
Полигон

Динамику изменения статистических данных во времени часто иллюстрируют с помощью полигона. Для построения полигона отмечают в координатной плоскости точки, абсциссами которых служат моменты времени, а ординатами – соответствующие им статистические данные. Соединив последовательно эти точки отрезками, получают ломаную, которую называют полигоном. Полигоны используют для наглядного изображения распределения данных, полученных в результате статистического исследования. Если данные представлены в виде таблицы частот или относительных частот, то для построения полигона отмечают в координатной плоскости точки, абсциссами которых служат статистические данные, а ординатами – их частоты или относительные частоты. Соединив эти точки отрезками, получают полигон распределения данных.

Рассмотрим пример. В результате опроса было установлено, сколько времени у них уходит на выполнение домашнего задания по математике в 5-х классах. После обработки данные были внесены в таблицу частот. Используя результаты построим полигон.

Время(в ч)	0,5	1	1,2	1,5
Частота	7	5	4	2

Время(в ч)	0,5	1	1,2	1,5
Частота	7	5	4	2

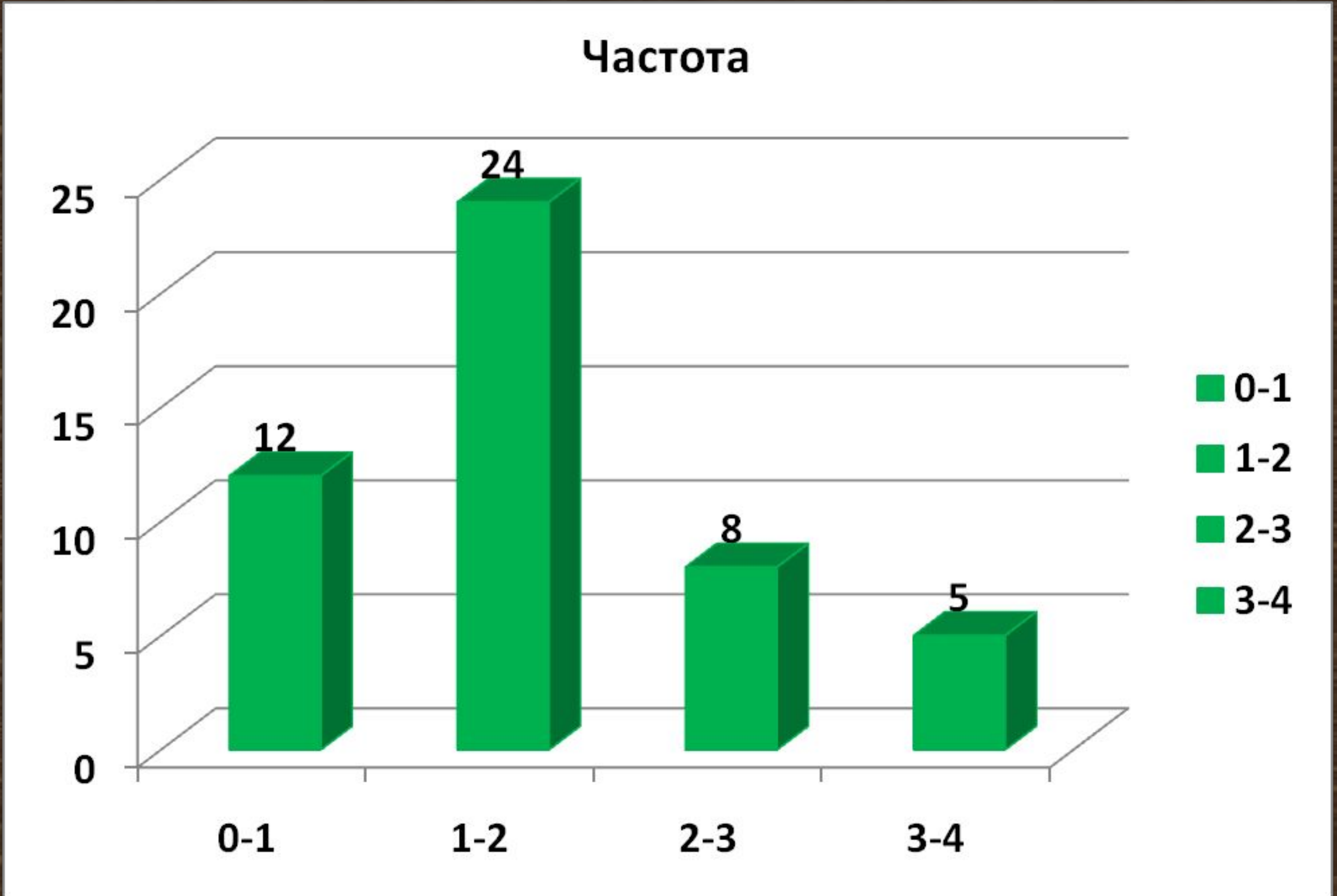


Гистограмма

Интервальные ряды данных изображают с помощью гистограмм. Гистограмма представляет собой ступенчатую фигуру, составленную из сомкнутых прямоугольников. Основание каждого прямоугольника равно длине интервала, высота – частоте или относительной частоте. Таким образом, в гистограмме, в отличии от обычной столбчатой диаграммы, основания прямоугольников выбираются не произвольно, а строго определены длиной интервала.

На основе опроса была составлена следующая таблица распределения учащихся по времени, которое они затратили в определенный учебный день на просмотр телепередач:

Время, ч.	Частота
0-1	12
1-2	24
2-3	8
3-4	5



Частота
25
20
15
10
5
0
24
12
8
5
0-1
1-2
2-3
3-4
0-1
1-2
2-3
3-4

Алгоритм вычисления дисперсии

Для нахождения дисперсии D данных $x_1, x_2, \ldots, x_n$ измерения следует вычислить:

1) среднее значение $M = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$;

2) отклонения данных от M, т. е. $x_1 - M$, $x_2 - M$, …, $x_{n-1} - M$, $x_n - M$;

3) квадраты $(x_i - M)^2$ отклонений $(x_i - M)$ ($i = 1, 2, 3, \ldots, n$), найденных на предыдущем шаге;

4) среднее значение всех квадратов отклонений — это и есть дисперсия D:

$$D = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \ldots + (x_{n-1} - M)^2 + (x_n - M)^2}{n};$$

$\sigma = \sqrt{D}$ — среднее квадратическое отклонение.

Чем меньше дисперсия D или среднее квадратичное отклонение σ, тем **ближе** данные измерения к своему среднему значению. Заметим, что вычисление этих характеристик достаточно трудоемкое занятие.

Пример 3

Для отбора почетного караула измерили рост (в см) двух групп солдат по пять человек и получили результаты – группа А: 178, 182, 180, 183, 177; группа Б: 183, 186, 180, 182, 184. Для каждой группы определим дисперсию D и среднее квадратичное отклонение σ и найдем группу, более однородную по росту.

Для группы А среднее арифметическое

$$M = \frac{178+182+180+183+177}{5} =$$

Солдат	1	2	3	4	5
Рост	178	182	180	183	177
Отклонение	–2	2	0	3	–3
Отклонение	–2	2	0	3	–3
Квадрат отклонения	4	4	0	9	9

$$M = \frac{178+182+180+183+177}{5} =$$

Найдем дисперсию $D = \dfrac{4+4+0+9+9}{5} =$

и среднес квадратичное отклонение $\sigma = \sqrt{5,2}$

Солдат	1	2	3	4	5
Рост	183	186	180	182	184
Отклонение	0	–3	3	–1	1
Квадрат отклонения	0	9	9	1	1

Аналогичные вычисления проделаем для группы Б.

Среднее арифметическое

$$M = \frac{183 + 186 + 180 + 182 + 184}{5}$$

Найдем дисперсию $D = \dfrac{0 + 9 + 9 + 1 + 1}{5}$

и среднее квадратичное отклонение $\sigma = \sqrt{4} = 2.$

Сравнивая полученные результаты, опредслим, что более однородной по росту является группа Б и ее целесообразно назначить в караул.

- Вариант 3. 1.Определение моды измерений, размаха, медианы, среднего арифметического.

- 2.Приведен рост (в см) семи человек:

183, 186, 183, 185,180,182, 184.

Найдите среднее, моду, медиану, дисперсию, среднее квадратичное отклонение.

Солдат	1	2	3	4	5	6	7
Рост							
Отклонение							
Квадрат отклонения							