

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В  
ЮРИДИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ.  
ЛЕКЦИЯ 4: «ОБРАБОТКА РЕЗУЛЬТАТОВ  
ЭКСПЕРИМЕНТА»**

# Статистическая обработка числовых данных (1)

Явления, происходящие в природе, в обществе, в человеке, очень сложны и разнообразны. Ученые изучают разные стороны этих явлений, причем каждая наука вырабатывает свои специфические методы исследования. Например, таким важным социальным явлением, как преступность, занимаются не только юристы, но и социологи, психологи, медики и иные специалисты. Есть тут серьезная работа и для математиков.

## Статистическая обработка числовых данных (2)

Их задача состоит в том, чтобы подвергнуть математической обработке огромный статистический материал: отчеты органов внутренних дел и другие документы, содержащие различные числовые данные. Цель этой работы — выделить наиболее существенные сведения об интересующем нас явлении.

Результаты обработки обычно представляют в виде таблиц, графиков, диаграмм и различных числовых характеристик, которые называют параметрами. Важнейшие из них — среднее арифметическое и дисперсия.

## Задача

УВД города Дрюково опубликовало сводку о количестве правонарушений, совершенных подростками за первые 20 дней сентября: 8, 6, 13, 4, 13, 13, 12, 9, 7, 6, 12, 14, 13, 12, 17, 6, 8, 12, 7, 12. По этим данным составлена следующая таблица:

$\tilde{x}_i$	4	6	7	8	9	12	13	14	17
$m_i$	1	3	2	2	1	5	4	1	1

Здесь  $m_i$  — число дней с одним и тем же количеством правонарушений,  $\tilde{x}_i$  — число правонарушений за день. Найдите среднее число правонарушений за день.

## Решение

Согласно приведенным данным, среднее число правонарушений за один день будет равно

$$\begin{aligned}\bar{x} &= \frac{1}{20} (4 \cdot 1 + 6 \cdot 3 + 7 \cdot 2 + 8 \cdot 2 + 9 \cdot 1 + 12 \cdot 5 + 13 \cdot 4 + 14 \cdot 1 + 17 \cdot 1) = \\ &= \frac{1}{20} \cdot 204 = 10,2.\end{aligned}$$

Мы использовали следующую формулу для подсчета среднего арифметического:

$$\bar{x} = \frac{1}{n} (\tilde{x}_1 m_1 + \tilde{x}_2 m_2 + \dots + \tilde{x}_k m_k).$$

## Замечание

Часто данные бывают представлены в иной форме. Заменим в предыдущей таблице вторую строку на новую, в которой вместо числа дней  $m_i$  поставим долю, которую это число составляет от числа всех дней. Эта доля называется **относительной частотой** или просто **частотой**.

$\tilde{x}_i$	4	6	7	8	9	12	13	14	17
$\tilde{p}_i$	1/20 = 0,05	3/20 = 0,15	2/20 = 0,10	2/20 = 0,10	1/20 = 0,05	5/20 = 0,25	4/20 = 0,20	1/20 = 0,05	1/20 = 0,05

Сумма чисел, стоящих во второй строке, равна единице. Это свойство следует из определения частоты.

# Новая формула для среднего арифметического

Среднее арифметическое  $\bar{x}$  равно сумме произведений чисел, взятых из первой строки таблицы, на их частоты. Обозначим частоты  $\tilde{p}_i$ :

$$\tilde{p}_1 = \frac{m_1}{n}, \quad \tilde{p}_2 = \frac{m_2}{n}, \quad \dots, \quad \tilde{p}_k = \frac{m_k}{n}.$$

В результате формула для среднего арифметического имеет вид

$$\bar{x} = \tilde{x}_1 \tilde{p}_1 + \tilde{x}_2 \tilde{p}_2 + \dots + \tilde{x}_k \tilde{p}_k.$$

## Пример

В городе Дрюково каждому пассажиру междугороднего автобуса вручают страховой полис на 5 000 рублей, взимая за это один рубль. Какова средняя прибыль страховой компании от продажи одного полиса, если страховые случаи происходят в среднем с одним пассажиром из 10 000? Учтите, что по обычаям города Дрюково страховка выплачивается только в случае гибели пассажира.



## Решение

Прибыль может принимать два значения: 1 рубль, если несчастного случая не произошло, и -4999 рублей при автокатастрофе (знак «минус» означает, что компания терпит убыток). Прибыль -4999 рублей появляется в одном случае из 10 000, следовательно, частота этого значения прибыли равна 0,0001. Частота другого значения (1 рубль) равна 0,9999. Получаем следующую таблицу:

<b>Прибыль</b>	1	-4999
<b>Частота</b>	0,9999	0,0001

Теперь подсчитаем среднее значение прибыли:

$$\bar{x} = 1 \cdot 0,9999 + (-4999) \cdot 0,0001 = 0,9999 - 0,4999 = 0,5 \text{ (руб.)}.$$

## Задача

За контрольную работу по математике 50 студентов первого курса получили следующие оценки: 4, 4, 2, 3, 5, 3, 5, 4, 3, 3, 4, 2, 4, 3, 5, 4, 3, 1, 3, 2, 2, 3, 4, 5, 4, 3, 3, 2, 4, 4, 3, 4, 3, 3, 4, 3, 2, 3, 3, 5, 3, 1, 3, 4, 5, 2, 4, 3, 3, 2. Найдите среднее арифметическое этих чисел.

## Решение

Расчетная таблица состоит из четырех строк и семи столбцов. В первой строке мы расположим значения оценок  $\tilde{x}_i$  в порядке их возрастания от 1 до 5. Во второй строке для каждой оценки укажем число ее повторений, то есть абсолютные частоты  $m_i$ . Сумма всех абсолютных частот равна количеству всех оценок, в нашем случае она равна 50. Запишем ее в последний столбец. В третьей строке таблицы будут записаны относительные частоты  $\tilde{p}_i$ . Так как у нас  $n=50$ , то общая формула для вычисления частот примет вид  $\tilde{p}_i = \frac{m_i}{50}$ .

## Продолжение решения

В последнем столбце третьей строки записывается сумма всех частот. Она всегда равна единице. Если получилось другое число, то в расчетах допущена ошибка. В последней строке записываются произведения оценок на их частоты — числа  $\tilde{x}_i \tilde{p}_i$ , сумма которых и равна среднему арифметическому  $\bar{x}=3,28$ . Она записывается в последнем столбце четвертой строки:

$\tilde{x}_i$	1	2	3	4	5	Сумма
$m_i$	2	8	20	14	6	50
$\tilde{p}_i$	0,04	0,16	0,40	0,28	0,12	1
$\tilde{x}_i \tilde{p}_i$	0,04	0,32	1,20	1,12	0,60	3,28

## Одна история

Двух студентов юридического факультета послали на практику, одного в город Дрюково, другого — в город Стуково. Практиканты узнали, что в это время года среднесуточная температура в этих городах равна нулю. Тот из них, кто поехал в Стуково, будучи человеком осторожным, взял с собой только теплые вещи. Другой, более легкомысленный, оделся по-летнему.

## Продолжение истории

Оказалось, что в течение всей практики в обоих городах температура была стабильной: в Дрюкове +2 днем и -2 ночью, в Стукове +15 днем и -15 ночью. В результате, несмотря на то, что среднесуточная температура действительно была нулевой, оба студента заболели, так как один постоянно перегревался, а другой — постоянно мерз.

## Комментарий

Только одно среднее арифметическое результатов измерений или наблюдений не всегда адекватно отражают ситуацию. Помимо средней величины, нужно знать еще и то, **как** заданные числа рассеяны около их среднего значения. Для этой цели вводятся **дисперсия** и **среднее квадратическое отклонение**.

# Определение

**Дисперсией** величин  $x_1, x_2, \dots, x_n$  называется число

$$D = \frac{1}{n} ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2).$$



## Пример

В следующей таблице указано время в минутах, затраченное на обследование каждого из десяти автомобилей. Символом  $x_i$  обозначено время, затраченное на обследование автомобиля с номером  $i$ . Найдите дисперсию величин  $x_i$ .

$i$	1	2	3	4	5	6	7	8	9	10
$x_i$	25	30	22	22	54	36	41	45	25	40

## Решение

Составим таблицу из трех столбцов: в первом столбце разместим числа  $x_i$ ; во втором — отклонения этих величин от их среднего значения; в третьем — квадраты этих отклонений. Второй и третий столбцы заполняются только после вычисления среднего времени обследования  $x=34$  (**МИН.**).

## Продолжение решения

$x_i$	25	30	22	22	54	36	41	45	25	40	$\Sigma$
$x_i - \bar{x}$	-9	-4	-12	-12	20	2	7	11	-9	6	<b>0</b>
$(x_i - \bar{x})^2$	81	16	144	144	400	4	49	121	81	36	<b>1076</b>

## Продолжение решения (2)

Последняя строка таблицы содержит суммы чисел по столбцам. В частности, в последней строке первого столбца записано общее время (340 минут) обследования всех автомобилей. Поделив ее на 10, мы и нашли среднее время обследования  $\bar{x}=34$  (мин.). В третьем столбце находятся квадраты отклонений, с помощью которых мы вычисляем дисперсию:

$$D = \frac{1}{10} \cdot 1076 = 107,6 \text{ (мин.}^2\text{)}.$$

# Частоты и дисперсия

Если известны частоты  $\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_k$ , то для вычисления дисперсии можно использовать формулу

$$D = \tilde{p}_1(\tilde{x}_1 - \bar{x})^2 + \tilde{p}_2(\tilde{x}_2 - \bar{x})^2 + \dots + \tilde{p}_k(\tilde{x}_k - \bar{x})^2,$$

где, как и выше,  $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k$  суть различные среди заданных чисел  $x_1, x_2, \dots, x_n$ .

# Определение

**Средним квадратическим отклонением** величин  $x_1, x_2, \dots, x_n$  от их среднего значения  $\bar{x}$  называется величина

$$S = \sqrt{\overline{D}}.$$

## Пример

В предыдущем примере

$$S = \sqrt{107,6} = 10,373 \dots \approx 10,4 \text{ (мин.)}.$$

# Случайная величина

Введем весьма важное понятие **случайной величины**. В рассмотренных ранее примерах случайными величинами являлись время обследования автомобиля, число дорожных происшествий, прибыль страховой компании и др. Математики в таких случаях говорят, что время обследования автомобиля, например, есть переменная (случайная) величина  $X$ , принимающая значения  $x_1, x_2, \dots, x_n$ .



# Проблема

Теперь допустим, что нужно обследовать **все** автомобили города Дрюково. Но число автомобилей так велико, что описать все значения величины  $X$  ( $X$  – время обследования автомобиля) практически невозможно. Однако мы можем, не проводя самого обследования, предсказать его результаты, используя результаты уже решенной задачи.

## Переход к решению

Переделаем немного ранее составленную таблицу:

$\tilde{x}_i$	22	25	30	36	40	41	45	54
$\tilde{p}_i$	0,2	0,2	0,1	0,1	0,1	0,1	0,1	0,1

Напомним, что среднее время обследования  $x=34$  (мин.), а среднеквадратическое отклонение  $S \approx 10,4$  (мин.).

## Идея решения

Результаты, полученные при обследовании десяти автомобилей, можно распространить на все автомобили города Дрюково. С определенной степенью достоверности можно считать, что в среднем на обследование каждого автомобиля в городе затрачивается 34 минуты; что время обследования примерно 60% автомобилей заключено в промежутке от 23,6 до 44.4 минут, и так далее. Такие приблизительные расчеты могут весьма пригодиться, например, тому, кто хочет открыть новую станцию техобслуживания, поскольку характеризует рынок в этой сфере бизнеса.

# Терминология

При обработке статистического материала используется специальная терминология. Совокупность всех рассматриваемых объектов называют **генеральной совокупностью**, а часть объектов, каким-либо способом выбранных для обследования, называют **выборкой**. В нашем примере с автомобилями генеральную совокупность образуют все автомобили города Дрюково, а выборку — те 10 автомобилей, которые рассматривались в примере.

## Замечание

Очень важно сделать выборку правильно. От этого зависит точность и достоверность прогноза. В математической статистике изучаются методы, позволяющие сделать выборку так, чтобы полученная с ее помощью информация давала достаточно полное и адекватное представление об интересующем нас признаке изучаемой генеральной совокупности. Тогда среднее арифметическое  $\bar{x}$  и дисперсия  $D$  будут близки к гипотетическим величинам — среднему арифметическому и дисперсии, которые могли бы быть получены при обработке **всей** генеральной совокупности.

# Статистические характеристики

Кроме среднего значения и среднеквадратического отклонения основными характеристиками являются: **весь диапазон значений** величины, **интервал ее наиболее вероятных значений** и **частоты больших и малых значений** исследуемой величины.

## Замечание 1

Чаще всего в качестве диапазона наиболее вероятных значений величины берут интервал  $(\bar{x} - S; \bar{x} + S)$ . Однако, если его частота оказывается меньше чем  $1/2$ , то его каким-нибудь образом расширяю; если же частота больше  $1/2$ , то интервал можно сузить.

## Замечание 2

Мы сами определяем, что такое **малые и большие значения** случайной величины. Если такие понятия вообще требуются для описания конкретного материала, то их вводят, сообразуясь со спецификой задачи.



## Пример

Комиссия изучала состояние борьбы с преступностью в регионе. Случайным образом выбрано 20 районов (вообще-то их больше двадцати) и по ним были представлены данные о числе раскрытых убийств: 11, 6, 12, 1, 3, 1, 6, 20, 10, 1, 1, 3, 3, 1, 23, 11, 3, 6, 10, 3. Составьте по этим данным прогноз для всего региона.

## Решение

Совокупность всех районов данного региона — это генеральная совокупность. Изучается величина  $X$ , которая представляет собой число раскрытых в районе убийств. Для изучения используется выборка из 20 объектов генеральной совокупности (районов региона). Для них получены значения величины  $X$  — выборочные данные. Мы обработаем выборочные данные по приведенной выше схеме и на этом основании составим прогноз для всего региона в целом. Иными словами, по выборочным данным мы опишем свойство всей генеральной совокупности, то есть составим представление о борьбе с преступностью во всех районах региона.

## Продолжение решения

$\tilde{x}_i$	$m_i$	$\tilde{p}_i$	$\tilde{x}_i \tilde{p}_i$	$\tilde{x}_i - \bar{x}$	$(\tilde{x}_i - \bar{x})^2$	$(\tilde{x}_i - \bar{x})^2 \tilde{p}_i$
1	5	0,25	0,25	-5,75	33,0625	8,265625
3	5	0,25	0,75	-3,75	14,0625	3,515625
6	3	0,15	0,90	-0,75	0,5625	0,084375
10	2	0,10	1,00	3,25	10,5625	1,056250
11	2	0,10	1,10	4,25	18,0625	1,806250
12	1	0,05	0,60	5,25	27,5625	1,378125
20	1	0,05	1,00	13,25	175,5625	8,778125
23	1	0,05	1,15	16,25	264,0625	13,203125
$\Sigma$	<b>20</b>	<b>1</b>	<b>6,75</b>	<b>–</b>	<b>–</b>	<b>38,09</b>

Найдем дисперсию и среднеквадратическое отклонение:

$$D = 38,0875, \quad S = \sqrt{38,0875} = 6,1715071 \dots \approx 6,17.$$

## Выводы (1)

- 1) Число убийств, раскрытых в одном районе, принимает значения от 1 до 23.
- 2) Наиболее вероятными значениями являются числа 1 и 3, так как они имеют наибольшую частоту 0,25.
- 3) Среднее значение числа убийств (среднее арифметическое) равно 6,75.
- 4) Среднеквадратическое отклонение составляет 6,17.

## Выводы (2)

5) В интервал  $(\bar{x} - S; \bar{x} + S) = (0,58; 12,92)$  попадают 7 значений случайной величины: 1, 3, 5, 6, 10, 11, 12; сумма из частот равна  $0,25 + 0,25 + 0,15 + 0,10 + 0,10 + 0,05 = 0,90$ ; этот интервал будет интервалом наивероятнейших значений величины  $X$ , так как он содержит 90% значений величины.

6) Частота малых значений числа  $X$ , то есть таких, которые меньше 0,58, равна нулю.

7) Частота больших значений (тех, которые больше 12,92) равна  $0,05 + 0,05 = 0,10$ ; иными словами, число случаев, когда раскрытых убийств больше 12, составляет всего 10% от их общего числа.

## Замечание

До сих пор мы предполагали, что все выкладки выполняются «вручную» или с помощью калькулятора. Для решения подобных задач можно использовать и компьютер.