Теория вероятностей и математическая статистика
Лектор   А.А. Лобузов
Семестр 6   Лекция 2

Дискретная статистическая модель

Пусть $\mathbf{X} = (X_1, X_2, ..., X_N)$ случайная выборка, полученная при измерении дискретной случайной величины ξ, принимающей значения

$x_1^* \quad x_2^* \quad \dots \quad \dots \quad \dots \quad x_m^*$.

По числовой выборке $\mathbf{x} = (x_1, x_2, ..., x_N)$ строим

Статистический ряд:

x_i	x_1^*	x_2^*	$\dots$	$\dots$	$\dots$	x_m^*
n_i	n_1	n_2	$\dots$	$\dots$	$\dots$	n_m
w_i	w_1	w_2	$\dots$	$\dots$	$\dots$	w_m

n_i – частота значения x_i^* , $\sum_{i=1}^{m} n_i = N$;

w_i – относительная частота (частость) значения x_i^*,

$w_i = \dfrac{n_i}{N}$, $\quad \sum_{i=1}^{m} w_i = 1$.

По статистическому ряду находят многие характеристики.

Пусть $X = (X_1, X_2, ..., X_N)$ - случайная выборка объёма N,
$x = (x_1, x_2, ..., x_N)$ - выборка (реализация) (числа)
При упорядочении x получаем числовой вариационный ряд $x_{(1)} \le x_{(2)} \le \ldots \le x_{(N-1)} \le x_{(N)}$
Случайная величина $X_{(k)}$, принимающая значение $x_{(k)}$, называется k-ой порядковой статистикой
?(или k-ый ~~элемент~~ член вариационного ряда)
Статистика - произвольная измеримая функция от X
Считаем, что нет повторов точек, т.е. $x_{(i)} \neq x_{(j)}$ при $i \neq j$
$x_{(1)}$ - минимальное значение

Эмпирическая функция распределения

$$F_N(x, X) = \frac{1}{N} \sum_{k=1}^{N} I_{(-\infty, x]}(X_k) = \frac{1}{N} \sum_{k=1}^{N} I_{(X_k \le x]}$$

(I - ~~идентификатор~~ индикатор множества, определяет значения x, ?не зав[исит] от вы[борки])
$F_N(x, X) = F_N(x, x_1, x_2, \ldots, x_N)$ - зависит от выборки

$$F_N(x, x) = \begin{cases} 0, & x < x_{(1)} \\ \frac{k}{N}, & x_{(k)} \le x < x_{(k+1)}, \quad 1 \le k \le N \\ 1, & x \ge x_{(N)} \end{cases}$$

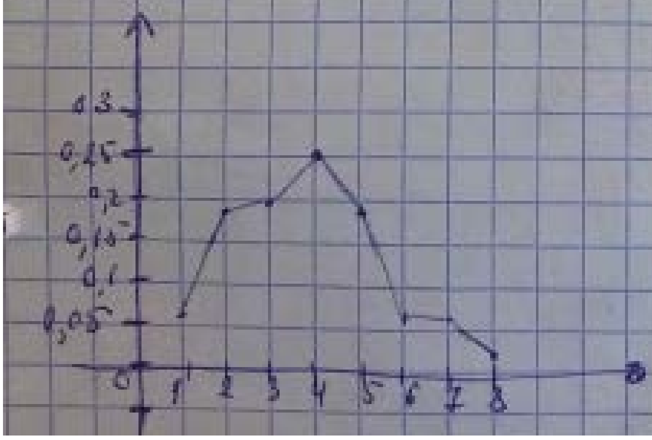
каждое число несет вес $\frac{1}{N}$
В результате эмпирическая функция зависит от количества испытаний

Полигон относительных частот

…истический ряд

x	1	2	3	4	5	6	7	8
n	6	19	20	25	17	6	6	1
w	0,06	0,19	0,2	0,25	0,17	0,06	0,06	0,01

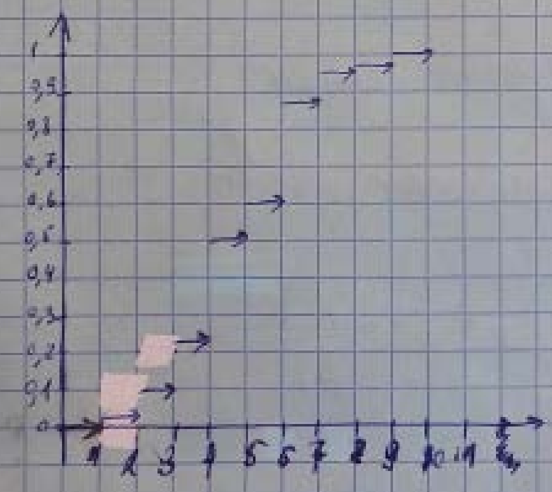
полигон относительных частот



Статистический ряд

x	1	2	3	4	5	6	7	8	9
n	1	9	12	27	11	28	7	3	2
w	0,01	0,09	0,12	0,27	0,11	0,28	0,07	0,03	0,02

Эмпирическая функция распределения



Числовые характеристики выборки

Выборочное среднее

$$\overline{x}=\frac{1}{N}\sum_{i=1}^{m}x_i^*\cdot n_i=\sum_{i=1}^{m}x_i^*\cdot w_i$$

Выборочный момент k-ого порядка

$$\overline{\mu}_k=\overline{x^k}=\sum_{i=1}^{m}(x_i^*)^k\cdot w_i\ ,\ \overline{\mu}_1=\overline{x}$$

Выборочная дисперсия

$$D_B=\sum_{i=1}^{m}(x_i^*-\overline{x})^2\cdot w_i=\sum_{i=1}^{m}(x_i^*)^2\cdot w_i-(\sum_{i=1}^{m}x_i^*\cdot w_i)^2=\overline{x^2}-(\overline{x})^2$$

Выборочное среднее квадратическое отклонение

$$\overline{\sigma}=\sqrt{D_B}$$

Выборочный центральный момент k-ого порядка

$$\overline{\mu}_k^0=\sum_{i=1}^{m}(x_i^*-\overline{x})^k\cdot w_i\ ,\ \overline{\mu}_1^0=0\ ,\ \overline{\mu}_2^0=D_B=\overline{\mu}_2-(\overline{\mu}_1)^2,$$

$$\overline{\mu}_3^0=\sum_{i=1}^{m}((x_i^*)^3-3(x_i^*)^2\overline{\mu}_1+3x_i^*\left(\overline{\mu}_1\right)^2-(\overline{\mu}_1)^3)\cdot w_i=\overline{\mu}_3-3\overline{\mu}_2\overline{\mu}_1+2(\overline{\mu}_1)^3,$$

$$\overline{\mu}_4^0=\sum_{i=1}^{m}((x_i^*)^4-4(x_i^*)^3\overline{\mu}_1+6(x_i^*)^2\left(\overline{\mu}_1\right)^2-4x_i^*(\overline{\mu}_1)^3+(\overline{\mu}_1)^4)\cdot w_i=$$
$$=\overline{\mu}_4-4\overline{\mu}_3\overline{\mu}_1+6\overline{\mu}_2(\overline{\mu}_1)^2-3(\overline{\mu}_1)^4$$

Выборочная мода $\overline{M}_0$

$\overline{M}_0=\{x_i^*\,|\,n_i=\max n_k\}$, если $n_i=\max n_k>n_j,\ i\neq j$;

если $n_i=n_{i+1}=...=n_{i+j}=\max n_k$, то $\overline{M}_0=\frac{1}{2}(x_i^*+x_{i+j}^*)$,

если $n_i=n_j=\max n_k>n_l\ ,\ i<l<j$, то $\overline{M}_0-$ не существует.

Выборочная медиана

$$\overline{M}_e = \begin{cases} x_i^*, & F_N^Э(x_{i-1}^*) < 0{,}5 < F_N^Э(x_i^*), \\ \frac{1}{2}(x_i^* + x_{i+1}^*), & F_N^Э(x_i^*) = 0{,}5\,. \end{cases}$$

Выборочный коэффициент асимметрии

$$\overline{a}_s = \frac{\overline{\mu}_3^0}{\overline{\sigma}^3}$$

Коэффициент асимметрии (**skewness**) положителен, если правый хвост распределения длиннее левого, и отрицателен в противном случае.

Если распределение симметрично относительно математического ожидания, то его коэффициент асимметрии равен нулю.

Выборочный коэффициент эксцесса

$$\overline{\varepsilon}_k = \frac{\overline{\mu}_4^0}{\overline{\sigma}^4} - 3$$

Коэффициент эксцесса (**kurtosis**, коэффициент островершинности)

– мера остроты пика распределения случайной величины.