ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

Пример 1. По выборке 5, 2, 2, 1, 6, 3, 1, 2, 3, 5

1) записать вариационный ряд;

2) составить статистический ряд;

3) записать эмпирическую функцию распределения и построить ее график;

4) построить полигон частот.

Решение. 1) Вариационный ряд: 1, 1, 2, 2, 2, 3, 3, 5, 5, 6. Размах выборки $\omega = 6 - 1 = 5$.

2) Подсчитав частоты различных значений в выборке, составим статистический ряд:

x_i	1	2	3	5	6
n_i	2	3	2	2	1

Объём выборки равен сумме частот наблюдаемых значений и равен $n = 2 + 3 + 2 + 2 + 1 = 10$.

3) Для построения эмпирической функции распределения удобно записать статистический ряд, заменив в таблице частоты на относительные частоты:

x_i	1	2	3	5	6
$\frac{n_i}{n}$	0,2	0,3	0,2	0,2	0,1

Наименьшее выборочное значение 1, поэтому эмпирическая функция распределения равна нулю для всех $x \le 1$. Дальше её значение изменяется каждый раз при переходе x через значения x_i, увеличиваясь на величину относительной частоты $\frac{n_i}{n}$. Наибольшее выборочное значение 6, поэтому эмпирическая функция распределения равна 1 при всех $x > 6$. Итак,

$$F^*(x) = \begin{cases} 0, & x \le 1, \\ 0{,}2, & 1 < x \le 2, \\ 0{,}5, & 2 < x \le 3, \\ 0{,}7, & 3 < x \le 5, \\ 0{,}9, & 5 < x \le 6, \\ 1, & x > 6. \end{cases}$$

Ее график изображен на рис. 1.

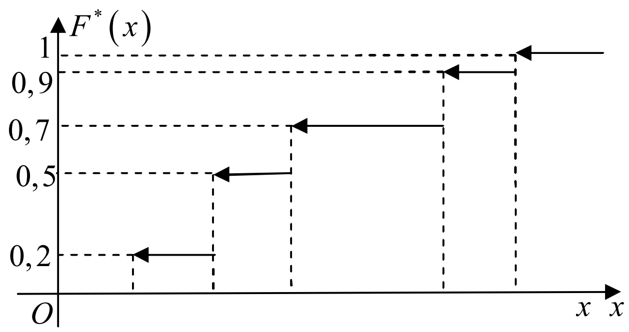

Рис. 1. График эмпирической функции распределения

Полигон частот изображен на рис. 2.

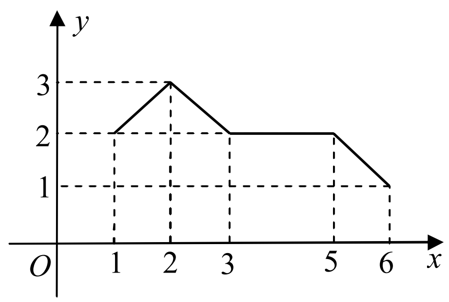

Рис. 2. Полигон частот

Пример 2. Построить гистограмму относительных частот и эмпирическую функцию распределения по данному интервальному ряду

Интервалы наблюдаемых значений СВ	5–10	10–15	15–20	20–25	25–30	30–35
Частоты n_i	6	7	17	36	24	10

Решение. Объём выборки $n = 6 + 7 + 17 + 36 + 24 + 10 = 100$. Длина каждого интервала h =5. Для построения эмпирической функции распределения найдём середины интервалов x_i^* и относительные частоты $\frac{n_i}{n}$; для построения гистограммы относительных частот найдём для каждого интервала значение $\frac{n_i}{nh}$. Дополним этими сведениями статистический ряд:

Интервалы наблюдаемых значений СВ	5–10	10–15	15–20	20–25	25–30	30–35
Частоты n_i	6	7	17	36	24	10
$\frac{n_i}{n}$	0,06	0,07	0,17	0,36	0,24	0,1
$\frac{n_i}{nh}$	0,012	0,014	0,034	0,072	0,048	0,02
Середины интервалов x_i^*	7,5	12,5	17,5	22,5	27,5	32,5

Гистограмма относительных частот и эмпирическая функция распределения представлены на рис. 3 и рис. 4 соответственно.

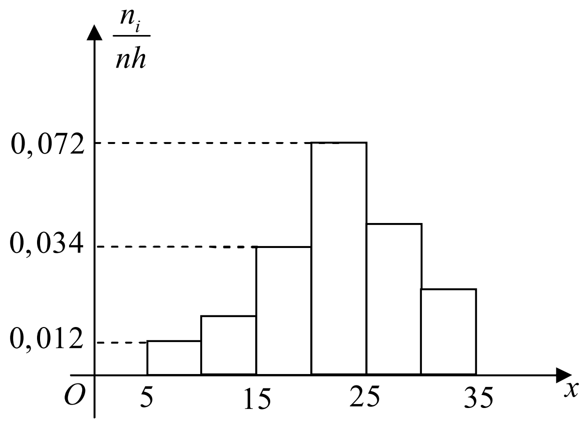

Рис. 3. Гистограмма частот

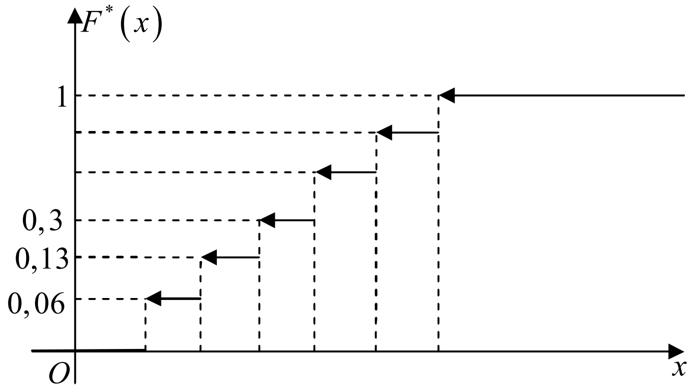

Рис. 4. Эмпирическая функция распределения

Пример 3. По выборке 5, 2, 2, 1, 6, 3, 1, 2, 3, 5 найти выборочное среднее, выборочную дисперсию, несмещённую оценку дисперсии.

Решение. Составим статистический ряд:

x_i	1	2	3	5	6
n_i	2	3	2	2	1

Найдем выборочное среднее

$$\bar{x} = \frac{1}{10}(1 \cdot 2 + 2 \cdot 3 + 3 \cdot 2 + 5 \cdot 2 + 6 \cdot 1) = 3$$

и выборочную дисперсию:

$$D_{в} = \frac{1}{10}(1^2 \cdot 2 + 2^2 \cdot 3 + 3^2 \cdot 2 + 5^2 \cdot 2 + 6^2 \cdot 1) - 3^2 = 2{,}8.$$

Тогда несмещённая оценка дисперсии равна:

$$s^2 = \frac{10}{9} \cdot 2{,}8 = 3{,}11;$$

несмещённая оценка σ (среднеквадратичного отклонения) $s = 1{,}7635$.

Пример 4. Найти выборочное среднее, выборочную дисперсию и несмещённую оценку дисперсии по данному интервальному статистическому ряду:

Интервалы наблюдаемых значений СВ	5–10	10–15	15–20	20–25	25–30	30–35
Частоты n_i	6	7	17	36	24	10

Решение. Для расчёта воспользуемся формулами числовых характеристик выборки в случае группированного статистического ряда. Итак,

$$\bar{x} = \frac{1}{100}(7{,}5 \cdot 6 + 12{,}5 \cdot 7 + 17{,}5 \cdot 17 + 22{,}5 \cdot 36 + 27{,}5 \cdot 24 + 32{,}5 \cdot 10) = 22{,}25;$$

$$D_{в} = \frac{1}{100}\Big((7{,}5 - 22{,}25)^2 \cdot 6 + (12{,}5 - 22{,}25)^2 \cdot 7 + (17{,}5 - 22{,}25)^2 \cdot 17 +$$

$$+ (22{,}5 - 22{,}25)^2 \cdot 36 + (27{,}5 - 22{,}25)^2 \cdot 24 + (32{,}5 - 22{,}25)^2 \cdot 10\Big) =$$

$$= 40{,}6875;$$

несмещённая оценка дисперсии равна $s^2 = \dfrac{100}{99} \cdot 40{,}6875 = 41{,}1.$

Пример 5. По результатам наблюдений установить вид эмпирической зависимости y от x и методом наименьших квадратов найти $y = f(x)$:

x_i	1	2	3	4
y_i	2	4	5	7

Решение. Наносим точки $(x_i; y_i)$ на плоскость Oxy (рис. 5) и по их расположению замечаем, что они группируются вокруг прямой линии. Поэтому сглаживать экспериментальные данные будем линейной функцией: $y = ax + b$.

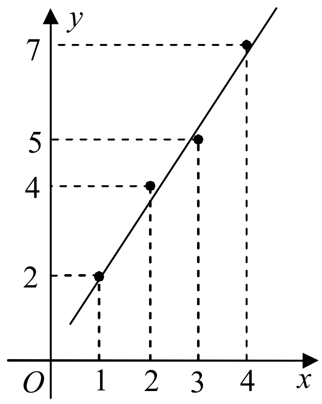

Рис. 5. Экспериментальные данные

Согласно методу наименьших квадратов, искомые коэффициенты a и b находим из системы

$$\begin{cases} nb + a\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \\ b\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases}$$

Предварительно необходимо найти указанные суммы. Для удобства все вычисления расположим в таблице:

Расчетная таблица

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	2	2	1	4
2	4	8	4	16
3	5	15	9	25
4	7	28	16	49
10	18	53	30	94

Подставляя найденные суммы в систему, имеем

$$\begin{cases} 4b + 10a = 18; \\ 10b + 30a = 53. \end{cases}$$

Решая систему, находим a = 1,6; b = 0,5 и уравнение сглаживающей линии y = 1,6x + 0,5, график которой изображён на рис. 5.