Лекция №11. Тема 8.8. Основные понятия математической статистики.

Генеральная совокупность и выборка; выборочная функция распределения. Сгруппированный статистический ряд, гистограмма.

§ 1. Задачи математической статистики

Математическая статистика возникла и создавалась параллельно с теорией вероятностей в XVII в. Дальнейшее развитие математической статистики (вторая половина XIX и начало XX в.) обязано, в первую очередь П. Л. Чебышеву, А. А. Маркову, А. М. Ляпунову и др.

Математическая статистика – это наука о способах получения выводов из *опытных данных* (результатов наблюдений). Итак, задача математической статистики состоит в создании методов сбора и обработки статистических данных для получения научных и практических выводов.

Отсюда вытекают основные разделы математической статистики.

1. Теория оценок.

Эта теория позволяет приближенно вычислить и оценить параметры случайных величин (математическое ожидание, дисперсию и т.д.) *по данным опыта.*

2. Статистическая проверка гипотез.

Эта теория позволяет проверить справедливость интересующих нас гипотез *по данным опыта*.

3. Дисперсионный анализ.

Эта теория позволяет найти слабые (статистические) зависимости между величинами, т. е. направлена на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях. В данном пособии будут рассмотрены только первые два раздела.

§ 2. Выборка из генеральной совокупности. Вариационный ряд. Гистограмма относительных частот

Пусть с испытанием связана случайная величина X и пусть в результате серии n независимых испытаний получен набор значений X :

$$x_1, \quad x_2, \ldots, \; x_n. \tag{3.1}$$

Пусть $G-$ *генеральная совокупность,* т.е. множество всех возможных значений случайной величины X . Набор чисел (3.1) называется *выборкой* из генеральной совокупности, число n называется *объемом выборки*, числа (3.1) называются *элементами выборки* или *вариантами*. Расположенные в порядке возрастания значения $x_1, \quad x_2, \ldots, \; x_n$ образуют ряд, называемый *вариационным рядом:*

$x_{\min}, \quad \ldots \quad , x_{\max}$ – вариационный ряд.

Число $R = x_{\max} - x_{\min}$ называется *размахом выборки*.

Для построения вариационного ряда выполним действия:

1. Разделим отрезок $[x_{\min},\ \ x_{\max}]$ на некоторое число l интервалов одинаковой длины $\varDelta = \left(\dfrac{R}{k}\right)$. Величину $к$ приближенно вычисляют по формуле $l = [1 + 3{,}2\lg n]$, где n – объём выборки.

$$x_i = x_{i-1} + i\varDelta.$$

2. Подсчитаем число элементов выборки, попадающих в каждый интервал:

$$n_1, n_2, \quad \ldots \quad , n_m. \tag{3.2}$$

Очевидно, $n_1 + n_2 + \ldots + n_m = n$.

Числа (3.2) называются *частотами попадания в интервал.* Составим табл. 6.

Таблица 6

$x_0 - x_1$	$x_1 - x_2$		$x_{n-1} - x_n$
n_i	n_1		n_m
$\omega_i = \dfrac{n_i}{n}$	$\dfrac{n_1}{n}$		$\dfrac{n_m}{n}$

Элементы третей строки называются *относительными частотами попадания в интервал.* Очевидно, $\dfrac{n_1}{n} + \dfrac{n_2}{n} + \ldots + \dfrac{n_m}{n} = 1$.

Полученная таблица называется *выборочным распределением случайной величины* X.

3) Изобразим выборочное распределение на графике (рис. 3.1). Для этого на оси OX отложим интервалы $x_i - x_{i+1}$ и на каждом из них как на основании построим прямоугольник, площадь которого ω_i, т.е. высота i-го прямоугольника $\dfrac{\omega_i}{\varDelta}$.

Построенный график называется *гистограммой относительных частот* и представляет собой *выборочный аналог* плотности вероятности случайной величины. Площадь гистограммы равна единице.

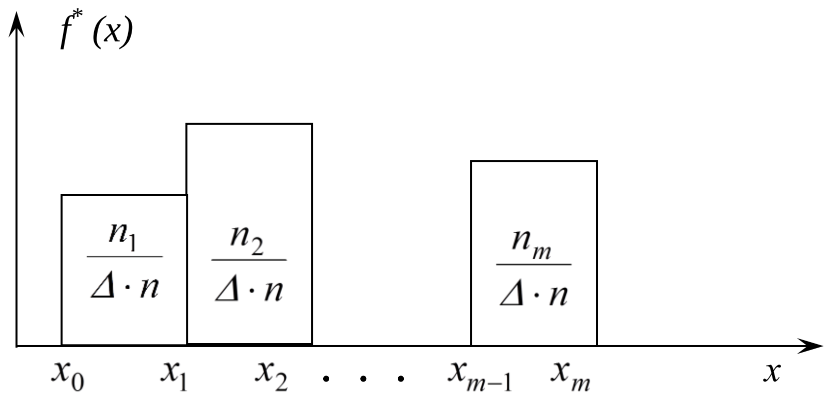

Рис. 3.1. Гистограмма относительных частот

Пример 1.

При проведении 20-ти серий из 10-ти бросков игральной кости число выпадений шести очков оказалось равным 1,1,4,0,1,2,1,2,2,0,5,3,3,1,0,2,2,3,4,1.Составим вариационный ряд: 0,1,2,3,4,5. Статистический ряд для абсолютных и относительных частот имеет вид табл. 7.

Таблица7

x_i	0	1	2	3	4	5
n_i	3	6	5	3	2	1
$\omega_i = \frac{n_i}{n}$	$\frac{3}{20}$	$\frac{6}{20}$	$\frac{5}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{1}{20}$

§ 3. Эмпирическая функция распределения

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

Для этого вначале на каждом интервале (рис. 3.2) выберем середину $\tilde{x}_i = \frac{x_i + x_{i+1}}{2} (i = 1,\ldots,m)$ и составим табл. 8.

Таблица 8

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

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

Таким образом, $F^*(x) = \frac{n_x}{n}$, 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.3.2).

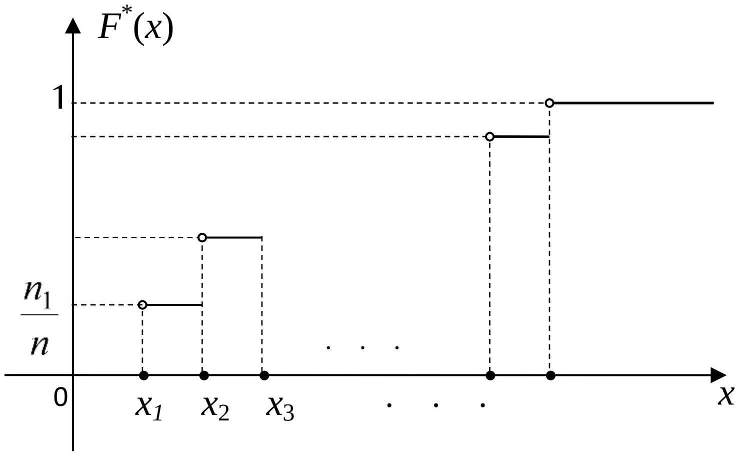

Рис. 3.2. Выборочная функция распределения

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