

Раздел. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Тема 1. Точечные оценки параметров распределения



План лекции

&1. Введение

**&2. Дискретные (точечные)
статистические распределения
выборки**

**&3. Эмпирическая функция
распределения**

&1. Введение

Математическая статистика
**изучает закономерности,
которые имеют место в
массовых совокупностях
однородных объектов.**

Основные задачи математической статистики:

1. Разработка методов сбора и группировки статистических сведений, полученных в результате наблюдений или в результате специально поставленных экспериментов.

2. Разработка методов анализа статистических данных в зависимости от целей исследования.

Сюда относятся:

а) оценка неизвестной вероятности события; оценка неизвестной функции распределения; оценка параметров распределения, вид которого известен; оценка зависимости случайной величины от одной или нескольких случайных величин и др.;

б) проверка статистических гипотез о виде неизвестного распределения или о величине параметров распределения, вид которого известен.

Итак, задача математической статистики состоит в создании методов сбора и обработки статистических данных для получения научных и практических выводов.

Современная математическая статистика разрабатывает способы определения числа необходимых испытаний до начала исследования (планирование эксперимента), в ходе исследования (последовательный анализ) и решает многие другие задачи. Современную математическую статистику определяют как науку о принятии решений в условиях неопределенности.

&2. Дискретные (точечные) статистические распределения выборки

ОПРЕДЕЛЕНИЕ_1

Генеральной совокупностью
называют совокупность
однородных объектов,
подлежащих изучению.

ОПРЕДЕЛЕНИЕ_2

Выборочной совокупностью
(выборкой) называется
совокупность объектов,
отобранных из генеральной
совокупности.

ОПРЕДЕЛЕНИЕ_3

***Объемом* совокупности
(выборной или генеральной)
называют число объектов этой
совокупности.**

$$n = \sum_{i=1}^l n_i \quad (1)$$

ОСНОВНЫЕ ПОНЯТИЯ

Пусть из генеральной совокупности
извлечена выборка,
причем x_i называют *вариантами*,
а последовательность вариантов,
записанных в возрастающем порядке,
– *вариационным рядом*.

Числа наблюдений называют частотами, а их отношения к объему выборки $\frac{n_i}{n} = W_i$ (2) — относительными частотами.

Определение_4

*Точечным статистическим
распределением выборки*
**называют перечень вариантов и
соответствующих им частот
или относительных частот.**

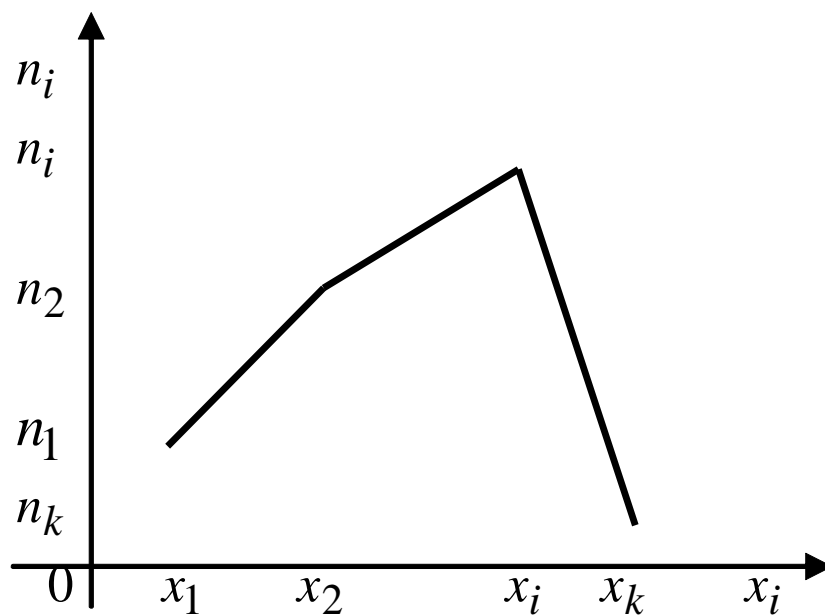
В случае дискретных случайных величин рассматривают *точечные* статистические распределения выборки, которые оформляются в виде следующей таблицы:

x_i	x_1	x_2	\simeq	x_k
n_i	n_1	n_2	\simeq	n_k
W_i	W_1	W_2	\simeq	W_k

$$\sum_{i=1}^k n_i = n; \quad \sum_{i=1}^k W_i = 1.$$

**Для графического представления
дискретного распределения
используют полигон частот
(полигон относительных частот).**

Полигоном частот называют ломаную,
отрезки которой соединяют точки
 $(x_1; n_1), (x_2; n_2), \dots, (x_k; n_k)$.



***Полигоном относительных частот* называют,
ломаную отрезки которой соединяют точки**

$$(x_1; W_1), (x_2; W_2), \dots, (x_k; W_k).$$

Определение_5

Несмещенной называют статистическую оценку θ^* , математическое ожидание которой равно оцениваемому параметру θ , т.е. $M(\theta^*) = \theta$ при любом объеме выборки, в противном случае оценка называется **смещенной**.

(4)

$$\overline{X}_e = \frac{\sum_{i=1}^k x_i \cdot n_i}{n}$$

Выборочная средняя

(несмещенная оценка математического ожидания)

(5)

$$D_B = \left(\frac{1}{n} \sum_{i=1}^m \overline{x}_i^2 n_i \right) - (\overline{x}_B)^2$$

Выборочная дисперсия

(несмещенная оценка выборочной дисперсии)

(6)

$$S^2 = \frac{n}{n-1} \cdot D_e$$

Исправленная выборочная дисперсия

(несмещенная оценка дисперсии)

(7)

$$\sigma_B = \sqrt{S^2}$$

Выборочное среднее квадратическое отклонение

(Среднее квадратическое отклонение)

Задача №1

Статистическое распределение выборки имеет вид:

x_i	1	2	3	4
n_i	6	5	4	3

Найти объем выборки.

РЕШЕНИЕ

Объем выборки вычисляется по формуле

$$n = \sum_{i=1}^k n_i, \text{ где } n_i - \text{частота варианты } x_i. \quad \text{Тогда}$$

$$n = 6 + 5 + 4 + 3 = 18.$$

Задача №2

Из генеральной совокупности извлечена выборка объема $n=100$:

x_i	1	3	5	7
n_i	15	16	17	n_4

Найти значение n_4 .

РЕШЕНИЕ

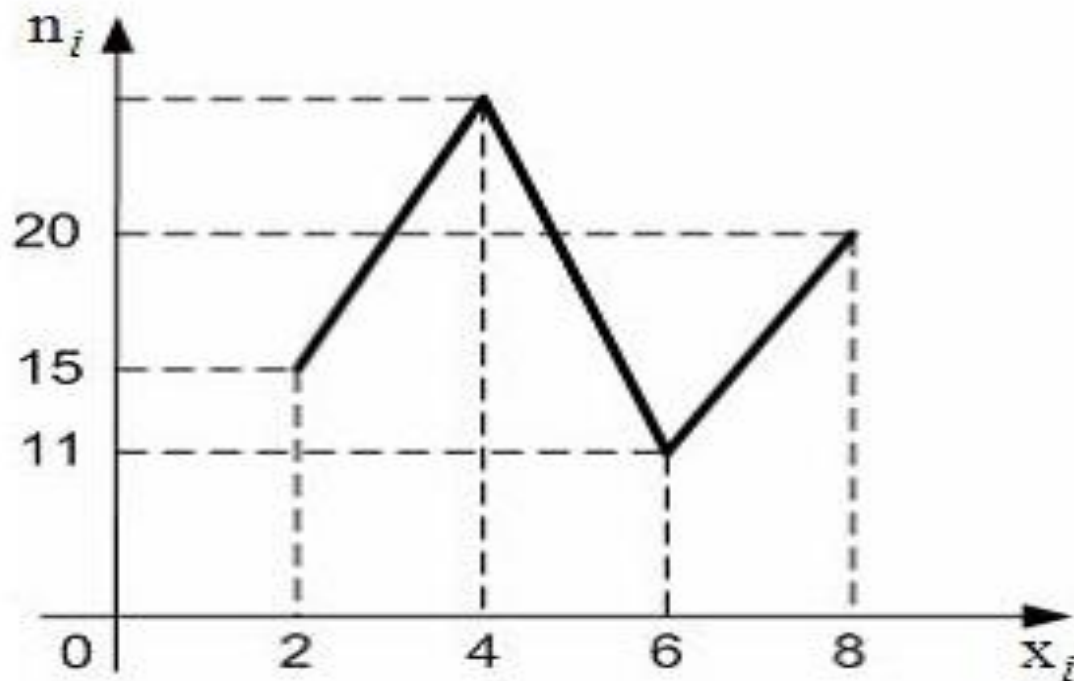
Объем выборки вычисляется по формуле

$n = \sum_{i=1}^k n_i$, где n_i - частота варианты x_i . Тогда

$$n_4 = 100 - 15 - 16 - 17 = 52$$

Задача №3

Из генеральной совокупности извлечена выборка объема $n=70$, полигон частот которой имеет вид:



Найти число вариант $x_i = 4$ в выборке.

РЕШЕНИЕ

Объем выборки вычисляется по формуле

$n = \sum_{i=1}^k n_i$, где n_i - частота варианты x_i . Тогда

$$n_2 = 70 - 20 - 15 - 11 = 24.$$

Задача 4:

Из генеральной совокупности извлечена
выборка объема $n=10$:

x_i	2	4	7
n_i	5	3	2

Найти несмещенную оценку математического
ожидания

РЕШЕНИЕ

$$\overline{X}_e = \frac{\sum_{i=1}^k x_i \cdot n_i}{n}$$

$$\bar{x}_B = \frac{2 \cdot 5 + 4 \cdot 3 + 7 \cdot 2}{10} = 3,6.$$

Задача 5:

Проведено четыре измерения (без систематических ошибок) некоторой случайной величины (в мм): 10, 11, 12, 15. Найти несмещенную оценку математического ожидания.

РЕШЕНИЕ

$$\overline{X}_e = \frac{\sum_{i=1}^k x_i \cdot n_i}{n}$$

$$\bar{x}_B = \frac{10 + 11 + 12 + 15}{4} = 12.$$

Задача 6:

В результате измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты (в мм): 5, 6, 7.
Найти несмещенную оценку дисперсии.

РЕШЕНИЕ

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x}_B)^2}{n-1}, \quad \text{где} \quad \bar{x}_B = \frac{\sum_{i=1}^n x_i}{n}.$$

$$\bar{x}_B = \frac{5+6+7}{3} = 6,$$

$$S^2 = \frac{(5-6)^2 + (6-6)^2 + (7-6)^2}{3-1} = 1.$$

Точность оценки

Точечной называют оценку, которая определяется одним числом. Все оценки, рассмотренные выше, точечные. При выборки малого объема точечная оценка может значительно отличаться от оцениваемого параметра, т.е. приводить к грубым ошибкам. По этой причине при наибольшем объеме выборки следует пользоваться интервальными оценками.

&3. Эмпирическая функция распределения

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

Итак, по определению,

$$F^*(x) = \frac{n_x}{n}, \quad (3)$$

где n_x - сумма частот вариантов, меньших x ; n - объем выборки.

Из определения функции $F^*(x)$ вытекают следующие ее свойства:

1. Значения эмпирической функции принадлежат отрезку $[0; 1]$.
2. $F^*(x)$ - неубывающая функция.
3. Если x_1 - наименьшая варианта, то $F^*(x) = 0$ при $x \leq x_1$; если x_k - наибольшая варианта, то $F^*(x) = 1$ при $x > x_k$.

Эмпирическая функция распределения выборки служит для оценки теоретической функции распределения генеральной совокупности.

Различие между эмпирической и теоретической функциями состоит в том, что теоретическая функция $F(x)$ определяет вероятность события $X < x$, а эмпирическая функция $F^*(x)$ определяет относительную частоту этого же события.

Задача №7

Из большой группы предприятий одной из отраслей промышленности случайным образом отобрано 30, по которым получены показатели основных фондов в млн. руб.:
2; 3; 2; 4; 5; 2; 3; 3; 6; 4; 5; 4; 6; 5; 3; 4; 2; 4; 3; 3; 5; 4; 6; 4; 5; 3;
4; 3; 2; 4.

1. Составить дискретное статистическое распределение выборки.
2. Найти объем выборки.
3. Составить распределение относительных частот.
4. Построить полигон частот.
5. Составить эмпирическую функцию распределения и построить ее график.
6. Найти несмещенные оценки числовых характеристик случайной величины.

РЕШЕНИЕ

1. Расположим различные значения признака в порядке их возрастания и под каждым из них запишем их частоты. Получим дискретное статистическое распределение выборки:

x_i	2	3	4	5	6
n_i	5	8	9	5	3

где x_i - варианты, n_i - частоты вариант x_i .

2. Сумма частот всех вариантов должна быть равной объему выборки. В данном примере объем выборки равен: $n=5 + 8 + 9 + 5 + 3=30$.

3. Найдем относительные частоты:

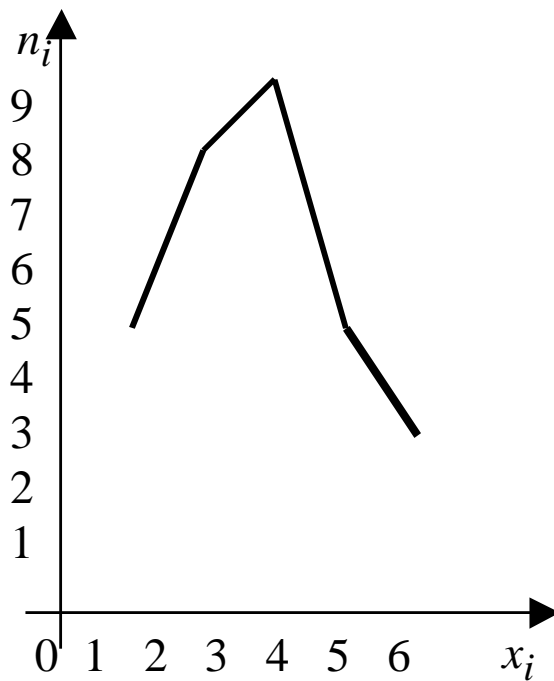
$$W_1 = \frac{5}{30} = \frac{1}{6}; \quad W_2 = \frac{8}{30} = \frac{4}{15}; \quad W_3 = \frac{9}{30} = \frac{3}{10}; \quad W_4 = \frac{5}{30} = \frac{1}{6}; \quad W_5 = \frac{3}{30} = \frac{1}{10}.$$

Запишем искомое распределение относительных частот

x_i	2	3	4	5	6
W_i	$\frac{1}{6}$	$\frac{4}{15}$	$\frac{3}{10}$	$\frac{1}{6}$	$\frac{1}{10}$

Контроль: $\frac{1}{6} + \frac{4}{15} + \frac{3}{10} + \frac{1}{6} + \frac{1}{10} = 1.$

4. Строим точки с координатами (x_i, n_i) и соединяем их последовательно отрезками. Полученная ломаная линия называется полигоном частот:



5. Согласно определению эмпирической функцией распределения называется функция вида

$$F^*(x) = \frac{n_x}{n},$$

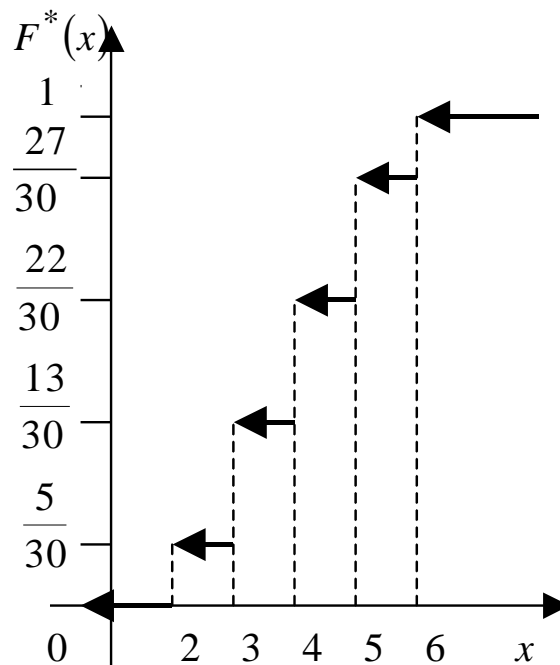
где n – объем выборки; n_x – сумма частот вариантов, меньших x .

Эмпирическая функция является оценкой функции распределения генеральной совокупности. Наименьшая варианта равна 2, поэтому при $x \leq 2$, $n_x = 0$ и $F^*(x) = 0$. Значение $X < 3$, а именно, $X = x_1 = 2$ наблюдалось 5 раз. Тогда для $2 < x \leq 3$ $n_x = 5$ и $F^*(x) = \frac{5}{30}$. Значение $X < 4$, а именно, $X = 2, X = 3$, наблюдалось $5 + 8 = 13$ раз. Поэтому для $3 < x \leq 4$ $n_x = 13$ и $F^*(x) = \frac{13}{30}$. Аналогично рассуждая, получаем: для $4 < x \leq 5$ $n_x = 5 + 8 + 9 = 22$ и $F^*(x) = \frac{22}{30}$, для $5 < x \leq 6$ $n_x = 5 + 8 + 9 + 5 = 27$ и $F^*(x) = \frac{27}{30}$ и при $x > 6$ $n_x = 5 + 8 + 9 + 5 + 3 = 30$ и $F^*(x) = \frac{30}{30} = 1$.

Таким образом,

$$F^*(x) = \begin{cases} 0 & \text{при } x \leq 2; \\ \frac{5}{30} & \text{при } 2 < x \leq 3; \\ \frac{13}{30} & \text{при } 3 < x \leq 4; \\ \frac{22}{30} & \text{при } 4 < x \leq 5; \\ \frac{27}{30} & \text{при } 5 < x \leq 6; \\ 1 & \text{при } x > 6. \end{cases}$$

График эмпирической функции имеет вид:



6. Несмещенной оценкой математического ожидания является средняя выборочная:

$$\bar{x}_B = \frac{\sum_{i=1}^k x_i n_i}{n} = \frac{2 \cdot 5 + 3 \cdot 8 + 4 \cdot 9 + 5 \cdot 5 + 6 \cdot 3}{30} = \frac{113}{30} \approx 3,77.$$

Несмещенная оценка дисперсии – исправления выборочная дисперсия:

$$s^2 = \frac{n}{n-1} D_B.$$

$$D_B = \overline{x_B^2} - (\bar{x}_B)^2 = \frac{2^2 \cdot 5 + 3^2 \cdot 8 + 4^2 \cdot 9 + 5^2 \cdot 5 + 6^2 \cdot 3}{30} - \left(\frac{113}{30}\right)^2 \approx 1,42.$$

$$s^2 = \frac{30}{29} \cdot 1,42 \approx 1,47.$$

ОСНОВНЫЕ ФОРМУЛЫ

Объем выборки

$$n = \sum_{i=1}^l n_i$$

Относительные частоты

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

Эмпирическая функция распределения

$$F^*(x) = \frac{n_x}{n},$$

**Несмещенная оценка
математического ожидания
(выборочное среднее)**

$$\overline{X}_e = \frac{\sum_{i=1}^k x_i \cdot n_i}{n}$$

**Несмещенная оценка дисперсии
(исправленная дисперсия)**

$$S^2 = \frac{n}{n-1} \cdot D_e$$

Выборочная дисперсия

$$D_B = \left(\frac{1}{n} \sum_{i=1}^m \overline{x_i^2} n_i \right) - (\overline{x_B})^2$$

**Среднее квадратическое
отклонение**

$$\sigma_B = \sqrt{S^2}$$

ВОПРОСЫ ДЛЯ САМОКОНТРОЛЯ

- 1. Что называется генеральной совокупностью?**
- 2. Что называется выборкой?**
- 3. Что называется объемом выборки?**
- 4. Что называется вариантом?**
- 5. Что называется частотой вариант?**
- 6. Что называется вариационным рядом?**
- 7. Что называется относительной частотой?**
- 8. Что называется полигоном частот?**
- 9. Что называется дискретным статистическим распределением выборки?**
- 10. Что называется эмпирической функцией распределения?**
- 11. Перечислите свойства эмпирической функции.**
- 12. Какие оценки называются точечными?**

ЗАДАЧИ ДЛЯ САМОКОНТРОЛЯ

№1. Из нескольких групп первого курса одной из университетов случайным образом отобрано 20 студентов, по которым получены показатели некоторого психологического признака: 8, 4, 4, 7, 5, 5, 5, 3, 10, 2, 8, 10, 4, 10, 10, 10, 8, 5, 3, 10

1. Составить дискретное статистическое распределение выборки.
2. Найти объем выборки.
3. Составить распределение относительных частот.
4. Построить полигон частот.
5. Составить эмпирическую функцию распределения и построить ее график.
6. Найти несмещенные оценки числовых характеристик случайной величины.