# 26. Введение в математическую статистику

## 26.1. Основные определения математической статистики

Значительная часть математической статистики связана с необходимостью описать большую совокупность объектов. Её называют ***генеральной совокупностью***. Если *генеральная совокупность* слишком многочисленна, или её объекты труднодоступны, или имеются другие причины, не позволяющие изучить все объекты, прибегают к изучению какой-то части объектов. Эта выбранная для полного изучения часть называется ***выборкой***. Необходимо, чтобы выборка наилучшим образом представляла генеральную совокупность, т.е. была ***репрезентативной*** (представительной). Если генеральная совокупность мала или совсем неизвестна, не удаётся предложить ничего лучшего, чем чисто случайный выбор.

***Определение*** 26.1. *Количество наблюдений $n$ называется* ***объёмом выборки****.*

***Определение*** 26.2. *Наблюдаемые значения $x_i$ называют вариантами, а их последовательность, записанную в возрастающем порядке —* ***вариационным рядом****. Числа наблюдений $m_1, m_2, \ldots, m_k$ называют частотами.*

*Разность $max(x_i) - min(x_i)$ называется* ***размахом вариационного ряда****.*

Статистическим распределением выборки называют перечень вариант и соответствующих им частот — табл. 26.1.

Таблица 26.1

| Статистическое распределение | | | |
|---|---|---|---|
| варианты $x_i$ | $x_1$ | ... | $x_k$ |
| частоты $m_i$ | $m_1$ | ... | $m_k$ |

***Определение*** 26.3. ***Эмпирической*** *(статистической)* ***функцией распределения*** *случайной величины $\xi$ называется функция $F^*(x)$, которая при каждом $x$ равна относительной частоте события $\xi < x$, т.е. отношению $m_x$ — числа наблюдений меньших $x$ к объёму выборки $n$:*

$$F^*(x) = P^*(\xi < x) = \frac{m_x}{n} \ .$$

***Определение*** 26.4. ***Медиана*** *— значение варианты, для которого количество элементов находящихся слева и справа, одинаково.*

Т.е., значение $M_e$, при котором $F^*(M_e) = 0{,}5$.

Для простой статистической совокупности медиана вычисляется следующим образом. Исследуемая выборка $\{x_i\}$ сортируется в порядке не убывания значений элементов. Далее, если объём выборки нечётное число, то $M_e = x_{(n+1)/2}$, иначе $M_e = \big(x_{n/2} + x_{n/2+1}\big)/2$.

Например, для вариационного ряда $\{1, 2, 5, 6, 8, 9, 15\}$ медиана равна четвёртому элементу $M_e = 6$, а для вариационного ряда $\{1, 2, 5, 6, 7, 9, 15, 16\}$ медиана равна полусумме четвёртого и пятого элементов $M_e = (6 + 7)/2 = 6{,}5$.

***Определение*** 26.5. ***Модой*** $M_0$ *называется варианта, которая имеет наибольшую частоту по сравнению с другими частотами.*

*В дискретно-вариационном ряду мода — это та варианта, которой соответствует наибольшая частота.*

Для простой статистической совокупности мода вычисляется простым подсчётом. Например, для вариационного ряда $\{1, 2, 3, 4, 4, 4, 4, 5, 6, 6, 6, 7\}$, $M_O = 4$, т.к. значение 4 встречается чаще других.

Статистические распределения, которые имеют несколько наиболее часто встречающихся значений, называются мультимодальными или полимодальными.

Например, для вариационного ряда: $\{1, 2, 3, 3, 4, 5, 6, 6, 7, 8, 8\}$, модами будут три значения $M_O = \{3, 6, 8\}$.

Простейшей характеристикой распределения является ***выборочное среднее***, которое для простой статистической совокупности вычисляется по формуле:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{26.1}$$

Если данные сгруппированы, то:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} m_i\, x_i. \tag{26.2}$$

Для характеристики разброса значений случайной величины относительно её среднего значения используется ***выборочная дисперсия***

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \overline{(x - \overline{x})^2} \tag{26.3}$$

для простой совокупности и

$$S^2 = \frac{1}{n}\sum_{i=1}^{k} m_i(x_i - \overline{x})^2 \qquad (26.4)$$

для сгруппированного распределения.

$$S = \sqrt{S^2} \qquad (26.5)$$

называется ***выборочным средним квадратическим отклонением*** (СКО).

На практике вместо формулы (26.3) бывает удобнее применять другую:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2 = \overline{x^2} - \overline{x}^2 \qquad (26.6)$$

для простой совокупности и

$$S^2 = \frac{1}{n}\sum_{i=1}^{k} m_i x_i^2 - (\overline{x})^2 \qquad (26.7)$$

для сгруппированного распределения.

При малых объёмах выборки $n$ для оценки дисперсии $\sigma^2$ используют **исправленную** выборочную дисперсию $S^{*^2}$:

$$S^{*^2} = \frac{n}{n-1} S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2. \qquad (26.8)$$

Оценка $S^{*^2}$ является ***несмещённой***, состоятельной оценкой дисперсии $\sigma^2$.

Формула (26.8) позволяет вычислять $S^{*^2}$ для простой совокупности. Для сгруппированных данных используют аналогичную формулу (26.9):

$$S^{*^2} = \frac{1}{n-1}\sum_{i=1}^{k} m_i(x_i - \overline{x})^2. \qquad (26.9)$$

***Замечание*** 26.1. *Исправленное выборочное СКО $S^*$ является несмещённой оценкой СКО $S$.*

***Пример*** 26.1. *Выборка задана в виде распределения частот:*

| $x_i$ | *3* | *5* | *8* | *10* | *11* |
|---|---|---|---|---|---|
| $m_i$ | *20* | *25* | *30* | *15* | *10* |

*Найти моду, медиану и эмпирическую функцию распределения.*

◀ Здесь объём выборки

$$n = 20 + 25 + 30 + 15 + 10 = 100.$$

Мода и медиана равна 8. Найдем относительные частоты:

$$p_1^* = 20/100 = 1/5, \quad p_2^* = 25/100 = 1/4, \quad , p_3^* = 30/100 = 3/10,$$

$$p_4^* = 15/100 = 3/20, \qquad p_5^* = 10/100 = 1/10.$$

Тогда распределение относительных частот примет вид:

| $x_i$ | 3 | 5 | 8 | 10 | 11 |
|---|---|---|---|---|---|
| $p_i^*$ | 0,2 | 0,25 | 0,3 | 0,15 | 0,1 |

Из этой таблицы нетрудно убедиться, что

$$\sum_{i=1}^{5} p_i^* = 1.$$

Получаем эмпирическую функцию распределения

$$F^*(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 3, \\ 0{,}2 & \text{при} \quad x \in (3, 5], \\ 0{,}45 & \text{при} \quad x \in (5, 8], \\ 0{,}75 & \text{при} \quad x \in (8, 10], \\ 0{,}9 & \text{при} \quad x \in (10, 11], \\ 1 & \text{при} \quad x > 11. \end{cases}$$

▶

***Пример*** 26.2. *Выборка задана в виде распределения частот:*

| $x_i$ | *3* | *5* | *8* | *10* | *11* | *15* |
|---|---|---|---|---|---|---|
| $m_i$ | *5* | *10* | *30* | *25* | *22* | *8* |

◀ Здесь объём выборки

$$n = 5 + 10 + 30 + 25 + 22 + 8 = 100.$$

Мода равна 8, а медиана равна 10. ►

***Пример*** 26.3. *Из генеральной совокупности извлечена выборка объёма* $n = 80$*:*

| $x_i$ | *0,9* | *1* | *1,2* | *1,4* | *1,5* |
|---|---|---|---|---|---|
| $m_i$ | *10* | *25* | *20* | *15* | *10* |

*Найти несмещённую оценку генерального среднего, выборочную дисперсию и выборочное среднее квадратическое отклонение.*

◄ Несмещённой оценкой генерального среднего является выборочное среднее. Тогда по формуле (26.2) найдем:

$$\overline{x} = \frac{1}{80}(10 \cdot 0{,}9 + 25 \cdot 1 + 20 \cdot 1{,}2 + 15 \cdot 1{,}4 + 10 \cdot 1{,}5) = 1{,}175.$$

Для нахождения выборочной дисперсии воспользуемся формулой (26.7):

$$S^2 = \frac{1}{80}(10 \cdot 0{,}9^2 + 25 \cdot 1^2 + 20 \cdot 1{,}2^2 + 15 \cdot 1{,}4^2 + 10 \cdot 1{,}5^2) - (1{,}175)^2 \approx$$

$$\approx 1{,}4225 - 1{,}3806 \approx 0{,}042.$$

Заметим, что отличная от нуля дисперсия является всегда положительной величиной.

Выборочное среднее квадратическое отклонение
$S = \sqrt{0{,}042} \approx 0{,}205.$ ►

**Ответ:** $\boxed{\overline{x} = 1{,}175;\ S^2 =\approx 0{,}042;\ S = \sqrt{0{,}042} \approx 0{,}205.}$

***Пример*** 26.4. *По выборке объёма* $n = 50$ *найдена смещённая оценка* $S^2 = 9{,}8$ *генеральной дисперсии. Найти несмещённую оценку дисперсии генеральной совокупности.*

◄ Согласно (26.8), исправленная выборочная дисперсия, является в то же время несмещённой оценкой

$$S^{*2} = \frac{n}{n-1} \cdot S^2 = \frac{50}{49} \cdot 9{,}8 = 10.$$ ►

**Ответ:** $\boxed{S^{*2} = 10.}$