ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

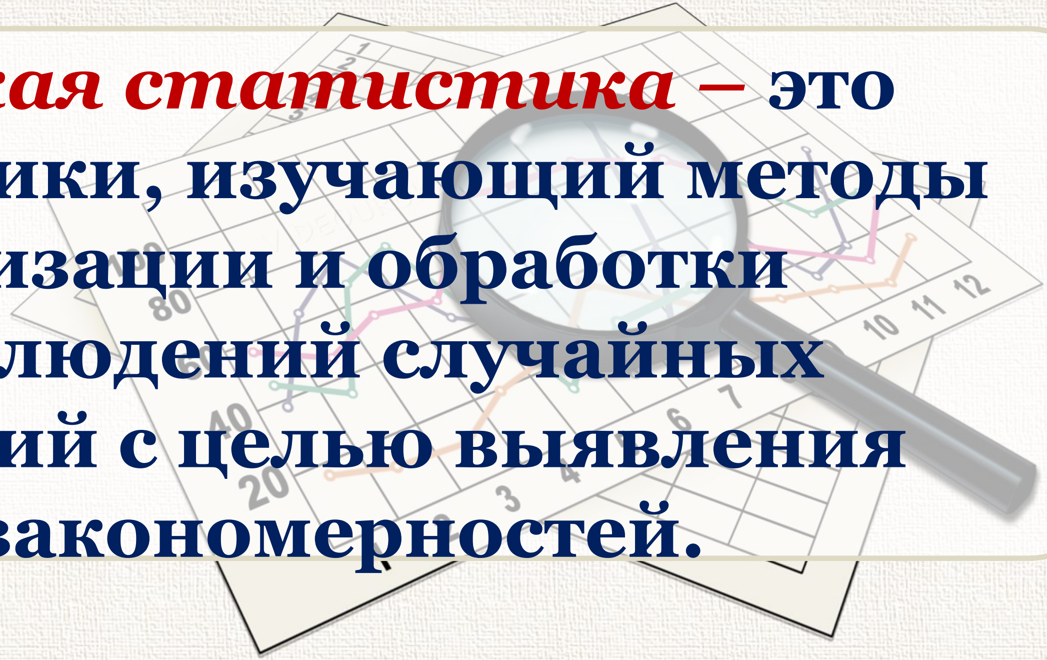
Статистика

(лат. «status») состояние дел

Статистика **— это точная наука, изучающая методы сбора, анализа и обработки данных, которые описывают массовые действия, явления и процессы.**

<u>Статистика</u>

(лат. «status») состояние дел



Математическая статистика **– это раздел математики, изучающий методы сбора, систематизации и обработки результатов наблюдений случайных массовых явлений с целью выявления существующих закономерностей.**

Статистика

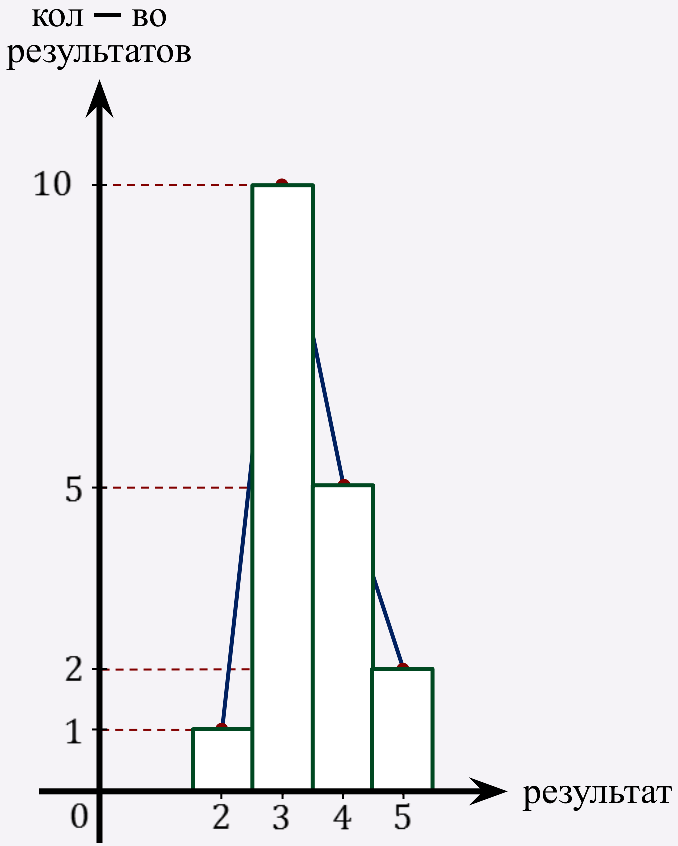
(лат. «status») состояние дел

Статистика ***имеет дело со случайными величинами.***

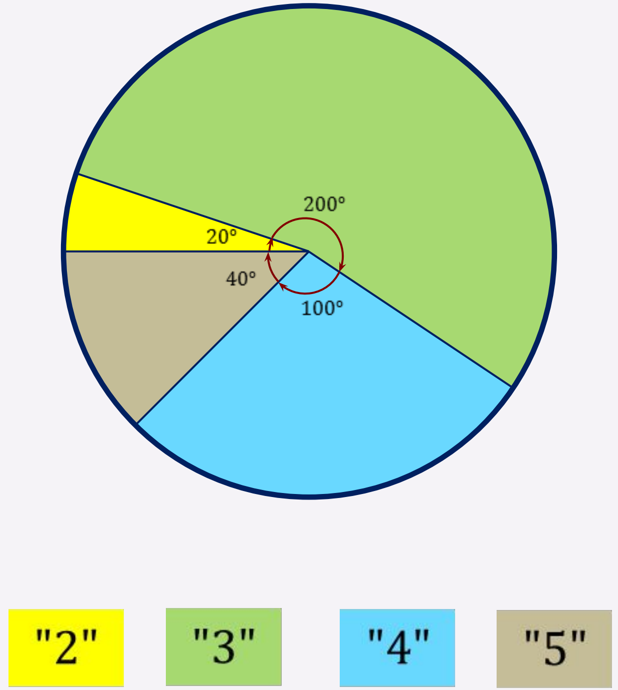
Случайные величины ***-величины, которые в ходе наблюдений или испытаний могут принимать различные значения. Можно сказать, чтоих значения зависят от случая.***

Многоугольник распределений

Гистограмма распределений



Круговая диаграмма



Этапы статистической обработки данных

1. Упорядочить и сгруппировать данные измерения

$$\underbrace{2}_{1},\underbrace{3,3,3,3,3,3,3,3,3,3}_{10},\underbrace{4,4,4,4,4}_{5},\underbrace{5,5}_{2}$$

2. Составить таблицу распределения данных

результат	2	3	4	5
количество результатов	1	10	5	2

3. Построить графики распределения данных

4. Получить паспорт измерения данных

объём, размах, мода измерения, среднее (или среднее арифметическое)

Таблица распределения значений случайной величины по их вероятностям

1-я кость	2-я кость					
	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	2	3	4	5	6	7	8	9	10	11	12
Количество	1	2	3	4	5	6	5	4	3	2	1
P	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Таблица распределения значений случайной величины по их относительным частотам

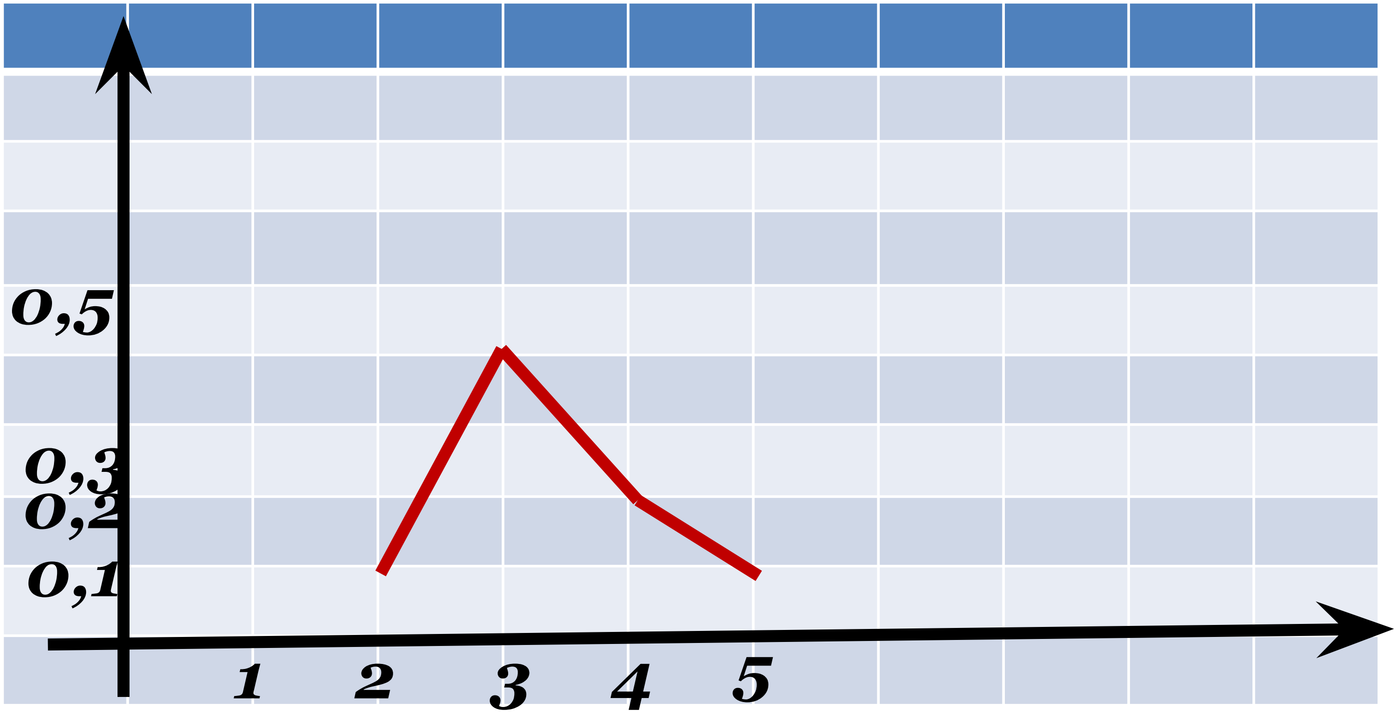
X	«2»	«3»	«4»	«5»
M	3	15	9	3
W	$\frac{1}{10}$	$\frac{1}{2}$	$\frac{9}{30}$	$\frac{1}{10}$

$$W = \frac{M}{N}$$

$$N = 3 + 15 + 9 + 3 = 30 = \sum M$$

$$\frac{1}{10} + \frac{1}{2} + \frac{9}{30} + \frac{1}{10} = 1 = \sum W$$

X	«2»	«3»	«4»	«5»
W	0,1	0,5	0,3	0,1

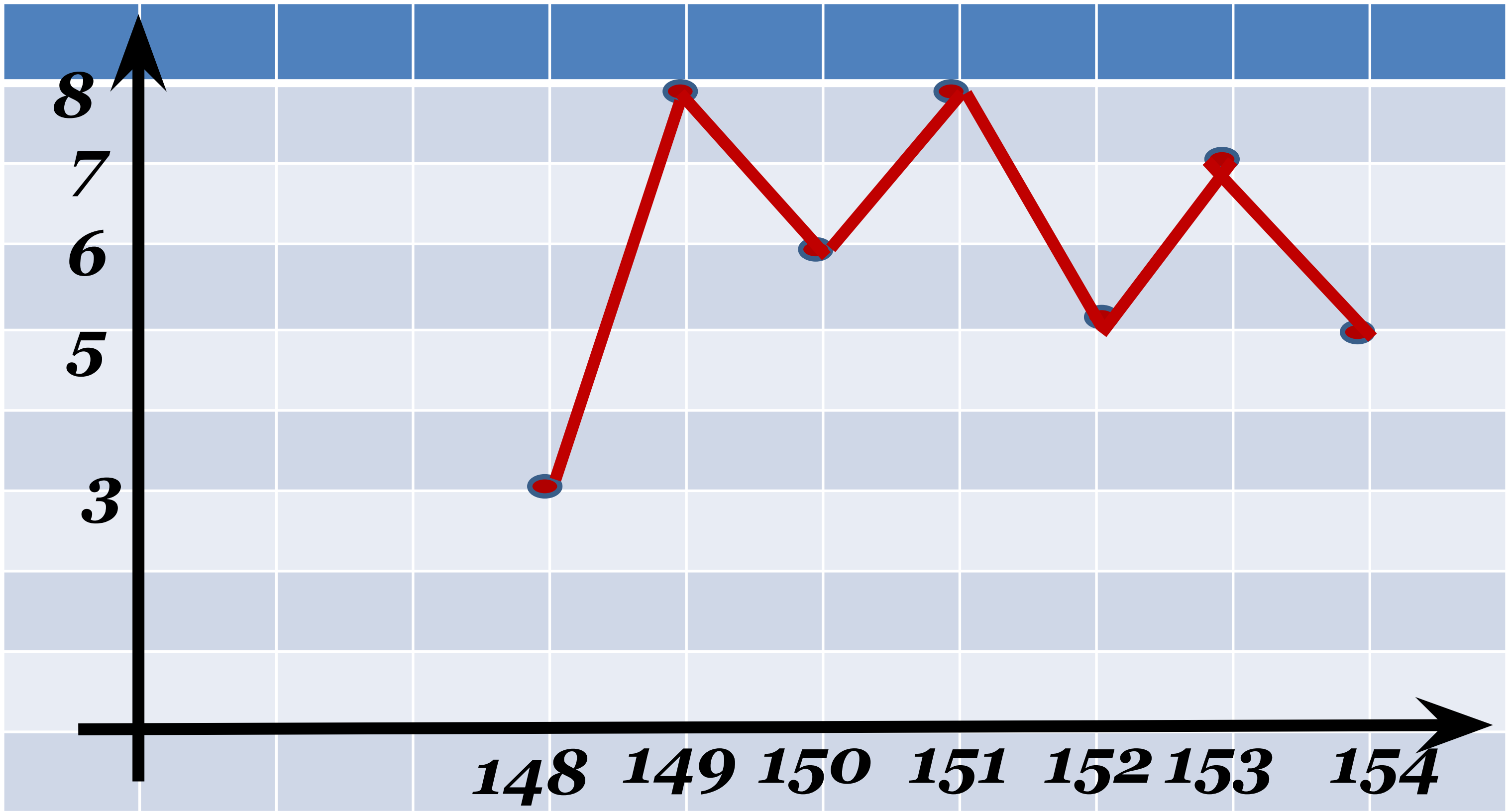


Рост 50 спортсменок занесён в таблицу:

148	148	148	149	149	149	149	149	149	149
149	150	150	150	150	150	150	150	150	150
150	151	151	151	151	151	151	151	151	152
152	152	152	152	152	152	152	152	153	153
153	153	153	153	153	154	154	154	154	154

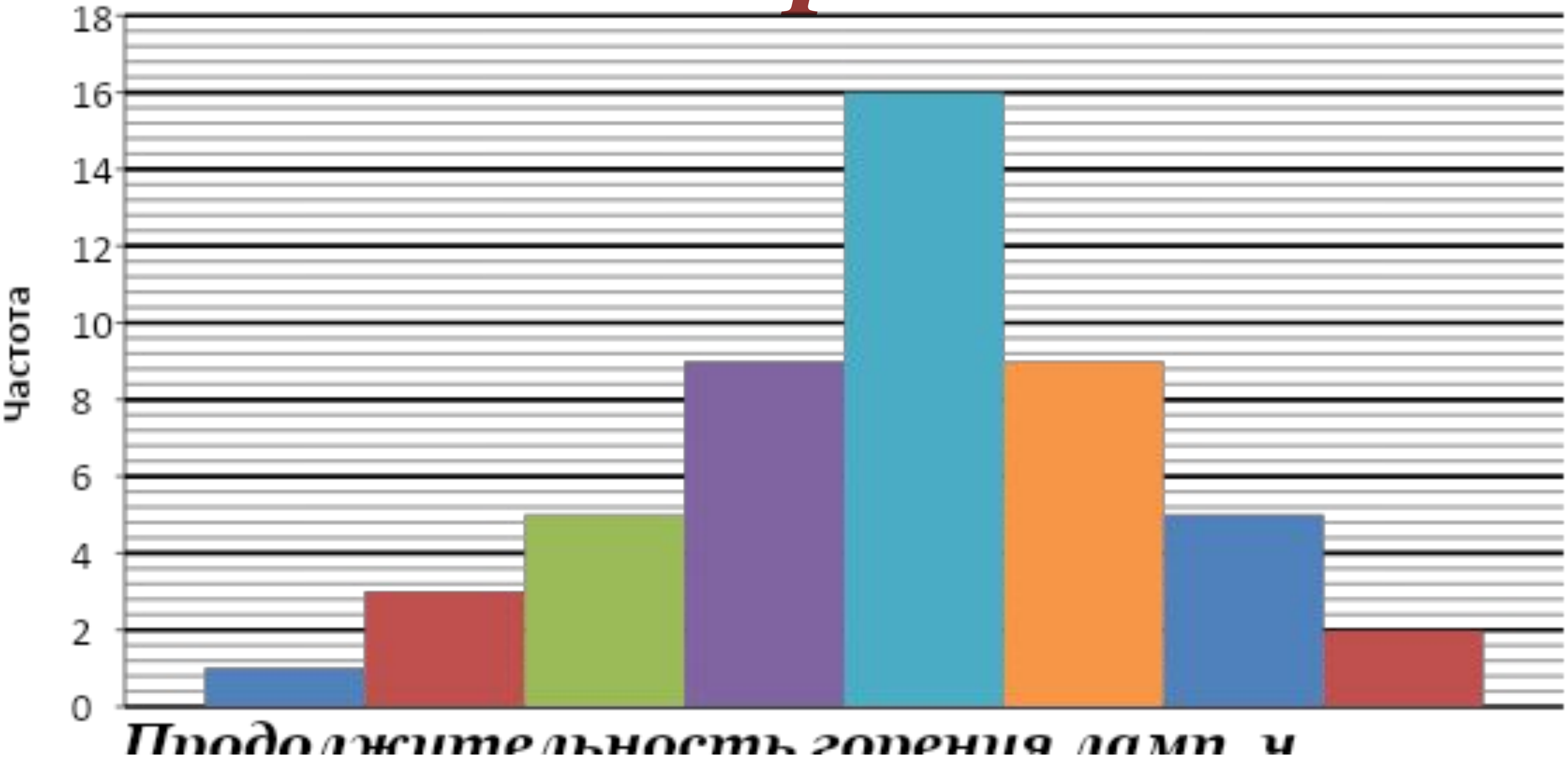
X	148	149	150	151	152	153	154
M	3	8	10	8	9	7	5
W	$\frac{3}{50}$	$\frac{4}{25}$	$\frac{1}{5}$	$\frac{4}{25}$	$\frac{9}{50}$	$\frac{7}{50}$	$\frac{5}{50}$

X	148	149	150	151	152	153	154
M	3	8	6	8	5	7	5



***Гистограммы** представляют собой ступенчатую фигуру, составленную из прямоугольников. Основание каждого прямоугольника равно длине интервала, а высота – частоте или относительной частоте. Таким образом, в гистограмме в отличие от обычной столбчатой диаграммы, основание прямоугольников выбирают не произвольно, а строго определенной длины интервала.*

Гистограмма



Число объектов генеральной совокупности и выборки называют соответственно *объемом* генеральной совокупности и *объемом* выборки.

Вместо изучения всех элементов совокупности, которую называют <u>генеральной совокупностью</u>, обследуют ее значительную часть, выбранную случайным образом, называемую <u>выборкой</u>.

Выборку называют <u>репрезентативной</u>, если в ней присутствуют все значения случайной величины примерно в тех же пропорциях, что и в генеральной совокупности.

На пример:

Генеральная совокупность – ***жители большого города.***

Репрезентативная выборка – ***жильцы многоквартирного дома, в котором примерно в тех же пропорциях, что и в самом городе, проживают люди разных возрастов.***

$\overline{\overline{S}}$ - ***объем генеральной совокупности***

N - ***объем репрезентативной выборки***

$M_1, M_2, \ldots, M_k$ - **частоты**

$$\boxed{\sum \boldsymbol{M} = \boldsymbol{N}}$$

$S_1, S_2, \ldots, S_k$ - ***частоты в генеральной совокупности***

$$\boxed{\sum \boldsymbol{S} = \overline{\overline{\boldsymbol{S}}}}$$

Для идеально составленной репрезентативной выборки должно выполняться равенство:

$$\frac{M_i}{N} = W_i = \frac{S_i}{\overline{\overline{S}}}, \qquad (1)$$

Где i – порядковый номер значения признака (1≤i≤k).

$$S_i = \overline{\overline{S}}\,\frac{M_i}{N} \text{ или } S_i = \overline{\overline{S}}W_i, \qquad (2)$$

$$\text{где } 1 \le i \le k.$$

Фабрика резиновых изделий выиграла тендер на изготовление $S = 10000$ армейских противогазов. Для определения того, сколько противогазов каждого из пяти существующих размеров следует изготовить, были сделаны замеры у N=100 случайным образом выбранных солдат ближайшей воинской части. Распределение размеров противогазов X по частотам M оказалось следующим:

X	0	1	2	3	4
M	5	21	47	22	5

Сколько противогазов каждого размера будет изготавливать фабрика?

N=100 солдат (объем репрезентативной выборки)

$\overline{\overline{S}}$ = 10000 - объем генеральной совокупности

Количество противогазов соответствующего размера можно найти по формуле (3)

Размер (Х)	0	1	2	3	4
Частота в выборке (М)	**5**	**21**	**47**	**22**	**5**
Относительная частота	**0,05**	**0,21**	**0,47**	**0,22**	**0,05**
Количество противогазов	**500**	**2100**	**4700**	**2200**	**500**

=100

=1

=10000

Центральные тенденции

В книгах по статистике моду, медиану и среднее арифметическое объединяют одним термином – меры центральной тенденции

(или, короче, центральные тенденции).

Мода

Модой ряда чисел называется число, которое встречается в данном ряду чаще других.

25

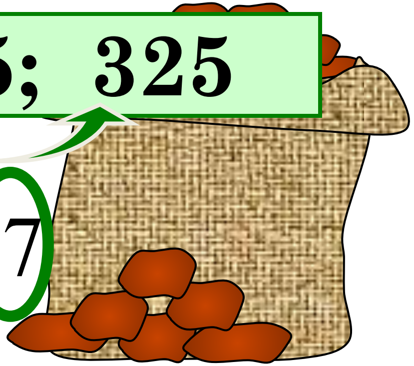
Медиана

Медианой упорядоченного ряда чисел с нечётным числом членов называется число, записанное посередине ряда.

250; 275; 286; $\boxed{290;}$ 296; 315; 325

1 2 3 4 5 6 7

290



Медиана

Медианой упорядоченного ряда чисел с чётным числом членов называется среднее арифметическое двух чисел, записанных посередине ряда.

18; 20; 23; 25; 25; $\boxed{25; 25}$; 26; 32; 34; 34; 37

$$\frac{25 + 25}{2} = 25$$

Средним значением случайной величины называется **среднее арифметическое** всех её значений

При опросе 12 учащихся узнали время затраченное на выполнение домашн

Получили такие данные:

23; 18; 25; 20; 25; 25; 32; 37; 34; 26; 34; 25

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

$$\overline{X} = \frac{23+18+25+20+25+25+32+37+34+26+34+25}{12} = \frac{324}{12} = 27$$

Среднее арифметическое: 27 минут

Произвели сбор данных о расходе электроэнергии в 9 квартирах. Получили следующие результаты:

№ кв.	1	2	3	4	5	6	7	8	9
расход	85	64	78	93	72	91	72	75	82

Составим из данных, приведенных в таблице, упорядоченный ряд:

64, 72, 72, 75, 78, 82, 85, 91, 93

Найдите размах, моду и медиану

29

72

78

В городе пять школ. В таблице приведен средний балл, полученный выпускниками каждой из этих школ за экзамен по математике. Найдите средний балл выпускного экзамена по математике по всему городу?

Номер школы	1	2	3	4	5
Количество выпускников	60	70	30	50	70
Средний балл	60	54	68	72	54

1. $60+70+30+50+70=280$
2. ***Если умножить количество учеников в школе на средний балл по школе , то получиться сумма баллов в этой школе, а если сложить все такие произведения, то сумма всех баллов по городу равна***

$$60 \cdot 60+70 \cdot 54+30 \cdot 68+50 \cdot 72+70 \cdot 54=$$

$$=3600+3780+2040+3600+3780 = 16800$$

3. ***Средний балл по городу равен*** $1685400:280=60$

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

23 18 25 20 25 25 32 37 34 26 34 25

Наибольшее -

Наименьшее -

$$37 - 18 = 19$$

Меры разброса

Отклонением от среднего ***называют разность между рассматриваемым значением случайной величины и средним значением выборки.***

Пример:

Задана выборка ***52,54,50,48,46.***

Пусть значение величины $X_1 = 52$*, а значение среднего*

$\bar{X} = (52+54+50+48+46):5 = 50,$

отклонение от среднего $X_1 - \bar{X} = 52 - 50 = 2.$

Меры разброса

Очевидно, отклонение от среднего может быть как положительным, так и отрицательным числом. Нетрудно понять, что сумма отклонений всех значений выборки от среднего значения равна нулю. Поэтому характеристикой стабильности элементов совокупности может служить сумма квадратов отклонений от среднего(чем меньше, тем лучше).

Меры разброса

Среднее арифметическое квадратов

Корень квадратный из дисперсии называют ***среднем квадратичным отклонением*** ***и обозначают*** $\sigma = \sqrt{D}$

N

Для оценки степени отклонения от среднего значения удобно иметь дело с величиной той же размерности, что и сама величина X. С этой целью используют значения корня квадратного из дисперсии

$\sigma = \sqrt{D}$

На испытательном стенде оружейного завода пристреливают готовые ружья, т.е. уточняют и корректируют их прицел.

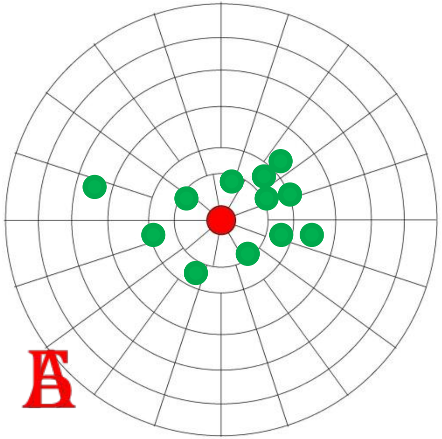
	Выстрелы									
Ружьё А										
Ружьё Б										

Среднее для ружья А:

$$\frac{1+1+2+1{,}5+2+2+1{,}5+1{,}5+0{,}5+1}{10}=1{,}4$$

Среднее для ружья Б:

$$\frac{1+0-1{,}5+1{,}5-0{,}5-1{,}5+2+1-1+2}{10}=0{,}3$$



$D = 0,24$ $(x_i - M)^2$										
$\sigma \approx 0,5$										
результат										
отклонение										
квадрат отклонения	0,16	0,16	0,36	0,01	0,36	0,36	0,01	0,01	0,81	0,16

$$D = 0,24 \quad \Longrightarrow \sigma = \sqrt{0,24} \approx 0,5$$

$D = 1,71$ $(x_i - M)^2$ $\sigma \approx 1,31$										
результат										
отклонение										
квадрат отклонения	0,49	0,09	3,24	1,44	0,64	3,24	2,89	0,49	1,69	2,89

$$D = 1,71 \quad \Rightarrow \sigma = \sqrt{1,71} \approx 1,31$$

$D = 0{,}24$ $\sigma \approx 0{,}5$										
результат										
отклонение										
квадрат отклонения										

$D = 1{,}71$ $\sigma \approx 1{,}31$										
результат										
отклонение										
квадрат отклонения										