116 117 107 105 119 107 119 111 112 129 113 106 104 106 98 105 106 139 108 109 93 107 117 107 118 99 108 108 109 109 128 128 127 121 118 122 116 124 125 114 126 113 109 96 103 145 104 109 116 109 117 106 101 113 87 123 108 93 119 98 108 101 131 141 143 112 98

построить:

а) интервальный вариационный ряд;

б) гистограмму частот;

в) кумулятивный ряд.

Занятие 2

ЭМПИРИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ И ЕЕ СВОЙСТВА. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ВЫБОРКИ

Краткие теоретические сведения

Эмпирической функцией распределения $F_n(x)$ называется относительная частота того, что признак (случайная величина X) примет значение, меньшее заданного x, т. е.

$$F_n(x) = W(X < x) = W_x^{\text{нак}}.$$

Для данного x эмпирическая функция распределения представляет накопленную частость

$$W_x^{\text{нак}} = \frac{n_x^{\text{нак}}}{n}.$$

Эмпирическая функция распределения обладает следующими свойствами:

1. Значение эмпирической функции принадлежит отрезку [0; 1].

2. Эмпирическая функция является неубывающей.

3. Если x_1 – наименьшее значение варианты, а x_k – наибольшее, то $F_n(x) = 0$ при $x \le x_1$ и $F_n(x) = 1$ при $x > x_k$.

В отличие от эмпирической функции распределения выборки $F_n(x)$ интегральную функцию $F(x)$ распределения генеральной совокупности называют ***теоретической функцией распределения***. Различие между эмпирической и теоретической функциями состоит в том, что теоретическая функция $F(x)$ определяет вероятность события $X < x$, а эмпирическая функция $F_n(x)$ определяет относительную частоту этого же события.

При больших n значения $F_n(x)$ и $F(x)$ мало отличаются одно от другого. Следовательно, $F_n(x)$ может служить для оценки $F(x)$.

Пример 2.1. По данному распределению выборки (табл. 2.1) найти эмпирическую функцию распределения.

Таблица 2.1

x_i	1	2	3	4	5
n_i	4	6	15	25	50

Решение. Определяем объем выборки: $n = \sum_{i=1}^{k} n_i = 100$.

Определяем относительные частоты вариант (табл. 2.2):

$$W_i = \frac{n_i}{n}.$$

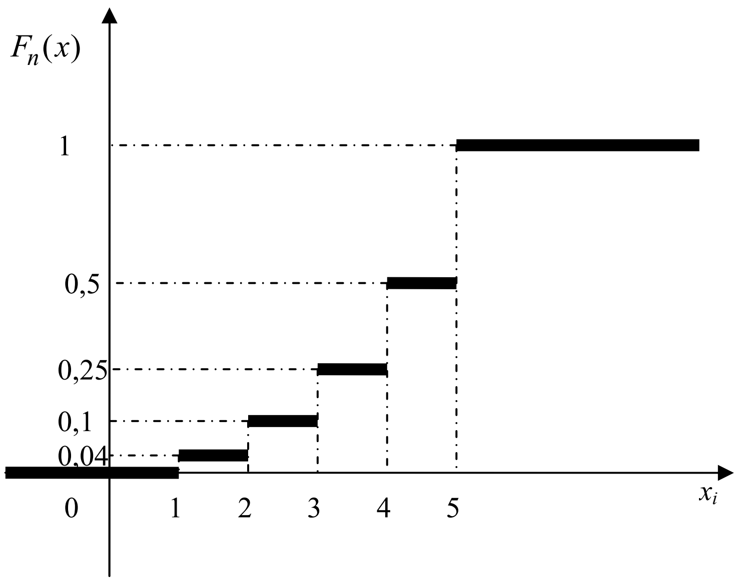
Таблица 2.2

x_i	1	2	3	4	5
W_i	0,04	0,06	0,15	0,25	0,5

Так как значение $F_n(x)$ есть сумма относительных частот вариант x_i, попадающих в интервал $(-\infty; x)$, то запишем эмпирическую функцию распределения:

$$F_n(x) = W(X < x) = \begin{cases} 0 & \text{при } x \le 1 \\ 0{,}04 & \text{при } 1 < x \le 2; \\ 0{,}1 & \text{при } 2 < x \le 3; \\ 0{,}25 & \text{при } 3 < x \le 4; \\ 0{,}5 & \text{при } 4 < x \le 5; \\ 1 & \text{при } x > 5. \end{cases}$$

График примет вид:



В случае дискретного вариационного ряда $F_n(x)$ представляет собой разрывную ступенчатую функцию.

Для интервального вариационного ряда имеем лишь значения функции $F_n(x)$ на концах интервала, поэтому для графического изображения этой функции целесообразно ее доопределить, соединив точки графика, соответствующие концам интервалов, отрезками прямой. В результате полученная ломаная совпадает с кумулятой.

Для описания выборки применяются числовые характеристики:

– характеристики положения (средняя арифметическая $\overline{x}$, мода Mo(X), медиана Me(X));

– характеристики рассеяния (выборочная дисперсия, выборочное среднее квадратическое отклонение);

– характеристики меры скошенности (коэффициент асимметрии) и эксцесс распределения.

Средней арифметической $\overline{\boldsymbol{x}}$ ($\overline{x}_{\text{в}}$ – выборочной средней) ***дискретного вариационного ряда*** называется отношение суммы произведений вариант на соответствующие частоты к объему выборки:

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i n_i}{n}.$$

Средняя арифметическая $\overline{\boldsymbol{x}}$ ***для интервального вариационного ряда*** находится по формуле

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i^* n_i}{n},$$

где x_i^* – середина соответствующего интервала.

Модой Mo(X) дискретного вариационного ряда называется варианта, которой соответствует наибольшая частота.

Медианой Me(X) дискретного вариационного ряда называется варианта, делящая ряд на две равные части.

Если дискретный вариационный ряд имеет $2m$ членов: $x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_{2m}$, то

$$\mathrm{Me}(X) = \frac{x_m + x_{m+1}}{2}.$$

Если дискретный вариационный ряд имеет $2m - 1$ членов: $x_1, x_2, \ldots, x_{m-1}, x_m, x_{m+1}, \ldots, x_{2m-1}$, то

$$\mathrm{Me}(X) = x_m.$$

Для интервальных вариационных рядов с равными интервалами применяют следующие формулы:

а) медианы

$$\mathrm{Me}(X) = x_{\mathrm{Me}} + h\frac{0,5n - S_{\mathrm{Me}-1}}{n_{\mathrm{Me}}},$$

где x_{Me} – начало медианного интервала, т. е. интервала $\left(x_l, x_{l+1}\right)$, где $l = \overline{1, k-1}$, номер которого определяется на основании неравенства

$$\sum_{i=1}^{l} n_i \le \frac{n}{2} < \sum_{i=1}^{l+1} n_i .$$

В случае выполнения равенства номер равен l, в противном случае – $l + 1$;

h – длина частичного интервала;

n – объем совокупности;

$S_{\mathrm{Me}-1}$ – накопленная частота интервала, предшествующего медианному;

n_{Me} – частота медианного интервала;

б) моды

$$\mathrm{Mo}(X) = x_{\mathrm{Mo}} + h\frac{n_{\mathrm{Mo}} - n_{\mathrm{Mo}-1}}{(n_{\mathrm{Mo}} - n_{\mathrm{Mo}-1}) + (n_{\mathrm{Mo}} - n_{\mathrm{Mo}+1})},$$

где x_{Mo} – начало модального интервала, т. е. интервала, которому соответствует наибольшая частота;

n_{Mo} – частота модального интервала;

$n_{\mathrm{Mo}-1}$ – частота предмодального интервала;

$n_{\mathrm{Mo}+1}$ – частота послемодального интервала.

Выборочная дисперсия вариационного ряда – средняя арифметическая квадратов отклонений вариант от их средней арифметической:

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \overline{x})^2.$$

Выборочная дисперсия равна разности между средним значением квадрата вариант и квадратом средней арифметической:

$$D_{\text{в}} = \overline{X}^2 - (\overline{x})^2,$$

где $\overline{X}^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_i.$

Выборочным средним квадратическим отклонением называется арифметическое значение корня квадратного из дисперсии:

$$\sigma_{\text{в}} = \sqrt{D_{\text{в}}}.$$

Коэффициентом асимметрии называется число

$$\overline{A} = \frac{\overline{\mu}_3}{\sigma_{\text{в}}^3} = \frac{\sum_{i=1}^{k}(x_i - \overline{x})^3 n_i}{n\sigma_{\text{в}}^3}.$$

Эксцессом называется число

$$\overline{E} = \frac{\overline{\mu}_4}{\sigma_{\text{в}}^4} = \frac{\sum\limits_{i=1}^{k}(x_i - \overline{x})^4 n_i}{n\sigma_{\text{в}}^4} - 3.$$

Задачи для аудиторной работы

Задача 2.1.

Построить эмпирическую функцию распределения по данным табл. 1.3 и вычислить числовые характеристики выборки.

Задача 2.2.

По данной выборке

12 6 8 6 10 11 7 10 12 8 7 7 6 7 8 6 11 9 11 9 10 11 9 10 7 8 8 8 11 9 8 7 5 9 7 7 14 11 9 8 7 11 15 6 4 7 5 5 10 7 7 5 8 10 10 15 10 10 13 12

построить:

а) интервальный вариационный ряд;

б) эмпирическую функцию распределения.

Найти числовые характеристики.

Задачи для внеаудиторной работы

Задача 2.3.

Построить эмпирическую функцию распределения по данным табл. 1.5 и вычислить числовые характеристики выборки.

Задача 2.4.

Вычислить по данным выборки числовые характеристики эмпирического распределения:

19 3 4 17 27 44 61 59 34 15 1 6 13 1 39 55 77 83 103 75 50 9 9 12 4 39 27 35 26 18 5 12 9 38 54 7 19 70 96 45 51 38 13 5 24 22 94 1 17 2 23 29 2 71 28 6 19 12 14 4 30 36 21 89 52 31 6 18 10 47 63 25 8 11 8 32 25 43 6 18 41 11 31 87 76 119 79.