Часть третья

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Глава девятая

ВЫБОРОЧНЫЙ МЕТОД

§ 1. Статистическое распределение выборки

Пусть для изучения количественного (дискретного или непрерывного) признака X из генеральной совокупности извлечена выборка $x_1, x_2, \ldots, x_k$ объема n. Наблюдавшиеся значения x_i признака X называют *вариантами*, а последовательность вариант, записанных в возрастающем порядке,—*вариационным рядом*.

Статистическим распределением выборки называют перечень вариант x_i вариационного ряда и соответствующих им частот n_i (сумма всех частот равна объему выборки n) или относительных частот w_i (сумма всех относительных частот равна единице).

Статистическое распределение выборки можно задать также в виде последовательности интервалов и соответствующих им частот (в качестве частоты интервала принимают сумму частот вариант, попавших в этот интервал).

439. Выборка задана в виде распределения частот:

x_i	2	5	7
n_i	1	3	6

Найти распределение относительных частот.

Решение. Найдем объем выборки: $n=1+3+6=10$. Найдем относительные частоты:

$$w_1=1/10=0{,}1; \quad w_2=3/10=0{,}3; \quad w_3=6/10=0{,}6.$$

Напишем искомое распределение относительных частот:

x_i	2	5	7
w_i	0,1	0,3	0,6

Контроль: $0{,}1+0{,}3+0{,}6=1$.

440. Выборка задана в виде распределения частот:

x_i	4	7	8	12
n_i	5	2	3	10

Найти распределение относительных частот.

§ 2. Эмпирическая функция распределения

Эмпирической функцией распределения (*функцией распределения выборки*) называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$:

$$F^*(x) = n_x/n,$$

где n_x — число вариант, меньших x; n — объем выборки.

Эмпирическая функция обладает следующими свойствами.

Свойство 1. *Значения эмпирической функции принадлежат отрезку* $[0; 1]$.

Свойство 2. $F^*(x)$ — *неубывающая функция*.

Свойство 3. *Если x_1 — наименьшая варианта, а x_k — наибольшая*, то $F^*(x) = 0$ при $x \leqslant x_1$ и $F^*(x) = 1$ при $x > x_k$.

441. Найти эмпирическую функцию по данному распределению выборки:

x_i	1	4	6
n_i	10	15	25

Решение. Найдем объем выборки: $n = 10 + 15 + 25 = 50$.

Наименьшая варианта равна единице, поэтому $F^*(x) = 0$ при $x \leqslant 1$.

Значение $X < 4$, а именно $x_1 = 1$, наблюдалось 10 раз, следовательно, $F^*(x) = 10/50 = 0{,}2$ при $1 < x \leqslant 4$.

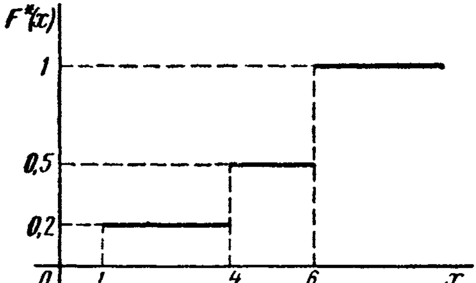

Рис. 11

Значения $x < 6$, а именно: $x_1 = 1$ и $x_2 = 4$, наблюдались $10 + 15 = 25$ раз; следовательно, $F^*(x) = 25/50 = 0{,}5$ при $4 < x \leqslant 6$.

Так как $x = 6$ — наибольшая варианта, то $F^*(x) = 1$ при $x > 6$.

Напишем искомую эмпирическую функцию:

$$F^*(x) = \begin{cases} 0 & \text{при} \quad x \leqslant 1, \\ 0{,}2 & \text{при} \quad 1 < x \leqslant 4, \\ 0{,}5 & \text{при} \quad 4 < x \leqslant 6 \\ 1 & \text{при} \quad x > 6. \end{cases}$$

График этой функции изображен на рис. 11.

442. Найти эмпирическую функцию по данному распределению выборки:

а)

x_i	2	5	7	8
n_i	1	3	2	4

б)

x_i	4	7	8
n_i	5	2	3

§ 3. Полигон и гистограмма

А. Дискретное распределение признака X. *Полигоном частот* называют ломаную, отрезки которой соединяют точки (x_1, n_1), (x_2, n_2), ..., (x_k, n_k), где x_i — варианты выборки и n_i — соответствующие им частоты.

Полигоном относительных частот называют ломаную, отрезки которой соединяют точки $(x_1;\ w_1)$, $(x_2;\ w_2)$, ..., $(x_k;\ w_k)$, где x_i — варианты выборки и w_i — соответствующие им относительные частоты.

Б. Непрерывное распределение признака X. При непрерывном распределении признака весь интервал, в котором заключены все наблюдаемые значения признака, разбивают на ряд частичных интервалов длины h и находят n_i — сумму частот вариант, попавших в i-й интервал. *Гистограммой частот* называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины h, а высоты равны отношению n_i/h (плотность частоты). Площадь частичного i-го прямоугольника равна $h\,(n_i/h)=n_i$ — сумме частот вариант, попавших в i-й интервал. Площадь гистограммы частот равна сумме всех частот, т. е. объему выборки n.

Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины h, а высоты равны отношению w_i/h (плотность относительной частоты). Площадь частичного i-го прямоугольника равна $h\,(w_i/h)=w_i$ — относительной частоте вариант, попавших в i-й интервал. Площадь гистограммы относительных частот равна сумме всех относительных частот, т. е. единице.

443. Построить полигон частот по данному распределению выборки:

x_i	1	4	5	7
n_i	20	10	14	6

Р е ш е н и е. Отложим на оси абсцисс варианты x_i, а на оси ординат — соответствующие им частоты n_i; соединив точки $(x_i,\ n_i)$ отрезками прямых, получим искомый полигон частот (рис. 12).

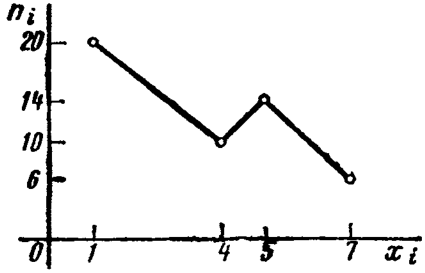

Рис. 12

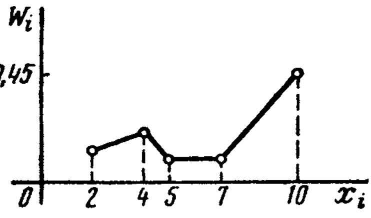

Рис. 13

444. Построить полигон частот по данному распределению выборки:

а)

x_i	2	3	5	6
n_i	10	15	5	20

б)

x_i	15	20	25	30	35
n_i	10	15	30	20	25

445. Построить полигон относительных частот по данному распределению выборки:

а)

x_i	2	4	5	7	10
w_i	0,15	0,2	0,1	0,1	0,45

б)

x_i	1	4	5	8	9
w_i	0,15	0,25	0,3	0,2	0,1

в) x_i 20 40 65 80

w_i 0,1 0,2 0,3 0,4

Решение. а) Отложим на оси абсцисс варианты x_i, а на оси ординат — соответствующие относительные частоты w_i. Соединив точки (x_i, w_i) отрезками прямых, получим искомый полигон относительных частот (рис. 13).

446. Построить гистограмму частот по данному распределению выборки объема $n = 100$:

Номер интервала i	Частичный интервал $x_i - x_{i+1}$	Сумма частот вариант интервала n_i	Плотность частоты n_i/h
1	1—5	10	2,5
2	5—9	20	5
3	9—13	50	12,5
4	13—17	12	3
5	17—21	8	2

Решение. Построим на оси абсцисс заданные интервалы длины $h = 4$. Проведем над этими интервалами отрезки, параллельные оси абсцисс и находящиеся от нее на расстояниях, равных соответствующим плотностям частоты n_i/h. Например, над интервалом (1, 5) построим отрезок, параллельный оси абсцисс, на расстоянии $n_i/h = 10/4 = 2,5$; аналогично строят остальные отрезки.

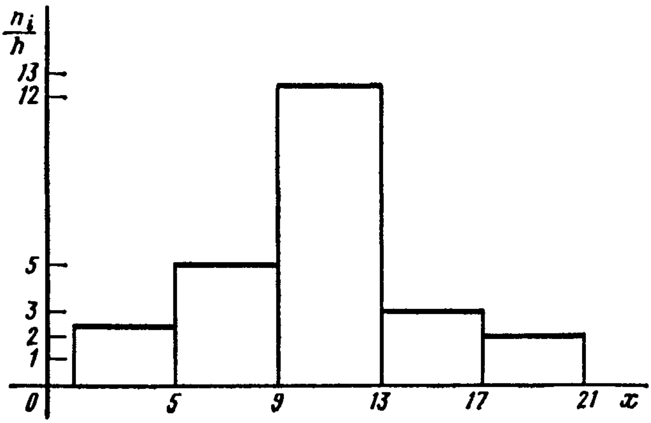

Рис. 14

Искомая гистограмма частот изображена на рис. 14.

447. Построить гистограмму частот по данному распределению выборки:

а)

Номер интервала i	Частичный интервал $x_i - x_{i+1}$	Сумма частот вариант интервала n_i	Плотность частоты n_i/h
1	2—7	5	
2	7—12	10	
3	12—17	25	
4	17—22	6	
5	22—27	4	

б)

Номер интервала i	Частичный интервал $x_i - x_{i+1}$	Сумма частот вариант интервала n_i	Плотность частоты n_i/h
1	3—5	4	
2	5—7	6	
3	7—9	20	
4	9—11	40	
5	11—13	20	
6	13—15	4	
7	15—17	6	

Указание. Найти предварительно плотность частоты n_i/h для каждого интервала и заполнить последний столбец таблицы.

448. Построить гистограмму относительных частот по данному распределению выборки:

Номер интервала i	Частичный интервал $x_i - x_{i+1}$	Сумма частот вариант частичного интервала n_i
1	0—2	20
2	2—4	30
3	4—6	50
		$n = \sum n_i = 100$

Решение. Найдем относительные частоты:

$$w_1 = 20/100 = 0{,}2,\quad w_2 = 30/100 = 0{,}3,\quad w_3 = 50/100 = 0{,}5.$$

Найдем плотности относительных частот, учитывая, что длина интервала $h=2$:

$$w_1/h=0{,}2/2=0{,}1,\ w_2/h=0{,}3/2=0{,}15,\ w_3/h=0{,}5/2=0{,}25.$$

Построим на оси абсцисс данные частичные интервалы. Проведем над этими интервалами отрезки, параллельные оси абсцисс и находящиеся от нее на расстояниях, равных соответствующим плотностям относительной частоты. Например, над интервалом (0, 2) проведем отрезок, параллельный оси абсцисс и находящийся от нее на расстоянии, равном 0,1; аналогично строят остальные отрезки.

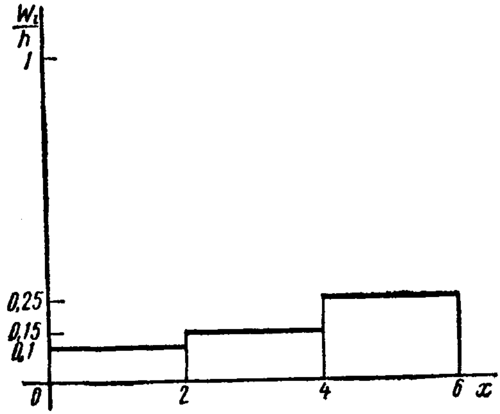

Рис. 15

Искомая гистограмма относительных частот изображена на рис. 15.

449. Построить гистограмму относительных частот по данному распределению выборки:

а)

Номер интервала i	Частичный интервал $x_i - x_{i+1}$	Сумма частот вариант частичного интервала n_i
1	10—15	2
2	15—20	4
3	20—25	8
4	25—30	4
5	30—35	2
		$n=\sum n_i=20$

б)

Номер интервала i	Частичный интервал $x_i - x_{i+1}$	Сумма частот вариант частичного интервала n_i
1	2—5	6
2	5—8	10
3	8—11	4
4	11—14	5
		$n=\sum n_i=25$

Указание. Найти сначала относительные частоты, соответствующие плотности относительной частоты для каждого интервала.

Глава десятая

СТАТИСТИЧЕСКИЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

§ 1. Точечные оценки

Статистической оценкой Θ^* неизвестного параметра Θ теоретического распределения называют функцию $f(X_1, X_2, \ldots, X_n)$ от наблюдаемых случайных величин $X_1, X_2, \ldots, X_n$.

Точечной называют статистическую оценку, которая определяется одним числом $\Theta^*=f(x_1, x_2, \ldots, x_n)$, где $x_1, x_2, \ldots, x_n$ — результаты n наблюдений над количественным признаком X (выборка).

Несмещенной называют точечную оценку, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки.

Смещенной называют точечную оценку, математическое ожидание которой не равно оцениваемому параметру.

Несмещенной оценкой генеральной средней (математического ожидания) служит выборочная средняя

$$\bar{x}_в=\left(\sum_{i=1}^{k} n_i x_i\right)\Big/n,$$

где x_i — варианта выборки, n_i — частота варианты x_i, $n=\sum_{i=1}^{k} n_i$ — объем выборки.

Замечание 1. Если первоначальные варианты x_i — большие числа, то для упрощения расчета целесообразно вычесть из каждой варианты одно и то же число C, т. е. перейти к условным вариантам $u_i=x_i-C$ (в качестве C выгодно принять число, близкое к выборочной средней; поскольку выборочная средняя неизвестна, число C выбирают «на глаз»). Тогда

$$\bar{x}_в=C+\left(\sum n_i u_i\right)/n.$$