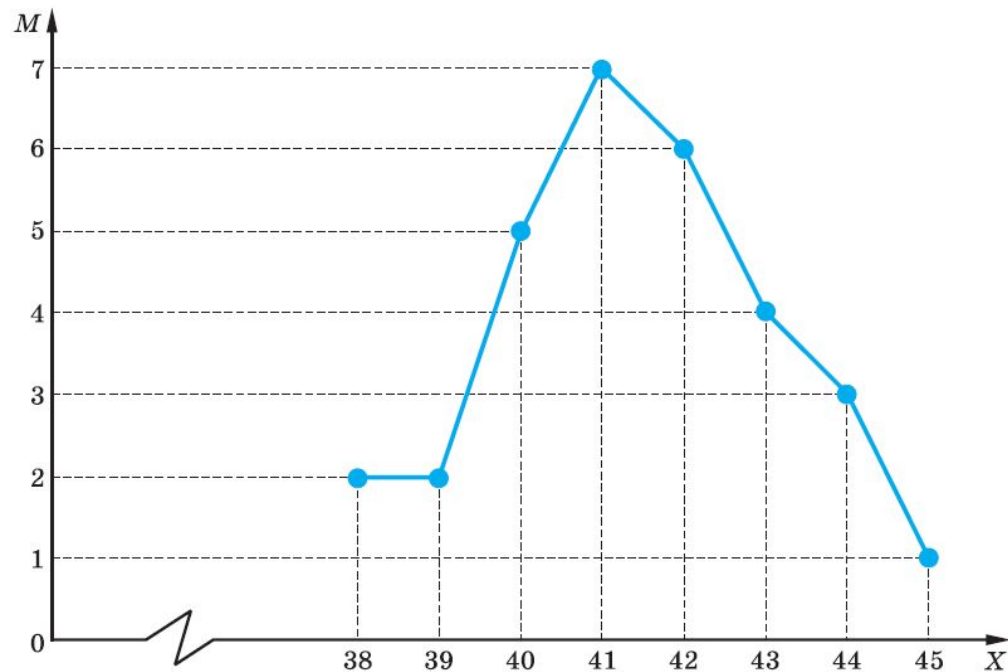


# Случайная величина

## Центральные тенденции



Преподаватель математики,  
кандидат физико-математических наук  
Рыикова Анна Анатольевна

# Случайные величины

Под *случайной величиной* в теории вероятностей понимают переменную величину, которая в данном случайном эксперименте может принимать те или иные числовые значения с определенной вероятностью. Обозначают случайные величины прописными буквами латинского алфавита:  $X, Y, Z, \dots$ , а их значения — соответствующими строчными буквами:  $x, y, z, \dots$ . Тот факт, что случайная величина  $X$  приняла значение  $x$ , записывают так:  $X = x$ .

Например, сумма чисел (очков), выпадающая при бросании двух игральных костей, — случайная величина. Обозначим её  $X$ , тогда  $X_1 = 2$ ,  $X_2 = 3$ ,  $X_3 = 4$ , ...,  $X_{11} = 12$  — значения этой случайной величины.

| I кость \ II 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 |

Значения случайной величины  $X$  и соответствующая вероятность ее появления ( $p_1, p_2, \dots, p_{10}, p_{11}^*$ ) приведены в таблице:

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

# Случайные величины

Отметим, что случайную величину можно задать в любом случайном эксперименте. Для этого достаточно каждому элементарному событию из пространства элементарных событий эксперимента поставить в соответствие некоторое число (в этом случае говорят, что задана числовая функция, областью определения которой является пространство элементарных событий).

*Случайной величиной называется числовая функция, областью определения которой является пространство элементарных событий.*

# Дискретная случайная величина

*Дискретная случайная величина* — это такая случайная величина, значения которой могут быть либо конечными, либо счетными. Под счетностью имеется в виду, что значения случайной величины можно занумеровать.

Приведем примеры дискретных случайных величин:

- а) число попаданий в мишень при  $n$  выстрелах, здесь возможные значения  $0, 1, \dots, n$ .
- б) число выпавших гербов при подкидывании монеты, здесь возможные значения  $0, 1, \dots, n$ .
- в) число вызовов, поступающих на АТС (счетное множество значений).

# Закон распределения дискретной случайной величины

Дискретная случайная величина  $X$  может принимать значения  $x_1, \dots, x_n$  с вероятностями  $p(x_1), \dots, p(x_n)$ . Соответствие между этими значениями и их вероятностями называется *законом распределения дискретной случайной величины*. Как правило, это соответствие задается с помощью таблицы, в первой строке которой указывают значения  $x_1, \dots, x_n$ , а во второй строке соответствующие этим значениям вероятности  $p_1, \dots, p_n$ .

|       |       |       |         |       |
|-------|-------|-------|---------|-------|
| $X_i$ | $x_1$ | $x_2$ | $\dots$ | $x_n$ |
| $p_i$ | $p_1$ | $p_2$ | $\dots$ | $p_n$ |

**Пример** Пусть случайная величина  $X$  — число выпавших очков при подбрасывании игрального кубика. Такая случайная величина  $X$  может принимать следующие значения 1, 2, 3, 4, 5, 6. Вероятности всех этих значений равны  $1/6$ . Тогда закон распределения вероятностей случайной величины  $X$ :

|       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|
| $X_i$ | 1     | 2     | 3     | 4     | 5     | 6     |
| $p_i$ | $1/6$ | $1/6$ | $1/6$ | $1/6$ | $1/6$ | $1/6$ |

*Замечание.* Поскольку в законе распределения дискретной случайной величины  $X$  события 1, 2, ..., 6 образуют полную группу событий, то в сумме вероятности должны быть равны единице, то есть  $\sum p_i = 1$ .

# Непрерывная случайная величина

Если случайная величина может принимать любое значение на некотором промежутке, то такая величина называется *непрерывной*. Например, время  $T$  ожидания автобуса на остановке является непрерывной случайной величиной в случае, если пассажир знает, что автобусы ходят через 10 мин, и приходит на остановку случайным образом. Эта случайная величина принимает любое числовое значение  $T \in [0; 10]$ .

Очевидно, что число значений непрерывной случайной величины бесконечно независимо от того, является ли промежуток значений ограниченным (отрезком) или неограниченным. Поэтому мы не можем для этой величины задать закон распределения так, как мы его задавали для дискретной случайной величины (с помощью таблицы, устанавливающей соответствие между каждым значением случайной величины и его вероятностью). Однако существует способ, с помощью которого можно задать распределение и непрерывной случайной величины. Для этого промежуток значений заданной непрерывной величины разбивают на части и считают вероятности попадания значений случайной величины в каждую из них.

# Непрерывная случайная величина

Например, пусть время горения  $X$  (в часах) электрической лампочки некоторого вида —  $X \in [0; 1000]$ . Промежуток  $[0; 1000]$  разделили на 5 одинаковых по длине частей и по результатам горения каждой из 100 экспериментальных лампочек составили следующую таблицу распределения случайной величины:

| $X$ | $[0; 200)$ | $[200; 400)$ | $[400; 600)$ | $[600; 800)$ | $[800; 1000]$ |
|-----|------------|--------------|--------------|--------------|---------------|
| $n$ | 1          | 3            | 10           | 18           | 68            |
| $P$ | 0,01       | 0,03         | 0,1          | 0,18         | 0,68          |

(проверка:  $\Sigma P = 0,01 + 0,03 + 0,1 + 0,18 + 0,68 = 1$ )

# Непрерывная случайная величина

В последнем примере для вычисления вероятностей принятия случайной величиной определенных значений было использовано статистическое определение вероятности (вероятности были оценены по результатам 100 экспериментов, в которых лампочки горели непрерывно до перегорания нити накаливания). В таких случаях удобно пользоваться расширенной таблицей распределения случайной величины, включая в нее *распределения рассмотренной величины по частотам и относительным частотам*. Тогда получим следующую таблицу распределений случайной величины  $X$ :

| $X$                                           | [0; 200) | [200; 400) | [400; 600) | [600; 800) | [800; 1000] |
|-----------------------------------------------|----------|------------|------------|------------|-------------|
| Частота $M = n(x)$                            | 1        | 3          | 10         | 18         | 68          |
| Относительная частота<br>$W = \frac{n(X)}{n}$ | 0,01     | 0,03       | 0,1        | 0,18       | 0,68        |
| $P$                                           | 0,01     | 0,03       | 0,1        | 0,18       | 0,68        |

Для проверки правильности заполнения такой таблицы используют то, что сумма относительных частот (как и сумма соответствующих вероятностей) равна 1 ( $\Sigma W = 1$ , или в процентах ( $\Sigma W = 100\%$ )), а сумма частот должна равняться количеству экспериментов  $n$  ( $\Sigma M = n$ ).

Рассмотрим составление такой таблицы по результатам экспериментов.

по закону больших чисел при значительном количестве экспериментов значения относительных частот близки к соответствующим вероятностям

# Пример

Результаты измерения роста 30 гимнасток одного спортивного клуба внесены в следующую таблицу:

|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 148 | 148 | 148 | 149 | 149 | 149 | 149 | 150 | 150 | 151 |
| 151 | 151 | 151 | 151 | 151 | 151 | 151 | 152 | 152 | 152 |
| 152 | 152 | 153 | 153 | 153 | 153 | 153 | 153 | 154 | 154 |

По этим данным составьте таблицу распределения значений случайной величины  $X$  — роста гимнасток клуба — по частотам ( $M$ ) и относительным частотам ( $W$ ).

## Решение

► Величина  $X$  принимает значения:

$$x_1 = 148, x_2 = 149, x_3 = 150, x_4 = 151, x_5 = 152, x_6 = 153, x_7 = 154.$$

Подсчитываем число  $M$  гимнасток каждого роста, заносим данные в частотную таблицу, а затем для каждого значения  $X$  находим значения относительной частоты  $W$ , зная, что  $n = 30$ . Получаем таблицу распределения значений случайной величины  $X$ :

|                   |                |                |                |                |               |               |                |
|-------------------|----------------|----------------|----------------|----------------|---------------|---------------|----------------|
| $X$               | 148            | 149            | 150            | 151            | 152           | 153           | 154            |
| $M$               | 3              | 4              | 2              | 8              | 5             | 6             | 2              |
| $W = \frac{M}{n}$ | $\frac{1}{10}$ | $\frac{2}{15}$ | $\frac{1}{15}$ | $\frac{4}{15}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{15}$ |

(проверка:  $\sum M = 3 + 4 + 2 + 8 + 5 + 6 + 2 = 30 = n$ ;

$$\sum W = \frac{1}{10} + \frac{2}{15} + \frac{1}{15} + \frac{4}{15} + \frac{1}{6} + \frac{1}{5} + \frac{1}{15} = \frac{30}{30} = 1) \triangleleft$$

# Полигон частот

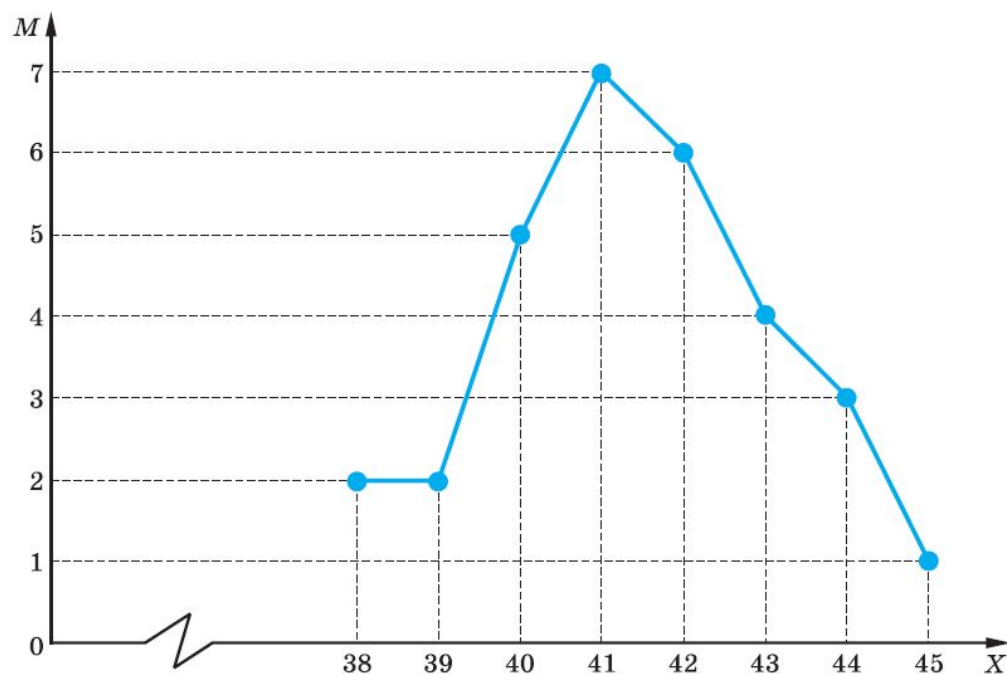
Распределение случайных величин можно задавать и иллюстрировать графически.

Пусть случайная величина  $X$  — размер обуви 30 мальчиков 11 класса одной из школ — имеет распределение по частотам, данное в таблице:

|     |    |    |    |    |    |    |    |    |
|-----|----|----|----|----|----|----|----|----|
| $X$ | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 |
| $M$ | 2  | 2  | 5  | 7  | 6  | 4  | 3  | 1  |

$n = \Sigma M = 30$

Отметим на координатной плоскости точки с координатами  $(x_1; m_1)$ ,  $(x_2; m_2)$ , ...,  $(x_8; m_8)$  и соединим их последовательно отрезками. Полученную ломаную линию называют *полигоном частот*.



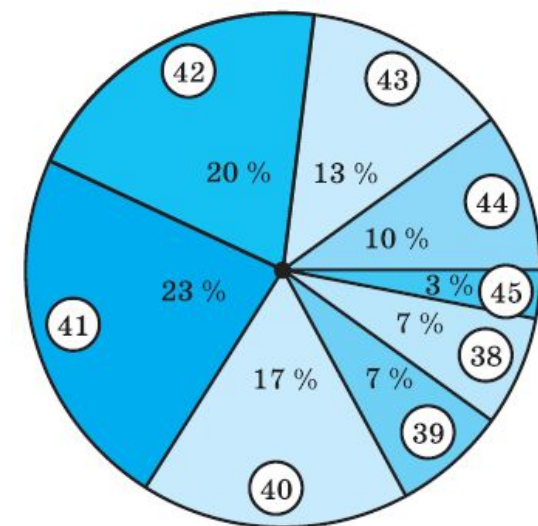
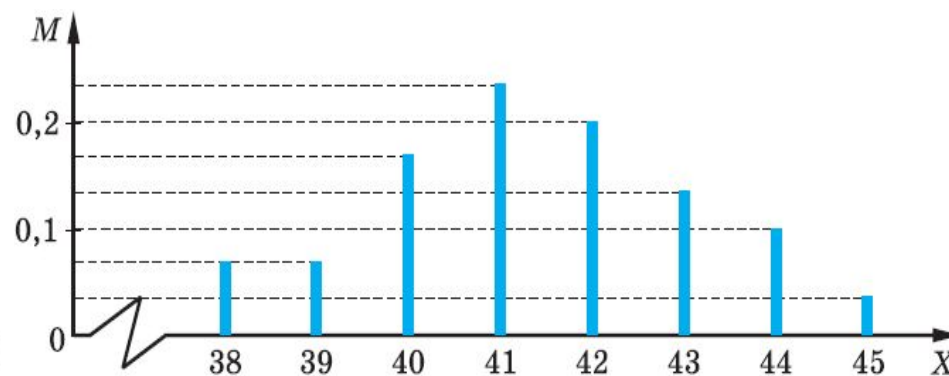
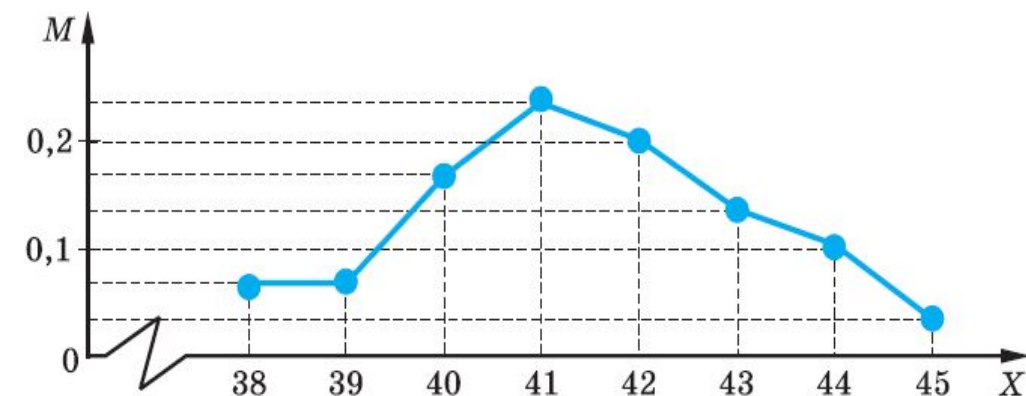
# Полигон относительных частот

Аналогично определяется и строится *полигон относительных частот* для случайной величины  $X$  (строятся точки с координатами  $(x_1; w_1)$ ,  $(x_2; w_2)$ , ...,  $(x_k; w_k)$ , где  $x_i$  — значения случайной величины, а  $w_i$  — соответствующие им относительные частоты.

Если вычислить относительные частоты для каждого значения случайной величины, рассмотренной в примере в начале этого пункта, то распределение величины  $X$  по относительным частотам можно задать таблицей:

| $X$ | 38                          | 39                          | 40                         | 41                          | 42                  | 43                          | 44                   | 45                          |                |
|-----|-----------------------------|-----------------------------|----------------------------|-----------------------------|---------------------|-----------------------------|----------------------|-----------------------------|----------------|
| $W$ | $\frac{1}{15} \approx 0,07$ | $\frac{1}{15} \approx 0,07$ | $\frac{1}{6} \approx 0,17$ | $\frac{7}{30} \approx 0,23$ | $\frac{1}{5} = 0,2$ | $\frac{2}{15} \approx 0,13$ | $\frac{1}{10} = 0,1$ | $\frac{1}{30} \approx 0,03$ | $\Sigma W = 1$ |

Также распределение случайной величины  $X$  по относительным частотам можно представить в виде полигона относительных частот, в виде линейной диаграммы или в виде круговой диаграммы, предварительно записав значения относительной частоты в процентах.



# Гистограмма частот

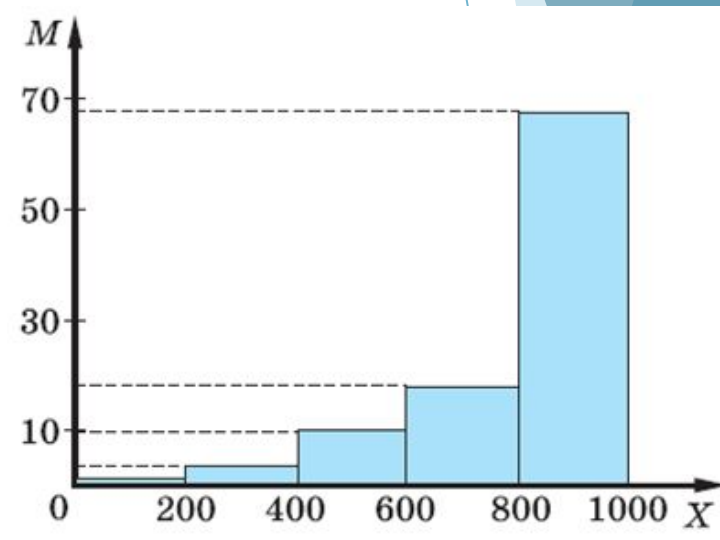
Распределение значений непрерывной случайной величины также можно представлять графически. Для этого промежуток значений заданной непрерывной величины разбивают на несколько равных частей и считают частоты попадания значений случайной величины в каждую из этих частей.

| $X$                | $[0; 200)$ | $[200; 400)$ | $[400; 600)$ | $[600; 800)$ | $[800; 1000]$ |
|--------------------|------------|--------------|--------------|--------------|---------------|
| Частота $M = n(x)$ | 1          | 3            | 10           | 18           | 68            |

Данные из этой таблицы можно представить с помощью так называемой *гистограммы частот* — ступенчатой фигуры.

Если основанием каждого столбца служит промежуток значений случайной величины длиной  $h$ , то высоту столбца берут равной  $\frac{M}{h}$  (это отношение называется *плотностью частоты* на рассмотренном промежутке), где  $M$  — частота значений величины  $X$  на соответствующем промежутке. Тогда площадь такого столбца будет равняться  $\frac{M}{h} \cdot h = M$ , а площадь фигуры под гистограммой равна

$$\Sigma M = n.$$



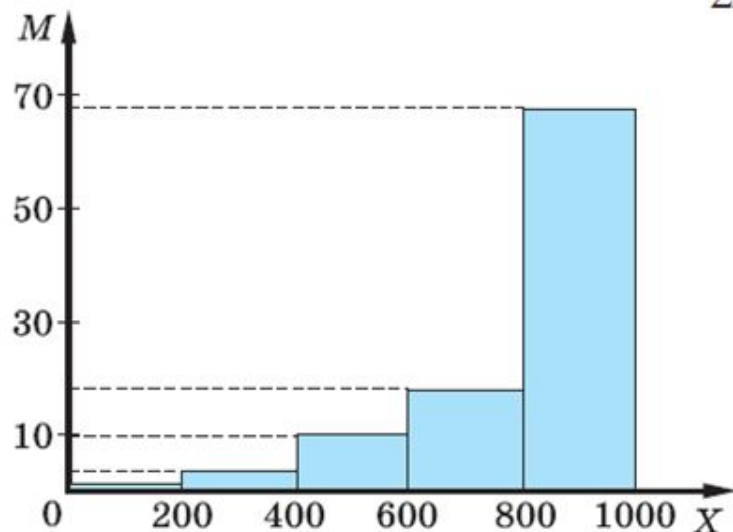
# Гистограмма частот

| $X$                | $[0; 200)$ | $[200; 400)$ | $[400; 600)$ | $[600; 800)$ | $[800; 1000]$ |
|--------------------|------------|--------------|--------------|--------------|---------------|
| Частота $M = n(x)$ | 1          | 3            | 10           | 18           | 68            |

Данные из этой таблицы можно представить с помощью так называемой *гистограммы частот* — ступенчатой фигуры.

Если основанием каждого столбца служит промежуток значений случайной величины длиной  $h$ , то высоту столбца берут равной  $\frac{M}{h}$  (это отношение называется *плотностью частоты* на рассмотренном промежутке), где  $M$  — частота значений величины  $X$  на соответствующем промежутке. Тогда площадь такого столбца будет равняться  $\frac{M}{h} \cdot h = M$ , а площадь фигуры под гистограммой равна

$$\Sigma M = n.$$



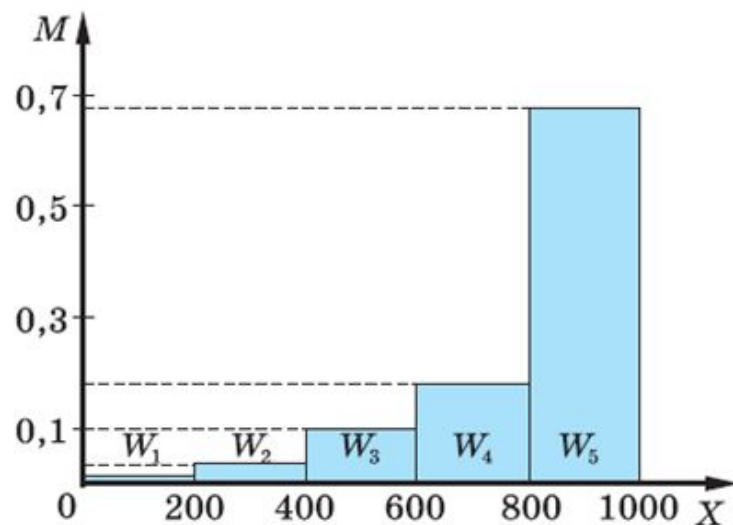
# Гистограмма относительных частот

| $X$ | $[0; 200)$ | $[200; 400)$ | $[400; 600)$ | $[600; 800)$ | $[800; 1000]$ |
|-----|------------|--------------|--------------|--------------|---------------|
| $W$ | 0,01       | 0,03         | 0,1          | 0,18         | 0,68          |

 $\Sigma W = 1$ 

Если договориться, что единица на горизонтальной оси соответствует величине  $h$  (в нашем примере  $h = 200$  ч), а единица на вертикальной оси — частоте, равной 1, то построенная на рисунке гистограмма удовлетворяет приведенным требованиям.

Если по данным предыдущей таблицы заполнить таблицу относительных частот, то построенную на ее основании ступенчатую фигуру называют *гистограммой относительных частот*



Гистограмму относительных частот строят обычно таким образом, чтобы площадь каждого столбца под ступенькой равнялась соответствующему значению  $W$ . Это делается аналогично построению гистограммы частот. Если основанием каждого столбца служит промежуток значений случайной величины длиной  $h$ , то высоту столбца берут равной  $\frac{W}{h}$  (это отношение называется *плотностью относительной частоты* на рассмотренном промежутке), где  $W$  — относительная частота значений величины  $X$  на соответствующем промежутке. Тогда площадь такого столбца будет равняться  $\frac{W}{h} \cdot h = W$ , а площадь фигуры под гистограммой равна единице ( $\Sigma W = 1$ ).

# Статистика

Статистика – наука, изучающая, обрабатывающая и анализирующая количественные данные о разнообразных массовых явлениях в жизни.

Экономическая статистика изучает изменение цен, спроса и предложения товаров, прогнозирует рост и падение производства и потребления.

Медицинская статистика изучает эффективность разных лекарств и методов лечения, вероятность возникновения некоторых заболеваний в зависимости от возраста, пола, наследственности, условий жизни, вредных привычек, прогнозирует распространение эпидемий.

Демографическая статистика изучает рождаемость, численность населения, его состав (возрастной, национальный, профессиональный).

А есть еще статистика финансовая, налоговая, биологическая, метеорологическая...

# Статистика

Статистика имеет многовековую историю. Уже в Древнем мире вели статистический учет населения. Однако случайное толкование статистических данных, отсутствие строгой научной базы статистических прогнозов даже в середине XIX в. еще не позволяли говорить о статистике как науке. Только в XX в. появилась математическая статистика — наука, опирающаяся на законы теории вероятностей. Выяснилось, что статистические методы обработки данных из самых разных областей жизни имеют много общего. Это позволило создать универсальные научно обоснованные методы статистических исследований и проверки статистических гипотез.

**математическая статистика — это раздел математики, изучающий математические методы обработки и использования статистических данных для научных и практических выводов.**

Результаты исследований, проводимых методами математической статистики, применяются для принятия решений. В частности, при планировании и организации производства, при контроле качества продукции, при выборе оптимального времени наладки или замены действующей аппаратуры (например, при определении времени замены двигателя самолета, отдельных частей станков и т. д.).

# Статистика

В математической статистике рассматриваются методы, которые дают возможность по результатам экспериментов (статистическим данным) делать определенные выводы вероятностного характера.

Среди основных задач математической статистики можно отметить следующие.

- 1. Оценка вероятности.** Пусть некоторое случайное событие имеет вероятность  $p > 0$ , но ее значение нам неизвестно. Требуется оценить эту вероятность по результатам экспериментов, то есть решить задачу об оценке вероятности через частоту.
- 2. Оценка закона распределения.** Исследуется некоторая случайная величина, точное выражение для закона распределения которой нам неизвестно. Необходимо по результатам экспериментов найти приближенное выражение для функции, задающей закон распределения.
- 3. Оценка числовых характеристик случайной величины** (математического ожидания, дисперсии).
- 4. Проверка статистических гипотез (предположений).** Исследуется некоторая случайная величина. Исходя из определенных рассуждений, выдвигается, например, гипотеза, что распределение этой случайной величины близко к нормальному. Необходимо по результатам экспериментов принять или отклонить эту гипотезу.

# Статистические наблюдения

Для изучения различных массовых явлений проводятся специальные статистические исследования. Любое статистическое исследование начинается с целенаправленного сбора информации об изучаемом явлении или процессе. Этот этап называют *этапом статистических наблюдений*. Для получения статистических данных в результате наблюдений похожие элементы некоторой совокупности сравнивают по разным признакам.

| Часто употребляемый термин | Смысл термина                                         | Научный термин                                    | Определение                                                                      |
|----------------------------|-------------------------------------------------------|---------------------------------------------------|----------------------------------------------------------------------------------|
| Общий ряд данных           | То, откуда выбирают                                   | <b>Генеральная совокупность</b>                   | <i>Множество всех возможных результатов наблюдения (измерения)</i>               |
| Выборка                    | То, что выбирают                                      | <b>Статистическая выборка, статистический ряд</b> | <i>Множество результатов, реально полученных в данном наблюдении (измерении)</i> |
| Варианта                   | Значение одного из результатов наблюдения (измерения) | <b>Варианта</b>                                   | <i>Одно из значений элементов выборки</i>                                        |
| Ряд данных                 | Значения всех результатов наблюдения (измерения)      | <b>Вариационный ряд</b>                           | <i>Упорядоченное множество всех вариантов</i>                                    |

# Выборка

При изучении реальных явлений часто бывает невозможно обследовать все элементы совокупности. Например, практически невозможно выяснить размеры обуви у всех людей планеты. А проверить, например, наличие листов некачественной фотобумаги в большой партии хотя и реально, но бессмысленно, потому что полная проверка приведет к уничтожению всей партии бумаги. В подобных случаях вместо изучения всех элементов совокупности, называемой *генеральной совокупностью*, обследуют ее значительную часть, выбранную случайным образом. Эту часть называют *выборкой*.

Если в выборке присутствуют все значения случайной величины в тех же пропорциях, что и в генеральной совокупности, то эту выборку называют *репрезентативной* (от французского *représentatif* — показательный).

Пусть  $S$  — объем генеральной совокупности,  $n$  — объем репрезентативной выборки, в которой  $k$  значений исследуемых признаков распределены по частотам  $M_1, M_2, \dots, M_k$ , где  $\sum M = n$ . Тогда в генеральной совокупности частотам  $M_1, M_2, \dots, M_k$  будут соответствовать частоты  $s_1, s_2, \dots, s_k$  тех же значений признака, что и в выборке ( $\sum s = S$ ). По определению репрезентативной выборки получаем:

$$\frac{M_i}{n} = W_i = \frac{s_i}{S}$$

где  $i$  — порядковый номер значения признака ( $1 \leq i \leq k$ ). Из этого соотношения находим

$$s_i = SW_i \left( \text{или } s_i = S \frac{M_i}{n} \right), \text{ где } 1 \leq i \leq k. \quad (1)$$

**Пример** Обувной цех должен выпустить 1000 пар кроссовок молодежного фасона. Для определения того, сколько кроссовок и какого размера необходимо выпустить, были выявлены размеры обуви у 50 случайным образом выбранных подростков. Распределение размеров обуви по частотам представлено в таблице:

|             |    |    |    |    |    |    |    |    |    |
|-------------|----|----|----|----|----|----|----|----|----|
| Размер (X)  | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 |
| Частота (M) | 2  | 5  | 6  | 12 | 11 | 7  | 4  | 2  | 1  |

 $\Sigma M = n = 50$ 

Сколько кроссовок разного размера будет изготавливать фабрика?

#### Решение

► Будем считать рассмотренную выборку объемом  $n = 50$  подростков репрезентативной. Тогда в генеральной совокупности (объемом  $S = 1000$ ) количество кроссовок каждого размера пропорционально количеству кроссовок соответствующего размера в выборке (и для каждого размера находится по формуле (1)). Результаты расчетов будем записывать в таблицу:

|                           |                |                |                |                |                 |                |                |                |                |
|---------------------------|----------------|----------------|----------------|----------------|-----------------|----------------|----------------|----------------|----------------|
| Размер (X)                | 36             | 37             | 38             | 39             | 40              | 41             | 42             | 43             | 44             |
| Частота (M)               | 2              | 5              | 6              | 12             | 11              | 7              | 4              | 2              | 1              |
| Относительная частота (W) | $\frac{1}{25}$ | $\frac{1}{10}$ | $\frac{3}{25}$ | $\frac{6}{25}$ | $\frac{11}{50}$ | $\frac{7}{50}$ | $\frac{2}{25}$ | $\frac{1}{25}$ | $\frac{1}{50}$ |
| Количество кроссовок (SW) | 40             | 100            | 120            | 240            | 220             | 140            | 80             | 40             | 20             |

 $\Sigma M = n = 50$   
 $\Sigma W = 1$   
 $\Sigma (SW) = S = 1000$ 

Ответ.

|                      |    |     |     |     |     |     |    |    |    |
|----------------------|----|-----|-----|-----|-----|-----|----|----|----|
| Размер               | 36 | 37  | 38  | 39  | 40  | 41  | 42 | 43 | 44 |
| Количество кроссовок | 40 | 100 | 120 | 240 | 220 | 140 | 80 | 40 | 20 |

# Статистические характеристики рядов данных

Иногда выборку случайных величин или всю генеральную совокупность этих величин приходится характеризовать одним числом. На практике это необходимо, например, для быстрого сравнения двух или больше совокупностей по общему признаку.

Рассмотрим конкретный пример.

Пусть после летних каникул провели опрос 10 девочек и 9 мальчиков одного класса о количестве книг, прочитанных ими за каникулы. Результаты были записаны в порядке опроса. Получили следующие ряды чисел:

для девочек: 4, 3, 5, 3, 8, 3, 12, 4, 5, 5;

для мальчиков: 5, 3, 3, 4, 6, 4, 4, 7, 4.

Чтобы удобнее было анализировать информацию, в подобных случаях числовые данные *ранжируют*, располагая их в порядке возрастания (когда каждое следующее число или больше, или не меньше предыдущего). В результате ранжирования получили следующие ряды.

Для девочек:

3, 3, 3, 4, 4, 5, 5, 5, 8, 12; (1)

для мальчиков:

3, 3, 4, 4, 4, 4, 5, 6, 7. (2)

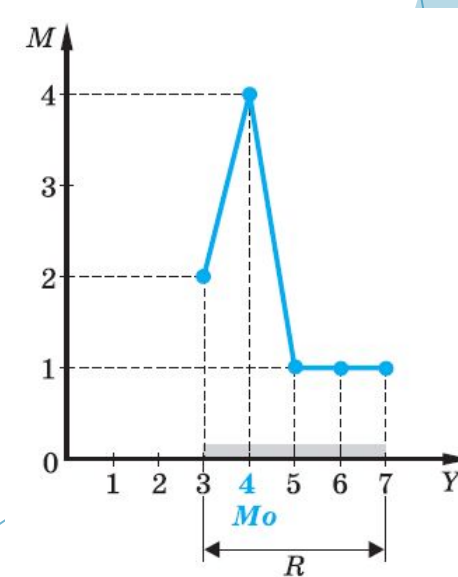
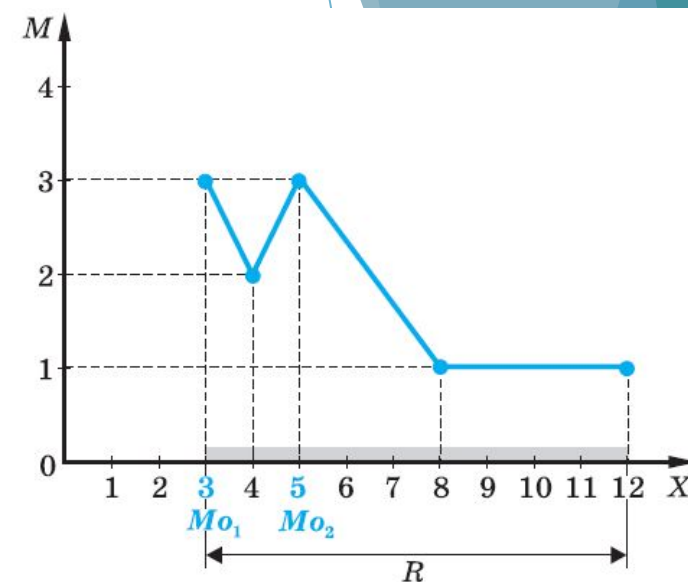
Тогда распределение по частотам  $M$  случайных величин:  $X$  — число книг, прочитанных за каникулы девочками, и  $Y$  — число книг, прочитанных за каникулы мальчиками, можно задать таблицами:

|     |   |   |   |   |    |
|-----|---|---|---|---|----|
| $X$ | 3 | 4 | 5 | 8 | 12 |
| $M$ | 3 | 2 | 3 | 1 | 1  |

$$\Sigma M = n = 10$$

|     |   |   |   |   |   |
|-----|---|---|---|---|---|
| $Y$ | 3 | 4 | 5 | 6 | 7 |
| $M$ | 2 | 4 | 1 | 1 | 1 |

$$\Sigma M = n = 9$$



# Ранжирование

| Определение                                                                                                                                                                               | Пример                                                                                                                                           |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| Ранжирование ряда данных                                                                                                                                                                  |                                                                                                                                                  |
| Под <i>ранжированием ряда данных</i> понимают расположение элементов этого ряда в порядке возрастания (имеется в виду, что каждое следующее число или больше, или не меньше предыдущего). | Если ряд данных выборки имеет вид<br>5, 3, 7, 4, 6, 4, 6, 9, 4,<br>то после ранжирования он превращается в ряд<br>3, 4, 4, 4, 5, 6, 6, 7, 9. (*) |

# Размах выборки

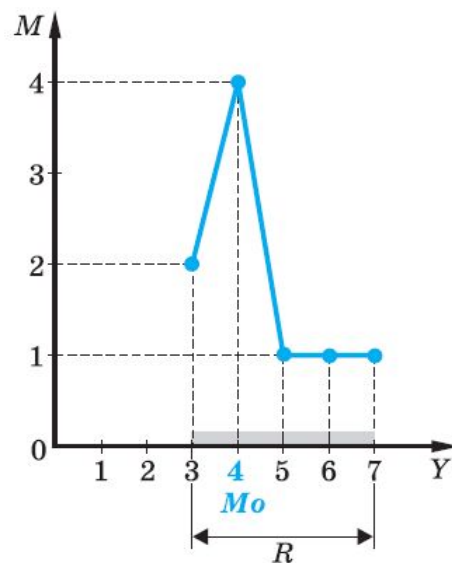
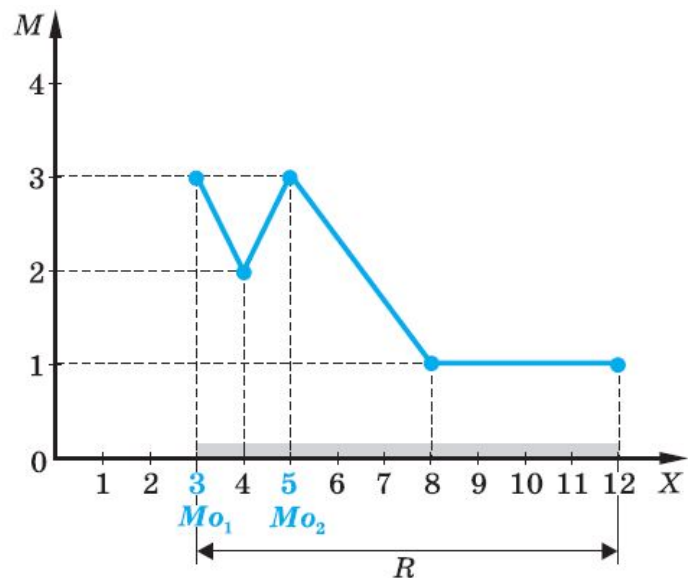
Размахом ряда чисел (обозначается  $R$ ) называют разность между наибольшим и наименьшим из этих чисел. Поскольку мы анализируем выборку случайных величин, то

**размах выборки — это разность между наибольшим и наименьшим значениями случайной величины в выборке.**

Для ряда (1) размах  $R = 12 - 3 = 9$ , а для ряда (2) размах  $R = 7 - 3 = 4$ . На графике размах — это длина области определения полигона частот.

3, 3, 3, 4, 4, 5, 5, 5, 8, 12; (1)

3, 3, 4, 4, 4, 4, 5, 6, 7. (2)



# Мода выборки

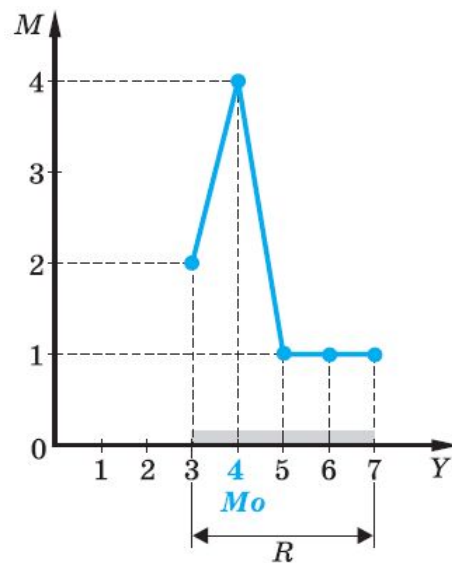
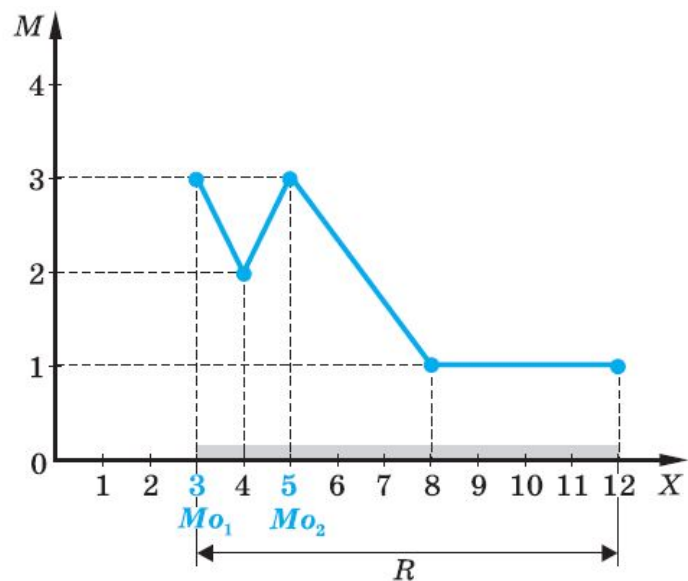
Важной статистической характеристикой ряда данных является его *мода* (обозначается  $Mo$ , от латинского слова *modus* — мера, правило).

**Мода — это значение случайной величины, встречающееся чаще остальных.**

Так, в ряду (1) две моды — числа 3 и 5:  $Mo_1 = 3$ ,  $Mo_2 = 5$ , а в ряду (2) одна мода — число 4:  $Mo = 4$ . На графике мода — это значение абсциссы точки, в которой достигается максимум полигона частот. Отметим, что моды может и не быть, если все значения случайной величины встречаются одинаково часто.

3, 3, 3, 4, 4, 5, 5, 5, 8, 12; (1)

3, 3, 4, 4, 4, 4, 5, 6, 7. (2)



# Мода выборки

Моду ряда данных обычно находят тогда, когда хотят выяснить некоторый типовой показатель. Например, когда изучают данные о моделях мужских рубашек, проданных в определенный день в универмаге, то удобно использовать такой показатель, как мода, который характеризует модель, пользующуюся наибольшим спросом (собственно, этим и объясняется название «мода»).

# Медиана выборки

*Медиана — это так называемое срединное значение упорядоченного ряда значений случайной величины (обозначается  $Me$ ).*

Медиана делит упорядоченный ряд данных на две равные по количеству элементов части.

**Если количество чисел в ряду нечетное, то медиана — это число, записанное посередине.**

Например, в ряду (2) нечетное количество элементов ( $n = 9$ ). Тогда его медианой является число, стоящее посередине, то есть на пятом месте:  $Me = 4$ .

3, 3, 4, 4, (4), 4, 5, 6, 7

**медиана**

Следовательно, о мальчиках можно сказать, что одна половина из них прочитала не больше 4 книг, а вторая — не меньше 4 книг. (Отметим, что в случае нечетного  $n$  номер среднего члена ряда равен  $\frac{n+1}{2}$ .)

**Если количество чисел в ряду четное, то медиана — это среднее арифметическое двух чисел, стоящих посередине.**

Например, в ряду (1) четное количество элементов ( $n = 10$ ). Тогда его медианой является число, равное среднему арифметическому чисел, стоящих посередине, то есть на пятом и шестом местах:  $Me = \frac{4+5}{2} = 4,5$ .

3, 3, 3, 4, 4, ↑ 5, 5, 5, 8, 12

**(4,5) — медиана**

Следовательно, о девочках можно сказать, что одна половина из них прочитала меньше 4,5 книг, а вторая — больше 4,5 книг. (Отметим, что в случае четного  $n$  номера средних членов ряда равны  $\frac{n}{2}$  и  $\frac{n}{2} + 1$ .)

# Среднее значение

*Средним значением случайной величины  $X$  (обозначается  $\bar{X}$ ) называется среднее арифметическое всех ее значений.*

Если случайная величина  $X$  принимает  $n$  значений  $x_1, x_2, \dots, x_n$ , то

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Если случайная величина  $X$  принимает значения  $x_1, x_2, \dots, x_k$  соответственно с частотами  $m_1, m_2, \dots, m_k$  (тогда  $\sum m = n$ ), то, заменяя одинаковые слагаемые в числителе на соответствующие произведения, получаем, что среднее арифметическое можно вычислять по формуле

$$\bar{X} = \frac{x_1 m_1 + x_2 m_2 + \dots + x_k m_k}{n}. \quad (3)$$

# Среднее значение

распределение по частотам  $M$  случайных величин:  $X$  — число книг, прочитанных за каникулы девочками, и  $Y$  — число книг, прочитанных за каникулы мальчиками, было задано такими таблицами:

|     |   |   |   |   |    |
|-----|---|---|---|---|----|
| $X$ | 3 | 4 | 5 | 8 | 12 |
| $M$ | 3 | 2 | 3 | 1 | 1  |

$$\Sigma M = n = 10$$

|     |   |   |   |   |   |
|-----|---|---|---|---|---|
| $Y$ | 3 | 4 | 5 | 6 | 7 |
| $M$ | 2 | 4 | 1 | 1 | 1 |

$$\Sigma M = n = 9$$

Тогда средние значения заданных случайных величин равны:

$$\bar{X} = \frac{3 \cdot 3 + 4 \cdot 2 + 5 \cdot 3 + 8 \cdot 1 + 12 \cdot 1}{10} = \frac{52}{10} = 5,2, \quad \bar{Y} = \frac{3 \cdot 2 + 4 \cdot 4 + 5 \cdot 1 + 6 \cdot 1 + 7 \cdot 1}{9} = \frac{40}{9} \approx 4,4.$$

Поскольку  $\bar{X} > \bar{Y}$ , то можно сказать, что за один и тот же промежуток времени девочки в классе читают книг больше, чем мальчики.

# Среднее значение

Среднее значение ( $\bar{X}$ ) случайной величины  $X$

*Средним значением случайной величины  $X$  называется среднее арифметическое всех ее значений.*

Если случайная величина  $X$  принимает  $n$  значений  $x_1, x_2, \dots, x_n$ , то

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}. \quad (**)$$

Если случайная величина  $X$  принимает значения  $x_1, x_2, \dots, x_k$  соответственно с частотами  $m_1, m_2, \dots, m_k$  (тогда  $\Sigma M = n$ ), то среднее арифметическое можно вычислить по формуле

$$\bar{X} = \frac{x_1 m_1 + x_2 m_2 + \dots + x_k m_k}{n}.$$

Пусть случайная величина  $X$  задана таблицей распределения по частотам  $M$ :

|     |   |   |   |   |
|-----|---|---|---|---|
| $X$ | 2 | 4 | 5 | 7 |
| $M$ | 3 | 1 | 2 | 2 |

$$\Sigma M = n = 8.$$

Тогда по формуле (\*\*)

$$\bar{X} = \frac{2 + 2 + 2 + 4 + 5 + 5 + 7 + 7}{8} = \frac{34}{8} = 4,25$$

или по другой формуле

$$\bar{X} = \frac{2 \cdot 3 + 4 \cdot 1 + 5 \cdot 2 + 7 \cdot 2}{8} = \frac{34}{8} = 4,25.$$

# Математическое ожидание

Пусть случайная величина  $X$  принимает значения  $x_1, x_2, \dots, x_k$  соответственно с вероятностями  $p_1, p_2, \dots, p_k$ , то есть имеет закон распределения:

|     |       |       |         |       |
|-----|-------|-------|---------|-------|
| $X$ | $x_1$ | $x_2$ | $\dots$ | $x_k$ |
| $P$ | $p_1$ | $p_2$ | $\dots$ | $p_k$ |

Сумма произведений всех значений случайной величины на соответствующие вероятности называется *математическим ожиданием величины  $X$* :

$$MX = x_1p_1 + x_2p_2 + \dots + x_kp_k. \quad (5)$$

*Математическое ожидание показывает, на какое среднее значение случайной величины  $X$  можно надеяться в результате эксперимента (при значительном количестве повторений эксперимента). С помощью математического ожидания можно сравнивать случайные величины, заданные законами распределения.*

# Математическое ожидание

Например, пусть количества очков, выбиваемых при одном выстреле каждым из двух ловких стрелков, имеют следующие законы распределения:

|     |     |     |     |
|-----|-----|-----|-----|
| $X$ | 8   | 9   | 10  |
| $P$ | 0,4 | 0,1 | 0,5 |

|     |     |     |     |
|-----|-----|-----|-----|
| $Y$ | 8   | 9   | 10  |
| $P$ | 0,1 | 0,6 | 0,3 |

Чтобы выяснить, какой из стрелков стреляет более метко, находят математическое ожидание для каждой случайной величины:

$$MX = 8 \cdot 0,4 + 9 \cdot 0,1 + 10 \cdot 0,5 = 9,1; MY = 8 \cdot 0,1 + 9 \cdot 0,6 + 10 \cdot 0,3 = 9,2.$$

Следовательно, среднее количество очков, выбиваемое при одном выстреле, у второго стрелка несколько больше, чем у первого. Это дает основание сделать вывод о том, что второй стрелок стреляет немного лучше, чем первый.

**при значительном количестве экспериментов среднее арифметическое всех значений случайной величины приближается к ее математическому ожиданию.**

# Центральные тенденции

Обратим внимание, что в пособиях по статистике моду, медиану и среднее значение объединяют одним термином — *меры центральной тенденции*, подчеркивая тем самым возможность охарактеризовать ряд выборки одним числом, к которому стремятся все ее значения.

Спасибо за внимание