Занятие №4.

1. Эмпирическая функция распределения

Построим *выборочный аналог* функции распределения $F(x)$.

Для этого вначале на каждом интервале (рис. 4.1) выберем середину $\tilde{x}_i = \dfrac{x_i + x_{i+1}}{2} \ (i = 1,\ldots,m)$ и составим табл. 1.

Таблица 1

$\tilde{x}_i$	$\tilde{x}_1$		$\tilde{x}_n$
n_i	n_1		n_m
$\omega_i = \dfrac{n_i}{n}$	$\dfrac{n_1}{n}$		$\dfrac{n_m}{n}$

Определение. Выборочной (эмпирической) функцией распределения называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$.

Таким образом, $F^*(x) = \dfrac{n_x}{n}$, где n_x – число вариант, меньших x, n– объем выборки. Из определения эмпирической функции распределения видно, что ее свойства совпадают со свойствами $F(x)$, а именно:

1. $0 \le F^*(x) \le 1$.
2. $F^*(x)$ – неубывающая функция.
3. Если x_1 – наименьшая варианта, то $F^*(x) = 0$ при $x \le x_1$; если $x_к$ – наибольшая варианта, то $F^*(x) = 1$ при $x > x_к$.

На оси ординат откладываем накопленные относительные частоты. Кружочки на графике означают, что соответствующие точки выброшены (рис.4.1).

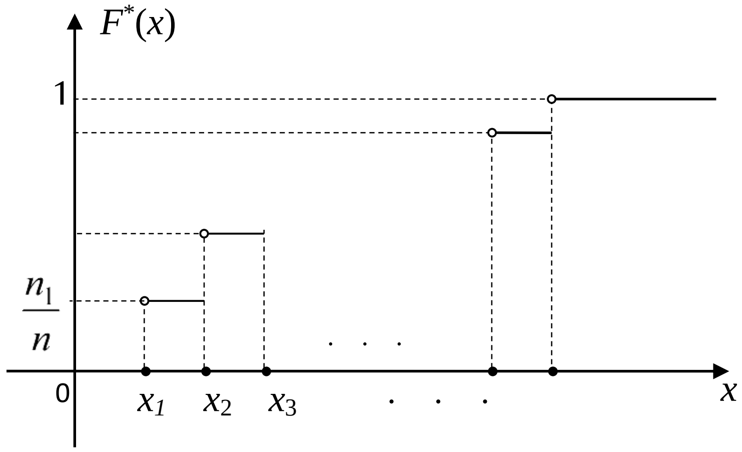

Рис. 4.1. Выборочная функция распределения

Можно доказать, что при достаточно большом объеме выборки и при достаточно мелком делении интервалов с практической достоверностью близка к истинной функции распределения $F(x)$.

2. Числовые характеристики статистического распределения: выборочное среднее, дисперсия

2.1. Параметры распределения

Одна из задач математической статистики: по имеющейся выборке оценить значения числовых характеристик исследуемой случайной величины. Основными числовыми характеристиками выборки являются выборочное среднее $\bar{x}_в$ и выборочная дисперсия $D_в$.

Определение. *Выборочным средним* называется среднее арифметическое значений случайной величины, принимаемых в выборке:

$$\bar{x}_в = \frac{x_1 + \ldots + x_n}{n} \text{ или } \bar{x}_в = \frac{\sum_{i=1}^{l} x_i n_i}{n}. \tag{4.1}$$

Выборочное среднее $\bar{x}_в$ служит для точечной оценки математического ожидания $M(X)$ исследуемой случайной величины

$$\bar{x} = \frac{x_1 + \ldots + x_n}{n}$$

Определение. *Выборочной дисперсией* называется

$$D_в(X) = \frac{\sum_{i=1}^{l} \left(x_i - \bar{x}_в\right)^2 \cdot n_i}{n} = \frac{\sum_{i=1}^{l} x_i{}^2 \cdot n_i}{n} - \bar{x}_в{}^2. \tag{4.2}$$

Выборочная дисперсия $D_в(X)$ служит для точечной оценки дисперсии $D(X)$.

2.2. Несмещенность, состоятельность, эффективность параметров распределения

Получив статистические оценки параметров распределения (выборочное среднее, выборочную дисперсию и т.д.), нужно убедиться, что они в достаточной степени служат приближением соответствующих характеристик генеральной совокупности. Определим требования, которые должны при этом выполняться.

Пусть θ^* - статистическая оценка неизвестного параметра θ теоретического распределения. Извлечем из генеральной совокупности несколько выборок одного и того же объема n и вычислим для каждой из них оценку параметра θ: $\theta_1^*, \theta_2^*, \theta_3^*, \ldots, \theta_n^*$.

Тогда оценку θ^* можно рассматривать как случайную величину, принимающую возможные значения $\theta_1^*, \theta_2^*, \theta_3^*, \ldots, \theta_n^*$.

1. Если математическое ожидание θ^* не равно оцениваемому

параметру, мы будем получать при вычислении оценок систематические ошибки одного знака (с избытком, если $M(\theta^*) > 0$, и с недостатком, если $M(\theta^*) < 0$). Следовательно, необходимым условием отсутствия систематических ошибок является требование $M(\theta^*) = 0$.

Оценка θ^* называется *несмещенной,* если ее математическое ожидание совпадает с истинным значением параметра $\theta : M\left[\theta^*\right] = \theta$.

2. *Определение.* Оценка θ^* называется *эффективной,* если она несмещенная и при этом имеет наименьшую дисперсию (наименьший разброс относительно θ) по сравнению с другими несмещенными оценками параметра θ.

3. *Определение*. Оценка θ^* называется *состоятельной,* если при неограниченном увеличении объема выборки θ^* сходится по вероятности к истинному значению параметра $\theta : \theta^* \to \theta$ при $n \to \infty$.

Состоятельной несмещенной оценкой математического ожидания *a* является

$$\bar{x}_в = \frac{\sum_{i=1}^{l} x_i n_i}{n}.$$

В отличие от выборочного среднего, выборочная дисперсия является смещенной оценкой дисперсии генеральной совокупности. Пользуясь оценкой $D_в$ вместо D, мы будем совершать некоторую систематическую ошибку, так как её математическое ожидание несколько меньше истинного значения. Чтобы её ликвидировать, достаточно ввести поправку, умножив $D_в$ на $\frac{n}{n-1}$. Таким образом, можно предложить другую оценку дисперсии – *исправленную дисперсию* s^2, вычисляемую по формуле $s^2 = \frac{n}{n-1} D_в = \frac{\sum_{i=1}^{l} (x_i - \bar{x}_в)^2 n_i}{n-1}$. Эта оценка является состоятельной несмещенной оценкой дисперсии *D*.

Такая оценка будет являться *несмещенной.* Ей соответствует *исправленное среднее квадратическое отклонение (СКО)*

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{l} (x_i - \bar{x}_в)^2 n_i}{n-1}}. \qquad (4.3)$$

При больших *n* поправка $\frac{n}{n-1}$ становится близкой к единице и её применение теряет смысл.

Пример 1. Оптический пирометр установлен на светящуюся нить накала, различными операторами было произведено несколько измерений температуры. Получены следующие результаты (табл.2).

Таблица 2

Температура, °C	925	950	975	1000	1025	500
Число измерений	1	9	6	18	10	2

Требуется найти среднюю квадратическую и вероятную ошибки в предположении, что эта выборка взята из нормально распределенной совокупности.

Решение. Найдем сначала выборочное среднее значение $\bar{x}_в$.

Проще находить среднее для 1000 – x, а не для x (табл. 3).

Таблица 3

$1000–x_i$	Число измерений n	$n\cdot(1000–x_i)$	$(1000–x_i)^2$	$n\cdot(1000–x_i)^2$
1	2	3	4	5
75	1	75	5625	5625
50	9	450	2500	22500
25	6	150	625	3750
0	18	0	0	0
-25	10	-250	625	6250
-50	2	-100	2500	5000
$\sum\limits_{i=1}^{l}$	46	325	–	43125

$$1000-\bar{x}_в=\frac{\sum\limits_{i=1}^{l}x_i n_i}{n}=\frac{325}{46}=7{,}1 \text{ или } \bar{x}_в=1000-7{,}1=992{,}9°.$$

$$D_в(X)=\frac{\sum\limits_{i=1}^{l}x_i^2\cdot n_i}{n}-\bar{x}_в^{\,2}=\frac{43125}{46}-7{,}1^2=887.$$

Найдем исправленную дисперсию s^2,

$$s^2=\frac{n}{n-1}D_в=\frac{46}{45}\cdot 887=906{,}7\,^{\circ}\text{C}.$$