

Раздел 3. Элементы математической статистики

Литература. [5], гл.15, гл.16

Математическая статистика занимается методами сбора и обработки статистического материала – результатов наблюдений над объектами и анализом полученных данных для выявления ряда закономерностей случайных явлений.

В статистике оперируют понятием статистическое распределение, то есть соответствием между количественными признаками наблюдаемых объектов и их частотами.

Назовем исследуемую совокупность однородных объектов генеральной совокупностью, а множество из n объектов, которые выбирают случайным образом из генеральной совокупности – выборочной совокупностью, или просто выборкой (n - объем выборки).

Подробнее о предмете и задачах математической статистики, о генеральной, выборочных совокупностях, способах отбора и требованиях к объектам, о статистическом распределении выборки см. [5] (глава 5 § 1-6). По статистическому распределению составляется эмпирическая функция распределения – оценка функции распределения признака в генеральной совокупности, строятся полигон и гистограмма частот [5] (§ 7 - 8).

Для параметров распределения признаков в генеральной совокупности находятся точечные и интервальные оценки. Оценка называется точечной, если она характеризуется одним числом. Точечными оценками параметров распределения являются выборочная средняя, выборочная дисперсия [5] (глава 16 § 1). Интервальная оценка характеризуется двумя числами – концами интервала.

Интервал $(0-b, 0+b)$, покрывает оцениваемый параметр с вероятностью γ , при этом вероятность γ называется доверительной. Подробнее об этой оценке, о связи между доверительным интервалом, доверительной вероятностью и

объемом выборки для случая нормального распределения признака в генеральной совокупности см. [5] (глава 16 § 14 – 19).

Основные вопросы к теме

1. Предмет и задачи математической статистики. Генеральная и выборочная совокупности. Репрезентативность выборки. Способы отбора статистического материала (повторный, случайный, механический).
2. Статистическое распределение. Варианты, частоты, относительные частоты. Графическое представление выборки – полигон частот, гистограмма частот.
3. Эмпирическая функция распределения. Определение, свойства.
4. Выборочные характеристики статистического распределения: выборочная средняя $\bar{X}_B$, выборочная D_B и исправленная выборочная s^2 дисперсии, выборочное σ_B и исправленное s выборочное среднее квадратическое отклонение, мода и медиана.
5. Статистические оценки параметров распределения. Основные определения. Несмещенная, состоятельная и эффективная оценка. Оценки математического ожидания и генеральной дисперсии.
6. Доверительный интервал. Доверительная вероятность (надежность). Доверительный интервал для оценки математического ожидания нормального распределения при неизвестном σ . Доверительный интервал для оценки среднего квадратического отклонения σ нормального распределения.

Пример. Дана выборка из 16 элементов некоторого признака X из генеральной совокупности 19, 15, 16, 17, 18, 19, 16, 16, 18, 18, 17, 17, 17, 16, 17, 18.

Считаем, что распределение дискретное – возможные значения признака изолированы друг от друга.

Требуется:

- 1) Записать в виде вариационного и статистического рядов эту выборку.
- 2) Построить полигон относительных частот выборки.
- 3) Найти и построить эмпирическую функцию распределения $F^*(x)$ по данному распределению выборки.
- 4) Найти выборочное среднее, выборочную дисперсию, исправленную дисперсию, исправленное среднее квадратическое отклонение, моду, медиану, среднее абсолютное отклонение, коэффициент вариации вариационного ряда.
- 5) В предположении, что X распределена по нормальному закону построить доверительный интервал для математического ожидания с надежностью $\gamma = 0,95$.

Решение:

Если упорядочим элементы выборки по величине получим вариационный ряд:

15, 16, 16, 16, 16, 17, 17, 17, 17, 17, 18, 18, 18, 18, 19, 19

Размах выборки $\omega = 19 - 15 = 4$.

Статистический ряд запишем в виде таблицы, где x_i - перечень вариант вариационного ряда, n_i - соответствующие частоты.

x_i	15	16	17	18	19
n_i	1	4	5	4	2

Контроль: $\sum_{i=1}^5 n_i = 16$

Найдем относительные частоты $\omega_i = \frac{n_i}{n}$.

ω_i	0,0625	0,25	0,3125	0,25	0,125
------------	--------	------	--------	------	-------

Построим полигон относительных частот выборки (рис. 1) – ломаную линию, соединяющие точки (X_i, ω_i) , где X_i - варианты выборки, ω_i - соответствующие им относительные частоты.

По данному распределению выборки найдем эмпирическую функцию распределения $F^*(x)$.

Функция $F^*(x)$ определяет для каждого X частоту события $X < x$, $F^*(x) = \frac{n_x}{n}$, где n_x - число вариант меньше X .

Меньшая варианта равна 15, следовательно, $F^*(x) = 0$, при $X \leq 15$, для наибольшей $X = 19$ имеем: $F^*(x) = 1$, при $X > 19$.

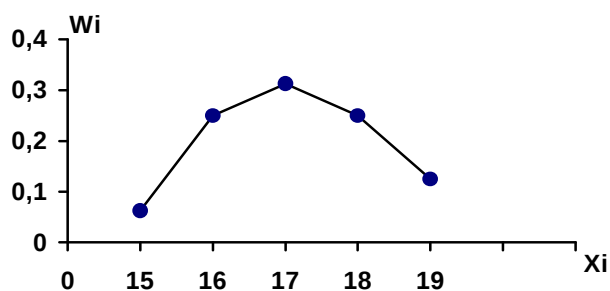


Рис.1

$$F^*(x) = \begin{cases} 0 & x \leq 15 \\ \frac{1}{16} = 0,0625 & 15 < x \leq 16 \\ \frac{1+4}{16} = 0,0625 + 0,25 = 0,3125 & 16 < x \leq 17 \\ \frac{1+4+5}{16} = 0,3125 + 0,3125 = 0,625 & 17 < x \leq 18 \\ \frac{1+4+5+4}{16} = 0,625 + 0,25 = 0,875 & 18 < x \leq 19 \\ 1 & x > 19 \end{cases}$$

График эмпирической функции распределения изображен на рис.2.

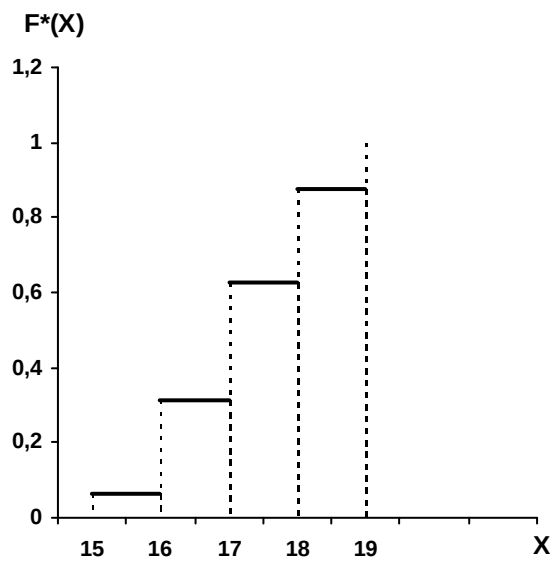


Рис. 2

Найдем выборочные числовые характеристики.

Выборочная средняя

$$\begin{aligned}\bar{x}_B &= \frac{1}{n} \sum_{i=1}^n x_i \cdot n_i = \frac{1}{16} (15 \cdot 1 + 16 \cdot 4 + 17 \cdot 5 + 18 \cdot 4 + 19 \cdot 2) = \\ &= \frac{15 + 64 + 85 + 72 + 38}{16} = \frac{274}{16} = 17,125\end{aligned}$$

Выборочная дисперсия – среднее арифметическое значение квадратов отклонения признака от выборочного среднего.

$$D_B = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_B)^2 n_i$$

Для подсчета D_B составим таблицу.

Первый способ подсчета:

$x_i - \bar{x}_B$	$(x_i - \bar{x}_B)^2$	$(x_i - \bar{x}_B)^2 n_i$
-2,125	4,516	4,515
-1,125	1,200	5,0625
0,125	0,0156	0,0781
0,875	0,766	3,0625
1,875	3,5156	7,031
		$\Sigma=19,75$

$$D_B = \frac{19,75}{16} \approx 1,2344$$

Второй способ подсчета:

x_i^2	225	256	289	324	361
n_i	1	4	5	4	2
$\cdot n_i x_i^2$	225	1024	1445	1296	722
	$\Sigma=4712$				

$$D_B = \overline{x^2} - (\bar{x}_B)^2 = \sum_{i=1}^n x_i^2 \frac{n_i}{n} - (\bar{x}_B)^2$$

$$(\bar{x}_B)^2 = (17,125)^2 \approx 293,2656$$

$$D_B = \frac{4712}{16} - 293,2656 \approx 294,56 - 293,2656 \approx 1,234$$

Вычислим исправленное среднее квадратическое выборочное

$$s^2 = \frac{n}{n-1} D_B = \frac{15}{16} \cdot 1,234 \approx 1,1317$$

и исправленное среднее квадратическое отклонение

$$s = \sqrt{s^2} \approx 1,147$$

Модой M_0 называют значение признака, которое имеет наибольшую частоту: $M_0 = 17$

Медианой m_e называют значение признака, который делит статистическое распределение на две равные части:

$$m_e = \begin{cases} x_{k+1}, & \text{если } n=2k+1 \\ \frac{x_k + x_{k+1}}{2}, & \text{если } n=2k \end{cases} \quad m_e = \frac{17+18}{2} = 17,5.$$

Коэффициент вариации:

$$V = \frac{\sigma_b}{x_b} \cdot 100\% = \frac{1,11}{17,125} \cdot 100\% \approx 6,49\%$$

Среднее абсолютное отклонение:

$$\theta = \sum_{i=1}^n \frac{n_i |x_i - \bar{x}_b|}{n} = \frac{|15 - 17,125| \cdot 1 + |16 - 17,125| \cdot 4 + |17 - 17,125| \cdot 5 + |18 - 17,125| \cdot 4 + |19 - 17,125| \cdot 2}{16} =$$

$$= \frac{14,5}{16} \approx 0,906$$

Доверительный интервал для оценки математического ожидания имеет вид:

$$\left(\bar{x}_B - t_\gamma \cdot \frac{\sigma_B}{\sqrt{n}}; \bar{x}_B + \frac{\sigma_B}{\sqrt{n}} \cdot t_\gamma \right)$$

Имеем: $n = 16$; $\sqrt{n} = 4$; $\sigma_B = \sqrt{D_B} = 1,111$; $\bar{x}_B = 17,125$.

Число t_γ определяем из условия: $\Phi(t_\gamma) = \frac{0,95}{2} = 0,475$, по таблице значений функции Лапласа $\Phi(x)$ (см. приложение 2 в [5]) отыщем аргумент $t_\gamma = 1,96$. Левый конец интервала равен

$$17,125 + 1,96 \cdot \frac{1,111}{4} = 17,687 \quad , \quad \text{правый конец} \quad -$$

$$17,125 - 1,96 \cdot \frac{1,111}{4} = 16,56.$$

Интервал (16,56; 17,687) покрывает неизвестное математическое математическое ожидание с надежностью $\gamma = 0,95$.

Контрольные задания

В задачах № 1241 – 1250 требуется по заданной выборке из n элементов некоторого признака x . Найти

1. Вариационный и статистический ряды;
2. Построить полигон относительных частот;

3. Эмпирическую функцию распределения $F^*(x)$ и построить ее график;

4. $\bar{X}_B$ - выборочное среднее; D_B - выборочную дисперсию; s^2 - исправленную дисперсию; σ_B , S - средние квадратические отклонения - выборочное и исправленное; M_0 - моду; m_e - медиану; θ - среднее абсолютное отклонение; V - коэффициент вариации вариационного ряда.

5 В предположении, что x распределена по нормальному закону построить доверительный интервал для неизвестного математического ожидания с данной надежностью γ .

№1241 10, 11, 10, 12, 12, 11, 13, 13, 12, 12, 11, 14, 14, 10, 12, 11, 13, 12 $\gamma=0,95$

№1242 21, 25, 23, 24, 24, 23, 21, 23, 22, 22, 22, 22, 23, 24, 25, 24, 21, 23; $\gamma=0,97$

№1243 1, 5, 3, 4, 4, 3, 1, 3, 2, 2, 2, 2, 3, 4, 5, 4, 1, 3; $\gamma=0,95$

№1244 36, 35, 36, 35, 37, 34, 38, 37, 38, 34, 34, 37, 36, 35, 36; $\gamma=0,99$

№1245 6, 2, 2, 8, 4, 4, 4, 2, 6, 2, 6, 8, 6, 8, 4, 4, 4, 6, 6, 8, 8, 2; $\gamma=0,95$

№1246 9, 9, 9, 8, 7, 8, 7, 6, 7, 8, 9, 8, 6, 6, 7, 7, 7, 8; $\gamma=0,99$

№1247 12, 12, 14, 14, 16, 12, 14, 14, 16, 16, 18, 18, 18, 16, 16, 14; $\gamma=0,97$

№1248 13, 15, 17, 13, 13, 15, 11, 11, 11, 9, 11, 13, 17, 15, 9, 9; $\gamma=0,95$

№1249 20, 21, 20, 20, 21, 21, 22, 22, 23, 24, 23, 22, 22, 21, 20, 21; $\gamma=0,97$

№1250 31, 32, 33, 34, 35, 34, 35, 33, 31, 33, 31, 32, 33, 34, 32, 32; $\gamma=0,95$