

РАЗДЕЛ 2

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Математическая статистика занимается изучением методов сбора и обработки опытных данных для получения научных и практических выводов и является разделом математики, очень близким к теории вероятностей. Лучше сказать, математическая статистика в решении своих специфических задач пользуется результатами теории вероятностей. Однако следует иметь в виду, что это другая, отдельная наука, решающая в каком-то смысле обратные задачи по сравнению с задачами теории вероятностей.

Типичная задача теории вероятностей. Задана вероятность p наступления случайного события A в одном опыте. Какова вероятность того, что в 200 опытах событие A наступит 3 раза?

Типичная задача математической статистики. Произвели 200 опытов, случайное событие A при этом наступило 3 раза. Какова вероятность наступления события A в одном опыте?

В математической статистике объектом исследования являются данные эксперимента.

§1. ВАРИАЦИОННЫЕ РЯДЫ И ИХ ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ

Установление статистических закономерностей, присущих случайным явлениям, основано на изучении статистических данных – сведений о том, какие значения принял в результате наблюдения интересующий нас признак (случайная величина) ξ . Например, исследуется спрос на определенные размеры мужской обуви по имеющимся статистическим данным о размерах 350 пар обуви, проданных магазином за неделю; исследуется точность измерительного прибора по результатам 50 независимых измерений.

Генеральной совокупностью называется совокупность объектов или наблюдений, все элементы которой подлежат изучению при статистическом анализе. Генеральная совокупность может быть конечной или бесконечной.

На практике изучение всего набора элементов генеральной совокупности в большинстве случаев оказывается невозможным. Часть объектов генеральной совокупности, используемая для исследования, называется **выборочной совокупностью** или **выборкой** (X). Число объектов в генеральной совокупности (N) или выборке (n) называют их объемами. В примере об исследовании спроса на размеры мужской обуви используется выборка объема $n = 350$, в примере об исследовании точности измерительного прибора $n = 50$. Сущность вы-

борочного метода состоит в том, что выводы, сделанные на основе изучения части совокупности (выборки), статистика позволяет распространять на всю генеральную совокупность.

Полученные в результате выборки значения x_i признака ξ называются **вариантами** или **элементами выборки**, $X = \{x_1, x_2, \dots, x_n\}$. Упорядоченная по возрастанию значений совокупность вариантов носит название **вариационного ряда**.

Дискретным вариационным (статистическим) рядом или статистическим распределением выборки называют таблицу, которая в первой строке содержит значения x_1, x_2, \dots, x_s ($x_1 < x_2 < \dots < x_s$), а во второй числа их повторений n_1, n_2, \dots, n_s , $n_1 + n_2 + \dots + n_s = n$:

X	x_1	x_2	\dots	x_s
n	n_1	n_2	\dots	n_s

Число n_i называют **частотой**, а отношение $w_i = \frac{n_i}{n}$ **относительной частотой** элемента x_i , $i = \overline{1, s}$. Дискретный вариационный ряд, как правило, представляет собой выборку значений дискретной случайной величины.

При изучении непрерывных признаков, для которых x_i могут принимать как угодно близкие значения, пользуются **интервальным вариационным (статистическим) рядом**. Также использование интервальных рядов целесообразно при изучении выборок большого объема. Интервал $J = [x_1, x_n]$ разбивают на s промежутков одинаковой длины Δ . При этом считают, что каждый промежуток содержит свой левый конец, но лишь последний промежуток содержит и свой правый конец. Далее для каждого промежутка J_i подсчитывают число элементов выборки, попавших в него – n_i , $i = \overline{1, s}$, в результате данные наблюдений представляют в виде таблицы:

X	$[x_1, x_2)$	$[x_2, x_3)$	\dots	$[x_s, x_{s+1}]$
n	n_1	n_2	\dots	n_s

Число промежутков s следует брать не очень большим, чтобы после группировки ряд не был громоздким, и не очень малым, чтобы не потерять особенности распределения признака. Согласно формуле Стерджеса, рекомендуемое число интервалов $s = 1 + 3,322 \lg n$, а величина частичного интервала $\Delta = \frac{x_{\max} - x_{\min}}{s}$.

Для графического изображения вариационных рядов наиболее часто используются полигон и гистограмма.

Полигон (относительных частот), как правило, служит для изображения дискретного вариационного ряда и представляет собой ломаную, отрезки которой последовательно соединяют точки с координатами $(x_i, n_i) \left(\left(x_i, \frac{n_i}{n} \right) \right)$, $i = \overline{1, s}$. При построении полигона для интервального вариационного ряда в качестве x_i используют середины интервалов.

Гистограмма (частот, относительных частот) служит только для изображения интервальных вариационных рядов и представляет собой ступенчатую фигуру из прямоугольников с основаниями равными интервалам значений признака $J_i = [x_i, x_{i+1})$ и высотами, равными плотности частоты $\frac{n_i}{n\Delta}$ (частоте n_i , относительной частоте $\frac{n_i}{n}$). Если соединить середины верхних оснований прямоугольников отрезками прямой, то можно получить полигон того же распределения. Суммарная площадь всех прямоугольников гистограммы равна единице:

$$\sum_{i=1}^s \frac{n_i}{n\Delta} \Delta = \sum_{i=1}^s \frac{n_i}{n} = \frac{1}{n} \sum_{i=1}^s n_i = \frac{n}{n} = 1.$$

Кроме того, площадь каждого прямоугольника $\frac{n_i}{n}$ есть относительная частота попадания элементов выборки в соответствующий интервал J_i статистического ряда. Гистограмма является статистическим аналогом кривой распределения (кривой плотности распределения $p_\xi(x)$), наблюдаемой случайной величины ξ . При большом объеме выборки n и достаточно малом Δ с вероятностью, близкой к 1, можно считать кривую распределения и гистограмму приблизительно совпадающими.

Весьма важным является понятие эмпирической функции распределения.

Эмпирической функцией распределения называется функция $F_n(x)$, определяющая для каждого значения x относительную частоту события $\xi < x$, то есть

$$F_n(x) = \frac{n_x}{n},$$

где n_x – накопленная частота, равная числу вариантов меньших, чем x , n – объем выборки.

Функция $F_n(x)$ является статистическим аналогом функции распределения $F_\xi(x)$ генеральной совокупности. Функцию распределения генеральной совокупности $F_\xi(x)$ в математической статистике называют теоретической функцией распределения. Различие между эмпирической и теоретической функциями распределения состоит в том, что $F_\xi(x)$ определяет вероятность события $\xi < x$, а $F_n(x)$ – относительную частоту этого события, в силу закона больших чисел $F_n(x) \xrightarrow[n \rightarrow \infty]{P} F_\xi(x)$.

Функция $F_n(x)$ обладает всеми свойствами функции распределения, отметим следующие:

1. $0 \leq F_n(x) \leq 1$;
2. $F_n(x)$ – неубывающая функция;
3. $F_n(x) = 0$ при $x \leq x_{\min}$, $F_n(x) = 1$ при $x > x_{\max}$.

Кумулятивная кривая (кумулята) – ломаная, соединяющая точки с координатами (x_i, n_{x_i}) или $\left(x_i, \frac{n_{x_i}}{n}\right)$, где n_{x_i} – накопленные частоты; для интервального ряда n_{x_i} – число вариант меньших значений вариант интервала J_i .

Итак, эмпирическая функция распределения служит для оценки вида теоретической функции распределения случайного признака, полигон и гистограмма – для оценки вида теоретической кривой распределения.

Пример 1.1. В течение суток измеряют напряжение тока в электросети в вольтах. В результате опыта получена выборка X объема $n = 30$:

107	108	110	109	110	111	109	110	111	107
108	109	110	108	107	111	109	111	111	110
109	112	113	110	106	110	109	110	108	112

Построить:

- 1) статистический ряд данной выборки,
- 2) полигон,
- 3) эмпирическую функцию распределения.

Решение.

1) Для построения статистического ряда различные значения признака располагаем в порядке их возрастания и под каждым из этих значений записываем его частоту

Таблица 1

x_i	106	107	108	109	110	111	112	113
n_i	1	3	4	6	8	5	2	1

2) Построим полигон этого распределения.

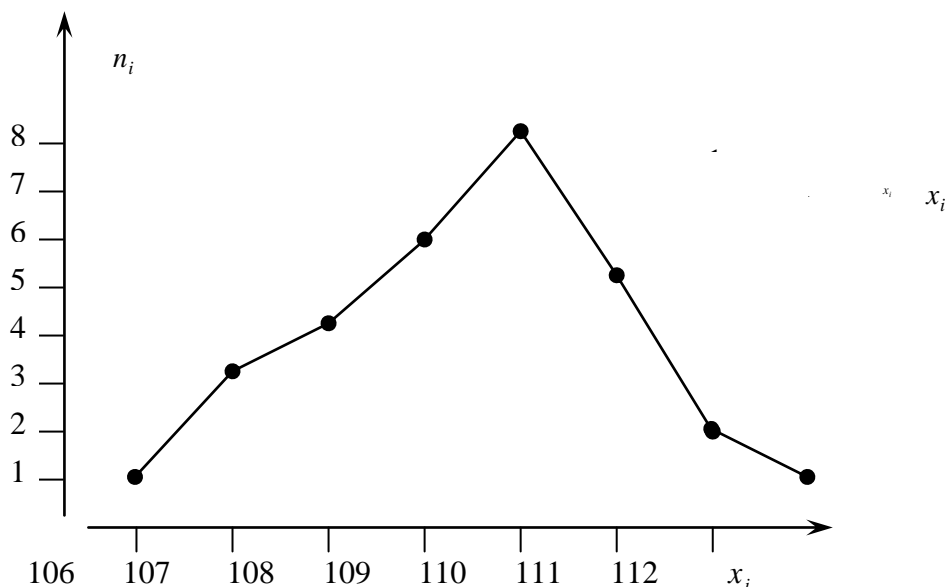


Рис. 1

3) Построим эмпирическую функцию распределения. Наименьшая варианта $x_{\min} = 106$, значит, $F_n(x) = 0$ при $x \leq 106$. Пусть $106 < x \leq 107$, меньше таких x только $x_1 = 106$, это значение наблюдалось 1 раз, следовательно, $F_n(x) = \frac{1}{30}$ при $106 < x \leq 107$. Пусть $107 < x \leq 108$, меньше значения x , удовлетворяющего этому интервалу $x_1 = 106$ и $x_2 = 107$, $n_1 + n_2 = 1 + 3 = 4$, следовательно, $F_n(x) = \frac{4}{30} = \frac{2}{15}$ при $107 < x \leq 108$. Аналогично находим $F_n(x)$ на остальных интервалах. Убеждаемся, что действительно $F_n(x) = 1$ при $x > x_{\max} = 113$.

$$F_n(x) = \begin{cases} 0 & \text{при } x \leq 106, \\ \frac{1}{30} & \text{при } 106 < x \leq 107, \\ \frac{1}{30}(1+3) = \frac{4}{30} = \frac{2}{15} & \text{при } 107 < x \leq 108, \\ \frac{1}{30}(4+4) = \frac{8}{30} = \frac{4}{15} & \text{при } 108 < x \leq 109, \\ \frac{1}{30}(8+6) = \frac{14}{30} = \frac{7}{15} & \text{при } 109 < x \leq 110, \\ \frac{1}{30}(14+8) = \frac{22}{30} = \frac{11}{15} & \text{при } 110 < x \leq 111, \\ \frac{1}{30}(22+5) = \frac{27}{30} & \text{при } 111 < x \leq 112, \\ \frac{1}{30}(27+2) = \frac{29}{30} & \text{при } 112 < x \leq 113, \\ \frac{1}{30}(29+1) = 1 & \text{при } x > 113, \end{cases} = \begin{cases} 0 & \text{при } x \leq 106, \\ \frac{1}{30} & \text{при } 106 < x \leq 107, \\ \frac{2}{15} & \text{при } 107 < x \leq 108, \\ \frac{4}{15} & \text{при } 108 < x \leq 109, \\ \frac{7}{15} & \text{при } 109 < x \leq 110, \\ \frac{11}{15} & \text{при } 110 < x \leq 111, \\ \frac{27}{30} & \text{при } 111 < x \leq 112, \\ \frac{29}{30} & \text{при } 112 < x \leq 113, \\ 1 & \text{при } x > 113. \end{cases}$$

Построим график эмпирической функции распределения.

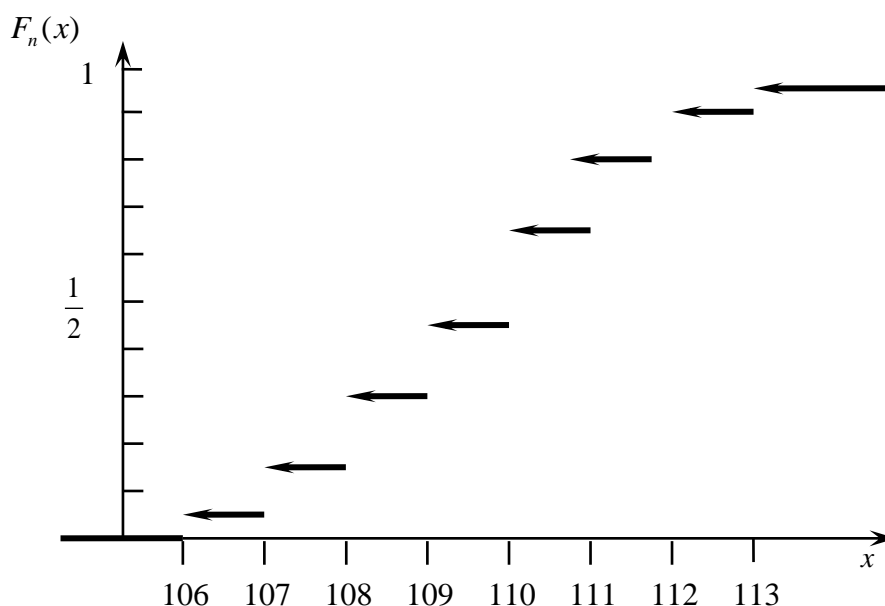


Рис. 2

Замечание. В рассмотренном примере функция $F_n(x)$ построена по дискретному вариационному ряду. Если результаты наблюдений представлены в виде интервального вариационного ряда, то эмпирическую функцию распределения построить в том виде, как это было сделано в примере 1.1, уже не представляется возможности. В данном случае можем определить лишь значения $F_n(x)$ на концах интервалов. Поэтому для графического изображения этой функции целесообразно ее доопределить, соединив точки $(x_i, F(x_i))$, $i = \overline{1, s+1}$, соответствующие концам интервалов, отрезками прямой. В результате полученная ломаная совпадает с кумулятой.

Пример 1.2. В таблице приведены значения промежутков времени τ (в минутах) между вызовами такси в городе Гродно.

0,000	0,000	0,000	0,003	0,011	0,042	0,191	0,405	0,002	0,432
0,517	0,456	0,047	0,162	0,067	0,261	0,168	0,028	0,324	0,125
0,438	0,136	0,019	0,269	0,092	0,653	0,376	0,069	0,702	0,438
0,092	0,134	0,307	0,181	0,327	0,338	0,539	0,250	0,350	0,096
0,125	0,174	0,159	0,091	0,229	0,468	0,283	0,051	0,244	0,261

Построить по этим данным:

- 1) интервальный вариационный ряд;
- 2) полигон;
- 3) гистограмму;
- 4) эмпирическую функцию распределения.

Решение.

1) По условию объем выборки $n = 50$. Определим оптимальную длину частичного интервала с помощью формулы Стерджеса:

$$\Delta = \frac{x_{\max} - x_{\min}}{s} = \frac{x_{\max} - x_{\min}}{1 + 3,322 \lg n} = \frac{0,702 - 0}{1 + 3,322 \lg 50} \approx 0,106.$$

За начало первого интервала можно выбрать величину $x_{нач} = x_{\min} - \Delta/3$ (или $x_{нач} = x_{\min} - \Delta/2$), при этом первый интервал должен покрывать x_{\min} , а последний – x_{\max} . В данном случае $x_{нач} = 0 - 0,106/3 = -0,035$. Сгруппированный ряд представим в виде таблицы 2, где n_i – число вариантов, принадлежащих i -му интервалу.

Таблица 2

i	Интервалы $x_i - x_{i+1}$	Частота n_i	Средины интервалов x_i^*	Относительная частота $w_i = \frac{n_i}{n}$	$\frac{n_i}{n\Delta}$	Накопленная частота n_{x_i}	Накопл. частость $w_i^{нак} = \frac{n_{x_i}}{n}$
1	-0,035 – 0,071	13	0,018	0,26	2,45	13	0,26
2	0,071 – 0,177	12	0,124	0,24	2,26	25	0,50
3	0,177 – 0,283	9	0,230	0,18	1,70	34	0,68
4	0,283 – 0,389	6	0,336	0,12	1,13	40	0,80
5	0,389 – 0,495	6	0,442	0,12	1,13	46	0,92
6	0,495 – 0,601	2	0,548	0,04	0,38	48	0,96
7	0,601 – 0,707	2	0,654	0,04	0,38	50	1
Σ		50	–	1	–	–	–

2) Согласно определению полигон имеет следующий вид:

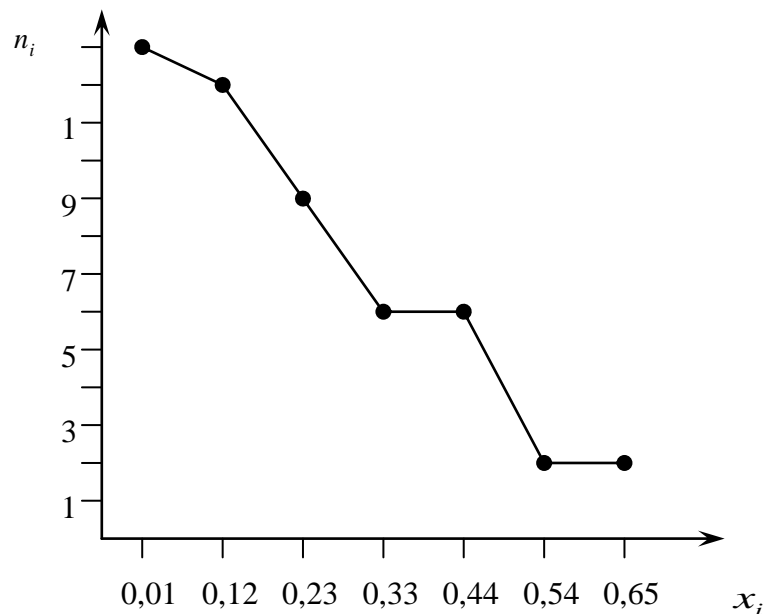


Рис. 3

3) Построим гистограмму.

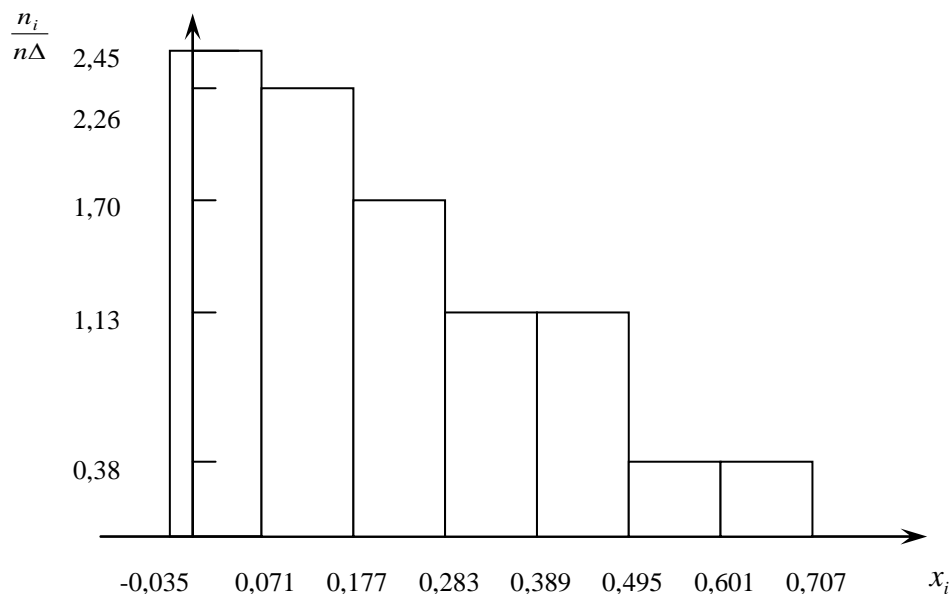


Рис. 11

5) В данном случае исследуется интервальный вариационный ряд для непрерывно распределенного случайного признака τ . Используя результаты, представленные в таблице 2, построим эмпирическую функцию распределения. Очевидно, что для всех $x \in (-\infty, -0,035]$ функция распределения равна нулю. Пусть теперь $x \in (-0,035, 0,071]$. В этом случае для $x \in (-0,035, 0,071)$ число n_x не определено, так как по таблице интервального вариационного ряда невозможно установить сколько выборочных значений случайной величины, принадлежащих этому интервалу, меньше x . Если $x = 0,071$, то $n_x = 13$. Следовательно $F_n(0,071) = 13/50 = 0,26$. Рассуждая аналогично, убеждаемся, что точками в которых значение функции $F_n(x)$ можно определить, являются правые концы интервалов и все точки интервала $(0,707, +\infty)$. Определим теперь значение функции $F_n(x)$ в указанных точках и запишем в виде таблицы.

x	$-\infty < x \leq -0,035$	0,071	0,177	0,283	0,389	0,495	0,601	$0,707 \leq x < +\infty$
$F_n(x)$	0	0,26	0,5	0,68	0,8	0,92	0,96	1

Так как эта таблица определяет функцию $F_n(x)$ не полностью (не для всех x известны ее значения), то при графическом изображении данной функции целесообразно ее доопределить, соединив точки графика, соответствующие концам интервалов, отрезками прямой (рис. 4). В результате график функции $F_n(x)$ будет представлять собой непрерывную линию.

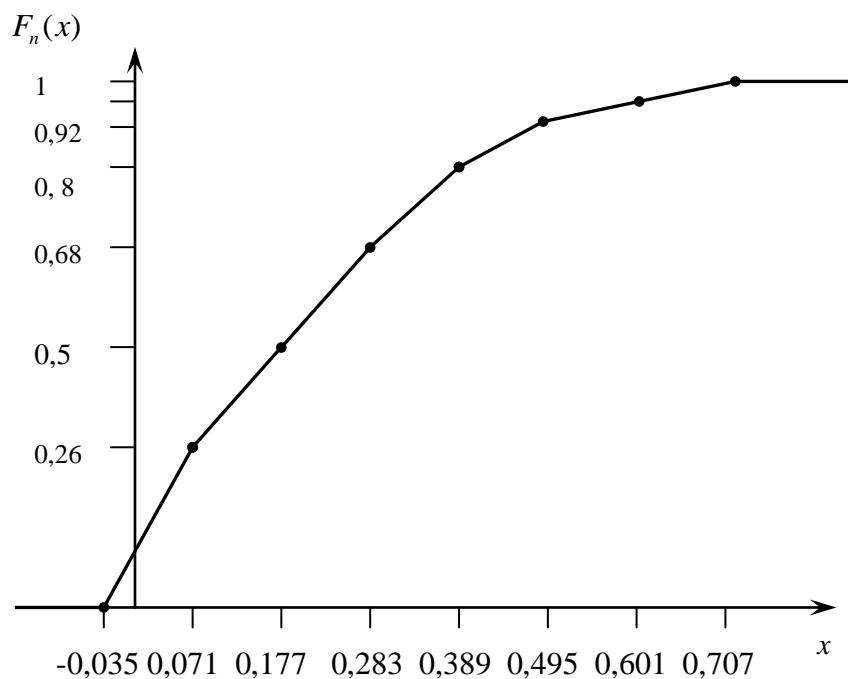


Рис. 4

Анализируя полученные результаты, можем предположить, интервалы времени между поступлениями вызовов такси распределены по показательному закону, так как полученная гистограмма схожа с кривой показательного распределения, графическое изображение эмпирической функции распределения также схоже с функцией показательного распределения.

§2. СРЕДНИЕ ВЕЛИЧИНЫ

Средние величины характеризуют значения признака, вокруг которых концентрируются наблюдения. Наиболее распространенной из средних величин является средняя арифметическая.

Средней арифметической (выборочной средней) вариационного ряда называется величина:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^s x_i n_i,$$

где x_i – варианты дискретного ряда или середины интервалов интервального вариационного ряда, n_i – соответствующие им частоты, $n = \sum_{i=1}^s n_i$.

Для несгруппированного ряда все частоты $n_i = 1$, а $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ есть «невзвешенная» средняя арифметическая.

Пример 2.1. Найти среднее напряжение тока в электросети для примера 1.1.

Решение.

$$\bar{x} = \frac{1}{30} (106 \cdot 1 + 107 \cdot 3 + 108 \cdot 4 + 109 \cdot 6 + 110 \cdot 8 + 111 \cdot 5 + 112 \cdot 2 + 113 \cdot 1) = 109,5.$$

Отметим основные свойства выборочной средней, аналогичные свойствам математического ожидания случайной величины:

1. $\bar{C} = C$, если $C = const$;
2. $\overline{Cx} = C\bar{x}$, $C = const$;
3. $\overline{x + C} = \bar{x} + C$, $C = const$;
4. $\overline{x + y} = \bar{x} + \bar{y}$.

Кроме рассмотренной средней арифметической, в статистическом анализе применяются структурные средние – медиана и мода.

Медианой Me вариационного ряда называется значение признака, приходящееся на середину ранжированного ряда наблюдений.

Для дискретного ряда с нечетным числом членов медиана равна срединному варианту, а для ряда с четным числом членов – полусумме двух срединных вариантов. Например,

для примера 1.1 $Me = \frac{1}{2} (x_{15} + x_{16}) = \frac{1}{2} (110 + 110) = 110.$

Для интервального ряда сначала находят медианный интервал $J_l = [x_l, x_{l+1})$, на который приходится середина ряда. Номер его будет соответствовать интервалу, кумулятивная частота которого равна или превышает половину суммы частот:

$$\sum_{i=1}^l n_i \leq \frac{n}{2} < \sum_{i=1}^{l+1} n_i.$$

В случае выполнения равенства в предыдущей формуле номер медианного интервала равен l , в противном случае – $l + 1$. Медиану вычисляют по формуле

$$Me = x_l + \Delta \frac{\frac{n}{2} - \sum_{i=1}^{l-1} n_i}{n_{Me}}.$$

Здесь l – порядковый номер интервала, где находится медиана, Δ – величина медианного интервала, $\sum_{i=1}^{l-1} n_i$ – накопленная частота домедианного интервала, n_{Me} – частота медианного интервала.

При получении медианы ряд разбивается на 2 равные части. Если ряд разбить на 4 части, то получатся **квартили** (q_1, q_2, q_3), на 10 частей – **децили**. Второй квартиль q_2 равен медиане, а q_1, q_3 вычисляются аналогично медиане с учетом разбиения.

Модой Mo вариационного ряда называется варианта, которой соответствует наибольшая частота. Например, в примере 1.1 $Mo = 110$.

Если распределение интервальное, то определяется модальный интервал $J_l = [x_l, x_{l+1})$, которому соответствует наибольшая частота n_l , мода вычисляется по формуле:

$$Mo = x_l + \Delta \frac{n_l - n_{l-1}}{(n_l - n_{l-1}) + (n_l - n_{l+1})},$$

где n_{l-1}, n_{l+1} – частоты предмодального и послемодального интервалов.

Пример 2.2. Обследование качества пряжи дало следующие результаты, представленные в таблице. Найти моду и медиану этого распределения.

Прочность нити, г	Частота	Накопленная частота
120 – 140	1	1
140 – 160	6	7
160 – 180	19	26
180 – 200	58	84
200 – 220	53	137
220 – 240	24	161
240 – 260	16	177
260 – 280	3	180
Σ	180	

Решение. Так как наибольшая частота $m_{Mo} = 58$ отвечает интервалу 180–200, то $x_l = 180$, $m_{l-1} = 19$, $m_{l+1} = 53$, $\Delta = 20$. Мода равна:

$$Mo = 180 + 20 \frac{58 - 19}{(58 - 19) + (58 - 53)} = 197,73.$$

Определим номер медианного интервала:

$$\sum_{i=1}^l n_i \leq \frac{180}{2} < \sum_{i=1}^{l+1} n_i, \quad \sum_{i=1}^4 n_i \leq 90 < \sum_{i=1}^5 n_i, \quad 84 \leq \frac{180}{2} < 137.$$

Следовательно, номер медианного интервала 5, а сам интервал 200–220. Тогда получаем

$$Me = 200 + 20 \frac{90 - 84}{53} = 202,26.$$

§3. ПОКАЗАТЕЛИ ВАРИАЦИИ, МОМЕНТЫ

Средние величины не отражают изменчивости (вариации) значений признака.

Простейшим показателем вариации является *вариационный размах* $R = x_{\max} - x_{\min}$.

Наибольший интерес представляет мера рассеяния наблюдений вокруг средней арифметической – дисперсия.

Дисперсией (выборочной дисперсией) вариационного ряда называется величина

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

При расчете дисперсии и других числовых характеристик интервальных рядов в качестве x_i также используют середины интервалов. Часто для вычисления дисперсии используют упрощенную формулу:

$$D_B = \overline{x^2} - \bar{x}^2 ,$$

где $\overline{x^2} = \frac{1}{n} \sum_{i=1}^s x_i^2 n_i .$

Если признак ξ измеряется в метрах, то, очевидно, его дисперсия – в метрах квадратных. Желательно в качестве меры вариации иметь характеристику, выраженную в тех же единицах, что и значения признака. Такой характеристикой является *среднее квадратическое отклонение*:

$$\sigma_B = \sqrt{D_B} .$$

Отношение среднего квадратического отклонения к средней величине признака, выраженное в процентах, называют *коэффициентом вариации*:

$$v = \frac{\sigma_B}{\bar{x}} \cdot 100 \% .$$

Если коэффициент вариации признака, принимающего только положительные значения, высок (например, более 100 %), то, как правило, это свидетельствует о неоднородности значений признака.

Дисперсия обладает следующими свойствами, аналогичными свойствам дисперсии случайной величины. Отметим следующие:

1. $D_B(C) = 0$, $C = const$;
2. $D_B(Cx) = C^2 D_B(x)$, $C = const$;
3. $D_B(C + x) = D_B(x)$, $C = const$;
4. Если выборки X и Y независимы, то $D_B(x + y) = D_B(x) + D_B(y)$.

Пример 3.1. Вычислить дисперсию, среднее квадратическое отклонение и коэффициент вариации напряжения тока в электросети для примера 1.1.

Решение.

Вычислим дисперсию по упрощенной формуле $D_B = \overline{x^2} - \bar{x}^2$.

$$\overline{x^2} = \frac{1}{30}(106^2 + 107^2 \cdot 3 + 108^2 \cdot 4 + 109^2 \cdot 6 + 110^2 \cdot 8 + 111^2 \cdot 5 + 112^2 \cdot 2 + 113^2) = 119929.$$

$$D_B = 11992,9 - 109,5^2 = 2,65.$$

Среднее квадратическое отклонение $\sigma_B = \sqrt{2,65} = 1,63$. Вариация $v = \frac{1,63}{109,5} \cdot 100 \% = 1,49 \%$.

Средняя арифметическая и дисперсия вариационного ряда являются частными случаями более общего понятия – выборочных (эмпирических) моментов.

Начальный момент \tilde{v}_k k -го порядка вариационного ряда определяется по формуле:

$$\tilde{v}_k = \frac{1}{n} \sum_{i=1}^s x_i^k n_i.$$

Центральный момент $\tilde{\mu}_k$ k -го порядка вариационного ряда определяется по формуле:

$$\tilde{\mu}_k = \frac{1}{n} \sum_{i=1}^s (x_i - \bar{x})^k n_i.$$

Очевидно, что $\tilde{v}_1 = \bar{x}$, $\tilde{\mu}_1 = 0$, $\tilde{\mu}_2 = D_B$.

Коэффициентом асимметрии вариационного ряда называется число

$$\tilde{A}_s = \frac{\tilde{\mu}_3}{\sigma_B^3} = \frac{1}{n \sigma_B^3} \sum_{i=1}^s (x_i - \bar{x})^3 n_i.$$

Если $\tilde{A}_s = 0$, то распределение имеет симметричную форму, то есть варианты, равноудаленные от \bar{x} , имеют одинаковую частоту. При $\tilde{A}_s > 0$ ($\tilde{A}_s < 0$) говорят о положительной (отрицательной) или правосторонней (левосторонней) асимметрии.

Эксцессом вариационного ряда называется число

$$\tilde{E}_x = \frac{\tilde{\mu}_4}{\sigma_B^4} - 3 = \frac{1}{n \sigma_B^4} \sum_{i=1}^s (x_i - \bar{x})^4 n_i - 3.$$

Эксцесс является показателем крутости кривой распределения вариационного ряда по сравнению с нормальным распределением, дисперсия которого равна σ_B^2 . При $\tilde{E}_x = 0$ распределение нормальное. Если $\tilde{E}_x > 0$, то кривая распределения имеет более острую вершину, чем при нормальном распределении, если $\tilde{E}_x < 0$ – более плоскую.

Пример 3.2. Вычислить коэффициент асимметрии и эксцесс распределения напряжения тока в электросети для примера 1.1.

Решение. Сначала находим центральные моменты третьего и четвертого порядков:

$$\tilde{\mu}_3 = \frac{1}{30} \sum_{i=1}^8 (x_i - 109,5)^3 n_i = -0,4; \quad \tilde{\mu}_4 = \frac{1}{30} \sum_{i=1}^8 (x_i - 109,5)^4 n_i = 18,06.$$

Тогда $\tilde{A}_s = \frac{\tilde{\mu}_3}{\sigma_B^3} = \frac{-0,4}{1,63^3} = -0,09$; $\tilde{E}_x = \frac{\tilde{\mu}_4}{\sigma_B^4} - 3 = \frac{18,06}{1,63^4} - 3 = -0,44$. Поскольку найденные показатели близки к нулю, то можно сделать вывод, что рассматриваемое в примере 1.1 распределение по асимметрии и крутости приближается к нормальной кривой.

Вычисление выборочной средней и дисперсии можно упростить, если использовать не первоначальные варианты x_i , а новые варианты

$$u_i = \frac{x_i - C}{k},$$

где C и k – специально подобранные постоянные. Тогда согласно свойствам средней арифметической и дисперсии

$$\bar{u} = \overline{\left(\frac{x-C}{k} \right)} = \frac{\bar{x}-C}{k}, \text{ следовательно } \bar{x} = k\bar{u} + C,$$

$$D_B(u) = D_B\left(\frac{x-C}{k}\right) = \frac{D_B(x)}{k^2}, \text{ следовательно } D_B(x) = k^2 D_B(u).$$

Данный метод дает существенное упрощение в случае больших значений x_i . В качестве постоянной k рекомендуется брать величину интервала по x , а в качестве C – варианту, имеющую наибольшую частоту (середину интервала, имеющего наибольшую частоту).