Министерство образования РФ

Санкт-Петербургский государственный
электротехнический университет “ЛЭТИ”

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

**Методические указания
к индивидуальным домашним заданиям
по дисциплине
“ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА”**

Санкт-Петербург
2004

Министерство образования РФ

Санкт-Петербургский государственный
электротехнический университет “ЛЭТИ”

ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

**Методические указания
к индивидуальным домашним заданиям
по дисциплине
“ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ
СТАТИСТИКА”**

Санкт-Петербург
Издательство СПбГЭТУ “ЛЭТИ”
2004

УДК 519.22

Элементы математической статистики: Методические указания к индивидуальным домашним заданиям по дисциплине “Теория вероятностей и математическая статистика” / Сост. А.В.Осетров, Н.А.Смурова, В.К.Шомесова; под ред. Н.А,Смуровой; СПбГЭТУ. СПб., 2004. 32 с.

Включают следующие элементы математической статистики: статистическую обработку экспериментальных данных и статистическое оценивание результатов измерений (точечное и интервальное), а также проверку гипотезы о законе распределения случайной величины по экспериментальным данным с использованием критерия согласия Пирсона. После краткого изложения теории рассмотрены типовые задания с подробными решениями.

Предназначены для студентов второго курса 4-го семестра ФПБЭИ, выполняющих домашние задания по курсу “Теория вероятностей и математическая статистика” по разделу “Математическая статистика”.

Утверждено
редакционно-издательским советом университета
в качестве методических указаний

1. Статистическая обработка экспериментальных данных и статистическое оценивание результатов измерений

1.1 Выборочный метод и основные характеристики выборки

Применение методов математической статистики к обработке экспериментальных данных оказывается возможным благодаря тому, что производство измерений полностью соответствует основной схеме статистических испытаний, называемой *выборочным методом*. Идея выборочного метода состоит в следующем. Имеется большая совокупность N объектов, называемая *генеральной совокупностью*. Из нее случайным образом извлекается n объектов ($n << N$), образующих *выборку объема* n. Эти n объектов подвергаются детальному исследованию, по результатом которого требуется определить те или иные свойства генеральной совокупности. *Выборочный метод* и состоит в определении характеристик генеральной совокупности на основании характеристик выборки.

Математически задача ставится следующим образом: имеется случайная величина X (возможные значения которой считаются генеральной совокупностью), и в результате n *независимых* испытаний получены n ее значений $(x_1, x_2, \ldots x_n)$, образующих выборку объема n. По этой выборке нужно определить те или иные характеристики случайной величины X (ее числовые характеристики или закон ее распределения).

Из случайного характера выборки следует, что любое суждение о генеральной совокупности (случайной величине X) на основании выборки, вообще говоря, является случайным. Но, пользуясь законами теории вероятностей, мы можем оценить степень достоверности того или иного суждения. На этом и основана математическая статистика.

Выборку (совокупность $x_1, x_2, \ldots x_n$ – n значений случайной величины X) принято записывать в таблицу. Различают две основные формы таких таблиц: вариационный ряд и сгруппированный вариационный ряд.

Вариационный ряд (табл. 1.1) представляет собой таблицу измеренных значений случайной величины x_i ($i = 1,2,\ldots,l$), расположенных в порядке их возрастания ($x_1 < x_2 < \ldots < x_l$), и соответствующих им частот m_i (числа опытов, в которых случайная величина X приняла значение x_i).

Сумма всех частот равна объему выборки $\sum_{i=1}^{l} m_i = n$.

Таблица 1.1

x_i	x_1	x_2	$\ldots$	x_l	$\sum$
m_i	m_1	m_2	$\ldots$	m_l	n

Из вариационного ряда легко составить *ряд распределения выборки* (табл. 1.2) – таблицу измеренных значений случайной величины x_i ($i = 1, 2, \ldots, l$), расположенных в порядке их возрастания, и соответствующих им относительных частот $p_i^* = m_i / n$. Сумма всех относительных частот равна единице $\sum_{i=1}^{l} p_i^* = 1$.

Таблица 1.2

x_i	x_1	x_2	$\ldots$	x_l	$\sum$
p_i^*	p_1^*	p_2^*	$\ldots$	p_l^*	1

В целях наглядности пользуются графическим изображением ряда распределения выборки – *полигоном относительных частот*, для построения которого по оси абсцисс откладывают значения x_i, а по оси ординат – соответствующие им относительные частоты p_i^*, и полученные точки соединяют отрезками прямых (рис. 1.1).

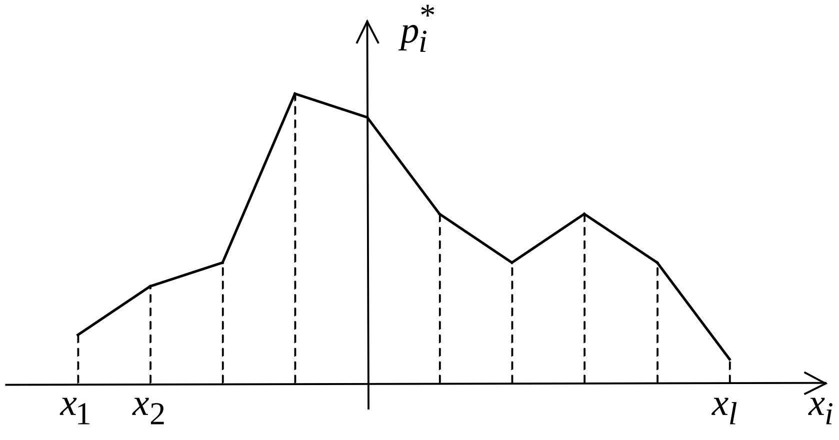

Рис. 1.1

Полигон относительных частот является аналогом многоугольника распределения дискретной случайной величины.

Имея вариационный ряд, можно построить график *эмпирической функции распределения* $F_n^*(x)$, которая определяется как относительная частота события $X < x$ в данной выборке: $F_n^*(x) = \nu_n(X < x)$. Эмпирическая функция распределения представляет собой по существу функцию распределения дискретной конечнозначной случайной величины X_n^*, принимающей с вероятностями p_i^* выборочные значения x_i. При достаточном объеме выборки n эмпирическая функция распределения $F_n^*(x)$ дает представление о теоретической функции распределения $F(x)$ случайной величины X. Для практического построения графика $F_n^*(x)$ удобно исходный вариационный ряд дополнить двумя графами – графой *накопленных частот* $N_i = m_1 + m_2 + \ldots + m_i$ и графой накопленных относительных частот N_i/n (табл. 1.3). Для получения накопленной частоты N_i надо к частоте вариационного ряда, стоящей слева от образуемой, прибавить накопленную частоту, стоящую непосредственно над образуемой: $N_i = N_{i-1} + m_i$. Последняя из накопленных частот равна объему выборки ($N_l = n$), последняя из накопленных относительных частот равна единице ($N_l/n = 1$). Это служит проверкой правильности составления табл. 1.3.

Таблица 1.3

x_i	m_i	$N_i = m_1 + m_2 + \ldots + m_i$	N_i/n
x_1	m_1	$N_1 = m_1$	N_1/n
x_2	m_2	$N_2 = N_1 + m_2$	N_2/n
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_l	m_l	$N_l = N_{l-1} + m_l = n$	$N_l/n = 1$
$\sum$	n		

Графы выборочных значений x_i и соответствующих им накопленных относительных частот N_i/n и служат для построения графика эмпирической функции распределения $F_n^*(x)$, которая выражается следующим образом

$$F_n^*(x) = \begin{cases} 0, & \text{при} \quad x \le x_1; \\ N_1/n, & \text{при} \quad x_1 < x \le x_2; \\ \dots\dots\dots\dots\dots\dots & \\ N_i/n, & \text{при} \quad x_i < x \le x_{i+1}; \\ \dots\dots\dots\dots\dots\dots & \\ 1, & \text{при} \quad x > x_l. \end{cases} \tag{1.1}$$

Функция $F_n^*(x)$ представляет собой разрывную ступенчатую функцию (рис. 1.2).

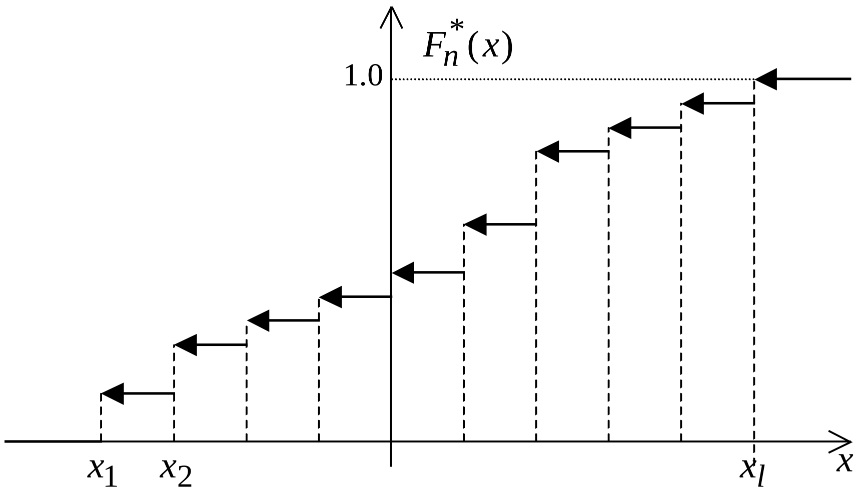

Рис. 1.2

Если объем выборки значителен и случайная величина X по своей природе непрерывная, то составляют так называемый *сгруппированный вариационный ряд* (табл. 1.4). Для этого весь интервал, в котором заключены измеренные значения случайной величины X, разбивают на ряд частичных интервалов $x_i \div x_{i+1}$ $(i = 1,2,\dots,l)$, для каждого из которых подсчитывают частоты m_i (числа опытов, в которые случайная величина X попала в данный интервал).

Таблица 1.4

$x_i \div x_{i+1}$	$x_1 \div x_2$	$x_2 \div x_3$	…	$x_l \div x_{l+1}$	$\sum$
m_i	m_1	m_2	…	m_l	n

Часто в результате измерений получается сразу сгруппированный вариационный ряд.

Для наглядной характеристики распределения выборки в этом случае строят так называемую *гистограмму выборки* – аналог кривой распределения непрерывной случайной величины. Для построения гистограммы предварительно составляют вспомогательную таблицу (табл. 1.5), содержащую следующие графы: интервал $x_i \div x_{i+1}$, ширина интервала $h_i = x_{i+1} - x_i$, относительная частота $p_i^* = m_i / n$, плотность относительной частоты p_i^* / h_i.

Таблица 1.5

$x_i \div x_{i+1}$	h_i	p_i^*	p_i^* / h_i
$x_1 \div x_2$	h_1	p_1^*	p_1^* / h_1
$x_2 \div x_3$	h_2	p_2^*	p_2^* / h_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_l \div x_{l+1}$	h_l	p_l^*	p_l^* / h_l

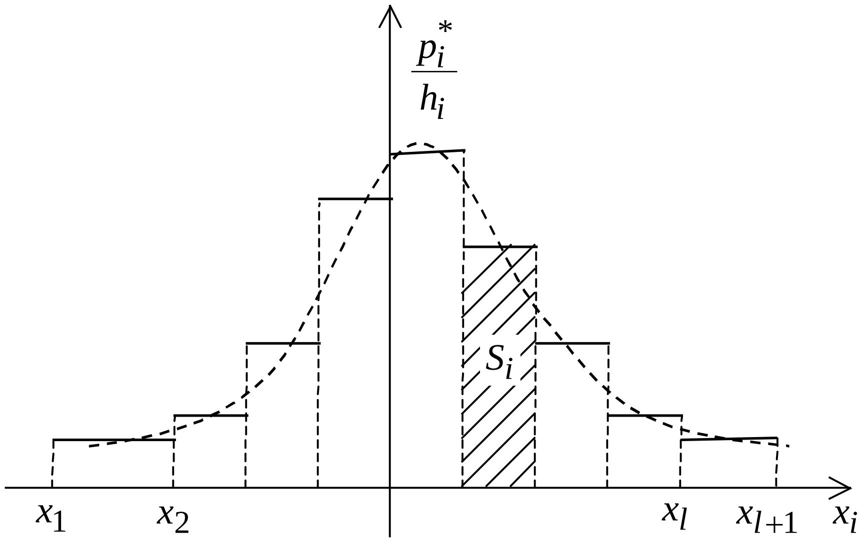

Рис. 1.3

Гистограмма строится следующим образом (рис. 1.3): по оси абсцисс откладываются граничные значения интервалов $(x_1, x_2, \ldots, x_{l+1})$, затем на каждом интервале, как на основании, строят прямоугольник высотой, равной плотности относительной частоты на данном интервале. Площадь каждого частичного прямоугольника равна соответствующей относительной частоте: $S_i = p_i^*$. Вся площадь, ограниченная гистограммой выборки, равна единице: $S = \sum_{i=1}^{l} S_i = \sum_{i=1}^{l} p_i^* = 1$.

Если объем выборки достаточно большой, то гистограмма выборки (если ее сгладить) дает примерное представление о кривой распределения случайной величины X (пунктир на рис. 1.3).

Для построения графика эмпирической функции распределения $F_n^*(x)$ по сгруппированному вариационному ряду предварительно составляется таблица, аналогичная табл. 1.3 для вариационного ряда, содержащая графы накопленных частот N_i и накопленных относительных частот N_i/n в каждом интервале (табл. 1.6).

Таблица 1.6

$x_i \div x_{i+1}$	m_i	$N_i = m_1 + m_2 + \ldots + m_i$	N_i/n
$x_1 \div x_2$	m_1	$N_1 = m_1$	N_1/n
$x_2 \div x_3$	m_2	$N_2 = N_1 + m_2$	N_2/n
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_l \div x_{l+1}$	m_l	$N_l = N_{l-1} + m_l = n$	N_l/n
$\sum$	n		

График эмпирической функции распределения $F_n^*(x)$ строят приближенно по крайним точкам интервалов, полагая

$$\left.\begin{array}{l} F_n^*(x) = 0 \quad \text{при } x \le x_1; \\ F_n^*(x_2) = N_1/n; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ F_n^*(x_{i+1}) = N_i/n; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ F_n^*(x) = 1 \quad \text{при } x \ge x_{l+1}; \end{array}\right\} \tag{1.2}$$

Полученные точки соединяют ломаной линией (рис. 1.4).

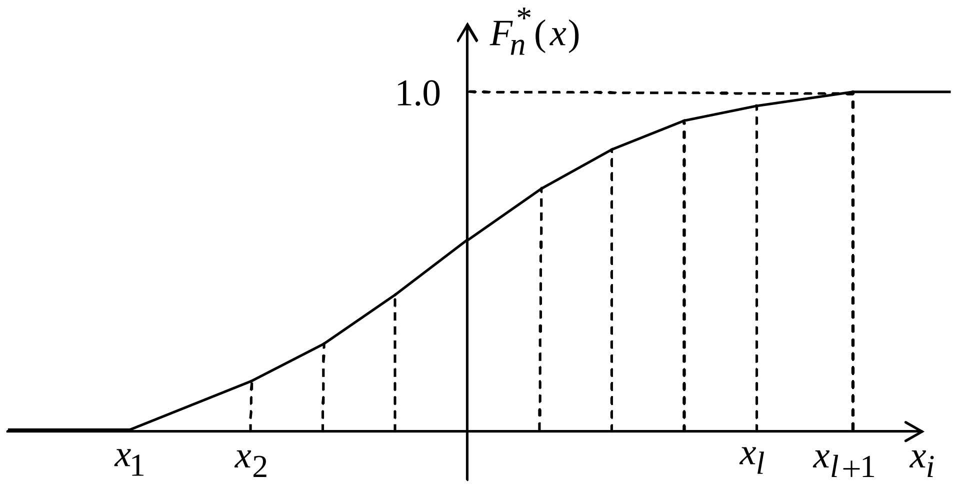

Рис. 1.4

Подобно тому, как для описания случайной величины вводят в рассмотрение различные числовые характеристики (начальные и центральные моменты), так и для описания выборки рассматривают аналогичные числовые характеристики выборки, называемые *выборочными моментами* (или моментами эмпирического распределения). В отличие от аналогичных числовых характеристик случайной величины X, называемых числовыми характеристиками теоретического распределения, числовые характеристики выборки снабжаются значком “*” наверху справа.

В зависимости от способа задания выборки (вариационный ряд или сгруппированный вариационный ряд) получаются различные формулы для вычисления выборочных моментов.

В случае **вариационного ряда** (см. табл. 1.1) выборочный начальный момент k-го порядка вычисляется по формуле:

$$m_k^*[\mathrm{X}] = \frac{1}{n}\sum_{i=1}^{l} x_i^k m_i \,. \tag{1.3}$$

В частности, выборочный начальный момент первого порядка, называемый *выборочным средним*,

$$\overline{x} = \mathrm{m}_1^*[X] = \frac{1}{n}\sum_{i=1}^{l} x_i\, m_i \,. \tag{1.4}$$

Выборочный центральный момент k-го порядка вычисляется по формуле:

$$\mu_k^*[X] = \frac{1}{n}\sum_{i=1}^{l}(x_i - \bar{x})^k m_i . \qquad (1.5)$$

В частности, *выборочная дисперсия* (второй центральный выборочный момент)

$$D^*[X] = D_X^* = \mu_2^*[X] = \frac{1}{n}\sum_{i=1}^{l}(x_i - \bar{x})^2 m_i . \qquad (1.6)$$

При практическом вычислении выборочной дисперсии удобно пользоваться следующей формулой:

$$D_X^* = \frac{1}{n}\sum_{i=1}^{l} x_i^2 m_i - \bar{x}^2 , \qquad (1.7)$$

аналогом которой является известная формула теории вероятностей: $D^*[X] = M[X^2] - M^2[X]$.

Положительное значение квадратного корня из выборочной дисперсии дает *выборочное среднеквадратическое отклонение*:

$$\sigma_X^* = \sqrt{D_X^*} . \qquad (1.8)$$

В случае **сгруппированного вариационного ряда** (см. табл. 1.4) для нахождения выборочных моментов его заменяют вариационным рядом (табл. 1.7), в котором частота i-го интервала m_i соответствует середине i-го интервала: $x_{i\,\text{ср}} = (x_i + x_{i+1})/2$, $i = 1, 2, \ldots, l$.

Таблица 1.7

$x_{i\,\text{ср}}$	$x_{1\text{ср}}$	$x_{2\,\text{ср}}$	…	$x_{l\,\text{ср}}$	$\sum$
m_i	m_1	m_2	…	m_l	n

Все выборочные моменты вычисляются по соответствующим формулам (1.3)–(1.8) для вариационного ряда, в частности

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{l} x_{i\,\text{ср}} m_i ; \qquad (1.9)$$

$$D_X^* = \frac{1}{n}\sum_{i=1}^{l}(x_{i\,\text{ср}} - \bar{x})^2 m_i = \frac{1}{n}\sum_{i=1}^{l} x_{i\,\text{ср}}^2 m_i - \bar{x}^2 . \qquad (1.10)$$

Для уменьшения ошибок, вносимых группировкой, применяют так называемые *поправки Шеппарда*. Если все интервалы имеют ширину h

($h_i = h = \text{const}$), то основные выборочные моменты с учетом поправок Шеппарда находятся по формулам:

$$\left.\begin{aligned} &\bar{x}_{\text{ш}} = \bar{x}; \\ &\mathrm{D}^*_{\mathrm{X}\,\text{ш}} = \mathrm{D}^*_{\mathrm{X}} - \frac{h^2}{12}. \end{aligned}\right\} \tag{1.11}$$

При выполнении условия $h < 0.5\sigma^*_X$ поправками Шеппарда можно пренебречь (так как при этом $\sigma^*_X - \sigma^*_{X\,\text{ш}} < 0.01\sigma^*_X$). При практическом вычислении выборочных моментов по приведенным выше формулам удобно использовать следующие два свойства выборочных моментов.

Свойство 1 (изменение начала отсчета). Если все выборочные значения x_i увеличить или уменьшить на одно и то же постоянное число x_0, то выборочное среднее увеличится или уменьшится на это же число x_0, а все центральные выборочные моменты не изменятся.

Это свойство используется на практике при вычислении выборочных моментов так называемым *методом ложного нуля*, который состоит в следующем. Если разброс выборочных значений незначителен, то из каждого значения вычитают величину, близкую к среднему (ложный нуль), что равносильно изменению начала отсчета. В результате получается вариационный ряд, который значительно проще обрабатывать, чем исходный.

Свойство 2 (изменение масштаба). Если все выборочные значения x_i умножить на одно и то же постоянное число c, то выборочный момент k-го порядка (начальный или центральный) умножится на c^k.

На основании этого свойства можно при вычислении выборочных моментов выражать выборочные значения в рабочих единицах, что в значительной степени упрощает вычисления. Для того, чтобы от моментов, выраженных в рабочих единицах, перейти к моментам, выраженным в единицах измерения, достаточно разделить момент k-го порядка на c^k.

На практике обычно оба свойства используются одновременно.

1.2. Статистические оценки параметров теоретического распределения

Одной из задач математической статистики является оценивание неизвестных параметров теоретического распределения (случайной величины X) на основании выборки. Приближенное значение параметра θ теоретического распределения, полученное из выборки, называется *статистической оценкой параметра* θ (или *статистикой*). *Точечной* называют статистическую оценку $\tilde{\theta}$, которая определяется одним числом (одной случайной величиной). Для данной конкретной выборки $\tilde{\theta} = \tilde{\theta}(x_1, x_2, \dots, x_n)$ – число, но от выборки к выборке оно меняется, так что в общем случае $\tilde{\theta} = \tilde{\theta}(X_1, X_2, \dots, X_n)$ – случайная величина, где $X_1, X_2, \dots, X_n$ – результаты 1-го, 2-го, ... , n-го измерений случайной величины X. В литературе по математической статистике принято случайную величину $\tilde{\theta}(X_1, X_2, \dots, X_n)$ и ее численное значение для данной выборки $\tilde{\theta}(x_1, x_2, \dots, x_n)$ обозначать одним и тем же символом $\tilde{\theta}$.

Итак, найти точечную статистическую оценку параметра θ теоретического распределения – это значит найти случайную величину $\tilde{\theta}(X_1, X_2, \dots, X_n)$, являющуюся функцией случайных величин $X_1, X_2, \dots, X_n$, которая для каждой конкретной выборки $(x_1, x_2, \dots, x_n)$ дает приближенное значение $\tilde{\theta}(x_1, x_2, \dots, x_n)$ оцениваемого параметра θ.

К точечным статистическим оценкам обычно предъявляют следующие требования:

1. *состоятельность* (сходимость по вероятности при $n \to \infty$ к оцениваемому параметру);
2. *несмещенность* (равенство математического ожидания оценки оцениваемому параметру);
3. *эффективность* (наименьшая возможная дисперсия оценки при данном объеме выборки) или асимптотическая эффективность (наименьшая возможная дисперсия при $n \to \infty$).

В математической статистике доказывается, что в качестве точечной статистической оценки $\tilde{\mathrm{M}}$ математического ожидания M теоретического распределения следует взять выборочное среднее $\overline{x}$, т. е. $\widetilde{\mathrm{M}} = \overline{x}$. Эта оценка является состоятельной, несмещенной, а для нормального закона также и эффективной.

Если в качестве точечной статистической оценки дисперсии D теоретического распределения взять выборочную дисперсию D_X^*, то можно доказать, что $\mathrm{M}[\mathrm{D}_X^*(X_1, X_2, \ldots, X_n)] = \dfrac{n-1}{n}\mathrm{D} \neq \mathrm{D}$, т. е. оценка D_X^* является *смещенной* оценкой дисперсии D. Чтобы ликвидировать этот недостаток, рассматривают *исправленную выборочную дисперсию*

$$\widetilde{\mathrm{D}} = \frac{n}{n-1}\mathrm{D}_X^*, \tag{1.12}$$

которая является состоятельной, несмещенной, а для нормального закона асимптотически эффективной оценкой дисперсии.

В случае обработки *сгруппированного вариационного ряда* выборочная дисперсия с учетом поправки Шеппарда (1.11) уже является несмещенной оценкой дисперсии (можно доказать), т. е. в этом случае

$$\widetilde{\mathrm{D}} = \mathrm{D}_{X\,\text{ш}}^* = \mathrm{D}_X^* - \frac{h^2}{12}, \tag{1.13}$$

и множитель $n/(n-1)$ вводить не надо.

Положительное значение квадратного корня из несмещенной оценки дисперсии дает несмещенную оценку *среднеквадратического отклонения*:

$$\widetilde{\sigma} = \sqrt{\widetilde{\mathrm{D}}}. \tag{1.14}$$

Кроме точечных статистических оценок, которые в ряде случаев могут сильно отличаться от оцениваемого параметра, в математической статистике рассматривают также *интервальные статистические оценки*. *Интервальной* называют статистическую оценку неизвестного параметра θ, которая определяется двумя числами $\widetilde{\theta}_1$ и $\widetilde{\theta}_2$ – концами *доверительного интервала* $\left(\widetilde{\theta}_1; \widetilde{\theta}_2\right)$. При определении интервальной статистической оценки необходимо задать *надежность* (*доверительную вероятность*) γ этой оценки. Обычно берут $\gamma = 0.9\ldots0.99$, реже 0.8 или 0.999. Тогда концы доверительного интервала $\widetilde{\theta}_1$ и $\widetilde{\theta}_2$ определяются из следующего условия:

$$P\left[\tilde{\theta}_1(X_1, X_2, \ldots, X_n) < \theta < \tilde{\theta}_2(X_1, X_2, \ldots, X_n)\right] = \gamma. \qquad (1.15)$$

Условие (1.15) следует понимать так: вероятность того, что интервал $\left(\tilde{\theta}_1; \tilde{\theta}_2\right)$ со случайными концами, меняющимися от выборки к выборке, покроет оцениваемый параметр θ, равна γ. Этот интервал и называют доверительным интервалом и обозначают $I_\gamma^{(\theta)} = \left(\tilde{\theta}_1; \tilde{\theta}_2\right)$. Границы интервала $\tilde{\theta}_1$ и $\tilde{\theta}_2$ называют *доверительными границами* (нижней и верхней соответственно). Для нормально распределенной случайной величины X в математической статистике выведены формулы, позволяющие находить на основании выборки по заданной надежности γ доверительные интервалы $I_\gamma^{(\mathrm{M})}$ и $I_\gamma^{(\sigma)}$ для основных параметров теоретического распределения – математического ожидания M и среднеквадратического отклонения σ.

Для выборок *малого объема* ($n-1 \le 30$) эти формулы имеют вид:

$$I_\gamma^{(\mathrm{M})} = \left(\bar{x} - t_{\gamma,n-1}\frac{\tilde{\sigma}}{\sqrt{n}}; \bar{x} + t_{\gamma,n-1}\frac{\tilde{\sigma}}{\sqrt{n}}\right), \qquad (1.16)$$

где $t_{\gamma,n-1}$ – коэффициенты Стьюдента (их значения для разных γ и n приведены в Приложении 1), определяемые из уравнения: $2 \cdot \int\limits_0^{t_{\gamma,n-1}} S_{n-1}(t)dt = \gamma$, в котором $S_{n-1}(t)$ – плотность вероятности случайной величины, распределенной по закону распределения Стьюдента с $n-1$ степенями свободы;

$$I_\gamma^{(\sigma)} = \left(\frac{\sqrt{n-1}}{\chi_2}\tilde{\sigma}; \frac{\sqrt{n-1}}{\chi_1}\tilde{\sigma}\right), \qquad (1.17)$$

где χ_1^2 и χ_2^2 определяются по таблицам распределения хи-квадрат (см. Приложение 2) для $r = n-1$ степеней свободы для вероятностей, соответственно равных

$$p_1 = (1+\gamma)/2 \text{ и } p_2 = (1-\gamma)/2. \qquad (1.18)$$

Для выборок *большого объема* ($n-1 > 30$)

$$I_\gamma^{(\mathrm{M})} = \left(\bar{x} - t_\gamma\frac{\tilde{\sigma}}{\sqrt{n}}; \bar{x} + t_\gamma\frac{\tilde{\sigma}}{\sqrt{n}}\right), \qquad (1.19)$$

$$I_\gamma^{(\sigma)} = \left(\frac{\tilde{\sigma}}{\sqrt{1 + t_\gamma \sqrt{2/n}}}; \frac{\tilde{\sigma}}{\sqrt{1 - t_\gamma \sqrt{2/n}}} \right), \tag{1.20}$$

где $\Phi_L(t_\gamma) = \gamma$; $\Phi_L(x) = \frac{2}{\sqrt{2\pi}} \int_0^x \exp\left(-\frac{t^2}{2} \right) dt$ – функция Лапласа.

Значение t_γ находится по таблице нормального распределения для заданного γ (последняя строка таблицы распределения Стьюдента в Приложении 1).

Иногда приводятся другие формулы для $I_\gamma^{(\sigma)}$, имеющие в виду использование других таблиц, но в данном пособии они не рассматриваются. Приведенные выше формулы (1.16), (1.19), (1.17), и (1.20) для доверительных интервалов $I_\gamma^{(\mathrm{M})}$ и $I_\gamma^{(\sigma)}$ широко используются при измерениях для оценки *истинного значения* измеряемой величины и *точности измерений*. Если производится n независимых равноточных измерений некоторой физической величины, истинное значение которой неизвестно, причем систематические ошибки измерений отсутствуют, а случайные (в силу центральной предельной теоремы) распределены нормально с нулевым математическим ожиданием и неизвестным нам средеквадратическим отклонением σ, характеризующим точность измерений, то мы можем рассматривать результаты отдельных измерений как случайные величины $X_1, X_2, \ldots, X_n$, которые независимы (измерения независимы), имеют одно и то же математическое ожидание (истинное значение измеряемой величины), одинаковые среднеквадратические отклонения σ (измерения равноточны) и распределены нормально. Таким образом, все предположения, при которых были выведены приведенные выше формулы для $I_\gamma^{(\mathrm{M})}$ и $I_\gamma^{(\sigma)}$, выполняются, следовательно, эти формулы можно применять для получения интервальной оценки истинного значения измеряемой величины (доверительный интервал для математического ожидания) и интервальной оценки точности измерений (доверительный интервал для среднеквадратического отклонения), что и используется в данной работе.

1.3. Порядок выполнения задания

Задание состоит в следующем.

По результатам n независимых равноточных измерений некоторой физической величины, которые сведены в таблицу, представляющую собой либо а) вариационный ряд (для выборки малого объема), либо б) сгруппированный вариационный ряд (для выборки большого объема), требуется:

1. построить для случая (а) полигон относительных частот, для случая (б) – гистограмму выборки;
2. построить график эмпирической функции распределения $F_n^*(x)$;
3. найти выборочное среднее $\overline{x}$;
4. найти выборочное среднеквадратическое отклонение σ_X^*;
5. найти несмещенную оценку $\tilde{\sigma}$ среднеквадратического отклонения теоретического распределения;
6. оценить с надежностью γ истинное значение измеренной величины (построить доверительный интервал $I_\gamma^{(\mathrm{M})}$ для математического ожидания);
7. оценить с надежностью γ точность измерений (построить доверительный интервал $I_\gamma^{(\sigma)}$ для среднеквадратического отклонения).

Задание содержит:

1. таблицу результатов измерений – случай (а) или (б);
2. n – число измерений (объем выборки);
3. γ – надежность (доверительную вероятность).

Пример 1. Выборка малого объема (случай (а)): n=24, γ=0.9, вариационный ряд представлен табл. 1.8

Таблица 1.8

x_i	540	550	560	570	580	590	600
m_i	1	1	4	9	6	2	1

Решение:

Для выполнения пп.1–4 задания заготавливаем таблицу (табл. 1.9), содержащую 9 вертикальных столбцов. Первые два столбца заполняем исходными данными, остальные – по мере выполнения пунктов задания.

Таблица 1.9

x_i	m_i	$p_i^* = \frac{m_i}{n}$	N_i	$\frac{N_i}{n}$	$y_i = x_i - 570$	$z_i = \frac{y_i}{10}$	$z_i m_i$	$z_i^2 m_i$
540	1	0.04	1	0.04	–30	–3	–3	9
550	1	0.04	2	0.08	–20	–2	–2	4
560	4	0.17	6	0.25	–10	–1	–4	4
570	9	0.38	15	0.63	0	0	0	0
580	6	0.25	21	0.88	10	1	6	6
590	2	0.08	23	0.96	20	2	4	8
600	1	0.04	24	1.00	30	3	3	9
$\sum$	24	1.00					4	40

1. Полигон относительных частот строится по данным 1-го и 3-го столбцов и имеет вид рис. 1.1.

2. Для построения графика эмпирической функции распределения $F_n^*(x)$ заполняем 4-й и 5-й столбцы табл. 1.9, находя накопленные частоты N_i и накопленные относительные частоты N_i/n (см. табл. 1.3). Затем по формулам (1.1) строим график $F_n^*(x)$, имеющий вид рис. 1.2.

3. Для нахождения выборочного среднего $\bar{x}$ заполняем 6-й, 7-й и 8-й столбцы табл. 1.9 и вместо исходного вариационного ряда обрабатываем более простой вариационный ряд (z_i и m_i). Пользуясь свойствами 1 и 2 выборочного среднего, осуществляем расчеты по следующим формулам: $\bar{z} = \frac{1}{n}\sum_{i=1}^{7} z_i m_i = \frac{1}{6}$, $\bar{y} = 10\bar{z} \approx 1.7$, $\bar{x} = 570 + \bar{y} \approx 571.7$. Итак, $\bar{x} \approx 571.7$.

4. Для нахождения выборочного среднеквадратического отклонения σ_X^* заполняем 9-й столбец табл. 1.9 и, используя свойства 1 и 2 выборочных моментов, производим расчет выборочной дисперсии D_X^* по сле-

дующим формулам: $D_Z^* = \frac{1}{n}\sum_{i=1}^{7} z_i^2 m_i - \bar{z}^2 = 1.639$, $D_X^* = D_Y^* = 10^2 D_Z^* \approx 163.9$. По формуле (1.8) находим выборочное среднеквадратическое отклонение $\sigma_X^* = \sqrt{D_X^*} \approx 12.8$. Итак, $\sigma_X^* \approx 12.8$.

5. Несмещенную оценку $\tilde{\sigma}$ среднеквадратического отклонения теоретического распределения определяем по формулам (1.12) и (1.14): $\tilde{\sigma} = \sqrt{\frac{n}{n-1}}\sigma_X^* \approx 13.1$. Итак, $\tilde{\sigma} \approx 13.1$

6. Оценку истинного значения измеренной величины дает доверительный интервал $I_\gamma^{(\mathrm{M})}$ для математического ожидания, который для случая малой выборки ($n-1 \le 30$) определяется по формуле (1.16). По заданной доверительной вероятности $\gamma = 0.9$ и числу степеней свободы $n-1 = 23$ по таблице распределения Стьюдента (см. Приложение 1) находим $t_{\gamma,n-1} = 1.714$. В результате подстановки найденных параметров в формулу (1.16) получаем следующий доверительный интервал $I_\gamma^{(\mathrm{M})}$, дающий интервальную оценку истинного значения измеренной величины: $I_\gamma^{(\mathrm{M})} = (567.1, 576.3)$.

7. Оценку точности измерений дает доверительный интервал $I_\gamma^{(\sigma)}$ для среднеквадратического отклонения, который для случая малой выборки ($n-1 \le 30$) определяется по формуле (1.17). Параметры χ_1^2 и χ_2^2 находим по таблице распределения хи-квадрат (см. Приложение 2) для числа степеней свободы $r = n-1 = 23$ и вероятностей p_1 и p_2, вычисленных для заданной $\gamma = 0.9$ по формуле (1.18): $r = 23$, $p_1 = 0.95 \Rightarrow \chi_1^2 = 13.09$; $r = 23$, $p_1 = 0.05 \Rightarrow \chi_2^2 = 35.2$. В результате подстановки найденных параметров в формулу (1.17) получаем следующий доверительный интервал $I_\gamma^{(\sigma)}$, дающий интервальную оценку точности измерений: $I_\gamma^{(\sigma)} = (10.6, 17.4)$.

Пример 2. Выборка большого объема (случай (б)): n=180, γ=0.98, сгруппированный вариационный ряд представлен табл. 1.10.

Таблица 1.10

$x_i \div x_{i+1}$	120÷130	130÷140	140÷150	150÷160	160÷170	170÷180	180÷190	190÷200
m_i	2	10	26	49	52	28	12	1

Решение:

Для выполнения пп. 1–4 заготавливаем таблицу (табл. 1.11), содержащую 12 вертикальных столбцов. Первые 2 столбца заполняем исходными данными, остальные – по мере выполнения задания.

Таблица 1.11

x_i	m_i	h_i	p_i^*	$\frac{p_i^*}{h_i}$	N_i	$\frac{N_i}{n}$	$x_{i\,\text{ср}}$	$y_i = x_{i\,\text{ср}} - 155$	$z_i = \frac{y_i}{10}$	$z_i m_i$	$z_i^2 m_i$
120÷130	2	10	0.01	0.001	2	0.01	125	–30	–3	–6	18
130÷140	10	10	0.05	0.005	12	0.06	135	–20	–2	–20	40
140÷150	26	10	0.14	0.014	38	0.21	145	–10	–1	–26	26
150÷160	49	10	0.27	0.027	87	0.48	155	0	0	0	0
160÷170	52	10	0.29	0.029	139	0.77	165	10	1	52	52
170÷180	28	10	0.16	0.016	167	0.93	175	20	2	56	112
180÷190	12	10	0.07	0.007	179	0.99	185	30	3	36	108
190÷200	1	10	0.01	0.001	180	1.00	195	40	4	4	16
$\sum$	180		1.00							96	372

1. Для построения гистограммы выборки в 3-й столбец табл. 1.11 помещаем ширину интервала $h_i = x_{i+1} - x_i$, в 4-й столбец – относительную частоту $p_i^* = m_i / n$, а в 5-й столбец – плотность относительной частоты p_i^* / h_i. По данным 1-го и 5-го столбцов строим гистограмму выборки, которая имеет вид рис. 1.3.

2. Для построения графика эмпирической функции распределения $F_n^*(x)$ заполняем 6-й и 7-й столбцы табл. 1.11, находя накопленные часто-

ты N_i и накопленные относительные частоты N_i/n (см. табл. 1.6). Затем по формулам (1.2) строим график $F_n^*(x)$, имеющий вид рис. 1.4.

3. Для нахождения выборочных моментов заменяем исходный сгруппированный вариационный ряд вариационным рядом, в котором частота i-го интервала m_i соответствует середине i-го интервала; значения $x_{i\,\text{ср}} = (x_i + x_{i+1})/2$ заносим в 8-й столбец табл. 1.11.

Для упрощения расчетов заполняем 9-й, 10-й и 11-й столбцы табл. 1.11, в результате вместо вариационного ряда ($x_{i\,\text{ср}}$ и m_i) обрабатываем более простой вариационный ряд (z_i и m_i). Пользуясь свойствами 1 и 2 для выборочного среднего, определяем $\bar{x}$: $\bar{z} = \frac{1}{n}\sum_{i=1}^{8} z_i m_i \approx 0.53$, $\bar{y} = 10\bar{z} \approx 5.3$, $\bar{x} = 155 + \bar{y} \approx 160.3$. Итак, $\bar{x} \approx 160.3$.

4. Для нахождения выборочного среднеквадратического отклонения σ_X^* сначала находим выборочную дисперсию D_X^*, заполняя 12-й столбец табл. 1.11 и пользуясь свойствами выборочной дисперсии: $\mathrm{D}_Z^* = \frac{1}{n}\sum_{i=1}^{8} z_i^2 m_i - \bar{z}^2 \approx 1.783$, $\mathrm{D}_X^* = \mathrm{D}_Y^* = 10^2\,\mathrm{D}_Z^* = 178.3$. Затем по формуле (1.8) находим выборочное среднеквадратическое отклонение $\sigma_X^* = \sqrt{\mathrm{D}_X^*} \approx 13.4$. Итак, $\sigma_X^* \approx 13.4$.

5. Для нахождения несмещенной оценки $\tilde{\sigma}$ среднеквадратического отклонения теоретического распределения сначала находим несмещенную оценку дисперсии $\tilde{\mathrm{D}}$, которая в данном случае равна выборочной дисперсии с учетом поправки Шеппарда (формула (1.13)): $\tilde{\mathrm{D}} = \mathrm{D}_X^* - \frac{h^2}{12} \approx 170.0$ ($h = 10$ – ширина интервала). Извлекая квадратный корень, получаем $\tilde{\sigma} = \sqrt{\tilde{D}} \approx 13.0$. Итак, $\tilde{\sigma} \approx 13.0$.

6. Оценку истинного значения измеренной величины дает доверительный интервал $I_\gamma^{(\mathrm{M})}$ для математического ожидания, который для случая большой выборки ($n - 1 > 30$) определяется формулой (1.19). По заданной доверительной вероятности $\gamma = 0.98$ по таблице нормального рас-

пределения (нижняя строка таблицы распределения Стьюдента, см. Приложение 1) находим $t_\gamma = 2.33$. В результате подстановки найденных параметров в формулу (1.19) получаем следующий доверительный интервал $I_\gamma^{(\mathrm{M})}$, дающий интервальную оценку истинного значения измеренной величины: $I_\gamma^{(\mathrm{M})} = (158.0, 162.6)$.

7. Оценку точности измерений дает доверительный интервал $I_\gamma^{(\sigma)}$ для среднеквадратического отклонения, который для случая большой выборки ($n-1>30$) определяется по формуле (1.20). Для $\gamma = 0.98$ $t_\gamma = 2.33$ (см. п. 6). В результате подстановки найденных параметров в формулу (1.20) получаем следующий доверительный интервал $I_\gamma^{(\sigma)}$, дающий интервальную оценку точности измерении: $I_\gamma^{(\sigma)} = (11.6, 15.0)$.

2. Определение закона распределения случайной величины по экспериментальным данным с использованием критерия согласия Пирсона

2.1. Постановка задачи о критерии согласия

Пусть X – случайная величина, возможные значения которой образуют генеральную совокупность. В результате n независимых испытаний получена выборка $(x_1, x_2, \ldots, x_n)$ объема n, элементы которой рассматриваются как значения независимых случайных величин $X_1, X_2, \ldots, X_n$ – результатов измерений случайной величины X.

Требуется определить закон распределения случайной величины X. В математической статистике эту задачу принято решать в два этапа.

I этап. Подбор теоретического распределения, сглаживающего данное эмпирическое распределение. Часто эту задачу решают в два приема.

А. Определение вида закона распределения (функция $F(x)$ подбирается по эмпирической функции распределения $F_n^*(x)$ или плотность вероятности $f(x)$ – по гистограмме выборки).

Б. Нахождение по выборке параметров теоретического закона распределения. При этом обычно используют так называемый *метод моментов*, согласно которому параметры теоретического распределения выбираются с таким расчетом, чтобы несколько важнейших моментов теоретического распределения совпали с их точечными статистическими оценками, полученными из выборки (например, для нормального закона $m_1 = \mathrm{M}[\mathrm{X}] = \bar{x}$, $\mu_2 = \mathrm{D}[\mathrm{X}] = \tilde{\mathrm{D}}$).

II этап. Проверка гипотезы H_0 о соответствии закона распределения, подобранного на первом этапе, данным выборки. Для этого используются критерии согласия.

Критерием согласия называется правило, позволяющее установить, являются ли расхождения между эмпирическим и предполагаемым теоретическим распределениями (т. е. между $F_n^*(x)$ и $F(x)$) случайными (не значимыми) или существенными (значимыми).

Идея любого критерия согласия состоит в следующем: выбирается некоторая случайная величина $U = U[X_1, X_2, \ldots X_n; F(x)]$, характеризующая степень расхождения между $F_n^*(x)$ и $F(x)$. Случайная величина U выбирается таким образом, чтобы при $n \to \infty$ был бы известен закон ее распределения, который не зависел бы от вида теоретического распределения $F(x)$. Для данной выборки $(x_1, x_2, \ldots, x_n)$ случайная величина U принимает определенное значение $u_{\text{в}} = U[x_1, x_2, \ldots x_n; F(x)]$. Чем больше значение $u_{\text{в}}$, тем хуже согласуется гипотеза H^0 с экспериментальными данными. При использовании критерия согласия задается *уровень значимости* критерия $\alpha = 1 - \gamma$, где γ – доверительная вероятность, или надежность (α – вероятность того, что отвергается верная гипотеза H_0). Из условия $P(U > u_\alpha) = \alpha$ можно найти u_α – так называемое пороговое значение случайной величины U, соответствующее уровню значимости α.

Схема применения критерия согласия сводится к следующему.

1. По выборке $(x_1, x_2, \ldots, x_n)$ вычисляется величина $u_{\text{в}} = U[x_1, x_2, \ldots x_n; F(x)]$.

2. Задается уровень значимости α.

3. По известному закону распределения случайной величины U по таблице находится пороговое (критическое) значения u_{α}, удовлетворяющее условию $P(U > u_{\alpha}) = \alpha$.

4. Сравниваются $u_{в}$ и u_{α}: если $u_{в} < u_{\alpha}$, то гипотеза H_0 принимается, если $u_{в} \geq u_{\alpha}$, то H_0 отвергается.

Имеют место следующие замечания.

1. Для критерия согласия обычно берут α=0.01, 0.05 или 0.1 (или в процентах 1%, 5%, 10%). Чем больше α, тем "жестче" критерий, так как большее число верных гипотез будет отвергнуто;

2. Любой критерий согласия предполагает выборку большого объема ($n \geq 100$);

3. С помощью критерия согласия нельзя доказать гипотезу H_0, можно лишь подтвердить, что она не противоречит экспериментальным данным или отвергнуть ее как маловероятное событие.

2.2. Критерий согласия Пирсона (критерий χ^2)

Критерий согласия Пирсона служит для проверки гипотезы H_0 о предполагаемом законе распределения любой случайной величины X (как дискретной, так и непрерывной). Пусть этот закон задан в виде функции распределения $F(x)$. Для применения критерия Пирсона выборка (объема $n \geq 100$) должна быть представлена в виде сгруппированного вариационного ряда (см. табл. 1. 4).

Зная закон распределения случайной величины X, можно найти вероятности попадания ее в интервалы (x_i, x_{i+1}): $p_i = P(x_i \leq x < x_{i+1}) = F(x_{i+1}) - F(x_i)$, $i = 1,2,\ldots,l$. Для того, чтобы $\sum_{i=1}^{l} p_i = 1$, крайние интервалы делают полубесконечными: $(x_1, x_2) \to (-\infty, x_2)$, $(x_l, x_{l+1}) \to (x_l, +\infty)$. Если проверяемая гипотеза H_0 верна, то частота m_i представляет собой число независимых испытаний, в которых данное событие (попадание случайной величины X в i-й интервал) произошло, если его вероятность в отдельном испытании из n равна

p_i. Следовательно, можно рассматривать частоту m_i как значение случайной величины (меняющееся от выборки к выборке), распределенной по биномиальному закону с математическим ожиданием np_i. Если число испытаний n велико, то на основании теоремы Муавра-Лапласа можно считать закон распределения частоты нормальным (на самом деле он является асимптотически нормальным).

В качестве меры расхождения данных выборки $m_1, m_2, \ldots, m_l$ с теоретическими данными $np_1, np_2, \ldots, np_l$ в критерии согласия Пирсона взята следующая случайная величина:

$$U(X_1, X_2, \ldots, X_n, F(x)) = \chi^2(r) = \sum_{i=1}^{l} \frac{(m_i - np_i)^2}{np_i}, \qquad (2.1)$$

(здесь m_i – случайные величины), которая для данной выборки принимает определенное значение

$$U(x_1, x_2, \ldots, x_n, F(x)) = \chi^2_{\text{в}} = \sum_{i=1}^{l} \frac{(m_i - np_i)^2}{np_i}. \qquad (2.2)$$

Очевидно, что чем больше согласуются эмпирическое и теоретическое распределения, тем меньше будут разности $m_i - np_i$ и меньше будет $\chi^2_{\text{в}}$. Пирсон доказал, что при $n \to \infty$ и в случае асимптотически нормального распределения частот m_i закон распределения случайной величины U (см. (2.1)) не зависит от вида функции $F(x)$ и приближается к закону хи-квадрат с r степенями свободы (если гипотеза H_0 верна). Число степеней свободы этого закона определяется формулой

$$r = l - s, \qquad (2.3)$$

где l – число интервалов, на которые разбита выборка, s – число независимых условий (связей), наложенных на частоты m_i. Примерами таких связей являются следующие: $\sum_{i=1}^{l} m_i = n$ (всегда имеет место), $\bar{x} = \mathrm{M}[X]$, $\tilde{\sigma} = \sigma[X]$ (если эти параметры находятся из выборки $(x_1, x_2, \ldots, x_n)$).

Так как при выводе критерия согласия Пирсона предполагалось, что частоты m_i распределены асимптотически нормально, то все рассуждения будут справедливы, если значения m_i достаточно велики. Поэтому на практике рекомендуется в каждом интервале иметь не менее 5...10 наблюдений. Если частоты отдельных интервалов малы ($m_i < 5$), то следует объединить соседние интервалы.

Результаты вычислений $\chi^2_{\text{в}}$ по формуле (2.2) удобно свести в таблицу (табл. 2. 1).

Таблица 2.1

i	$x_i \div x_{i+1}$	m_i	p_i	np_i	$m_i - np_i$	$(m_i - np_i)^2$	$\dfrac{(m_i - np_i)^2}{np_i}$
1	$-\infty \div x_2$	m_1	p_1	np_1	$m_1 - np_1$	$(m_1 - np_1)^2$	$\ldots$
2	$x_2 \div x_3$	m_2	p_2	np_2	$m_2 - np_2$	$(m_2 - np_2)^2$	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
l	$x_l \div +\infty$	m_l	p_l	np_l	$m_l - np_l$	$(m_l - np_l)^2$	$\ldots$
$\sum$		n	1	n			$\chi^2_{\text{в}}$

Схема применения критерия согласия Пирсона сводится к следующему.

1. По выборке по формуле (2.2) вычисляется $\chi^2_{\text{в}}$ (см. табл. 2. 1).

2. Задается уровень значимости α.

3. По таблице распределения хи-квадрат $\chi^2(r)$ (см. Приложение 2) для числа степеней свободы r, найденного по формуле (2.3), и заданного уровня значимости α находится пороговое (критическое) значение $\chi^2(\alpha, r)$, удовлетворяющее условию $P\left(\chi^2(r) > \chi^2(\alpha, r)\right) = \alpha$.

4. Сравниваются $\chi^2_{\text{в}}$ и $\chi^2(\alpha, r)$. Если $\chi^2_{\text{в}} < \chi^2(\alpha, r)$, то <u>гипотеза H_0 принимается</u>, т. е. можно считать, что предполагаемая функция распределения $F(x)$ согласуется с экспериментальными данными; если $\chi^2_{\text{в}} \geq \chi^2(\alpha, r)$, то <u>гипотеза</u> H_0 <u>отвергается</u> как маловероятная.

2.3. Порядок выполнения задания

Пример задания: Результаты 200 измерений некоторой физической величины представлены сгруппированным вариационным рядом (табл. 2.2). Проверить с помощью критерия согласия Пирсона гипотезу о нормальном законе распределения этой физической величины (случайной величины X) с параметрами, найденными из выборки, взяв 5% уровень значимости (т. е. $\alpha = 0.05$).

Таблица 2.2

i	1	2	3	4	5	6	7	8	$\sum$
$x_i \div x_{i+1}$	180÷190	190÷200	200÷210	210÷220	220÷230	230÷240	240÷250	250÷260	
m_i	3	7	26	56	64	30	12	2	$n = 200$

Решение.

Проверяемая гипотеза H_0: X распределена по закону $N(a, \sigma)$, т. е $F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left(-\frac{(t-a)^2}{2\sigma^2} \right) dt$ – функция распределения X, где параметры $a = \mathrm{M}[X] = \bar{x}$ и $\sigma = \sigma[X] = \tilde{\sigma}$ определяются из выборки. Решение поставленной задачи разобьем на две части.

1. Обработаем выборку (табл. 2.2) и определим выборочное среднее $\bar{x}$ и несмещенную оценку среднеквадратического отклонения $\tilde{\sigma}$. Для этого найдем середины интервалов $x_{i\,\text{ср}} = (x_i + x_{i+1})/2$ и сгруппированный вариационный ряд (табл. 2.2) заменим соответствующим вариационным рядом (см. табл. 1.7). Пользуясь формулами (1.9) и (1.10), вычислим выборочное среднее $\bar{x} = \frac{1}{200}\sum\limits_{i=1}^{8} x_{i\,\text{ср}} m_i = 221$ и выборочную дисперсию $D_X^* = \frac{1}{200}\sum\limits_{i=1}^{8}(x_{i\,\text{ср}} - \bar{x})^2 m_i = 167.1$. По формуле (1.13) найдем несмещенную оценку дисперсии $\tilde{D} = D_X^* - h^2/12 \approx 158.77$ (здесь $h = \Delta x_i = x_{i+1} - x_i = 10$ –

ширина интервалов), а по формуле (1.14) – несмещенную оценку среднеквадратического отклонения $\tilde{\sigma} = \sqrt{\tilde{D}} \approx 12.6$. Итак, $\bar{x} = 221$, $\tilde{\sigma} \approx 12.6$.

2. Проверим с помощью критерия Пирсона гипотезу H_0 о том, что X распределена по закону $N(221; 12.6)$. Для этого представим исходную выборку в виде, необходимом для использования критерия (см. 3 левых столбца табл. 2.3). Так как $m_1 = 3 < 5$ и $m_8 = 2 < 5$, объединим первый со вторым и седьмой с восьмым интервалами из табл. 2.2, а крайние интервалы (верхний и нижний) расширим до бесконечности.

Таблица 2.3

i	$x_i \div x_{i+1}$	m_i	p_i	np_i	$m_i - np_i$	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
1	$-\infty \div 200$	10	0.047	9.4	0.6	0.36	0.038
2	200÷210	26	0.145	29.0	–3.0	9.00	0.310
3	210÷220	56	0.276	55.2	0.8	0.64	0.012
4	220÷230	64	0.293	58.6	5.4	29.16	0.498
5	230÷240	30	0.173	34.6	–4.6	21.16	0.612
6	$240 \div +\infty$	14	0.066	13.2	0.8	0.64	0.048
$\sum$		200	1	200			$\chi^2_{\text{в}} = 1.52$

Результаты вычислений $\chi^2_{\text{в}}$ по формуле (2.2) приведены в следующих столбцах табл. 2.3. Расчет теоретических вероятностей p_i попадания случайной величины X в интервалы (x_i, x_{i+1}) осуществляем по известной формуле для нормального закона $N(\bar{x}, \tilde{\sigma})$:

$$p_i = P(x_i < X < x_{i+1}) = \frac{1}{2}\left[\Phi_L\left(\frac{x_{i+1} - \bar{x}}{\tilde{\sigma}}\right) - \Phi_L\left(\frac{x_i - \bar{x}}{\tilde{\sigma}}\right)\right],\ i = 1,2,\ldots 6,$$

где $\Phi_L(x) = \frac{2}{\sqrt{2\pi}}\int\limits_0^x \exp\left(-\frac{t^2}{2}\right)dt$ – функция Лапласа. Ее свойства: $\Phi_L(-x) = -\Phi_L(x)$, $\Phi_L(\pm\infty) = \pm 1$. Значения функции $\Phi_L(x)$ приведены в Приложении 3. Суммируя данные последнего столбца табл. 2.3, находим $\chi^2_{\text{в}} \approx 1.52$.

Полученное значение $\chi_{\text{в}}^2$ надо сравнить с пороговым $\chi^2(\alpha, r)$. Уровень значимости $\alpha = 0.05$ (задан в условии), число степеней свободы r распределения хи-квадрат определяем по формуле (2.3) В данной задаче число интервалов $l = 6$, число связей $s = 1 + 2 = 3$ (так как параметры теоретического распределения находятся по выборке), следовательно $r = 6 - 3 = 3$. По таблице распределения хи-квадрат (см. Приложение 2) находим пороговое значение $\chi^2(\alpha, r) = \chi^2(0.05, 3) = 7.82$. Так как $\chi_{\text{в}}^2 \approx 1.52 < 7.82$, т. е. найденное по выборке значение $\chi_{\text{в}}^2$ меньше порогового, то нет оснований отвергать гипотезу H_0, а можно считать ее согласующейся с экспериментальными данными.

Ответ: гипотеза о том, что данная физическая величина распределена по нормальному закону $N(221; 12.6)$, согласуется с результатами измерений.

Приложение 1. Коэффициенты Стьюдента $t_{\gamma,n-1}$

$n-1$	γ				
	0.8	0.9	0.95	0.98	0.99
5	1.476	2.02	2.57	3.36	4.03
6	1.440	1.943	2.45	3.14	3.71
7	1.415	1.895	2.36	3.00	3.50
8	1.397	1.860	2.31	2.90	3.36
9	1.383	1.833	2.26	2.82	3.25
10	1.372	1.812	2.23	2.76	3.17
11	1.363	1.796	2.20	2.72	3.11
12	1.356	1.782	2.18	2.68	3.06
13	1.350	1.771	2.16	2.65	3.01
14	1.345	1.761	2.14	2.62	2.98
15	1.341	1.753	2.13	2.60	2.95
16	1.337	1.746	2.12	2.58	2.92
17	1.333	1.740	2.11	2.57	2.90
18	1.330	1.734	2.10	2.55	2.88
19	1.328	1.729	2.09	2.54	2.86
20	1.325	1.725	2.09	2.53	2.84
21	1.323	1.721	2.08	2.52	2.83
22	1.321	1.717	2.07	2.51	2.82
23	1.319	1.714	2.07	2.50	2.81
24	1.318	1.711	2.06	2.49	2.80
25	1.316	1.708	2.06	2.48	2.79
26	1.315	1.706	2.06	2.48	2.78
27	1.314	1.703	2.05	2.47	2.77
28	1.313	1.701	2.05	2.47	2.76
29	1.311	1.699	2.04	2.46	2.76
30	1.310	1.697	2.04	2.46	2.75
∞	1.282	1.645	1.96	2.33	2.58

ПРИЛОЖЕНИЕ 2. ЗНАЧЕНИЯ χ^2 В ЗАВИСИМОСТИ ОТ p И r

r \ p	0.99	0.98	0.95	0.90	0.80	0.20	0.10	0.05	0.02	0.01
1	0.000	0.001	0.004	0.016	0.064	1.642	2.71	3.84	5.41	6.64
2	0.020	0.040	0.103	0.211	0.446	3.22	4.60	5.99	7.82	9.21
3	0.115	0.185	0.352	0.584	1.005	4.64	6.25	7.82	9.84	11.34
4	0.297	0.429	0.711	1.064	1.649	5.99	7.78	9.49	11.67	13.28
5	0.554	0.752	1.145	1.610	2.34	7.29	9.24	11.07	13.39	15.09
6	0.872	1.134	1.635	2.20	3.07	8.56	10.64	12.59	15.03	16.81
7	1.239	1.564	2.17	2.83	3.82	9.80	12.02	14.07	16.62	18.48
8	1.646	2.03	2.73	3.49	4.59	11.03	13.36	15.51	18.17	20.1
9	2.09	2.53	3.32	4.17	5.38	12.24	14.68	16.92	19.68	21.7
10	2.56	3.06	3.94	4.86	6.18	13.44	15.99	18.31	21.2	23.2
11	3.05	3.61	4.58	5.58	6.99	14.63	17.28	19.68	22.6	24.7
12	3.57	4.18	5.23	6.30	7.81	15.81	18.55	21.0	24.1	26.2
13	4.11	4.76	5.89	7.04	8.63	16.98	19.81	22.4	25.5	27.7
14	4.66	5.37	6.57	7.79	9.47	18.15	21.1	23.7	26.9	29.1
15	5.23	5.98	7.26	8.55	10.31	19.31	22.3	25.0	28.3	30.6
16	5.81	6.61	7.96	9.31	11.15	20.5	23.5	26.3	29.6	32.0
17	6.41	7.26	8.67	10.08	12.00	21.6	24.8	27.6	31.0	33.4
18	7.02	7.91	9.39	10.86	12.86	22.8	26.0	28.9	32.3	34.8
19	7.63	8.57	10.11	11.65	13.72	23.9	27.2	30.1	33.7	36.2
20	8.26	9.24	10.85	12.44	14.58	25.0	28.4	31.4	35.0	37.6
21	8.90	9.92	11.59	13.24	15.44	26.2	29.6	32.7	36.3	38.9
22	9.54	10.60	12.34	14.04	16.31	27.3	30.8	33.9	37.7	40.3
23	10.20	11.29	13.09	14.85	17.19	28.4	32.0	35.2	39.0	41.6
24	10.86	11.99	13.85	15.66	18.06	29.6	33.2	36.4	40.3	43.0
25	11.52	12.70	14.61	16.47	18.94	30.7	34.4	37.7	41.7	44.3
26	12.20	13.41	15.38	17.29	19.82	31.8	35.6	38.9	42.9	45.6
27	12.88	14.12	16.15	18.11	20.7	32.9	36.7	40.1	44.1	47.0
28	13.56	14.85	16.93	18.94	21.6	34.0	37.9	41.3	45.4	48.3
29	14.26	15.57	17.71	19.77	22.5	35.1	39.1	42.6	46.7	49.6
30	14.95	16.31	18.49	20.6	23.4	36.2	40.3	43.8	48.0	50.9

Приложение 3. Таблица значений функции Φ_L

x	$\Phi_L(x)$	x	$\Phi_L(x)$	x	$\Phi_L(x)$
0.00	0.0000	1.00	0.6827	2.00	0.9545
0.05	0.0399	1.05	0.7063	2.05	0.9596
0.10	0.0797	1.10	0.7287	2.10	0.9643
0.15	0.1192	1.15	0.7499	2.15	0.9684
0.20	0.1585	1.20	0.7699	2.20	0.9722
0.25	0.1974	1.25	0.7887	2.25	0.9756
0.30	0.2358	1.30	0.8064	2.30	0.9786
0.35	0.2737	1.35	0.8230	2.35	0.9812
0.40	0.3108	1.40	0.8385	2.40	0.9836
0.45	0.3473	1.45	0.8529	2.45	0.9857
0.50	0.3829	1.50	0.8664	2.50	0.9876
0.55	0.4177	1.55	0.8789	2.55	0.9892
0.60	0.4515	1.60	0.8904	2.60	0.9907
0.65	0.4843	1.65	0.9011	2.65	0.9920
0.70	0.5161	1.70	0.9109	2.70	0.9931
0.75	0.5467	1.75	0.9199	2.75	0.9940
0.80	0.5763	1.80	0.9281	2.80	0.9949
0.85	0.6047	1.85	0.9357	2.85	0.9956
0.90	0.6319	1.90	0.9426	2.90	0.9963
0.95	0.6579	1.95	0.9488	2.95	0.9968
1.00	0.6827	2.00	0.9545	3.00	0.9973

Содержание

1. Статистическая обработка экспериментальных данных и статистическое оценивание результатов измерений 3
   1.1 Выборочный метод и основные характеристики выборки 3
   1.2. Статистические оценки параметров теоретического распределения 12
   1.3. Порядок выполнения задания 16
2. Определение закона распределения случайной величины по экспериментальным данным с использованием критерия согласия Пирсона 21
   2.1. Постановка задачи о критерии согласия 21
   2.2. Критерий согласия Пирсона (критерий χ^2) 23
   2.3. Порядок выполнения задания 26

Приложение 1. Коэффициенты Стьюдента $t_{\gamma,n-1}$ 29

Приложение 2. Значения χ^2 в зависимости от p и r 30

Приложение 3. Таблица значений функции Φ_L 32

Редактор Э.К.Долгатов
Лицензия ЛР № 020617 от 10.08.92

Подписано в печать 13.09.04. Формат 60х84 1/16. Бумага тип. № 2.
Печать офсетная. Усл. печ. л. 1,86. Уч.–изд. л. 2,0.
Тираж 250 экз. Заказ 96.
Издательско–полиграфический центр СПбГЭТУ

197376, С.–Петербург, ул. Проф. Попова, 5