2. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

2.1. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ

2.1.1. Случайные величины и их числовые характеристики

Планирование эксперимента и обработка экспериментальных результатов опираются на методы, разработанные теорией вероятностей и математической статистикой. Поэтому уместно напомнить основные понятия этих дисциплин.

Основным объектом теории вероятностей является **случайная величина** – величина, которая в результате экспериментов принимает различные, заранее не известные значения. Случайные величины могут быть дискретными и непрерывными.

Дискретными называют такие случайные величины, которые принимают отдельные, большей частью целочисленные значения. Дискретные случайные величины задаются набором отдельных значений. Например, число деталей, обработанных на станке, есть дискретная случайная величина.

Непрерывной случайной величиной называется такая, которая может принимать любые численные значения из непрерывного ряда возможных значений. Например, действительные размеры деталей, обработанных на станке, являются непрерывными случайными величинами, так как они могут принимать любые численные значения в определенных границах. Непрерывные случайные величины задаются в виде функции.

Дискретная случайная величина X может принимать значения x_1, $x_2, \ldots x_i$. Отношение числа опытов m, при проведении которых случайная величина X приняла значения x_i, к общему числу опытов n называется частотой повторения события. Частота m / n является случайной величиной и меняется в зависимости от числа проведенных опытов. При большом числе опытов она имеет тенденцию стабилизироваться около некоторого значения P_i , называемого вероятностью события. Сумма вероятностей всех возможных значений дискретной случайной величины равна единице, так как вероятность того, что случайная величина в результате опыта примет одно из своих значений, есть достоверное событие.

Вероятность события $X < x_i$ является функцией от $P(X < x_i) = F(x)$ и называется **функцией распределения случайной величины**. Функция распределения – самая уникальная характеристика любой случайной величины (как дискретной, так и непрерывной) – полностью характеризует ее с вероятностной точки зрения. Функция распределения дискретной случайной величины – всегда разрывная ступенчатая, скачки в ней происходят в точках, соответствующих возможным значениям случайной величины, и равны вероятностям этих значений. Сумма всех скачков равна 1: $\sum_i P_i = 1$.

Непрерывную случайную величину X, кроме функции распределения $F(x)$, характеризуют еще и производной от функции распределения – **плотностью распределения** случайной величины X. Для непрерывной и дифференцируемой функции $F(x)$ плотность распределения $f(x) = F`(x)$.

Задание $f(x)$ тоже полностью определяет случайную величину. Связь между функцией распределения $F(x)$ и плотностью распределения определяется выражением

$$F(x) = P(-\infty \le X \le x) = \int_{-\infty}^{x} f(x)dx. \qquad (2.1)$$

Отсюда выводится одно из важных свойств плотности распределения

$$\int_{-\infty}^{+\infty} f(x)dx = 1, \tag{2.2}$$

так как попадание случайной величины в интервал $-\infty < x < +\infty$ есть достоверное событие.

Для количественной оценки случайных величин пользуются числовыми характеристиками. К числовым характеристикам, определяющим положение центра группировки, относятся среднее арифметическое, медиана и мода; к характеристикам, определяющим меру рассеяния, – математическое ожидание, дисперсия и среднее квадратичное отклонение.

Среднее арифметическое ($\bar{X}$) есть частное от деления суммы значений случайных величин на число наблюдений:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}, \tag{2.3}$$

где X_i – значения случайных величин, n – число наблюдений случайных величин.

Медианой (M_e) непрерывной случайной величины называется такое ее значение, которое делит весь ряд значений пополам.

Модой (M_o) называется такое значение случайной величины, которое наблюдается наиболее часто.

Математическим ожиданием случайной величины X называется число $M[X]$, которое вычисляется по формулам

$$M[X] = \sum_i x_i p_i \qquad \text{(дискретная величина)}, \tag{2.4}$$

$$M[X] = \int_{-\infty}^{+\infty} x \cdot f(x)dx \qquad \text{(непрерывная величина),} \qquad (2.5)$$

где $f(x)$ – плотность распределения X.

Дисперсией случайной величины X называется число $Д[X]$, которое определяется формулой

$$Д[X] = M[X - M(X)]^2 . \qquad (2.6)$$

Из формулы (2.4) следует:

$$Д[X] = \sum_i (x_i - M[X])^2 p_i \quad \text{(дискретная величина),} \qquad (2.7)$$

$$Д[X] = \int_{-\infty}^{+\infty} (x - M[X])^2 \qquad \text{(непрерывная величина).} \qquad (2.8)$$

Дисперсия характеризует разброс значений X относительно ее математического ожидания.

Среднее квадратичное отклонение (σ) случайной величины X определяют по формуле

$$\sigma = \sqrt{D[X]}, \qquad (2.9)$$

$$\sigma^2[X] = D[X].$$

2.1.2. Генеральная совокупность

На практике исследователь располагает ограниченным числом значений случайной величины. **Генеральная совокупность** – это полный набор всех возможных значений, которые может принимать случайная величина в ходе эксперимента. Генеральная совокупность может быть конечной и реально существующей или бесконечной, гипотетической.

Генеральная совокупность обладает некоторыми неслучайными свойствами, которые надо выявит в результате эксперимента. Исчерпывающей характер случайной величины *X*, а следовательно, и генеральной совокупности является ее функция распределения *F(x)*, равная вероятности того, что в результате эксперимента эта случайная величина примет значение меньшее, чем *x*. Таким образом, $F(x) = P(X < x)$. Если функция распределения имеет производную, то функция $f(x) = F'(x)$ называется плотностью распределения.

Как функция распределения *F(x)*, так и плотность распределения *f(x)* характеризуют всю генеральную совокупность и являются детерминированными, неслучайными функциями, имеющими вполне конкретный аналитический (или графический) вид.

2.1.3. Важнейшие функции распределения

В теории вероятностей и математической статистике важную роль играют следующие функции распределения.

1. Нормальный закон распределения случайной величины.

Случайная величина *X* распределена по нормальному закону (или закону Гаусса), если ее плотность вероятности имеет вид

$$f(x) = -\frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \exp\left[-(x-\mu)^2 / \left(2\sigma^2\right)\right], \qquad (2.10)$$

где $\mu = MX$, $\sigma = \sqrt{DX}$.

Удобно ввести случайную величину z:

$$z = \frac{x - \mu}{\sigma},$$

подстановка которой в предыдущую формулу позволяет представит ее в виде

$$f(z) = \left(\sqrt{2\pi}\right)^{-1} \exp\left(-z^2/2\right) . \tag{2.11}$$

Кривая нормального распределения (график функции *f(z)*) представлена на рис. 2.1.

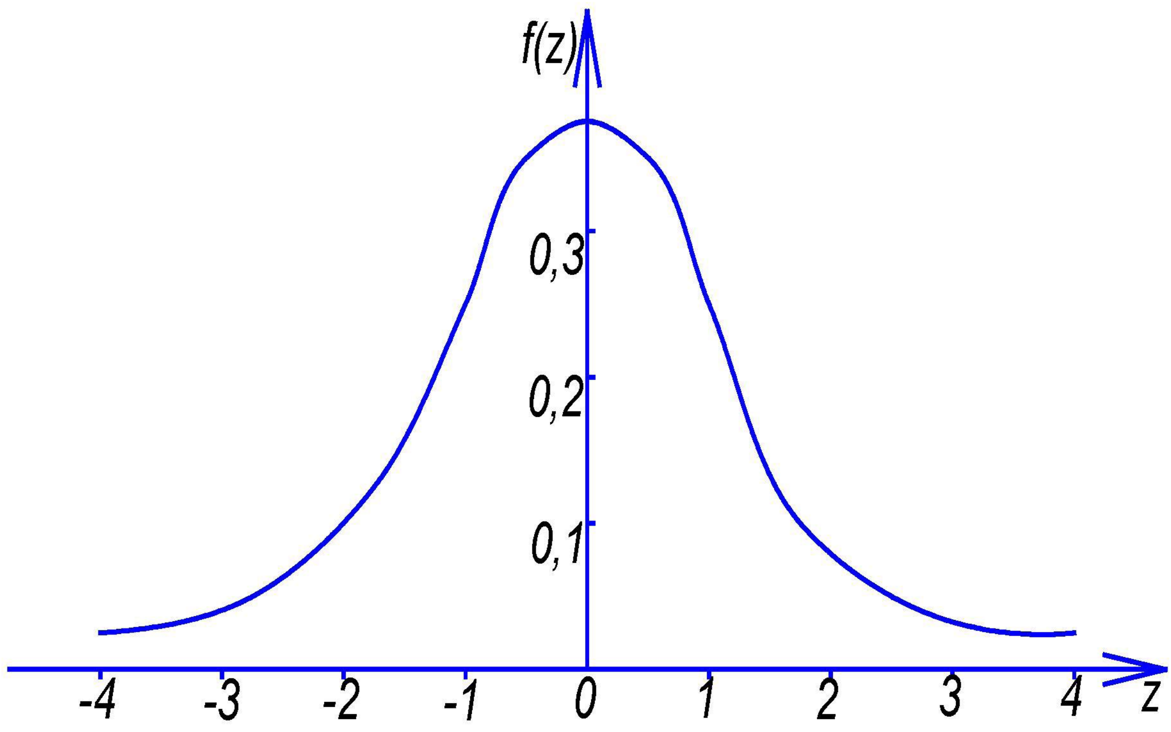

Рис. 2.1. Кривая нормального распределения (график функции *f(z)*)

Произведенное преобразование сохраняет закон распределения, но приводит его к частному виду, соответствующему случаю $\mu = 0$ и $\sigma = 1$.

Говорят, что плотность нормального распределения $f(z)$ является нормированной (площадь под кривой $f(z)$ равна единице), центрированной (максимум находится при $z = 0$), стандартизированной ($\sigma = 1$).

2. Функция распределения Пирсона (χ^2-распределение).

Рассмотрим n независимых случайных величин $u_1, \dots u_n$, каждая из которых распределена нормально с параметрами (0,1). Сумма квадратов этих случайных величин называется χ_f^2-суммой:

$$\chi_f^2 = \sum_{i=1}^{n} u_i^2 \qquad (0 \le \chi^2 < \infty), \tag{2.12}$$

а $f = n$ называется числом степеней свободы χ^2. Функция χ^2 обладает распределением, которое в силу нормированности u_i зависит только от f. Плотность χ^2-распределения, характеризующая вероятность обнаружения значений в интервале [χ^2, $\chi^2 + d\chi^2$], имеет вид (Пирсон)

$$\rho(\chi^2) = \frac{1}{2^{\frac{f}{2}}\Gamma(\frac{f}{2})}(\chi^2)^{\frac{f}{2}-1}\exp(-\frac{\chi^2}{2}), \tag{2.13}$$

где Γ – гамма-функция.

Кривые плотности χ^2-распределения для различных f указаны на рис. 2.2. Согласно (2.13) $\overline{\chi}^2 = f$; $D(\chi^2) = 2f$. При $f > 1$ распределение переходит в нормальное. Очевидно, величина $\frac{\chi^2}{f} \xrightarrow[f \to \infty]{} 1$, так как ее дисперсия стремится к нулю.

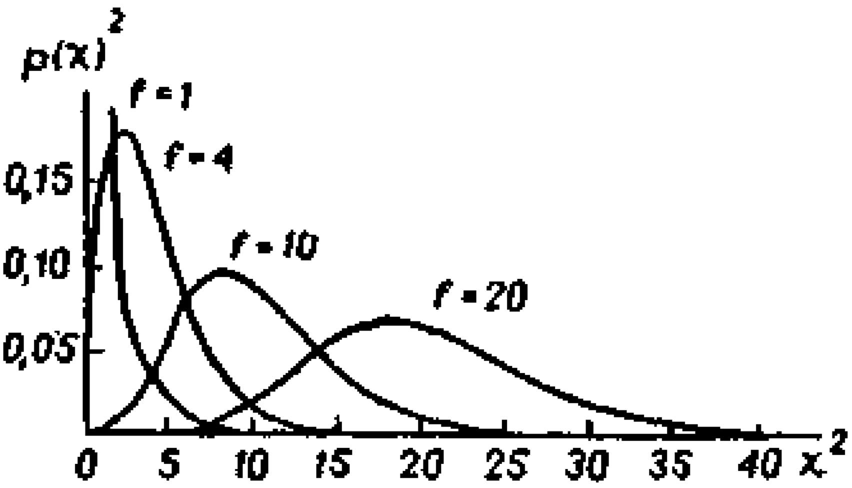

Рис. 2.2. χ^2-распределения для различных f

Справедлива теорема сложения, согласно которой сумма k стохастически независимых $\chi^2_{f_i}$ величин имеет χ^2-распределение с числом степеней свободы, равным сумме чисел степеней свободы каждой из величин:

$$\sum_{i=1}^{k} \chi^2_{f_i} = \chi^2_f \qquad \text{при} \qquad f = \sum_{i=1}^{k} f_i . \tag{2.14}$$

Существует также обратная теорема – теорема разложения χ-распределения на суммы квадратов величин, линейно связанных со случайными переменными u_i, которая формулируется следующим образом.

Пусть имеются суммы квадратов: $Q = \sum_{j=1}^{\lambda} l_j^2$, где λ случайных величин l_j являются линейными функциями стохастически независимых переменных u_i, распределенных нормально с параметрами (0,1) каждая. Если имеются m линейных соотношений (связей) между u_i, то числом степеней свободы Q-суммы является число $f = \lambda - m$.

Тогда, если в результате линейных преобразований сумма квадратов n случайных величин u_1, ..., u_n разбита на k сумм квадратов Q_1, ..., Q_k с f_1, …, f_k степенями свободы, т. е.

$$\chi^2 = \sum_{i=1}^{n} u_i^2 = Q_1 + ... + Q_k, \tag{2.15}$$

то необходимым и достаточным условием того, чтобы величины Q_1, ..., Q_k оказались стохастически независимыми и описывались распределениями с f_1, .., f_k епенями свободы, является выполнение равенства

$$f_1 + f_2 + ... + f_k = n. \tag{2.16}$$

Теорема разложения оказывается чрезвычайно полезной для различных статистических оценок.

3. Функция распределения Фишера (F-распределение). Рассмотрим случайную величину:

$$F(f_1, f_2) = \frac{\chi_1^2}{f_1} \Big/ \frac{\chi_2^2}{f_2} \qquad (0 \le F < \infty), \tag{2.17}$$

где $\chi_{1,2}^2$ определяются согласно формулам (2.12) и (2.13).

Плотность распределения F описывается соотношением Фишера

$$\rho(F) = \frac{\Gamma(\frac{f_1+f_2}{2})}{\Gamma(\frac{f_1}{2})\Gamma(\frac{f_2}{2})} \cdot f_1^{\frac{f_1}{2}} \cdot f_2^{\frac{f_1}{2}} \cdot \frac{(F)^{\frac{f_1}{2}-1}}{(f_2 + f_1 F)^{\frac{f_1+f_2}{2}}}. \tag{2.18}$$

Кривые плотности F-распределения для различных f_1 и f_2 представлены на рис. 2.3.

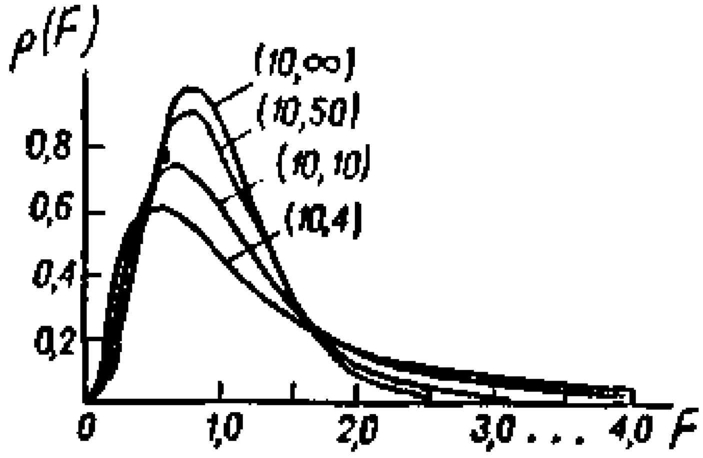

Рис. 2.3. F-распределение при различных f_1 и f_2

Согласно (2.18) среднее значение $\overline{F} = f_2/(f_2 - 2)$, если $f_2 > 2$.

4. Распределение Стьюдента (t-распределение).

Случайная величина

$$t = u/\sqrt{\chi^2/f} \qquad (- \infty < 1 < \infty), \tag{2.19}$$

где u и χ^2 определяются согласно (2.12), (2.13), имеет распределение Стьюдента

$$\rho(t) = \frac{1}{\sqrt{\pi f}} \cdot \frac{\Gamma\left(\frac{f+1}{2}\right)}{\Gamma\left(\frac{f}{2}\right)} \cdot \left(1 + \frac{t^2}{f}\right)^{-\frac{f+1}{2}}, \tag{2.20}$$

которое может быть получено непосредственно из (2.18) преобразованием $t = \pm\sqrt{F(1, f)}$,

где t – распределение симметрично относительно нуля. При $f \to \infty$ распределение стремится к нормальному с параметрами (0,1) (рис. 2.4).

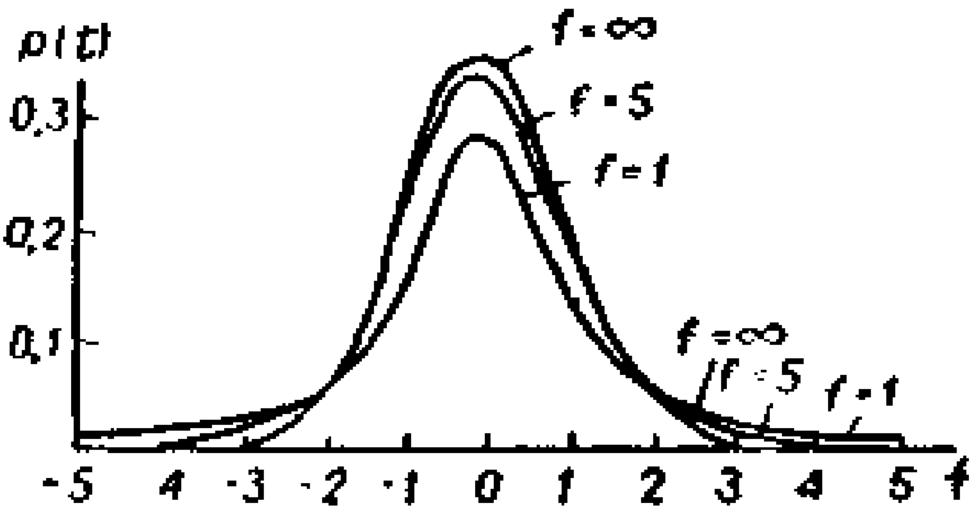

Рис. 2.4. Кривые плотности t-распределения

2.2. ВЫБОРКА И ЕЕ ХАРАКТЕРИСТИКИ

На практике исследователь имеет дело с конечным набором реализаций случайной величины, полученным в результате эксперимента. В очень редких случаях в распоряжении исследователя находится вся генеральная совокупность значений, при этом случайная величина фактически превращается в детерминированную.

Некоторый набор значений случайной величины $(x_1, x_2, x_3, \ldots, x_N)$ называют **выборкой.**

Число полученных экспериментальных результатов N называется объемом выборки. Основная задача математической статистики заключается в том, чтобы по результатам эксперимента (по данным выборки) высказать обоснованное суждение о свойствах генеральной совокупности.

Выборка должна достаточно полно характеризовать генеральную совокупность, она должна быть представительной. Чтобы представительность выборки была обеспечена, необходимо выполнение двух важных условий: во-первых, все элементы генеральной совокупности должны появляться в выборке с одинаковой вероятностью; во-вторых, наблюдения должны быть независимыми, т. е. появление каждого из элементов выборки не должно влиять на вероятность появления других элементов. Так как элементы выборки случайные, все заключения и результаты, полученные на основе выборочных данных, носят вероятностный характер.

2.2.1. Оценки числовых характеристик по выборке

Выборка содержит лишь часть генеральной совокупности, по которой можно попытаться оценить числовые характеристики всей генеральной совокупности. Для этого предложены различные формулы. Следовательно, возможны различные оценки одной и той же числовой характеристики (например, математического ожидания) по результатам одной и той же выборки. Конечно, хотелось бы, чтобы оценки были «хорошими», близкими к той величине, которую они оценивают. Но что это означает? Существуют три меры близости оценки к оцениваемому параметру:

несмещенность – равенство математического ожидания оценки значению соответствующей числовой характеристики при любом объеме выборки;

состоятельность – сходимость по вероятности оценки к оцениваемому параметру (выражающаяся в том, что при увеличении объема выборки оценка приближается к оцениваемому значению числовой характеристики);

эффективность – мера рассеяния оценки в окрестности оцениваемого параметра. Эффективная оценка обладает наименьшей дисперсией по сравнению с другими оценками данной числовой характеристики. Чаще всего в качестве оценки математического ожидания используется среднее арифметическое значение $\bar{x}$.

$$\widehat{M} = \overline{x} = \frac{x_1 + x_2 + ... + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad , \qquad (2.21)$$

где $\widehat{M}[X]$ обозначает оценку $M[X]$. Эта оценка является несмещенной, состоятельной, а для нормального распределения – эффективной.

Другой возможной оценкой математического ожидания является полусумма максимального и минимального значений случайной величины, полученных в выборке.

$$\widehat{M}[X] = \frac{x_{\min} + x_{\max}}{2} \quad . \qquad (2.22)$$

Эта оценка несмещена, состоятельна для симметричных распределений, но эффективна только для равномерного распределения. На практике нередко именно $x_{\min}$ и $x_{\max}$ испытывают наибольший разброс от выборки к выборке, что делает последнюю оценку очень чувствительной к такой неравномерности распределения.

Для нахождения оценок дисперсии $\widehat{\sigma}^2[X]$ (S^2)чаще всего используются следующие формулы:

$$\widehat{\sigma}^2[x] = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2 \qquad (2.23)$$

и

$$\widehat{\sigma}^2[X] = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2 . \qquad (2.24)$$

Оценка (2.23) состоятельна, эффективна для нормального распределения, но смещена. Она, следовательно, носит некоторую систематическую ошибку.

Оценка (2.24) лишена этого недостатка. Она несмещена, состоятельна, но для нормального распределения неэффективна. Обычно предпочитают пользоваться несмещенной оценкой (2.24).

Приведенные выше оценки числовых характеристик генеральной совокупности называются точечными оценками. Получение точечных оценок дает важную информацию, но не позволяет судить о степени их близости к оцениваемой числовой характеристике. Более содержательные процедуры оценивания связаны с построением интервала, который накрывает оцениваемое значение с известной степенью достоверности. Такие процедуры являются интервальными оценками.

2.2.2. Построение доверительного интервала

Примером интервальной оценки является доверительный интервал. **Доверительный интервал** – это отрезок, центром которого является точечная оценка числовой характеристики, включающий истинное значение данной числовой характеристики с заданной вероятностью. Эта вероятность называется **доверительной вероятностью**. Таким образом, доверительный интервал является мерой точности оценки, а доверительная вероятность характеризует достоверность оценки. Размер доверительного интервала зависит от того, каким значением доверительной вероятности задается экспериментатор. Чем выше доверительная вероятность, тем шире должен быть интервал, чтобы с заданной вероятностью включать в себя истинное значение числовой характеристики. Практически часто выбирают значение доверительной вероятности P = 0,95, полагая, что это значение достаточно велико, чтобы считать, что доверительный интервал «практически всегда» накрывает истинное значение. Только иногда, в случае ответственных и очень ответственных исследований полагают P = 0,99 и 0,999 соответственно.

Процедура построения доверительного интервала включает в себя два этапа:

– записывается вероятностное утверждение относительно некоторой случайной функции, включающей в себя разность или отношение оценки и числовой характеристики. Такая функция несет информацию о степени близости этих величин. Необходимо, чтобы закон распределения упомянутой функции был известен;

– вероятностное утверждение преобразуется к виду, при котором границы доверительного интервала числовой характеристики представлены в явном виде.

Примерами функций с известным распределением, которые удовлетворяют необходимым требованиям, являются следующие:

$$1)\ \frac{\bar{x}-M(X)}{\sigma[X]/\sqrt{N}}, \tag{2.25}$$

имеющая нормальное распределение;

$$2)\ \frac{\bar{x}-M(X)}{\sigma[X]/\sqrt{N}}, \tag{2.26}$$

имеющая распределение Стьюдента с числом степеней свободы $m = N - 1$;

$$3)\ \frac{m\hat{\sigma}[X]}{\sigma[X]}, \tag{2.27}$$

имеющая распределение Пирсона с числом степеней свободы $m = N - 1$.

Здесь использованы обозначения: $\bar{x}$ – среднее арифметическое значение; $\sigma^2[X]$ – дисперсия X; $\hat{\sigma}^2[X]$ – несмещенная оценка дисперсии; вычисляемая по формуле (2.24); N – объем выборки.

Построим доверительный интервал для математического ожидания при известной дисперсии. Поскольку функция (2.25) распределена нормально, можно использовать соответствующую таблицу для определения значения Z_α такого, что за пределами $-Z_\alpha$ и $+Z_\alpha$ остается часть площади, равная α, тогда как в пределах $[-Z_\alpha, +Z_\alpha]$ заключена часть площади, равная $1 - \alpha$ (рис. 2.5).

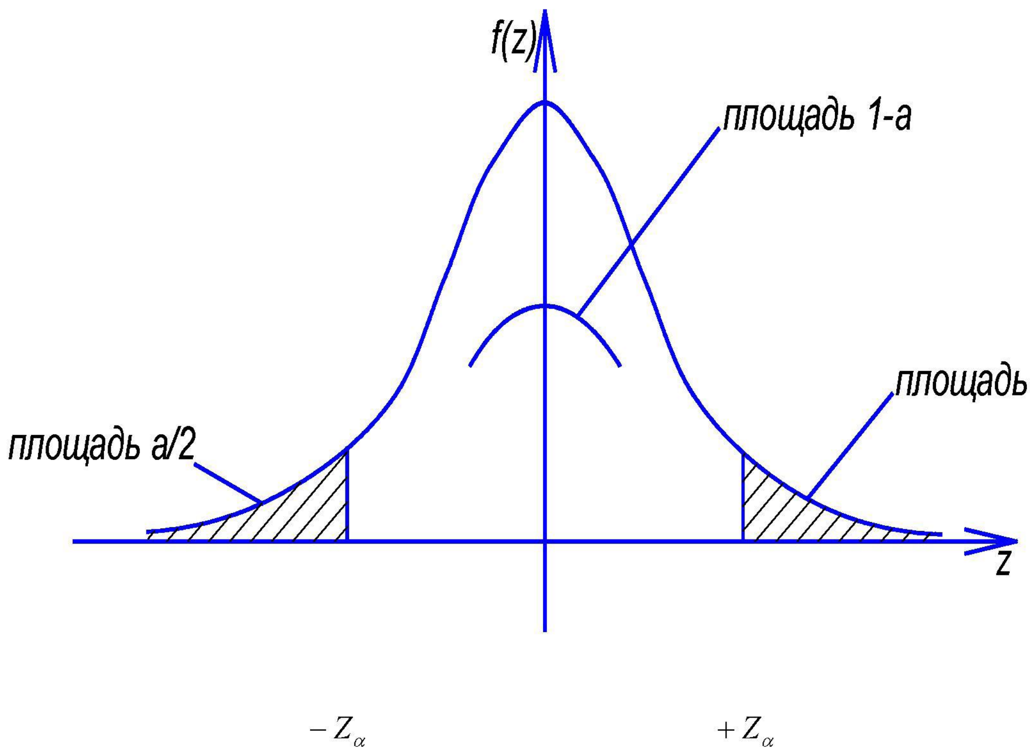

Рис. 2.5. Кривая нормального распределения

Можно вышесказанное записать в виде следующего вероятностного утверждения:

$$P\left\{-Z_\alpha \le \frac{\overline{x} - M(X)}{\sigma[X]/\sqrt{N}} \le Z_\alpha\right\} = 1 - \alpha \qquad (2.28)$$

(вероятность выполнения неравенства, заключенного в фигурных скобках, равна $1 - \alpha$).

Преобразуем выражение в скобках:

$$P\left\{-Z_\alpha \frac{\sigma[X]}{\sqrt{N}} \le \bar{x} - M[X] \le Z_\alpha \frac{\sigma[X]}{\sqrt{N}}\right\} = 1 - \alpha \qquad (2.29)$$

или

$$P\left\{\bar{x} - Z_\alpha \frac{\sigma[X]}{\sqrt{N}} \le M[X] \le \bar{x} + Z_\alpha \frac{\sigma[X]}{\sqrt{N}}\right\} = 1 - \alpha\ . \qquad (2.30)$$

Назовем величину $1 - \alpha = P_Д$ доверительной вероятностью $P_Д$. При этой доверительной вероятности доверительный интервал $M[X]$ задается пределами $\left[\overline{X} - Z_\alpha \frac{\sigma[X]}{\sqrt{N}}, \overline{X} + Z_\alpha \frac{\sigma[X]}{\sqrt{N}}\right]$.

При построении доверительного интервала для математического ожидания $M[X]$ при неизвестной дисперсии используются функция (2.26) и распределение Стьюдента.

При построении доверительного интервала для дисперсии $\sigma^2[X]$ (S^2) используются функция (2.27) и распределение Пирсона.

2.2.3. Проверка статистических гипотез

Статистической гипотезой называется некоторое утверждение о свойствах генеральной совокупности. Например, утверждение о том, что некоторая случайная величина распределена нормально, является статистической гипотезой. Можно также высказывать гипотезы о значениях, числовых характеристиках, их соотношениях и т. д. Любая гипотеза формулируется до опыта и проверяется на основе предыдущего эксперимента. Основная проверяемая гипотеза (нулевая гипотеза) обычно обозначается H_0 . Одновременно формулируется альтернативная гипотеза H_1.

Процедура проверки статистической гипотезы производится при помощи критерия – функции результатов наблюдения, включающей в себя также параметры, по поводу которых высказаны предположения. Необходимо, чтобы случайная функция-критерий имела известный закон распределения. Критерий имеет некоторую область возможных значений, которые он принимает в зависимости от конкретной выборки. В соответствии с характером распределения одни значения критерия являются более вероятными, другие – менее. При проверке статистических гипотез ситуацию огрубляют тем, что область маловероятных значений критерия считают совсем запрещенной: если критерий принял столь маловероятное значение, то гипотеза считается неверной, отвергается. Если же критерий принял достаточно вероятное значение, то считается, что гипотеза не противоречит опыту и может быть принята. Таким образом, область возможных значений делится на две части. Одна называется областью принятия гипотезы, другая (где гипотеза должна быть отвергнута) – критической областью. Чтобы проверить гипотезу, надо вычислить критерий и посмотреть, в какую область попадает вычисленное значение.

Принимая или отклоняя гипотезу H_0, можно допустить ошибки двух видов. Ошибка 1-го рода состоит в том, что гипотеза H_0 отвергается, в то время как в действительности она верна, ошибка 2-го рода – гипотеза H_0 принимается, в то время как верна гипотеза H_1.

Примерами критериев для проверки гипотез могут служить функции (2.25)–(2.27), которые раньше использовались для построения доверительных интервалов.

2.2.4. Ошибки, допускаемые при проверке гипотез

Поскольку вычисления критерия производятся по случайной выборке, то и результат может быть различным. Всегда имеется некоторая вероятность получить как значение, отвергающее гипотезу, так и значение, требующее ее принятия. Иными словами, может быть совершена ошибка.

Итак, высказана некоторая гипотеза, которая может быть объективно верна или неверна. Затем берется выборка, вычисляется критерий и выносится решение принять гипотезу или ее отвергнуть. Возможные варианты приведены в табл. 2.1.

Вероятность совершить ошибку 1-го рода α называется **уровнем значимости**. Вероятность не совершить ошибку 2-го рода $(1 - \beta)$ называется **мощностью критерия**.

Таблица. 2.1

Последствия решений

Гипотеза / Решение	Объективно верна	Объективно неверна
Принять гипотезу	Ошибки нет	Ошибка 2-го рода β верна
Отвергнуть гипотезу	Ошибка 1-го рода α верна	Ошибки нет

Вероятности ошибок 1-го и 2-го рода зависят от выбора размеров области принятия гипотезы и критической области. При фиксированном объеме выборки N стремление уменьшить ошибку 1-го рода с необходимостью приводит к увеличению ошибки 2-го рода. Обычно, исходя из конкретной ситуации, задаются приемлемой вероятностью ошибки 1-го рода (уровнем значимости). Типичное значение выбираемого уровня значимости $\alpha = 0{,}05$. Задание уровня значимости однозначно определяет критическую область, а следовательно, и область принятия гипотезы.

Распишем по этапам процесс проверки статистической гипотезы.

1. Формулировка гипотезы H_0: обычно гипотеза высказывается в форме, отрицающей наличие каких-либо видимых эффектов. Поэтому основная гипотеза H_0 называется нулевой.

2. Формулировка альтернативной гипотезы H_1: нулевой гипотезе может быть много альтернатив. Нужно выбрать одну, которая является более существенной.

3. Выбор уровня значимости α – выбор значения вероятности ошибки 1-го рода, которое представляется приемлемым.

4. Выбор критерия для проверки гипотезы H_0.

5. Нахождение функции распределения критерия при условии, что нулевая гипотеза H_0 справедлива.

6. Нахождение критической области: при этом используются сведения о функции распределения критерия, альтернативной гипотезе H_1 и уровне значимости α.

7. Извлечение выборки $(x_1, x_2, \ldots, x_N)$.

8. Вычисление по результатам выборки значения критерия.

9. Принятие решения: если вычисленное значение попадает в критическую область, то гипотеза H_0 отвергается, если же оно попадает в область принятия гипотезы, то она считается допустимой, т. е. данные выборки не противоречат гипотезе H_0.

2.2.5. Примеры проверки некоторых гипотез

I. Проверка гипотезы о равенстве математического ожидания некоторому числу при известной дисперсии

Проведем проверку на конкретном примере.

Осуществляется литье изделий из алюминия. Случайная величина X – масса изделия. На основании предыдущего опыта известно, что X – нормально распределенная величина с $\sigma^2[X]$ = 25 г^2. Номинальное значение массы изделия должно быть равным 100 г. Необходимо извлечь выборку объемом N = 20 изделий и решить вопрос о том, действительно ли в среднем значение изделий равно 100 г и процесс не нуждается в переналадке.

Решение.

1. Нулевая гипотеза H_0: $M(X)$ = 100 г.

2. Альтернативная гипотеза H_1: $M(X) \neq$ 100 г (в принципе могли быть альтернативные гипотезы $M(X)$ < 100 г или $M(X)$ > 100 г, но нас интересует, нет ли отклонения любого знака от номинального значения).

3. Выберем уровень значимости α = 0,05. Это значит, что мы готовы в 5 % случаев ошибочно рекомендовать переналадку оборудования. Если мы не хотим мириться с таким процентом ошибок 1-го рода, то можем выбрать, например, α = 0,01, но тогда мы должны быть готовы к увеличению числа ошибок 2-го рода, вследствие чего возрастут жалобы потребителей на несоответствие массы указанному номиналу.

4. Критерием для проверки гипотезы может служить функция $Z = \dfrac{\overline{X} - M(X)}{\sigma\ [X]/\sqrt{N}}$, поскольку в нее входят $M(X)$ и $\sigma[X]$.

5. Функция распределения критерия $\dfrac{\overline{X} - M(X)}{\sigma[X]/\sqrt{N}}$ при условии выполнения нулевой гипотезы является нормальной.

6. Для построения критической области следует воспользоваться таблицей нормального распределения и найти Z_α при α = 0,05. Оно равно, согласно таблице приложения 1, значению 1,96. Критическая область будет расположена на крыльях распределения, попадание куда маловероятно.

Площадь кривой плотности распределения в заштрихованных крыльях, согласно выбранному уровню значимости, должна составлять 0,05. Критическими значениями $Z_{\alpha=0,05}$ = 1,96 и – $Z_{\alpha=0,05}$ = – 1,96 отделяются критические области от области принятия гипотезы. Альтернативная гипотеза H_1 была использована при определении расположения критических областей. Поскольку она не различает отклонений математического ожидания в большую или меньшую сторону, критические области должны включать оба крыла распределения, как это показано на рис. 2.6. Если бы альтернативная гипотеза имела, например, вид $M(X) < 100$ г, критическая область располагалась бы целиком слева.

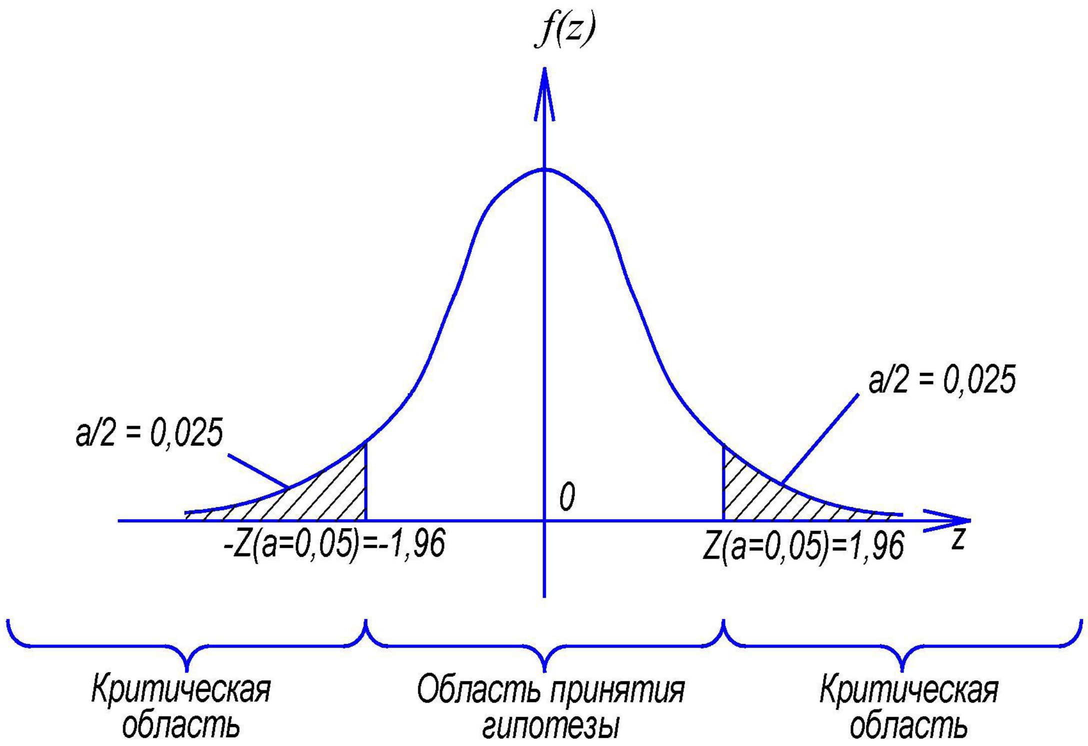

Рис. 2.6. Кривая нормального распределения, расположение области применения гипотезы и критических областей

7. Пусть извлечена выборка объемом N = 20, элементы которой представлены в таблице.

Масса, г	102,3	100,3	107,4	105,1	107,0	104,5	105,8	92,5	96,6	106,8
	97,6	93,1	95,0	100,0	107,0	91,6	100,5	93,4	94,2	101,3

Вычисляем сначала $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = 100{,}1$ г, а затем значение критерия

$$\frac{\bar{x} - M(X)}{\sigma[X]/\sqrt{N}} = \frac{100{,}1 - 100}{5/\sqrt{20}} = 0{,}09 .$$

8. Значение критерия 0,09 попадает в область принятия гипотезы $\left[-Z_\alpha = -1{,}96; Z_\alpha = 1{,}96\right]$. Следовательно, нет основания отвергать гипотезу H_0 и переналаживать производство.

Таким образом, сравнение двух математических ожиданий сводится к проверке гипотезы:

$$H_0 : \quad M[Y] = \mu_0,$$

при H_1 : $M[Y] \neq \mu_0$ (μ_0 – некоторое значение).

Выборка наблюдений объемом m для проверки нулевой гипотезы осуществляется из генеральной совокупности значений случайной величины Y, распределенных по нормальному закону.

Оценкой математического ожидания по выборке будет среднее арифметическое результатов параллельных измерений $\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i$, а оценкой дисперсии – $S^2 = \frac{1}{m-1}\sum_{i=1}^{m}(y_i - \bar{y})^2$. Для проверки нулевой гипотезы вычисляется наблюдаемое (расчетное) значение критерия Стьюдента

$$t_{расч} = \frac{|\,\bar{y} - \mu_0\,|}{S\sqrt{m}} .$$

Критическое значение критерия $t_{табл}$ определяется для выбранной доверительной вероятности $P = 1 - \alpha$ (или уровня значимости) и заданного объема выборки (числа степеней свободы) из таблицы приложения 2. При $t_{расч} < t_{табл}$ можно считать, что данные выборки не противоречат нулевой гипотезе; если $t_{расч} > t_{табл}$, гипотеза отвергается.

II. Проверка гипотезы о равенстве средних значений

Пусть выборки наблюдений объемом m_1 и m_2 берутся из двух генеральных совокупностей с нормальным распределением, причем дисперсии генеральных совокупностей σ_1^2 и σ_2^2 равны. Необходимо проверить гипотезу H_0 о равенстве математических ожиданий $M(Y_1) = M(Y_2)$.

По данным выборок определяются оценки математических ожиданий и дисперсий. Объединенная оценка дисперсии генеральных совокупностей

$$S^2 = [\,S_1^2\,(m_1 - 1) + S_2^2\,(m_2 - 1)]/(\,m_1 + m_2 - 1). \qquad (2.31)$$

Для проверки нулевой гипотезы вычисляется наблюдаемое (расчетное) значение критерия Стьюдента

$$t_{расч} = \frac{|\,\bar{y}_1 - \bar{y}_2\,|}{S}\sqrt{\frac{m_1 m_2}{m_1 + m_2}} \qquad (2.32)$$

Критическое значение критерия определяется для данной доверительной вероятности P и $m = m_1 + m_2$. При $t_{расч} < t_{табл}$ гипотеза принимается; если $t_{расч} > t_{табл}$, она отвергается.

Пример. В одинаковых условиях было обработано по 25 втулок развертками диаметром 6 и 10 мм. Результаты измерений показали, что средняя величина разбивки отверстий (разность диаметра развертки и диаметра отверстия) для разверток диаметром 6 мм составляет 10,9 мкм, а диаметром 10 мм – 9,8 мкм. Оценки дисперсий в обоих случаях соответственно S_1^2 = 3,8 мкм^2; S_2^2 = 4,76 мкм^2. Необходимо установить, влияет ли диаметр развертки на разбивку отверстия.

Согласно формуле (2.31)

$$S^2 = \frac{3{,}8 \cdot 24 + 4{,}76 \cdot 24}{49} = 4{,}19,$$

согласно формуле (2.32)

$$t_{расч} = \frac{10,9 - 9,8}{\sqrt{4,19}} \sqrt{\frac{25 \cdot 25}{25 + 25}} = 1,9.$$

По таблице приложения 2 при $\alpha = 0,05$ и $m = 50$ находим $t_{табл} = 2,01$. Поскольку $t_{расч} < t_{табл}$ (1,9 < 2,01), можно считать, что изменение диаметра развертки в пределах от 6 до 10 мм не оказывает существенного влияния на величину разбивки развернутого отверстия.

III. Проверка гипотезы о равенстве дисперсий

При обработке экспериментальных данных часто требуется выяснить вопрос об однородности выборочных дисперсий, т. е. их равенстве дисперсии генеральной совокупности.

Пусть для двух независимых выборок из нормальной генеральной совокупности с объемами m_1 и m_2 вычислены оценки S_1^2 и S_2^2 (согласно $S^2 = \frac{\sum_{i=1}^{m}(Y_i - \bar{Y})^2}{f}$ $f = m - 1$, Y_i – текущее значение случайной величины, f – число степеней свободы). Требуется проверить нулевую гипотезу (H_0) $\sigma_1^2 = \sigma_2^2$ относительно альтернативной гипотезы (H_1) $\sigma_1^2 > \sigma_2^2$. Проверка проводится при помощи критерия Фишера (F-критерия).

Наблюдаемое (расчетное) значение критерия

$$F_{расч} = S_1^2 \ / \ S_2^2, \qquad S_1^2 > S_2^2 \tag{2.33}$$

сравнивается с критическим $F_{табл}$, которое определяется из таблицы приложения 3 для выбранных уровней значимости (доверительной вероятности) и m_1 и m_2. Если $F_{расч} < F_{табл}$, то выборочные данные не противоречат нулевой гипотезе.

При анализе выборочных данных могут выдвигаться гипотезы об однородности дисперсий в нескольких выборках. В этом случае можно использовать критерий Кохрена. Наблюдаемое (расчетное) значение критерия $G_{расч}$ определяется по формуле

$$G_{расч} = S^2_{i\,max} / \sum_{i=1}^{n} S_i^2 \ , \qquad (2.34)$$

где $S^2_{i\,max}$ – максимальная оценка дисперсии среди n сравниваемых дисперсий (все n выборок имеют одинаковый объем m).

Критическое значение критерия определяется из таблицы приложения 4 в зависимости от принятых доверительной вероятности (уровня значимости), объема выборок m и их числа n.

Пример. Проверить гипотезу о равенстве дисперсий в примере, рассмотренном выше.

По формуле (2.33) определяем наблюдаемое значение критерия Фишера $F_{расч}$ = 4,76/3,8 =1,25. Для P = 0,95, $m_1 = m_2$ = 20 находим $F_{табл}$ = 2,16, а для $m_1 = m_2$ =30 – $F_{табл}$ = 1,84. Искомое значение $F_{табл}$ для $m_1 = m_2$ = 25 находим линейной интерполяцией: $F_{табл}$ = (1,84+2,16)/2=2.

Так как $F_{расч} < F_{табл}$ (1,25 < 2), нулевая гипотеза о равенстве дисперсий принимается.

Пример. С четырех автоматов, настроенных на обработку одних и тех же деталей, взято по одной текущей выборке объемом m = 10. Оценки дисперсий их размеров имели следующие значения:

S_1^2 = 106 мкм^2; S_2^2 = 294 мкм^2; S_3^2 =216 мкм^2; S_4^2 = 410 мкм^2. Требуется определить, одинакова ли точность автоматов, т. е. можно ли принять гипотезу об однородности дисперсий.

Согласно (2.34) имеем

$$G_{расч} = \frac{410}{106 + 294 + 216 + 410} = 0{,}3996.$$

По таблице приложения 4 для *m(f)* = 10, *n(k)* = 4 находим $G_{табл}$ = 0,4884. Поскольку $G_{расч} < G_{табл}$ (0,3996 < 0,4884), то можно считать точность автоматов практически одинаковой.

IV. Проверка случайности и независимости результатов измерений в выборке

До статистической обработки результатов измерения отклика необходимо убедиться в том, что они являются стохастически независимыми. Альтернативной гипотезой может быть предположение о наличии монотонного или циклического смещения (дрейфа) значения отклика, вызванного неконтролируемым фактором. Подобный случай может иметь место при анализе размеров деталей, обрабатываемых на настроенном станке, когда вследствие изнашивания инструмента или нагрева станка центр группирования размеров постепенно смещается при неизменной стандартной погрешности. Наиболее мощным критерием проверки нулевой гипотезы будет критерий последовательных разностей τ.

Наблюдаемое (расчетное) значение критерия

$$\tau_{расч} = c^2/S^2,$$

где $c^2 = \frac{1}{2(m-1)}\sum_{i=1}^{m-1}(Y_{i+1} - Y_i)^2$; *m* – объем выборки; *i* – порядковый номер измерения отклика в выборке; S^2 – оценка дисперсии.

Критическое значение $\tau_{табл}$ определяется из таблицы приложения 6 в зависимости от принятой доверительной вероятности P и объема выборки m. Если $\tau_{расч} < \tau_{табл}$, то гипотеза о независимости и случайности измерений в выборке отвергается.

Пример. По результатам измерения деталей, обработанных на токарно-револьверном автомате, необходимо проверить наличие или отсутствие дрейфа размеров. Объем выборки $m = 40$.

$$\bar{y} = \frac{1}{40}\sum_{i=1}^{40} Y_i = 45 \text{ мкм},$$

$$S^2 = \frac{1}{40-1}\sum_{i=1}^{40}(Y_i - \bar{y})^2 = 267{,}95 \text{ мкм}^2,$$

$$c^2 = \frac{1}{2(40-1)}\sum_{i=1}^{39}(Y_{i+1} - Y_i)^2 = 114{,}74 \text{ мкм}^2,$$

$\tau_{расч} = 114{,}74 / 267{,}95 = 0{,}428.$

По таблице приложения 6 для $m = 40$ и $P = 0{,}95$ получаем $\tau_{табл} = 0{,}742$. Так как $\tau_{расч} < \tau_{табл}$ ($0{,}428 < 0{,}742$), гипотезу об отсутствии дрейфа следует отклонить. Следовательно, размер обрабатываемых деталей зависит от неучтенного фактора, вызвавшего циклическое смещение центра группирования размеров.