ББК В172

Методические указания содержат краткое изложение основных понятий и сведений теоретического характера по математической статистике, необходимых для выполнения студентами двух лабораторных работ. В эту часть лабораторного практикума по статистике включены задачи, относящиеся к построению точечных и интервальных оценок параметров распределений, а также к методам проверки статистических гипотез. Приведены варианты заданий, задачи. Даны примеры их решения. В приложении содержатся статистические таблицы, необходимые в этом лабораторном цикле. При выполнении данных лабораторных работ рекомендуется использовать возможности современных систем автоматизации математических расчетов, таких как **MatLAB** [11],[12] или **MATHCAD** [5],[10]. Это сделает работу приятной, так как избавит от утомительных рутинных вычислений.

1. ОСНОВНЫЕ ПОНЯТИЯ ВЫБОРОЧНОГО МЕТОДА

Исходным материалом любого исследования в математической статистике служат числовые данные, полученные в результате эксперимента. Обычно предполагается, что исследователь провел n опытов, эти опыты производились независимо, а условия, в которых они производились, оставались неизменными. Если мы вспомним исходные конструкции теории вероятностей, то увидим, что снова имеем дело с моделью *случайного эксперимента*, с которым можно связать *вероятностное пространство* (Ω,F,P), где Ω - множество элементарных исходов эксперимента; F - заданное на этом множестве поле событий; P - вероятностная мера, определенная на поле F и сопоставляющая каждому событию $A \in F$ его вероятность $\mathrm{P}(A)$. Предполагается, что в каждом из n опытов исследователь производил измерения и что результаты измерений - это значения, которые принимает в итоге эксперимента наблюдаемая случайная величина (с.в.) ξ (возможно, векторная). Вероятностное пространство (Ω,F,P) и опре-

деленная на этом пространстве с.в. ξ называются *генеральной совокупностью.*

Конкретный элементарный исход эксперимента $\omega \in \Omega$ и соответствующее этому исходу значение с.в. $\xi(\omega)$ (результат эксперимента) называют *единицей генеральной совокупности.* Результат одного опыта (наблюдения случайной величины ξ) трактуют как выбор единицы генеральной совокупности, а весь полученный набор наблюдений называют *выборкой* из генеральной совокупности.

Определение 1.1. Набор $X_1, X_2,\ldots,X_n$, состоящий из n независимых наблюдений с.в. ξ, называется *выборкой* объемом n из генеральной совокупности с.в. ξ.

Набор $X_1,X_2,\ldots,X_n$ можно трактовать, таким образом, как совокупность значений, принятых n независимыми с.в. $\xi_1,\xi_2,\ldots,\xi_n$, имеющими тот же закон распределения $F_\xi(x)=\mathrm{P}(\xi<x)$, что и наблюдаемая с.в. ξ, называемый *теоретическим распределением* или *законом распределения генеральной совокупности.* Значения $X_1,X_2,\ldots,X_n$ называют *выборочными значениями.*

Замечание 1.1. В зависимости от задачи $X_1,X_2,\ldots,X_n$ будут обозначать либо n независимых одинаково распределенных с.в., либо n числовых значений, полученных в результате наблюдений (реализаций с.в.ξ).

Закон распределения генеральной совокупности, как правило, неизвестен (хотя бы частично), и одна из основных задач математической статистики состоит в получении на основе результатов наблюдений выводов о характере этого закона распределения и о его параметрах.

Все задачи математической статистики делятся на две большие группы: *параметрические* и *непараметрические*. Поясним, что означают эти понятия. Функция распределения $F_\xi(x)$ генеральной совокупности неизвестна, но на основе имеющейся априорной информации часто можно утверждать, что эта функция принадлежит некоторому множеству $\mathcal{F}$. Чем "шире" это множество, тем больше неопределенность и тем меньшую точность можно обеспечить при решении задач статистики. Поэтому при постановке задачи желательно "сузить" это множество на основе априорной информации. На практике часто можно считать, что неизвестная функция распределения принадлежит некоторому конкретному семейству распределений, зависящих от одного или нескольких параметров.

Примерами таких семейств распределений являются нормальный закон $N(a,\sigma)$, зависящий от двух параметров: a и $\sigma>0$ (где a - математическое ожидание, σ - среднее квадратическое отклонение), показательное распределение, зависящее от одного параметра $\lambda>0$ и другие модельные, изучаемые обычно в теории вероятностей распределения (биноминальное, пуассоновское и др.). Задача называется *параметрической*, если известен вид распределения, зависящего от параметра $\theta=(\theta_1,\theta_2,\ldots,\theta_k)$ и область изменения параметра θ. Если представления о виде распределения нет, то говорят о *непараметрических задачах.*

Замечание 1.2. О непараметрической ситуации часто говорят как о случае с бесконечномерным параметром, где в роли параметра θ может выступать сама неизвестная функция распределения F_ξ.

Одной из основных задач математической статистики является оценка параметров распределений.

Определение 1.2. Оценкой параметра θ (в том числе и в случае бесконечномерного параметра $\theta=F_\xi(x)$) называется функция выборочных значений $\hat\theta_n=\hat\theta_n(X_1,X_2,\ldots,X_n)$, которая близка к истинному значению параметра θ в определенном статистическом смысле.

Близость оценки к истинному значению параметра определяют следующие основные статистические свойства: *состоятельность*, *несмещенность* и *эффективность*.

Определение 1.3. Оценка $\hat\theta_n=\hat\theta_n(X_1,X_2,\ldots,X_n)$ называется *состоятельной* оценкой параметра θ, если при $n \longrightarrow \infty$ имеет место сходимость по вероятности

$$\hat\theta_n \xrightarrow{\mathrm{P}} \theta, \tag{1.1}$$

$\left[\text{что означает, что для } \forall\ \varepsilon > 0 \ \ \mathrm{P}\{|\hat\theta_n-\theta| < \varepsilon\} \to 1 \text{ при } n \to \infty\right]$.

Определение 1.4. Оценка $\hat\theta_n=\hat\theta_n(X_1,X_2,\ldots,X_n)$ называется *несмещенной*, если

$$\mathrm{E}_\theta\, \hat\theta_n=\theta, \tag{1.2}$$

где E_θ - знак взятия математического ожидания по распределению генеральной совокупности со значением параметра, равным истинному значению. Несмещенность означает, что оценка не дает систематического занижения или завышения результата.

Определение 1.5. Несмещенная оценка $\hat{\theta}_n$ называется эффективной, если она имеет минимальную дисперсию в определенном классе оценок.

Замечание 1.3. К числу желательных свойств оценки при современной статистической обработке относят также *робастность* (robust, англ. - устойчивый, крепкий, сильный). Под робастностью понимают устойчивость основных статистических свойств оценки к присутствию в выборке небольшого количества некачественных данных (резко выделяющихся наблюдений). Робастные процедуры оценки параметров [14] сводятся обычно к отбрасыванию части выборочных значений (урезанию выборки). Рекомендации по методам нейтрализации влияния на оценку посторонних наблюдений имеются в литературе по робастной статистике [14],[15].

В заключение этого раздела введем еще некоторые основные понятия.

Определение 1.6. Упорядоченная по величине последовательность выборочных значений

$$X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$$

называется *вариационным рядом*. (Равные между собой члены вариационного ряда нумеруются в произвольном порядке.) Элементы вариационного ряда $X_{(i)}$, $i=1,2,\ldots,n$, называются *порядковыми статистиками*.

Замечание 1.4. В случае непрерывного распределения генеральной совокупности вероятность совпадения выборочных значений, очевидно, равна нулю, поэтому в определении вариационного ряда можно в этом случае ставить знаки строгого неравенства:

$$X_{(1)} < X_{(2)} < \ldots < X_{(n)}.$$

Определение 1.7. Величина

$$W = X_{(n)} - X_{(1)}$$

называется *размахом выборки*.

Определение 1.8. Порядковая статистика $X_{([np]+1)}$, где $0 < p < 1$, $[np]$ - целая часть числа np, называется *выборочной квантилью уровня* p (или p 100% - ой выборочной квантилью).

Определение 1.9. *Выборочной медианой* называется статистика, определяемая равенством:

$$\xi_{0.5} = \begin{cases} X_{([n/2]+1)}, & \text{если } n \text{ - нечетное,} \\ \left[X_{(n/2)} + X_{(n/2+1)}\right] / 2, & \text{если } n \text{ - четное.} \end{cases}$$

2. ОЦЕНКА ФУНКЦИИ РАСПРЕДЕЛЕНИЯ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Функция распределения (ф.р.) является наиболее полной характеристикой случайной величины. Пусть $X_1, X_2, \ldots, X_n$ - выборка из генеральной совокупности с.в. ξ, принимающей значения в множестве вещественных чисел R. Рассмотрим задачу оценивания ф.р. $F_\xi(x)$. По определению $F_\xi(x)$ есть вероятность события $\left\{\xi < x\right\}$. Статистическим аналогом вероятности события является относительная частота наступления события в эксперименте. Обозначим через $\nu_n(x)$ число выборочных значений X_i, меньших x. Тогда относительная частота наступления события $\left\{\xi < x\right\}$ равна $\frac{\nu_n(x)}{n}$.

Определение 2.1. *Эмпирической функцией распределения* с.в. ξ называется функция $F_n(x)$, которая при всех значениях $x \in$ R определяется равенством

$$\hat{F}_n(x) = \frac{\nu_n(x)}{n} \quad . \tag{2.1}$$

Если $X_{(i)}$, $i=1,2\ldots n$, обозначают элементы вариационного ряда, причем все $X_{(i)}$ различны, то, учитывая определение (2.1), можно записать $\hat{F}_n(x)$ в виде

$$\hat{F}_n(x) = \begin{cases} 0, & x \leqslant X_{(1)} \\ \frac{k}{n}, & X_{(k)} < x \leqslant X_{(k+1)}, \quad k=1,\ldots,n-1 \\ 1, & X_{(n)} < x . \end{cases} \tag{2.2}$$

График эмпирической ф.р. $\hat{F}_n$ - это график ступенчатой функции со скачками, кратными $1/n$ в точках, определяемых элементами вариационного ряда (см. рис.2.1).

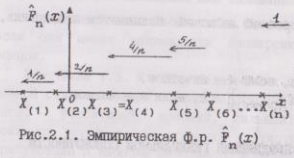


Рис.2.1. Эмпирическая ф.р. $\hat{F}_n(x)$

Эмпирическая ф.р. обладает всеми свойствами ф.р., так как она и есть ф.р.дискретной с.в., принимающей значения $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ с вероятностями $1/n$ каждое. Эмпирическая ф.р. называется еще выборочной ф.р. и представляет собой оценку неизвестной ф.р. генеральной совокупности $F_\xi(x)$. Справедливо следующее утверждение.

Теорема 2.1. При всех $x \in$ R

$$\mathrm{E}(\hat{F}_n(x)) = F_\xi(x), \qquad \mathrm{D}(\hat{F}_n(x)) = \frac{F_\xi(x)(1-F_\xi(x))}{n} \tag{2.3}$$

и имеет место сходимость по вероятности:

$$\hat{F}_n(x) \xrightarrow{P} F_\xi(x) \tag{2.4}$$

при $n \to \infty$.

Первое равенство в (2.3) означает, что $F_n(x)$ есть несмещенная оценка $F_\xi(x)$:

$$\mathrm{E}(\hat{F}_n(x)) = \frac{1}{n}\ \mathrm{E}\left(\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\ n\, F_\xi(x) = F_\xi(x).$$

Соотношение (2.4) означает, по-существу, состоятельность эмпирической ф.р. $\hat{F}_n(x)$ как оценки ф.р. $F_\xi(x)$ при всех x.

Можно доказать, что в действительности эмпирическая ф.р. сходится к $F_\xi(x)$ в более сильном смысле: равномерно по $x \in$ R с вероятностью единица. Справедливо следующее утверждение.

Теорема 2.2. (Гливенко-Кантелли)

$$\mathrm{P}\left\{\lim_{n\to\infty} \sup_{x\in R} |\hat{F}_n(x) - F_\xi(x)| = 0 \right\} = 1.$$

Величину отклонения $\hat{F}_n(x)$ от $F_\xi(x)$ измеряют обычно при помощи одного из следующих двух расстояний (метрик):
метрики Колмогорова

$$D_n = \sup_{x\in R} |\ \hat{F}_n(x) - F_\xi(x)\ | \tag{2.7}$$

или метрики фон Мизеса - Смирнова

$$\omega_n^2 = \int_{-\infty}^{\infty} \left[\hat{F}_n(x) - F_\xi(x)\right]^2 \mathrm{d}F_\xi(x) =$$

$$= \frac{1}{12n^2} + \frac{1}{n}\sum_{k=1}^{n}\left[F_\xi(X_{(k)}) - \frac{2k-1}{2n}\right]^2. \tag{2.8}$$

Метрика D_n называется также равномерной, т.к. приписывает отклонениям $\hat{F}_n(x)$ от $F_\xi(x)$ одинаково важное значение при всех значениях x, тогда как ω_n^2 учитывает с большим весом отклонения в наиболее вероятных точках и с меньшим весом - в точках маловероятных значений.

Теорема 2.3. (А.Н.Колмогоров). Если $F_\xi(x)$ непрерывна, то при $n \to \infty$

$$\mathrm{P}\ \{\ \sqrt{n}\ D_n < z\ \} \to K(z) = \begin{cases} \sum\limits_{k=-\infty}^{\infty} (-1)^k\, e^{-2k^2 z^2}, & z > 0, \\ 0, & z \leqslant 0. \end{cases}$$

Функция $K(z)$ называется ф.р. Колмогорова. Значения $K(z)$ табулированы (см. приложения). Теорема позволяет определить доверительную полосу для ф.р. $F_\xi(x)$: если z_α - решение уравнения $K(z) = \alpha$, то с вероятностью, близкой к α при достаточно больших n и при всех $x \in$ R, справедливо неравенство

$$|\hat{F}_n(x) - F_\xi(x)| < \frac{z_\alpha}{\sqrt{n}} \ ,$$

или

$$\max\left\{0,\ \hat{F}_n(x) - \frac{z_\alpha}{\sqrt{n}}\right\} < F_\xi(x) < \min\left\{1, \hat{F}_n(x) + \frac{z_\alpha}{\sqrt{n}}\right\}. \tag{2.9}$$

Так, для $\alpha=0.95$ из таблиц $K(z)$ находим $z_{0.95} = 1.358$ и, следовательно, с вероятностью, близкой к 0.95,

$$\hat{F}_n(x) - \frac{1.358}{\sqrt{n}} < F_\xi(x) < \hat{F}_n(x) + \frac{1.358}{\sqrt{n}} \ . \tag{2.10}$$

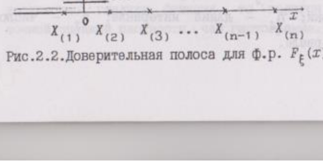


Рис.2.2.Доверительная полоса для ф.р. $F_\xi(x)$

Расстояние ω_n^2 при $n\to\infty$ также сходится по распределению к пределу. Если $F_\xi(x)$ непрерывна, то при $z > 0$

$\mathrm{P}\{\omega_n^2 < z\} \to S(z)$, где $S(z)$ – ф.р. Смирнова, значения которой также табулированы.

3. ОЦЕНКА ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

Пусть $X_1, X_2, \ldots, X_n$ – выборка из генеральной совокупности с.в. ξ, $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ – соответствующий ей вариационный ряд. Плотность распределения $p(x) = \frac{d}{dx}(F(x))$ оценивается обычно двумя способами: посредством гистограммы или полигона частот. Для их построения интервал значений от $X_{(1)}$ до $X_{(n)}$, содержащий все выборочные значения, разбивается на k непересекающихся интервалов точками $a_0 = X_{(1)} < a_1 < a_2 < \ldots < X_{(n)} = a_k$ (рис.3.1).

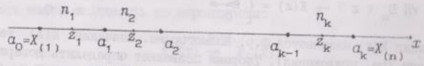


Рис.3.1. Группировка наблюдений

Обычно строят интервалы одинаковой длины $h = \frac{X_{(n)} - X_{(1)}}{k}$, хотя это не обязательно. Обозначим $h_j = a_j - a_{j-1}$, $j = 1,2\ldots k$, длину j-го интервала, $z_j = \frac{1}{2}(a_{j-1}+a_j)$ – середину j-го интервала. Пусть n_j – число выборочных значений X_i, попавших в j-й интервал (элемент, совпадающий с нижней границей j-го интервала, относится к j-му интервалу).

Определение 3.1. Гистограммой относительных частот называется кусочно-постоянная функция $g(x)$, $x \in [X_{(1)}, X_{(n)}]$, постоянная в интервалах $(a_{j-1}, a_j]$, $j=1,2,\ldots,k$:

$$g(x) = g_j = \frac{n_j}{nh_j} \quad \text{для } x \in (a_{j-1}, a_j], \tag{3.1}$$

где n – объем выборки; h_j – длина интервала; n_j – число наблюдений в j-ом интервале.

Графиком гистограммы является ступенчатая линия (рис. 3.2). Площадь S_g ступенчатой фигуры под графиком гистограммы (как и площадь фигуры под графиком плотности распределения) равна единице:

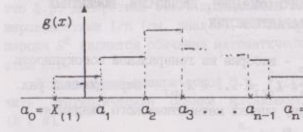


Рис.3.2. График гистограммы

$$S_g = \sum_{j=1}^{k} \frac{n_j}{nh_j} h_j = \frac{1}{n} \sum_{j=1}^{k} n_j = 1.$$

Другой способ оценки плотности – полигон частот.

Определение 3.2. *Полигоном относительных частот* называется кусочно-линейная функция $\hat{P}_n(x)$, $x \in [z_1, z_k]$, линейная в интервалах $[z_j, z_{j+1}]$, $j=1,2,\ldots,k-1$, причем

$$\hat{P}_n(z_j) = \frac{n_j}{nh_j}, \quad j = 1,2\ldots k. \tag{3.2}$$

Графиком $\hat{P}_n(x)$ является ломаная линия с вершинами в точках (z_j, g_j), $j=1,\ldots,k$ (рис.3.3).

Если при $n \to \infty$ $\max\limits_{1 \leqslant j \leqslant k} h_j \to 0$, а $n \min\limits_{1 \leqslant j \leqslant k} h_j \to \infty$, то гистограмма и полигон частот сходятся по вероятности к плотности распределения

$$p_\xi(x) = F'_\xi(x).$$

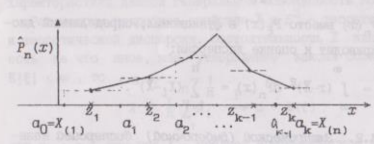


Рис.3.3. График полигона относительных частот

Причем, если плотность распределения $p_\xi(x)$ – достаточно гладкая функция, то полигон $\hat{P}_n(x)$ является более точной оценкой плотности, чем гистограмма.

4. ОЦЕНКА МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ, ДИСПЕРСИИ, МОМЕНТНЫХ ХАРАКТЕРИСТИК

Пусть $X_1, X_2, \ldots, X_n$ – выборка из генеральной совокупности с.в. ξ с ф.р. $F_\xi(x)$, $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ – вариационный ряд.

Рассмотрим вначале задачу оценки математического ожидания:

$$a = \mathrm{E}\xi = \int_{-\infty}^{\infty} x \, dF_\xi(x). \tag{4.1}$$

Подставим в правую часть (4.1) вместо неизвестной ф.р. $F_\xi(x)$ ее оценку, т.е. эмпирическую ф.р. $\hat{F}_n$. Тогда по свойству интеграла Стилтьеса (см. дополнение 3) для ступенчатой функции $\hat{F}_n(x)$ получаем:

$$\int_{-\infty}^{\infty} x \, d\hat{F}_n(x) = \sum_{i=1}^{n} X_{(i)} \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} X_{(i)} = \frac{1}{n} \sum_{i=1}^{n} X_i .$$

Подстановка эмпирической ф.р. $\hat{F}_n(x)$ в функционал, определяющий математическое ожидание, приводит к оценке теоретического среднего значения посредством среднего арифметического элементов выборки.

Определение 4.1. *Эмпирическим (выборочным) средним* значением называется величина

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i . \tag{4.2}$$

Подстановка $\hat{F}_n(x)$ вместо $F_\xi(x)$ в функционал, определяющий дисперсию с.в. ξ приводит к оценке дисперсии:

$$S^2 = \int_{-\infty}^{\infty} (x-\bar{X})^2 \, d\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} (X_i-\bar{X})^2.$$

Определение 4.2. *Эмпирической (выборочной) дисперсией* называется величина

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i-\bar{X})^2, \tag{4.3}$$

где $\bar{X}$ – эмпирическое среднее.

Замечание 4.1. Отметим, что так как эмпирическая ф.р. $\hat{F}_n(x)$ – это ф.р. дискретной с.в., принимающей значения X_i, $i = 1,\ldots,n$, с вероятностями $1/n$ (см. разд.2), то эмпирические среднее $\bar{X}$ и дисперсия S^2 являются обычными математическим ожиданием и дисперсией этой с.в.

Подстановкой эмпирической ф.р. $\hat{F}_n(x)$ вместо $F_\xi(x)$ можно получить статистические оценки начального момента k-го порядка ($k \in N$)

$$a_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k ; \tag{4.4}$$

центрального момента к-го порядка

$$\mu_k = \frac{1}{n} \sum_{i=1}^{n} (X_i-\bar{X})^k ; \tag{4.5}$$

параметра асимметрии

$$\hat{\beta}_n = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{S^3} ; \tag{4.6}$$

параметра эксцесса

$$\hat{\gamma}_n = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\mu_4}{S^4} - 3. \tag{4.7}$$

По закону больших чисел все введенные эмпирические моментные характеристики являются состоятельными оценками соответствующих теоретических моментных характеристик, т.е. сходятся к ним по вероятности при $n \to \infty$ (если только соответствующие теоретические характеристики данной генеральной совокупности конечны).

Рассмотрим подробнее свойства $\bar{X}$ и S^2 – теоретического среднего и теоретической дисперсии. Состоятельность $\bar{X}$ как оценки $a = \mathrm{E}\xi$ есть не что иное, как утверждение закона больших чисел. Если $\mathrm{E}|\xi| < \infty$, то

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{P} a = \mathrm{E}(\xi) \tag{4.8}$$

при $n \to \infty$, так как $\bar{X}$ – среднее арифметическое независимых одинаково распределенных с.в., $\bar{X}$ – несмещенная оценка параметра a.

Действительно,

$$\mathrm{E}(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}X_i = \frac{1}{n} \, n a = a. \tag{4.9}$$

Кроме того, можно показать [6],[8], что для многих распределений $\overline{X}$ является эффективной (т.е. имеющей минимальную дисперсию) оценкой среднего в классе линейных несмещенных оценок.

Важным свойством эмпирического среднего является асимптотическая нормальность. По центральной предельной теореме теории вероятностей с.в.

$$\frac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{D}(\overline{X})}} = \frac{\overline{X} - a}{\sigma/\sqrt{n}}$$

сходится по распределению к стандартному нормальному закону, т.е.

$$\mathrm{P}\left\{\sqrt{n}\ \frac{\overline{X} - a}{\sigma} < x\right\} \to \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2} dt \qquad (4.10)$$

при $n \to \infty$ и всех $x \in \mathrm{R}$, а с.в. $\frac{1}{\sqrt{n}}\ \overline{X}$ сходится по распределению к нормальному закону с параметрами a и σ. Свойство асимптотической нормальности позволяет находить не только точечные, но и интервальные оценки для математического ожидания (см. разд.6).

Теперь рассмотрим свойства S^2. Если дисперсия $\sigma^2 = \mathrm{D}(\xi)$ конечна, то $S^2 \to \sigma^2$ при $n \to \infty$ по закону больших чисел, т.е. S^2 – состоятельная оценка дисперсии. Будет ли она обладать свойством несмещенности ? Перепишем S^2 в виде, удобном для вычисления среднего:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - a)^2 = \frac{1}{n}\sum_{i=1}^{n} [(X_i - a) - (\overline{X} - a)]^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n} (X_i - a)^2 - \frac{2}{n}(\overline{X} - a)\sum_{i=1}^{n} (X_i - a) + (\overline{X} - a)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n} (X_i - a)^2 - (\overline{X} - a)^2 .$$

Теперь находим

$$\mathrm{E}S^2 = \frac{1}{n}\,\mathrm{E}\sum_{i=1}^{n} (X_i - a)^2 - \mathrm{E}(\overline{X} - a)^2 = \frac{1}{n}\,n\sigma^2 - \frac{\sigma^2}{n} =$$

$$= \sigma^2 - \frac{1}{n}\sigma^2 = \frac{n-1}{n}\,\sigma^2 \neq \sigma^2. \qquad (4.11)$$

Из (4.11) следует, что эмпирическая дисперсия – смещенная оценка. Однако введением поправочного коэффициента $\frac{n}{n-1}$ легко устранить это смещение. Действительно, величина

$$\hat{\sigma}^2 = \frac{n}{n-1}\cdot S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X})^2 \qquad (4.12)$$

является уже несмещенной оценкой дисперсии параметра σ^2 (если $\sigma^2 < \infty$) и ее обычно предпочитают использовать в качестве оценки дисперсии.

5. ТОЧЕЧНЫЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ

Пусть $X_1, X_2, \dots , X_n$ – выборка из генеральной совокупности с.в. ξ, θ – числовой параметр, характеризующий ф.р. генеральной совокупности: $F_\xi(n) = F(x,\theta)$, $\theta \in \Theta \subset \mathrm{R}$, где Θ – интервал, конечный или бесконечный. Имеется два способа оценки параметров: точечный и интервальный. Точечные методы указывают точку, вблизи которой находится неизвестный оцениваемый параметр θ. Интервальные методы позволяют получить интервал, в котором с вероятностью α, близкой к единице (α выбирается самим исследователем) находится неизвестное значение параметра. В этом разделе мы рассмотрим два основных метода точечной оценки параметров.

5.1. Метод моментов

В разд. 2 было показано, что с ростом n эмпирическая ф.р. $F_n(x)$ сколь угодно мало отличается от теоретической $F_\xi(x)$, поэтому при больших n эмпирические и теоретические моменты также близки друг другу. Если теоретическая ф.р. зависит от параметров, то для их определения можно воспользоваться эмпирическими моментами. Предположим, что для любого $\theta \in \Theta$ величина

$$\mathrm{E}_\theta(\xi) = \int\limits_{-\infty}^{\infty} x\, dF(x,\theta) = a(\theta)$$

конечна. Приравниваем первый теоретический и первый эмпирический

моменты:

$$a(\theta) = \overline{X}, \qquad (5.1)$$

где $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Если уравнение (5.1) имеет единственное решение $\hat{\theta}(X_1,\dots,X_n)$, то это решение есть одна из возможных оценок параметра θ.

Пример 5.1. Пусть имеется показательное распределение с неизвестным параметром $\lambda > 0$, т.е. $F_\xi(x) = 1 - e^{-x/\lambda}$ $(x > 0)$. Известно, что $\mathrm{E}\xi = a(\lambda) = (\frac{1}{\lambda})^{-1} = \lambda$ для данного показательного распределения. Приравнивая эмпирический и теоретический первые моменты, получаем оценку параметра λ по методу моментов: $\hat{\lambda} = \overline{X}$, причем эта оценка является несмещенной, состоятельной и эффективной.

Пример 5.2. Если с.в. ξ имеет распределение Релея с плотностью $p(x,\theta) = \frac{x}{\theta}\exp\left[-\frac{x^2}{2\theta}\right]$, $x > 0$, то

$$a(\theta) = \mathrm{E}\xi = \int\limits_{0}^{\infty} \frac{x^2}{\theta}\exp(-\frac{x^2}{2\theta})dx = \sqrt{\frac{\pi\theta}{2}}.$$

Из уравнения $\sqrt{\frac{\pi\theta}{2}} = \overline{X}$ находим оценку параметра θ: $\hat{\theta} = \frac{2}{\pi}(\overline{X})^2$.

Если ф.р. генеральной совокупности характеризуется не одним, а несколькими параметрами, которые нужно оценить по выборке, $F_\xi(x) = F(x,\theta)$, $\theta \in \Theta \subset \mathrm{R}^k$, т.е. θ – векторная величина: $\theta = (\theta_1,\theta_2,\dots,\theta_k)$, можно поступить следующим образом: найти первые k теоретических момента (при условии, что они существуют)

$$a_j(\theta_1,\theta_2,\dots,\theta_k) = \int\limits_{-\infty}^{\infty} x^j dF(x,\theta_1,\theta_2,\dots,\theta_k)\ ,\ j = 1,\dots,k,$$

и приравнять их к соответствующим эмпирическим моментам: $\alpha_j = \frac{1}{n}\sum_{i=1}^{n} X_i$. Если получаемая при этом система k уравнений с k неизвестными

$$a_j(\theta_1,\theta_2, \dots,\theta_k) = \alpha_j,\ j = 1,2,\dots k, \qquad (5.2)$$

имеет единственное решение, то это решение $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots , \hat{\theta}_k)$ можно принять в качестве оценки для неизвестного параметра $\theta = (\theta_1, \theta_2, \dots , \theta_k)$.

Пример 5.3. Если ξ – гауссовская с.в. с параметрами $a = \mathrm{E}\xi$ и $\sigma = \sqrt{\mathrm{D}\xi}$, то параметр $\theta = (a,\sigma)$ распределения – векторный, с двумя компонентами. По методу моментов получаем оценку $\hat{\theta} = (\hat{a},\hat{\sigma})$, где $\hat{a} = \overline{X}$, $\hat{\sigma}^2 = S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2$, т.е. $\hat{a}$ и $\hat{\sigma}^2$ – эмпирические моменты и, следовательно, $\hat{\theta}$ – асимптотически эффективная, асимптотически несмещенная и состоятельная оценка параметра θ.

Замечание 5.1. Преимущество метода моментов перед другими методами состоит в сравнительной простоте получения оценок. Однако этот способ оценивания содержит элемент произвола: кроме начальных моментов, можно приравнивать и другие характеристики, например центральные моменты. При этом получаются, вообще говоря, другие значения оценок параметров. Оценки по методу моментов нередко бывают смещенными и неэффективными, но они имеют и ряд положительных свойств. В частности, они состоятельны и асимптотически нормальны (при достаточно широких предположениях).

5.2. Метод максимального правдоподобия

В основу метода максимального правдоподобия (ММП) положен принцип, согласно которому реализуется чаще всего то, что более правдоподобно (имеет наибольшую вероятность). Для получения оценки максимального правдоподобия (ОМП) параметра θ ищется такое значение θ^*, при котором вероятность реализации полученной в ходе эксперимента выборки $X_1, X_2,\dots, X_n$ была бы максимальной. Такой способ выбора оценок обеспечивает им ряд оптимальных свойств.

При нахождении ОМП различают два случая: дискретный и непрерывный. Если с.в. ξ имеет дискретное распределение с возможными значениями $x_0,x_1,x_2,\dots$ и с вероятностями этих значений соответственно $p_0(x_0,\theta)$, $p_1(x_1,\theta),\dots$, то вероятность при n независимых наблюдениях получить выборку $X_1,X_2,\dots,X_n$ равна

$$L(\theta) = p_0^{n_0}(x_0,\theta)\ p_1^{n_1}(x_1,\theta)\ \dots\ p_j^{n_j}(x_j,\theta), \qquad (5.3)$$

где n_i – число повторений значения x_i в выборке, $i = 0,1,\dots,J$, J – номер максимального (среди реализовавшихся в выборке) значения x_J. При этом $\sum_{i=1}^{J} n_i = n$, где n – объем выборки. Функция $L(\theta)$, определенная формулой (5.3), называется *функцией правдоподобия*, а значение θ^*, при котором $L(\theta)$ достигает максимального значения, есть ОМП.

В случае непрерывного распределения генеральной совокупности, имеющего плотность $p_\xi(x,\theta)$, зависящую от неизвестного параметра θ, ОМП ищется из условия достижения максимума плотности распределения выборки $p_\xi(x_1,\dots,x_n) = \prod_{i=1}^{n} p_\xi(x_i,\theta)$, вычисленной в точке $(X_1,X_2,\dots,X_n)$:

$$L(\theta) = \prod_{i=1}^{n} p_\xi(X_i,\theta) , \qquad (5.4)$$

являющейся в этом случае функцией правдоподобия. Известно, что точка максимума не изменится, если вместо функции $L(\theta)$ рассматривать $\ln(L(\theta))$ (т. к. логарифм – монотонная функция). Функция $\ln(L(\theta))$ называется *логарифмической функцией правдоподобия*. Переход к ней зачастую облегчает поиск ОМП. Если функция $L(\theta)$ дифференцируема по θ при всех $X_1,X_2,\dots,X_n$ и достигает максимума внутри интервала возможных значений θ, то ОМП находится из уравнения $\frac{d}{d\theta}\left[L(\theta)\right] = 0$ или из уравнения $\frac{d}{d\theta}\left[\ln(L(\theta))\right] = \frac{1}{L(\theta)}\frac{dL(\theta)}{d\theta} = 0$. В случае векторного параметра $\theta = (\theta_1, \theta_2,\dots, \theta_k)$ ОМП находится из системы уравнений $\frac{\partial L(\theta)}{\partial\theta_j} = 0$, $j = 1,2,\dots,k$, или из системы уравнений

$$\frac{\partial}{\partial\theta_j}\left[\ln(L(\theta))\right] = \frac{1}{L(\theta)}\frac{\partial}{\partial\theta_j}\left[L(\theta)\right], \quad j=1,2,\dots,k . \qquad (5.5)$$

Замечание 5.1. Решение системы (5.5) может соответствовать не только максимуму функции $\ln(L(\theta))$, но и минимуму или точке перегиба, поэтому требуется проверять, действительно ли найденная стационарная точка является точкой максимума $L(\theta)$.

5.3. Свойства ОМП

При выполнении некоторых (достаточно широких) условий регулярности [1],[6],[8], налагаемых на плотность распределения (в

непрерывном случае) или на вероятности значений (в дискретном случае) ОМП θ^* обладают следующими свойствами:

1) решение θ^* уравнений правдоподобия единственно;

2) θ^* – состоятельная оценка параметра θ;

3) величина $\sqrt{n}(\theta^*-\theta)$ имеет асимптотически нормальное предельное распределение с математическим ожиданием 0 и дисперсией $1/I(\theta)$, где $I(\theta) = E_\theta\left[\frac{\partial \ln p(\xi,\theta)}{\partial\theta}\right]^2$ – информационное количество Фишера;

4) ОМП θ^* – асимптотически эффективная оценка параметра θ при $n \to \infty$.

Пример 5.4. Рассмотрим последовательность n испытаний Бернулли с неизвестной вероятностью *"успеха"*. Пусть в этой последовательности *"успех"* осуществился m раз, а *"неудача"* – $n-m$ раз. Если *"успех"* соответствует $\xi = 1$, а *"неудача"* – $\xi = 0$, то выборка $X_1,X_2,\dots,X_n$ из генеральной совокупности ξ состоит из чисел 0 и 1, причем 1 в ней встречается m раз. Функция правдоподобия имеет вид $L(p)= p^m(1-p)^{n-m}$, уравнение правдоподобия

$$\frac{\partial}{\partial p}\left[\ln(L(p))\right] = \frac{m}{p} - \frac{n-m}{1-p} = 0$$

имеет единственное решение $p^* = \frac{m}{n}$, которое является несмещенной ($E(\frac{m}{n}) = \frac{np}{n} = p$), состоятельной (по закону больших чисел Бернулли) и асимптотически нормальной (по теореме Муавра – Лапласа) оценкой параметра p. Можно показать [1],[6],[8], что эта оценка эффективна.

Пример 5.5. Пусть ξ – гауссовская с.в. с параметрами a и σ. Найдем ОМП для неизвестного параметра $\theta = (a,\sigma)$. В этом случае

$$L(\theta) = \frac{1}{\left(\sqrt{2\pi\sigma}\right)^n}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-a)^2\right],$$

$$\ln(L(\theta)) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i- a)^2.$$

Дифференцируя $\ln(L(\theta))$ по a и σ^2 и приравнивая производные к нулю, получаем систему уравнений правдоподобия:

$$\begin{cases} \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - a) = 0 \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i - a)^2 = 0 , \end{cases}$$

из которой находим ОМП $\theta^* = (a^*, \sigma^{*2})$,

где $a^* = \frac{1}{n}\sum_{i=1}^{n}X_i = \bar{X}$; $\sigma^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2$. При выполнении условий регулярности [6],[8] эта оценка является асимптотически нормальной и асимптотически эффективной оценкой параметра θ.

6. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Пусть $X = (X_1,X_2,\dots,X_n)$ – выборка из генеральной совокупности с.в. ξ, θ – числовой параметр, характеризующий ф.р. генеральной совокупности: $F_\xi(x) = F(x,\theta)$, $\theta \in \Theta \subset R$. В разд.5 были рассмотрены некоторые основные методы точечного оценивания параметра θ. Однако знание точечной оценки для представления об истинных значениях параметра θ оказывается недостаточным. Важно знать еще и разброс этой оценки. На практике часто более полезным бывает не столько знание точечных оценок неизвестных параметров, сколько знание интервалов, в которых значения этих параметров содержатся с большой вероятностью.

Определение 6.1. Интервал $I(X) = (\hat{\theta}_1(X),\hat{\theta}_2(X))$, где $X = (X_1,\dots,X_n)$ – вектор выборочных значений, называется *доверительным интервалом надежности* α для неизвестного параметра θ, если

$$P\left[\hat{\theta}_1(X) < \theta < \hat{\theta}_2(X)\right] = \alpha$$

для любого $\theta \in \Theta$.

6.1. Асимптотические доверительные интервалы

При построении доверительных интервалов обычно используют свойство асимптотической нормальности, которое имеет место для многих применяемых на практике оценок. Так, известно, что при достаточно большом объеме выборки (при $n \to \infty$) ОМП и оценки метода моментов (см. разд. 5) асимптотически нормальны (если выполнены условия регулярности), т.е. имеют близкое к нормальному распределение. Если $\hat{\theta}_n$ – такая асимптотически несмещенная оценка параметра θ, то есть

$$\sqrt{n}\left[\hat{\theta}_n - \theta\right] \xrightarrow{d} N(0,\sigma), \text{ при } n\to\infty, \text{ где } \sigma = \sigma(\theta),$$

то при достаточно больших n

$$P\left[\left|\frac{\sqrt{n}}{\sigma}(\hat{\theta}_n - \theta)\right| < u\right]= 2\,\Phi(u) - 1 , \qquad (6.1)$$

где $\Phi(u)$ – ф.р. стандартного нормального закона (см.дополнение 1). Обозначим u_α корень уравнения $2\Phi(u)-1 = \alpha$, то есть $\Phi(u_\alpha) = \frac{1+\alpha}{2}$. Значение квантили u_α находят по таблицам функции $\Phi(u)$. Тогда из (6.1) следует, что с вероятностью, близкой к α, выполняются неравенства

$$\hat{\theta}_n - u_\alpha\frac{\sigma}{\sqrt{n}} < \theta < \hat{\theta}_n + u_\alpha\frac{\sigma}{\sqrt{n}} . \qquad (6.2)$$

Значение $\sigma = \sigma(\theta)$ в (6.2), как правило, неизвестно, и при достаточно больших n его можно заменить оценкой $\hat{\sigma} = \sigma(\hat{\theta}_n)$. Неравенства (6.2) обеспечивают асимптотический доверительный интервал $\left(\hat{\theta}_n - u_\alpha\frac{\hat{\sigma}}{\sqrt{n}} , \hat{\theta}_n + u_\alpha\frac{\hat{\sigma}}{\sqrt{n}}\right)$ надежности α. Например, если $\hat{\theta}_n$ – оценка максимального правдоподобия, то асимптотический доверительный интервал имеет вид:

$$\left(\hat{\theta}_n - u_\alpha\frac{1}{\sqrt{nI(\hat{\theta}_n)}} , \hat{\theta}_n + u_\alpha\frac{1}{\sqrt{nI(\hat{\theta}_n)}}\right), \qquad (6.3)$$

где $I(\hat{\theta}_n)$ – информационное количество Фишера, вычисленное при $\theta = \hat{\theta}_n$ (см.разд.5).

6.2. Интервальные оценки параметров по малой выборке

В практических расчетах часто приходится оценивать параметры на основе выборки сравнительно небольшого объема (n порядка 100). В этом случае замена распределения оценки $\hat{\theta}_n$ ее асимптотическим нормальным распределением при построении доверительного интервала может привести к заметной погрешности. Недостаток экспериментальных данных в этом случае приходится компенсировать некоторыми дополнительными предположениями о характере распределения генеральной совокупности. Часто можно предполагать, что $\xi \sim N(a,\sigma)$, т.е. что выборка получена из нормально распределенной генеральной совокупности с неизвестными параметрами a и $\sigma > 0$. Доверительные интервалы для этих двух параметров наиболее хорошо исследованы.

6.2.1. Доверительные интервалы для параметра a (среднего нормально распределенной генеральной совокупности)

Доверительный интервал для параметра a построим на основе несмещенных и асимптотически эффективных оценок:

$$\hat{a} = \overline{X} = \sum_{i=1}^{n} X_i \ , \qquad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 .$$

В случае нормального распределения выборочных данных величина

$$t = \frac{\sqrt{n}\,(\overline{X}-a)}{\hat{\sigma}}$$

имеет распределение Стьюдента с $(n-1)$ степенями свободы (см. дополнение 2).

Зададим доверительную вероятность α и найдем величину t_α как корень уравнения

$$\int_{-t_\alpha}^{t_\alpha} S_{n-1}(x)\,dx = \alpha \ ,$$

где $S_{n-1}(x)$ – плотность распределения Стьюдента с $(n-1)$ степенями свободы (рис. 6.1). Величина t_α находится как квантиль уровня $\frac{1+\alpha}{2}$ по таблицам распределения Стьюдента (при $n \leqslant 30$) и по таблицам стандартного нормального распределения $\Phi(x)$ (при $n > 30$) (см. приложение, табл. 1). Для этого t_α имеем:

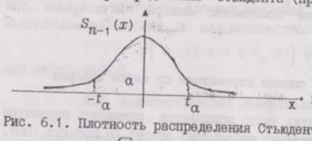


Рис. 6.1. Плотность распределения Стьюденте

$$P\left(\left|\frac{\sqrt{n}\,(\overline{X}-a)}{\hat{\sigma}}\right| < t_\alpha\right) = \int_{-t_\alpha}^{t_\alpha} S_{n-1}(x)\,dx = \alpha,$$

следовательно, с вероятностью α выполняются неравенства

$$\overline{X} - t_\alpha \frac{\hat{\sigma}}{\sqrt{n}} < a < \overline{X} + t_\alpha \frac{\hat{\sigma}}{\sqrt{n}} \ , \qquad (6.4)$$

означающие, что интервал

$$\left[\overline{X} - t_\alpha \frac{\hat{\sigma}}{\sqrt{n}} \ , \ \overline{X} + t_\alpha \frac{\hat{\sigma}}{\sqrt{n}}\right] \qquad (6.5)$$

является доверительным интервалом надежности α для значения параметра a.

Замечание 6.1. Доверительный интервал (6.5) совпадает с асимптотическим доверительным интервалом для математического ожидания ($a = \mathrm{E}\xi$) генеральной совокупности с.в. ξ в том случае, когда ее распределение не является нормальным. Действительно, в этом случае в силу центральной предельной теоремы (справедливой, если $\mathrm{D}\xi = \sigma^2 < \infty$) величина $\dfrac{\overline{X} - \mathrm{E}\overline{X}}{\sqrt{\mathrm{D}\overline{X}}} = \sqrt{n}\,\dfrac{\overline{X}-a}{\sigma}$ имеет асимптотическое стандартное нормальное распределение, откуда вытекает, что с вероятностью,близкой к α при достаточно больших n для параметра $a = \mathrm{E}\xi$ имеют место неравенства (6.4), в которых t_α – квантиль уровня $\frac{1+\alpha}{2}$ стандартного нормального закона ($2\Phi(t_\alpha) - 1 = \alpha$).

6.2.2. Доверительные интервалы для дисперсии нормально распределенной генеральной совокупности

Для построения доверительных интервалов для параметра σ снова воспользуемся оценками $\hat{a} = \overline{X}$, $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. В случае нормального распределения генеральной совокупности статистика

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad (6.6)$$

имеет распределение χ^2 с $(n-1)$ степенями свободы (см.дополн. 2). Зададим доверительную вероятность α и выберем значения χ_1^2 и χ_2^2 так, чтобы

$$\int_{\chi_1^2}^{\chi_2^2} k_{n-1}(x)\,dx = \alpha \ ,$$

где $k_{n-1}(x)$ – плотность распределения χ^2 с $n-1$ степенями свободы. Значения χ_1^2 и χ_2^2 можно выбрать многими способами. Обычно их находят при дополнительном условии $\int_0^{\chi_1^2} k_{n-1}(x)\,dx = \frac{1-\alpha}{2}$. В этом случае $\int_{\chi_2^2}^{\infty} k_{n-1}(x)dx = \frac{1-\alpha}{2}$. Тогда квантили χ_1^2 и χ_2^2 определяются однозначно и находятся по таблицам квантилей распределения χ^2

(см. приложение, табл. 3) при $n \leqslant 30$ и с использованием таблиц стандартной нормальной ф.р. $\Phi(x)$ ввиду асимптотической нормальности с.в. χ_n^2 (см. дополнение 1) при $n > 30$. Для найденных таким образом значений χ_1^2 и χ_2^2 имеем

$$P\left(\chi_1^2 < \frac{(n-1)\hat{\sigma}^2}{\sigma^2} < \chi_2^2\right) = \int_{\chi_1^2}^{\chi_2^2} k_{n-1}(x)\,dx = \alpha.$$

Последнее означает, что с вероятностью α справедливы неравенства:

$$\frac{(n-1)\hat{\sigma}^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)\hat{\sigma}^2}{\chi_1^2} \ ; \qquad (6.7)$$

$$\frac{\sqrt{n-1}\ \hat{\sigma}}{\chi_2} < \sigma < \frac{\sqrt{n-1}\ \hat{\sigma}}{\chi_1} \ , \qquad (6.8)$$

которые означают, что интервалы $\left[\dfrac{\sqrt{n-1}\ \hat{\sigma}}{\chi_2} , \dfrac{\sqrt{n-1}\ \hat{\sigma}}{\chi_1}\right]$ и $\left[\dfrac{(n-1)\hat{\sigma}^2}{\chi_2^2} , \dfrac{(n-1)\hat{\sigma}^2}{\chi_1^2}\right]$ являются доверительными интервалами надежности α для параметров σ и σ^2 соответственно.

7. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ. КРИТЕРИИ СОГЛАСИЯ

7.1. Постановка задачи

Пусть $X_1,\ldots,X_n$ – выборка из генеральной совокупности с.в. ξ, причем ф.р. $F(x) = F_\xi(x)$ неизвестна. Часто можно выдвинуть различные предположения о неизвестной ф.р. F. Любое такое предположение называется *статистической гипотезой* H и задается соотношением

$$H : F_\xi \in \mathrm{F} \ ,$$

где F – заданное множество ф.р. Если множество F состоит из одного элемента $\mathrm{F} = \left\{ F \right\}$, то гипотеза H называется *простой*:

$$H : F_\xi = F.$$

Если множество F состоит более чем из одного элемента, то гипотеза H называется *сложной*.

Задачи проверки гипотез возникают при наличии двух конкурирующих гипотез:

$$H_0 : F_\xi \in \mathrm{F}_0 \ \text{ и } \ H_1 : F_\xi \in \mathrm{F}_1 \ ,$$

причем множества F_0 и F_1 не пересекаются. Гипотезу H_0 часто называют *основной* , а H_1 – *альтернативой*. Основная гипотеза, так же как альтернатива, может быть простой и сложной. При проверке гипотез часто есть основания предполагать, что ф.р. F_ξ принадлежит заданному k-мерному параметрическому семейству ф.р.:

$$F_\xi(x) = F(x,\theta) \ ; \ \theta \in \Theta \ ; \ \Theta \subset \mathrm{R}^k,$$

причем основной гипотезе и альтернативе соответствуют непересекающиеся подмножества $\Theta_0 \subset \Theta$ и $\Theta_1 \subset \Theta$ (рис.7.1). Такие гипотезы называют *параметрическими* и записывают обычно в виде

$$H_0 : \theta \in \Theta_0 \quad \text{и} \quad H_1 : \theta \in \Theta_1 .$$

Если такого параметрического семейства нет, то гипотезы называют *непараметрическими*.

Часто в задаче формулируется лишь основная гипотеза $H_0 : F_\xi \in \mathrm{F}_0$. При этом, однако, имеется в виду, что неизвестная ф.р. F_ξ может принадлежать и более широкому множеству ф.р. F, $\mathrm{F}_0 \subset \mathrm{F}$. В этом случае альтернатива состоит в том, что $F_\xi \notin \mathrm{F}_0$, т.е.

$$H_1 : F_\xi \in \mathrm{F}_1 = \mathrm{F} \setminus \mathrm{F}_0 \ .$$

Такие задачи проверки гипотез называют задачами *проверки согласия* (рис.7.2).

Параметрическая гипотеза H_0 часто состоит в том, что часть параметров $\theta_1,\ldots,\theta_s$ из общей совокупности $\theta=(\theta_1,\ldots,\theta_k)$, $s \leqslant k$, принимает заданные значения: $\theta_1=\theta_{1,0},\ldots, \theta_s=\theta_{s,0}$. В этом случае векторный параметр $\theta^{(s)}=(\theta_1,\ldots,\theta_s)$ называют полезным, а $\theta^{(k-s)}=(\theta_{s+1},\ldots,\theta_k)$ – мешающим. Нулевую гипотезу можно представить в виде H_0: $\theta^{(s)}= \theta_0^{(s)}= (\theta_{1,0},\ldots,\theta_{s,0})$. При этом множество $\Theta_0 = \left\{ \theta = \left(\theta^{(s)}, \theta^{(k-s)}\right) \in \Theta, \ \theta^{(s)}= \theta_0^{(s)}\right\}$ есть подмножество размерности k-s в Θ. При одномерном полезном параметре альтернативы $H_1 : \theta \in \Theta$, $\theta^{(1)} > \theta_0^{(1)}+ \delta$ или $H_1 : \theta \in \Theta$, $\theta^{(1)} < \theta_0^{(1)}- \delta$, $\delta \geqslant 0$, называют *односторонними*, а альтернативы $H_1 : \theta \in \Theta$, $|\theta^{(1)}-\theta_0^{(1)}| > \delta$ – *двусторонними*.

При $\delta=0$ двусторонняя альтернатива соответствует задаче проверки согласия с H_0.

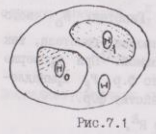


Рис.7.1

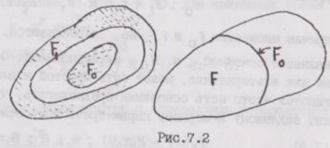


Рис.7.2

В задачах проверки гипотез требуется на основе наблюдаемой выборки $X_1,\dots,X_n$ выбрать одно из двух возможных решений: принять основную гипотезу H_0 (в этом случае говорят, что *данные наблюдений не противоречат гипотезе* H_0) или отвергнуть H_0 (т.е. принять альтернативу H_1). Правила (алгоритмы) принятия таких решений называют *критериями (или тестами) проверки гипотез*.

Задачи проверки гипотез возникают в самых разнообразных областях науки и техники: технической и медицинской диагностике, дефектоскопии, радиотехнике, передаче информации и т.д. В задачах проверки согласия основная гипотеза часто соответствует наиболее простым возможным предположениям о характере наблюдаемых данных, использование которых может существенно упростить дальнейшее исследование задачи, но нуждается в экспериментальной проверке. Так в автоматике, радиотехнике многие выводы и рекомендации основаны на предположении о нормальности измеряемых случайных величин; их использование в конкретных задачах требует проверки *гипотезы нормальности*, которая (в простейшем случае одномерных данных) имеет вид:

$$H_0 : F_\xi(x) \in \left\{\Phi\left[(x-a)/\sigma\right];\ a \in R^1,\ \sigma > 0\right\},$$

где $\Phi(x)$ – ф.р. стандартного нормального закона. В теории систем с очередями многие результаты получены для показательного закона распределения элементов управляющих последовательностей; для использования этих результатов необходимо проверить *гипотезу показательности*:

$$H_0 : F_\xi(x) \in \mathrm{F}_0,\quad \mathrm{F}_0=\left\{ F(x,\lambda);\ \lambda > 0;\ F(x,\lambda) = 1 - e^{-\lambda x},\ x > 0\right\}.$$

Множество F при этом характеризует возможные в данной задаче

отклонения от основной гипотезы; это может быть более широкое параметрическое семейство ф.р. или непараметрическое семейство ф.р. (например, множество всех ф.р., непрерывных ф.р., ф.р. с гладкими плотностями и т.д.). Отметим, что чем "шире" множество F, тем обычно менее достоверными являются принимаемые решения.

В теории проверки гипотез разрабатываются методы построения критериев проверки гипотез и изучаются различные характеристики достоверности принимаемых решений. Как правило, исследования проводятся в рамках *асимптотического подхода* при больших объемах выборки ($n \to \infty$).

7.2. Критерии проверки гипотез

Обозначим через S пространство наблюдений, т.е. множество возможных значений $\bar{x}$ случайной выборки $X_1,\dots,X_n$. Критерий проверки гипотез сопоставляет каждому значению $\bar{x} \in S$ число $\phi(\bar{x})=0$ (принятие H_0) или $\phi(\bar{x})=1$ (принятие H_1), т.е. критерий может рассматриваться как функция $\phi(\bar{x})$, заданная на пространстве наблюдений S и принимающая значения 0 или 1. Подмножество пространства наблюдений $S_0 = \{\bar{x} : \phi(\bar{x})=0\}$, на котором принимается основная гипотеза, называется *допустимым*, а подмножество $S_1 = \{\bar{x} : \phi(\bar{x}) = 1\}$, на котором принимается альтернатива, называется *критическим* множеством критерия ϕ. Ясно, что $S_1 = S \setminus S_0$ и каждое из этих подмножеств однозначно определяет критерий ϕ. Обычно критерий задается с помощью *статистики критерия* $L_n = L_n(\bar{x})$, $\bar{x} \in S$ и числового *порога критерия* $T = T_n$:

$$\phi(\bar{x}) = \begin{cases} 0 \text{ при } L_n(\bar{x}) < T, \\ 1 \text{ при } L_n(\bar{x}) \geqslant T. \end{cases}$$

Алгоритм принятия решения в этом случае состоит из двух этапов:

1) вычисление по выборке $X_1, \dots, X_n$ значения статистики $L_n=L_n(X_1,\dots,X_n)$;

2) сравнение значения L_n с порогом T; основная гипотеза принимается при $L_n < T$ и отвергается при $L_n \geqslant T$.

7.3. Ошибки первого и второго рода

Решения, принимаемые на основе случайных данных с помощью того или иного критерия ϕ, могут быть как правильными, так и ошибочными. В задачах проверки гипотез различают ошибки двух типов.

1. *Ошибки I рода: отклонение основной гипотезы* H_0, *в то время как она справедлива*, т.е. $\phi(\bar{x}) = 1$ при $F_\xi \in \mathrm{F}_0$.

2. *Ошибки II рода: принятие основной гипотезы* H_0, *в то время как имеет место альтернатива*, т.е. $\phi(\bar{x}) = 0$ при $F \in \mathrm{F}_1$.

Количественными характеристиками ошибок I и II рода являются их вероятности.

Пусть основная гипотеза и альтернатива являются простыми: $\mathrm{F}_0=\{F_0\}$, $\mathrm{F}_1 = \{F_1\}$. Тогда для критерия ϕ *вероятность ошибок I рода*

$$\alpha(\phi) = \alpha(\phi,F_0) = P_{F_0}(S_1),$$

вероятность ошибок II рода

$$\beta(\phi) = \beta(\phi,F_1) = P_{F_1}(S_0).$$

Здесь $P_F(A)$ обозначает вероятность события $A \subset S$, вычисленную в предположении $F_\xi = F$:

$$P_F(A) = \int\underset{A}{\dots}\int dF(x_1)\dots dF(x_n).$$

Введем также величину

$$\gamma(\phi) = \gamma(\phi,F_1) = P_{F_1}(S_1) = 1 - \beta(\phi,F_1),$$

называемую *мощностью* критерия ϕ: она равна вероятности отвергнуть основную гипотезу H_0, в то время как имеет место альтернатива, т.е. $F_\xi = F_1$. Для критерия ϕ, заданного с помощью статистики L_n и порога T_n, вероятности ошибок и мощность имеют вид

$$\alpha(\phi) = 1 - G_{n,F_0}(T_n);\ \beta(\phi) = G_{n,F_1}(T_n),\ \gamma(\phi) = 1 - G_{n,F_1}(T_n),$$

где $G_{n,F}(t) = P_F\left[L_n < t\right]$ – ф.р. статистики L_n. (рис.7.3).

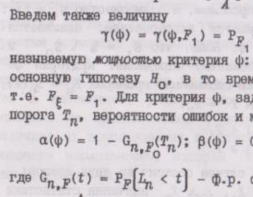


Рис. 7.3

В задачах проверки гипотез желательно построить такие критерии, у которых вероятности ошибок как I рода, так и II рода были бы минимальны. Это требование, однако, противоречиво: обычно уменьшение вероятностей ошибок I рода влечет увеличение вероятностей ошибок II рода (уменьшение мощности) и наоборот. В этой связи обычно используют

подход Неймана – Пирсона, состоящий в следующем. Задается малая величина $\alpha > 0$ – максимально допустимая величина вероятностей ошибок I рода (ее называют также *допустимым уровнем значимости*), и ищется критерий $\phi = \phi_\alpha$, минимизирующий вероятность ошибок II рода (т.е. максимизирующий мощность) при ограничении

$$\alpha(\phi) \leqslant \alpha .$$

Такой критерий называется *наиболее мощным* (НМ).

Выбор величины $\alpha > 0$ зависит от конкретной задачи. Обычно выбирают $\alpha = 0.1$, 0.05 или 0.01; если ошибочное отклонение основной гипотезы сопряжено с большими потерями, выбирают $\alpha = 0.001$ и менее.

Пусть основная гипотеза H_0 является сложной. Тогда вероятность ошибок I рода $\alpha(\phi,F)$ зависит от $F \in \mathrm{F}_0$; при параметрической гипотезе обозначают $\alpha(\phi,F(x,\theta)) = \alpha(\phi,\theta)$, $\theta \in \Theta_0$. Величина $\alpha^*(\phi) = \sup_{F\in\mathrm{F}} \alpha(\phi,F)$ называется *уровнем значимости* критерия ϕ. Обычно на критерии накладывается ограничение:

$$\alpha^*(\phi) \leqslant \alpha .$$

Критерий ϕ называется *подобным*, если $\alpha(\phi,F) = \alpha^*(\phi)$ для всех $F \in \mathrm{F}_0$.

Пусть альтернатива H_1 является сложной. Тогда вероятность ошибок II рода $\beta(\phi,F)$ и мощность $\gamma(\phi,F)$ зависят от $F \in \mathrm{F}_1$. При параметрической альтернативе обозначают:

$$\beta(\phi,F(x,\theta)) = \beta(\phi,\theta),\quad \gamma(\phi,F(x,\theta)) = \gamma(\phi,\theta),\quad \theta \in \Theta_1 .$$

Критерий ϕ называется *несмещенным*, если $\gamma(\phi,F) \geqslant \alpha^*(\phi)$ для всех $F \in \mathrm{F}_1$.

Критерий $\phi = \phi_\alpha$ уровня значимости α называется *равномерно наиболее мощным* (РНМ), если $\gamma(\phi_\alpha,F) \geqslant \gamma(\phi_\alpha',F)$ для любого критерия ϕ_α' уровня значимости α и любой $F \in \mathrm{F}_1$.

Несмещенный критерий $\phi = \phi_\alpha$ уровня значимости α называется *равномерно наиболее мощным несмещенным* (РНМН), если аналогичное неравенство выполнено для любого несмещенного критерия ϕ_α' уровня значимости α.

При асимптотическом подходе аналогично определяются асимптотически наиболее мощные (АНМ), асимптотически подобные (АП), асимптотически равномерно наиболее мощные (АРНМ), асимптотически равномерно наиболее мощные несмещенные (АРНМН) последовательности критериев $\psi_{n,\alpha}$. Смысл определений состоит в выполнении аналогичных неравенств в пределе при $n \to \infty$ (точные определения см. [1], [9]). Обычно говорят просто о критерии, обладающем тем или иным асимптотическим свойством, имея в виду последовательность критериев.

При одномерном полезном параметре РНМ или АРНМ критерии существуют обычно лишь для односторонних альтернатив, в то время как РНМН или АРНМН критерии – для двусторонних альтернатив.

Если размерность полезного параметра больше 1, то таких критериев обычно не существует.

Критерий ψ_n называется *состоятельным*, если $\beta(\psi_n, F) \to 0$ при $n \to \infty$ для любой ф.р. $F \in \mathrm{F}_1$. Критерий ψ_n называется *состоятельным для последовательности ф.р.* $F^{(n)} \in \mathrm{F}_1$, если $\beta(\psi_n, F^{(n)}) \to 0$ при $n \to \infty$.

Различные критерии могут быть состоятельными для различных последовательностей альтернатив $F^{(n)}$, могут иметь разную скорость убывания вероятностей ошибок *II* рода при $n \to \infty$. Этим объясняется многообразие различных критериев, используемых в статистике.

7.4. Методы построения критериев

7.4.1. Проверка гипотез и оценивание

Задачи проверки гипотез тесно связаны с задачами оценивания. Пусть $\hat{\theta}_n$– точечная оценка неизвестного параметра $\theta \in \Theta$. Тогда критерий проверки гипотезы H_0: $\theta \in \Theta_0$ против альтернативы H_1: $\theta \in \Theta_1$ часто можно построить, сравнивая меры близости $d(\hat{\theta}_n, \Theta_0)$ и $d(\hat{\theta}_n, \Theta_1)$ между значением оценки $\hat{\theta}_n$ и параметрическими множествами Θ_0 и Θ_1. В задаче проверки согласия это обычно сводится к сравнению с порогом $T_{n,\alpha}$ меры близости $d(\hat{\theta}_n, \theta_0)$.

Пусть $I_n = (\theta_{1,n}, \theta_{2,n})$ – доверительный интервал надежности $(1-\alpha)$ для полезного параметра $\theta^{(1)}$, $\theta_0^{(1)}$ – заданное значение полезного параметра. Тогда можно построить критерий проверки согласия с гипотезой H_0: $\theta^{(1)} = \theta_0^{(1)}$ уровня значимости α: гипотеза H_0 принимается (или отвергается), в зависимости от того, попадает

(или не попадает) значение $\theta^{(1)}$ в интервал I.

При использовании достаточно хороших оценок (обычно оценок максимального правдоподобия), при выборе надлежащей меры близости или наилучших доверительных интервалов такой подход часто приводит к критериям, обладающим различными свойствами оптимальности). Этот подход распространяется и на многие непараметрические задачи проверки гипотез.

Далее мы опишем некоторые методы построения оптимальных (или асимптотически оптимальных) критериев проверки гипотез.

Пусть с.в. ξ является непрерывной или дискретной $\Big[$с конечным или счетным множеством $\mathrm{T} = \left\{t^{(1)}, \ldots, t^{(k)}, \ldots\right\}$ возможных значений$\Big]$. Будем обозначать через $p(x)$ функцию плотности распределения $F_\xi(x)$ или функцию вероятностей значений с.в. ξ, заданную на множестве T, т.е. $p(t^{(j)}) = \mathrm{P}\left[\xi = t^{(j)}\right]$.

7.4.2. Простая гипотеза и простая альтернатива

Пусть H_0: $F_\xi = F_0$, H_1: $F_\xi = F_1$, $p_0(x)$ и $p_1(x)$ – функции плотности распределения или вероятностей значений для F_0 и F_1.

Статистикой отношения правдоподобия называется статистика

$$L_n = L_n(X_1, \ldots, X_n) = \frac{\prod_{i=1}^{n} p_1(X_i)}{\prod_{i=1}^{n} p_0(X_i)} .$$

Здесь числитель характеризует величину вероятности или плотности вероятностей получить имеющуюся выборку $X_1, \ldots, X_n$ в случае справедливости альтернативы H_1, а знаменатель – в случае справедливости основной гипотезы H_0. Поэтому при больших значениях L_n альтернатива H_1 более вероятна, чем основная гипотеза H_0, а при малых значениях L_n – наоборот. Это позволяет использовать L_n как статистику критериев проверки H_0 против H_1. Критерии, основанные на статистике L_n, являются наилучшими.

Предположим, что существует такая величина $T_n = T_n(\alpha)$, что $G_{n,F_0}(T_n) = 1 - \alpha$, т.е. $T_n(\alpha)$ есть $(1-\alpha)$ – квантиль ф.р. G_{n,F_0} статистики L_n. Это условие выполнено для любого $\alpha \in (0,1)$, если ф.р. $G_{n,F_0}(t)$ непрерывна.

Тогда *наиболее мощный критерий проверки* H_0 *против* H_1 *основан на статистике отношения правдоподобия* L_n при пороге $T_n = T_n(\alpha)$, т.е. основная гипотеза отвергается при $L_n > T_n$ и принимается в противном случае. Ясно, что вместо L_n можно использовать любую возрастающую функцию $h(L_n)$ от L_n при пороге $h(T_n)$. При значении порога $T_n = 1$ достигается минимально возможная сумма $\alpha(\psi, F_0) + \beta(\psi, F_1)$ вероятностей ошибок. Асимптотические свойства критерия отношения правдоподобия приведены в дополнении.

Критерий максимума отношения правдоподобия

Рассмотрим параметрическую задачу проверки гипотезы H_0: $F_\xi(x) = F(x,\theta)$, $\theta \in \Theta_0$, против альтернативы H_1: $F_\xi(x) = F(x,\theta)$, $\theta \in \Theta_1$, причем задано множество ф.р. $\mathrm{F} = \left\{F(x,\theta);\ \theta \in \Theta\right\}$ и $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$, $\Theta \subset \mathrm{R}^k$. Пусть все ф.р. $F(x,\theta)$, $\theta \in \Theta$ имеют функции плотности распределения или вероятностей значений $p(x,\theta)$. *Критерий максимума отношения правдоподобия* в этой задаче основан на статистике *максимума отношения правдоподобия*

$$L_n = L_n(X_1, \ldots, X_n) = \frac{\sup_{\theta_1 \in \Theta_1} \prod_{i=1}^{n} p_1(X_i, \theta_1)}{\sup_{\theta_0 \in \Theta_0} \prod_{i=1}^{n} p_0(X_i, \theta_0)}.$$

В частности, при простой основной гипотезе H_0: $F_\xi(t) = F(t, \theta_0)$, $\theta_0 \in \Theta$, статистика максимума отношения правдоподобия имеет вид:

$$L_n = L_n(X_1, \ldots, X_n) = \sup_{\theta \in \Theta_1} \prod_{i=1}^{n} \frac{p_1(X_i, \theta)}{p_0(X_i, \theta_0)} .$$

Качественный смысл статистики максимума отношения правдоподобия – в замене неизвестных параметров $\theta_0 \in \Theta_0$ и $\theta_1 \in \Theta_1$ оценками максимального правдоподобия. Критерий максимума отношения правдоподобия обладает многими замечательными асимптотическими свойствами. Пусть проверяется согласие с гипотезой H_0: $\theta_0 \in \Theta_0$, $\theta^{(s)} = \theta_0^{(s)}$, о значении полезного параметра $\theta^{(s)}$, $1 \leqslant s \leqslant k$. В этом случае при выполнении достаточно общих предположений регулярности имеют место следующие свойства.

1. При справедливости основной гипотезы H_0 статистика $l_n = 2\log(L_n)$ имеет предельное распределение хи-квадрат с s степенями

свободы (см. дополнения, п.1):

$$G_{n,F_0}(t) \to P_s(t) = \mathrm{P}(\chi_s^2 < t), \quad n \to \infty .$$

Это позволяет при больших n использовать в качестве порога критерия максимума отношения правдоподобия $\psi_{n,\alpha}$ величину $T_{n,\alpha} = \exp\left[2t_{s,\alpha}\right]$, где $t_{s,\alpha}$ есть $(1-\alpha)$ – квантиль ф.р. $P_s(t)$: $P_k(t_{s,\alpha}) = 1 - \alpha$. При этом $\alpha(\psi_{n,\alpha}, \theta) \to \alpha$, $n \to \infty$ для любого $\theta \in \Theta_0$, т.е. этот критерий является асимптотически подобным. Величины $t_{s,\alpha}$ определяются из таблиц распределения хи-квадрат.

2. Критерий максимума отношения правдоподобия асимптотически не смещен и состоятелен, а также состоятелен для последовательностей F_{θ_n} таких, что $\sqrt{n}\,\|\theta_n - \theta_0\| \to 0$ при любой $\theta_0 \in \Theta$.

При s=1 этот критерий является асимптотически равномерно наиболее мощным несмещенным (АРНМН) критерием.

7.4.3. Критерий хи-квадрат. Дискретная случайная величина

Пусть с.в. ξ является дискретной с конечным множеством значений $\bar{t} = \{t_1, \ldots, t_k\}$, $\bar{p} = \{p_1, \ldots, p_k\}$ – набор неизвестных вероятностей значений. Пусть основная гипотеза является простой и характеризуется набором $\bar{p}_0 = \{p_{1,0}, \ldots, p_{k,0}\}$; $p_{j,0} > 0$. Обозначим через P множество векторов:

$$\mathrm{P} = \left\{\bar{p} = (p_1, \ldots, p_k) : p_j \geqslant 0 \ ;\ \sum_j p_j = 1\right\} .$$

Тогда $\bar{p}_0$ можно рассматривать как параметр θ из $(k-1)$-мерного множества $\Theta = \mathrm{P}$ и рассматривать параметрическую задачу проверки согласия с H_0: $\bar{p} = \bar{p}_0$. Пусть n_j – количество элементов выборки $X_1, X_2, \ldots, X_n$, равных t_j, $j = 1, 2, \ldots, k$.

Критерий хи-квадрат основан на статистике

$$\chi_{n,k}^2 = \sum_{j=1}^{k} \frac{(n_j - np_{j,0})^2}{np_{j,0}} , \tag{7.1}$$

которая называется *статистикой хи-квадрат*.

Критерий хи-квадрат асимптотически эквивалентен критерию максимума отношения правдоподобия в рассматриваемой задаче и обладает аналогичными свойствами. В частности, в качестве порога статистики хи-квадрат можно выбрать величину $t_{k-1,\alpha}$, т.е. $(1-\alpha)$ - ю квантиль распределения хи-квадрат с k-1 степенью свободы.

7.4.4. Критерий хи-квадрат. Дискретизация

Рассмотрим теперь случайную величину ξ общего вида. Пусть основная гипотеза является простой, т.е. H_0: $F_\xi(x) = F_0(x)$. Чтобы применить критерий хи-квадрат в задаче проверки гипотезы H_0, используют *дискретизацию* данных: множество возможных значений случайной величины ξ разбивается на k непересекающихся подмножеств $I_1,\ldots,I_k$ (обычно это интервалы $I_j = [a_{j-1}, a_j)$, $j = 1,\ldots,k$, где $a_1<\ldots<a_{k-1}$ - точки деления; $a_0 = -\infty$, $a_k = +\infty$ (рис. 7.4).

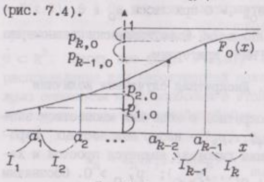


Рис.7.4

Для каждого элемента выборки X_i регистрируется номер $j = j(X_i)$ интервала I_j, в который попадает X_i, т.е. исходная случайная величина ξ заменяется новой дискретной случайной величиной ξ_k, принимающей значения 1,..., k, и при основной гипотезе:

$$p_{j,0} = P(\xi_k=j) = P(\xi\in I_j) = F_0(a_j) - F_0(a_{j-1}), \quad j=1,\ldots,k. \qquad (7.2)$$

Для случайной величины ξ_k основная гипотеза имеет вид H_0: $\bar{p} = \bar{p}_0 = (p_{1,0},\ldots,p_{k,0})$, и для ее проверки можно использовать критерий хи-квадрат, описанный в 7.4.3 и основанный на статистике $\chi^2_{n,k}$ вида (7.1).

Пусть основная гипотеза является сложной параметрической, т.е. H_0: $F_\xi(x) \in \left\{F(x,\theta),\ \theta \in \Theta\right\}$, где Θ - область в R^m, $m \geqslant 1$. Аналогично тому, как это делается в критерии максимума отношения правдоподобия для сложной основной гипотезы, при выполнении достаточно общих предположений "регулярности" в статистике хи-квадрат неизвестные вероятности $p_{j,0}$ можно заменить их оценками, вычисленными по формуле:

$$\hat{p}_{j,0} = F(a_j,\hat{\theta}_n) - F(a_{j-1},\hat{\theta}_n), \quad j=1,\ldots,k, \qquad (7.3)$$

где $\hat{\theta}_n$ - оценки максимального правдоподобия параметра $\theta \in \Theta$ или асимптотически эквивалентные им оценки. При этом пороговое значение статистики хи-квадрат заменяется на $t_{k-m,\alpha}$.

Таким образом, процедура проверки простой или параметрической гипотезы с помощью критерия хи-квадрат состоит из следующих этапов.

1. *При сложной основной гипотезе по выборке $X_1,\ldots,X_n$ строятся оценки максимального правдоподобия (или асимптотически эквивалентные им оценки) параметра $\theta \in \Theta$ для гипотезы.*

2. *Выбирается допустимый уровень значимости α и определяется порог $t_{k-m,\alpha}$ из таблиц распределения хи-квадрат.*

3. *Выбираются точки деления $a_1 < \ldots < a_{k-1}$ для разбиения $I_1,\ldots I_k$. Подсчитываются количества n_j элементов выборки, попадающих в подмножества I_j, $j=1,\ldots,k$.*

4. *Рассчитываются вероятности $p_{j,0}$ или $\hat{p}_{j,0}$ по формулам (7.2) или (7.3); разбиение выбирается так, чтобы эти вероятности были положительны (желательно, близкими к $1/k$) для всех $j=1,\ldots,k$.*

5. *Рассчитываются значения статистики $\chi^2_{n,k}$ по формуле (7.1) (с заменой $p_{j,0}$ на $\hat{p}_{j,0}$ для сложной гипотезы) и сравниваются с порогом $t_{k-m,\alpha}$, определяемым из таблиц для заданного α. Основная гипотеза принимается при $\chi^2_{n,k} < t_{k-m,\alpha}$ и отвергается в противном случае.*

Иногда рассчитывают "достигаемый уровень значимости α_n" из соотношения: $P_{k-m}(\chi^2_{n,k}) = 1 - \alpha_n$, характеризующий "степень надежности" решения: если α_n мало (существенно меньше α), то решение отвергнуть гипотезу является надежным. Отметим, что значения α_n, близкие к 1, *не говорят в пользу справедливости гипотезы:* для достаточно больших k-m при справедливости гипотезы наиболее вероятны значения α_n, близкие к 1/2.

Пример решения задачи проверки гипотезы нормальности с помощью критерия хи-квадрат приведен в примере к лабораторной работе №1.

7.5. Непараметрические критерии согласия,основанные на эмпирических мерах отклонения

7.5.1. Метод построения критериев

Пусть основная гипотеза является простой, т.е. H_0: $F_\xi(x)=F_0(x)$ и F - множество всех ф.р. Многие критерии согласия строятся

следующим образом. Пусть $\rho(F,F_0)$ - некоторая мера отклонения ф.р. $F(x)$ от $F_0(x)$, заданная для всех ф.р. $F(x)$. Построим по выборке X_1, ... , X_n эмпирическую ф.р. $F_n(x)$. Тогда статистику $\rho_n=\rho(F_n,F_0)$ можно рассматривать как статистику некоторого критерия. Если, кроме того, ф.р. $G_{n,F_0}(t)$ статистики $L_n = \sqrt{n}\ \rho(F_n,F_0)$ при гипотезе H_0 сходится при $n \to \infty$ к некоторой ф.р. $G(t)$ при любых t, то при заданном $\alpha \in (0,1)$ в качестве порога T_n критерия $\phi_{n,\alpha}$, основанного на статистике L_n, можно использовать $(1-\alpha)$- ю квантиль предельной ф.р., т.е. величину t_α, определяемую равенством $G(t_\alpha) = 1 - \alpha$.

Пусть мера отклонения $\rho(F,F_0)$ обладает следующими свойствами.

1) $\rho(F,F_0) \geqslant 0$; $\rho(F,F_0) = 0$, лишь если $F(x)=F_0(x)$ для всех x.

2) Если $n \to \infty$ и $F^{(n)}(x)$ - такая последовательность ф. р., что $D(F^{(n)},F) \to 0$, то $\rho(F^{(n)},F_0) \to \rho(F,F_0)$. Здесь

$$D(F,G) = \sup_x |F(x) - G(x)|. \qquad (7.4)$$

Тогда критерий $\phi_{n,\alpha}$, основанный на статистике $\rho(F_n,F_0)$, состоятелен , т.е. $\beta_n(\phi_{n,\alpha},F) \to 0$ для любой ф.р. $F \neq F_0$.

Свойствами 1 и 2 обладают меры отклонения $\rho(F,F_0)$, рассматриваемые ниже, что обеспечивает состоятельность соответствующих критериев.

7.5.2. Критерий Колмогорова

Пусть ф.р. F_0 непрерывна и $\rho(F,F_0) = D(F,F_0)$, где мера отклонения $D(F,F_0)$ определяется (7.4). Тогда по теореме Колмогорова ф.р. $G_{n,F_0}(t)$ статистики $D_n = \sqrt{n}\ D(F_n,F_0)$ при гипотезе H_0 сходится при $n \to \infty$ к ф.р. Колмогорова $K(t)$, не зависящей от F_0. Критерий, основанный на этой статистике, назывется *критерием Колмогорова*.

При вычислении меры отклонения $D(F_n,F_0)$ достаточно ограничиться точками $t_i = X_{(i)}$ вариационного ряда $X_{(1)},\ldots,X_{(n)}$ выборки, в которых эмпирическая ф.р. $F_n(t)$ имеет скачки со значения $(i-1)/n$ на i/n. Поэтому

$$D_n=D(F_n,F_0)=\max_{1\leqslant i\leqslant n}\left\{\max\left\{|F_0(X_{(i)})-i/n|,\ |F_0(X_{(i)})-(i-1)/n|\right\}\right\}. \qquad (7.5)$$

Процедура проверки гипотезы H_0: $F_\xi(t) = F_0(t)$ для непрерывной ф.р. $F_0(t)$ с помощью критерия Колмогорова состоит из следующих этапов.

1. *Выбирается допустимый уровень значимости α и определяется порог t_α из таблиц ф.р. Колмогорова из равенства $K(t_\alpha) = 1 - \alpha$.*
2. *Строится вариационный ряд $X_{(1)},\ldots,X_{(n)}$ выборки.*
3. *Рассчитывается величина $D_n = D(F_n,F_0)$ по формуле (7.5).*
4. *Величина $\mathrm{D}_n = \sqrt{n}\ D_n$ сравнивается с порогом t_α. Основная гипотеза принимается при $\mathrm{D}_n < t_\alpha$ и отвергается в противном случае.*

Пример. Пусть ξ есть случайное время, требующееся для обработки состава на грузовой горке, $\bar{u}$ - среднее нормативное время обработки. При изучении работы грузовой горки с помощью методов теории массового обслуживания удобно считать, что с.в. ξ имеет показательное распределение с параметром $\lambda = 1/\bar{u}$, т.е. H_0:$F_\xi(x) = 1 - \exp(x/\bar{u})$ при $x > 0$. Пусть $X_1,\ldots,X_n$ - значения, полученные при измерении времени обработки n составов, n = 100. Для проверки гипотезы H_0 с помощью критерия Колмогорова зададим допустимый уровень значимости α = 0.05. Тогда из таблиц ф.р. Колмогорова находим: t_α = 1.36. Далее по вариационному ряду $X_{(1)},\ldots,X_{(n)}$ выборки вычисляем значение статистики D_n по формуле (7.5), которая в данном случае имеет вид:

$$D_n = \max_{1\leqslant i\leqslant n}\{|\exp(X_{(i)}/\bar{u}) - (n-i)/n|,\ |\exp(X_{(i)}/\bar{u}) - (n-i+1)/n|\}.$$

Пусть получено значение D_n = 0.1. Тогда $\mathrm{D}_n = \sqrt{n}D_n = 1 < t_\alpha = 1.36$, т.е. результаты измерений не противоречат гипотезе H_0.

Если же получено значение D_n = 0.2, то D_n = 2 > t_α и гипотеза H_0 отвергается. Это решение можно считать весьма надежным, т.к. достижимый уровень значимости при этом $\alpha_n = 1 - K(\mathrm{D}_n) < 0.0001$.

7.5.3. Критерий ω^2 Мизеса - Смирнова

Критерий Мизеса - Смирнова основан на мере отклонения $\rho(F,F_0) = \omega(F,F_0)$, где мера отклонения $\omega(F,F_0)$ определяется соотношением

$$\omega^2(F,F_0) = \int_{-\infty}^{+\infty} (F(t) - F_0(t))^2\, dF_0(t)\ . \qquad (7.6)$$

Пусть ф.р. F_0 непрерывна. Тогда ф.р. $G_{n,F_0}(t)$ статистики

$L_n^2 = n\,\omega^2(F_n, F_0)$ при гипотезе H_0 сходится при $n \to \infty$ к ф.р. Смирнова $S(t)$, не зависящей от F_0. Значения ф.р. $S(t)$ табулированы. Поэтому в качестве статистики критерия Мизеса – Смирнова удобнее рассматривать не L_n, а L_n^2, и величину порога определять из таблиц ф.р. $S(t)$. Для с.в. ω^2, имеющей ф.р. $S(t)$, среднее и дисперсия есть $E(\omega^2) = 1/6$; $D(\omega^2) = 1/45$.

Так же как и для критерия Колмогорова, значения статистики $\omega_n^2 = \omega^2(F_n, F_0)$ можно вычислить по вариационному ряду $X_{(1)}, \ldots, X_{(n)}$ выборки (см. дополнение 5):

$$\omega_n^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^{n} \left[F_0(X_{(i)}) - \frac{2i-1}{2n} \right]^2. \tag{7.7}$$

Процедура проверки гипотез H_0: $F_\xi(t) = F_0(t)$ для непрерывной ф.р. $F_0(x)$ с помощью критерия Мизеса – Смирнова аналогична процедуре для критерия Колмогорова и основана на вычислениях по формуле (7.7).

ЛАБОРАТОРНАЯ РАБОТА № 1

Статический анализ данных в случае нормального распределения генеральной совокупности

Задание к лабораторной работе

В предположении, что данные эксперимента есть выборка из генеральной совокупности случайной величины (с.в.) $\xi \sim N(a, \sigma)$:

а) найти точечные оценки для математического ожидания a, дисперсии σ^2, параметров асимметрии β и эксцесса γ;

б) найти интервальные оценки для математического ожидания:

1) по правилу "3σ";

2) с доверительными вероятностями $\alpha_1 = 0.95$, $\alpha_2 = 0.99$;

в) получить интервальные оценки для параметра σ с доверительными вероятностями $\alpha_1 = 0.90$, $\alpha_2 = 0.95$;

г) пользуясь точечными оценками параметров a и σ^2, оценить значения функции распределения (ф.р.) $F_\xi(x)$ и плотности распределения $p_\xi(x)$ в точках, являющихся серединами интервалов группировки;

д) вычислить значения эмпирической ф.р. $\hat{F}_n(x)$ и гистограммы в тех же точках;

е) построить графики эмпирической ф.р., гистограммы, полигона частот;

ж) на чертежи из пункта *е* нанести результаты оценок ф.р. и плотности распределения в соответствующих точках, полученные в пункте *г*. Если наблюдения действительно распределены нормально, то результаты оценки этими двумя способами (*г* и *д*) обычно близки. Сравнить их визуально.

Пример выполнения задания в случае малой выборки

В таблице приведены результаты контрольного замера 30 деталей (микросхем для ЭВМ), изготовленных на одном станке.

Границы интервала, мм	2.9 – 3.9	3.9 – 4.9	4.9 – 5.9	5.9 – 6.9	6.9 – 7.9
Частота	3	5	14	6	2

Экспериментальные данные представлены в виде группированного статистического ряда. В этом случае элементами выборки считают середины интервалов группировки z_i ($i = 1, \ldots, k$, k – число интервалов), причем полагают, что z_i встречается n_i раз среди выборочных значений, n_i – частота попадания наблюдений в i-й интервал ($n_1 + n_2 + \ldots + n_k = n$). Отсюда вытекают формулы для вычисления выборочного среднего $\bar{X}$ и выборочной дисперсии S^2:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{k} z_i n_i , \quad S^2 = \frac{1}{n} \sum_{i=1}^{k} (z_i - \bar{X})^2 n_i ,$$

а также для оценок параметров асимметрии β и эксцесса γ:

$$\hat{\beta}_n = \frac{1}{n}\Big[\sum_{i=1}^{k} (z_i - \bar{X})^3 n_i \Big] / S^3, \quad \hat{\gamma}_n = \frac{1}{n}\Big[\sum_{i=1}^{k} (z_i - \bar{X})^4 n_i \Big] / S^4 - 3.$$

Выполнение пункта *а* задания состоит в оценке среднего, дисперсии и характеристик формы распределения: параметров асимметрии и эксцесса. Вычисления по приведенным выше формулам приводят к следующим результатам:

$$\hat{a} = \bar{X} = 5.3667, \quad \hat{\sigma}^2 = \frac{n}{n-1} S^2 = 1.0678,$$

$$\hat{\sigma} = (\hat{\sigma}^2)^{1/2} = 1.033,$$

$$\hat{\beta}_n = -0.1240, \quad \hat{\gamma}_n = -0.1758.$$

Перейдем к п. *б*. Построим вначале доверительный интервал для параметра $a = E\xi$ по правилу "3σ". С.в. $\bar{X} \sim N(a, \sigma/\sqrt{n})$ (см. дополн., п. 2), причем $\sigma_{\bar{X}} = \sigma/\sqrt{n}$; заменяя значение параметра σ его оценкой $\hat{\sigma} = 1.033$, находим, что неравенство

$$\bar{X} - 3\,\frac{\hat{\sigma}}{\sqrt{n}} < a < \bar{X} + 3\,\frac{\hat{\sigma}}{\sqrt{n}}, \quad \text{т. е. неравенство}$$

$4.801 < a < 5.933$, выполняется с вероятностью, близкой к 0.9973 (вероятности *"практической достоверности"*).

Теперь построим доверительные интервалы, содержащие значения параметра a с вероятностями $\alpha_1 = 0.95$ и $\alpha_2 = 0.99$. Найдем t_{α_1} по таблицам распределения Стьюдента как корень уравнения

$$\int_{-t_{\alpha_1}}^{t_{\alpha_1}} s_{n-1}(x)\,dx = \alpha_1 ,$$

где $s_{n-1}(x)$ – плотность распределения Стьюдента с $n - 1 = 29$ степенями свободы (см. приложение, табл.6). Мы имеем $t_{\alpha_1} = 2.045$. Подставляя это значение вместо t_α в выражение для границ доверительного интервала

$$\bar{X} - t_\alpha \frac{\hat{\sigma}}{\sqrt{n}} < a < \bar{X} + t_\alpha \frac{\hat{\sigma}}{\sqrt{n}}$$

(см. разд. 6.2), находим, что интервал (4.981, 5.753) содержит истинное значение неизвестного параметра $a = E\xi$ с вероятностью $\alpha_1 = 0.95$. Аналогично для $\alpha_2 = 0.99$ находим $t_\alpha = 2.756$ и, повторяя вычисления для этого значения, получаем интервал (4.847, 5.887), содержащий внутри себя значение параметра a с вероятностью $\alpha_2 = 0.99$.

Выполним п. *в* задания. Для уровня доверия $\alpha = 0.90$ находим χ_1^2 и χ_2^2 как корни уравнений

$$\int_0^{\chi_1^2} k_{n-1}(x)\,dx = \frac{1 - \alpha_1}{2} ; \quad \int_0^{\chi_2^2} k_{n-1}(x)\,dx = \frac{1 + \alpha_1}{2} ,$$

где $k_{n-1}(x)$ – плотность распределения χ^2 с $n-1 = 29$ степенями свободы соответственно по таблицам значений ф.р. χ^2 с 29 степенями свободы. Находим, что $\chi_1^2 = 17.7$ и $\chi_2^2 = 42.6$. Подставляя эти значения в выражения для границ доверительных интервалов

$$\frac{(n-1)\hat{\sigma}^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)\hat{\sigma}^2}{\chi_1^2} \quad \text{и} \quad \frac{(n-1)^{1/2}\,\hat{\sigma}}{\chi_2} < \sigma < \frac{(n-1)^{1/2}\,\hat{\sigma}}{\chi_1}$$

(см.разд. 6.2), находим, что с вероятностью $\alpha_1 = 0.90$ имеют место неравенства $0.727 < \sigma^2 < 1.75$, $0.853 < \sigma < 1.323$.

Аналогичные вычисления для уровня доверия $\alpha_2 = 0.99$, при котором $\chi_1^2 = 13.1$, $\chi_2^2 = 52.3$, приводят к неравенствам

$$0.592 < \sigma^2 < 2.364, \quad 0.769 < \sigma < 1.537,$$

которые выполняются с вероятностью $\alpha_2 = 0.99$.

Перейдем к п. *г* задания. В случае нормального распределения генеральной совокупности, естественной оценкой плотности распределения является функция:

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}}\, e^{-\frac{1}{2}\left[\frac{x - \hat{a}}{\hat{\sigma}}\right]^2},$$

получаемая заменой параметров a и σ в формуле плотности нормального закона их оценками $\hat{a}$ и $\hat{\sigma}$. Аналогично оценкой для ф.р. $F_\xi(x)$ может служить функция:

$$\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(t)\,dt = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}} \int_{-\infty}^{x} e^{-\frac{(t-\hat{a})}{2\hat{\sigma}^2}}\,dt.$$

Поскольку $\hat{a} \xrightarrow{P} a$ и $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ при $n \to \infty$ (см.разд. 4), то в силу непрерывности функций $\hat{f}$ и $\hat{F}$ как функций $\hat{a}$ и $\hat{\sigma}^2$ при $n \to \infty$ имеется сходимость по вероятности $\hat{f}(x) \xrightarrow{P} f(x)$ и $\hat{F}(x) \xrightarrow{P} F(x)$ при всех вещественных x.

Мы вычислим значения $\hat{f}(x)$ и $\hat{F}(x)$ в точках z_i (серединах интервалов группировки). Имеем

$$\hat{f}(z_i) = \frac{1}{\hat{\sigma}}\,\varphi\left[\frac{z_i - \hat{a}}{\hat{\sigma}}\right],$$

где $\varphi(t) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}}$ – плотность распределения стандартного нормального закона. Значения $\hat{f}(z_i)$ можно найти, пользуясь таблицами функции $\varphi(t)$ (см. приложение,табл.2) или с помощью компьютера.

Оценки для значений ф.р. $\hat{F}_\xi(x) = \Phi\left(\frac{x-a}{\sigma}\right)$ найдем, используя соотношение:

$$\hat{F}_\xi(z_i) = \Phi\left[\frac{z_i - \hat{a}}{\hat{\sigma}}\right],$$

где $\Phi(x)$ – стандартная нормальная ф.р., по таблицам функции $\Phi(x)$

(см. приложение, табл.1). Результаты вычислений внесем в таблицу.

Выполним пункт *д* задания. Более универсальный, справедливый не только в параметрической ситуации (как здесь) способ оценки ф.р. и плотности распределения состоит в построении эмпирической ф.р., гистограммы и полигона частот. Эмпирическая ф.р. $\hat{F}_n(x)$ – это неубывающая ступенчатая функция, определенная при каждом вещественном x равенством $\hat{F}_n(x) = \dfrac{\nu_n(x)}{n}$, где $\nu_n(x)$ – число выборочных значений X_i, меньших x (см. разд.2). Однако в случае гладкой ф.р. $F_\xi(x)$ (какой является нормальная ф.р.) ее значения в серединах интервалов группировки точнее оцениваются, если линейно сгладить эмпирическую ф.р. внутри интервалов группировки (т.е. считать, что выборочные значения равномерно распределены внутри каждого из интервалов группировки, что означает, что ф.р. $F_\xi(x)$ заменяется линейной функцией на каждом из них). Таким образом, если для первого интервала (в нашем примере) значения $\hat{F}_n(x)$ на его концах равны 0 и 0.1 соответственно, то в точке z_1 (середине первого интервала) мы полагаем $\hat{F}_n(z_1) = \frac{1}{2}\,(0 + 0.1) = 0.05$. Для второго интервала группировки значения $\hat{F}_n(x)$ на концах равны соответственно 3/30 и 8/30, и мы полагаем $\hat{F}_n(z_2) = \frac{1}{2}\,(3/30 + 8/30) = 11/30 = 0.367$. Далее, $\hat{F}_n(z_3) = \frac{1}{2}\,\frac{1}{30}\,(8 + 30) = 0.5$, $\hat{F}_n(z_4) = \frac{1}{2}\,\frac{1}{30}\,(22 + 28) = \frac{5}{6} = 0.833$, $\hat{F}_n(z_5) = \frac{1}{2}\,\frac{1}{30}\,(28 + 30) = \frac{29}{30} = 0.967$. Результаты вычислений внесены в таблицу.

Границы интервалов	Частоты n_i	z_i	$\dfrac{z_i - \hat{a}}{\hat{\sigma}}$	$\hat{F}_\xi(z_i)$	$\hat{f}(z_i)$	$\hat{F}_n(z_i)$	$g(z_i)$
2.9 - 3.9	3	3.4	-1.903	0.0287	0.063	0.05	0.1
3.9 - 4.9	5	4.4	-0.935	0.1762	0.249	0.367	0.167
4.9 - 5.9	14	5.4	0.032	0.5120	0.386	0.5	0.467
5.9 - 6.9	6	6.4	1.000	0.8413	0.234	0.833	0.2
6.9 - 7.9	2	7.4	1.968	0.9750	0.056	0.967	0.067

Значения гистограммы $g(x)$ (и полигона частот) в точках z_i определяются по формулам:

$$g(x) = \frac{n_i}{nh}, \quad \text{где } h = \frac{W}{k} = \frac{X_n - X_1}{k}$$

(см.разд. 3). Мы имеем $h = \dfrac{7.9 - 2.9}{5} = 1$, то есть $g(z_i) = \dfrac{n_i}{n}$.

Результаты вычислений внесены в последний столбец таблицы.

Выполнение п. *ж* состоит в построении (и сравнении) графиков функций $\hat{F}_\xi(x)$ и $\hat{F}_n(x)$ и функций $\hat{f}_\xi(x)$ и $g(x)$.

40

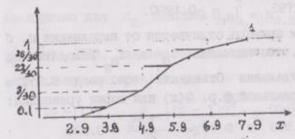


Рис.1. Графики оценок ф.р.

На рис.1 изображен график функции $\hat{F}_\xi(x)$ (расчетные точки $(z_i, \hat{F}_\xi(z_i))$ соединены плавной кривой) и график эмпирической (линейно сглаженной) ф.р. $\hat{F}_n(x)$.

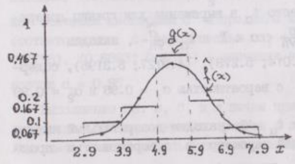


Рис.2. Графики оценок плотности

На рис.2 изображены графики гистограммы, полигона частот и функции $\hat{f}(x)$ – оценки плотности (расчетные точки $(z_i, \hat{f}(z_i))$ соединены гладкой кривой), полученной подстановкой оценок математического ожидания $\hat{a}$ и дисперсии $\hat{\sigma}^2$ в формулу плотности нормального закона.

Пример выполнения задания в случае выборки большого объема

В таблице приведены значения времени опозданий 130 пассажирских поездов.

Границы интервалов, мин	3.0–3.6	3.6–4.2	4.2–4.8	4.8–5.4	5.4–6.0	6.0–6.6	6.6–7.2
Частота	2	8	35	43	22	15	5

Выполним п. *а* задания. При $k = 7$, $n = 130$ и данных частотах находим точечные оценки для математического ожидания, дисперсии

среднего квадратического отклонения генеральной совокупности:

$$\hat{a} = \bar{X} = \frac{1}{n}\sum_{i=1}^{k} z_i n_i = 5.146,$$

$$\hat{\sigma}^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(z_i - \bar{X})^2\, n_i = 0.589,$$

$$\hat{\sigma} = (\hat{\sigma}^2)^{1/2} = 0.768.$$

Оценки параметров асимметрии и эксцесса:

$$\hat{\beta}_n = 0.2793, \quad \hat{\gamma}_n = -0.1990.$$

Выполнение п.*б* в этом примере отличается от выполнения п. *б* в предыдущем примере тем, что значения t_{α_1} и t_{α_2} находятся не по таблице квантилей распределения Стьюдента (см. разд. 6.2), а по таблицам стандартной нормальной ф.р. $\Phi(x)$ как корни уравнений:

$$\Phi(t) = \frac{1 + \alpha_1}{2} = 0.975 \text{ и } \Phi(t) = \frac{1 + \alpha_2}{2} = 0.995$$

соответственно. Находим $t_{\alpha_1} = 1.96$, $t_{\alpha_2} = 2.576$.

Подставляя эти числа вместо t_α в выражение для границ доверительного интервала $\bar{X} - t_\alpha \frac{\hat{\sigma}}{\sqrt{n}} < a < \bar{X} + t_\alpha \frac{\hat{\sigma}}{\sqrt{n}}$, находим доверительные интервалы (5.014; 5.278), (4.9727; 5.3196), содержащие значения параметра a с вероятностью $\alpha_1 = 0.95$ и $\alpha_2 = 0.99$ соответственно.

В результате подстановки $t_\alpha = 3$, находим доверительный интервал (4.944; 5.348), содержащий параметр a с вероятностью *практической достоверности*.

Выполнение п.*в* задания также подразумевает использование других таблиц. Значения ф.р. χ^2 не табулируют при большом числе степеней свободы, имея в виду возможность аппроксимации ф.р. нормированной и центрированой с.в. χ^2 стандартной нормальной ф.р. $\Phi(x)$, т.е. приближенное равенство (при достаточно больших n)

$$P\{\chi_n^2 < x\} \approx \Phi\left[\frac{x - n}{\sqrt{2n}}\right],$$

которое используют на практике уже при $n > 30$. Таким образом, для доверительной вероятности α значения χ_1^2 и χ_2^2 находят из уравнений:

$$\Phi\left[\frac{x-(n-1)}{\sqrt{2(n-1)}}\right] = \frac{1 - \alpha_1}{2}, \quad \Phi\left[\frac{x-(n-1)}{\sqrt{2(n-1)}}\right] = \frac{1 + \alpha_1}{2}.$$

Отсюда, учитывая что $\Phi(-x) = 1 - \Phi(x)$, получаем формулы:

$\chi_1^2 = (n-1) - u_{\frac{1+\alpha}{2}}\sqrt{2(n-1)}$, $\chi_2^2 = (n-1) + u_{\frac{1+\alpha}{2}}\sqrt{2(n-1)}$, где $u_{\frac{1+\alpha}{2}}$ – квантиль уровня $\frac{1+\alpha}{2}$ стандартного нормального закона. Для α_1 по таблицам функции $\Phi(x)$ (см.приложение, табл.1) находим $u_{\frac{1+\alpha_1}{2}} = u_{0.95} = 1.65$. Соответственно получаем $\chi_1^2 = 102.497$ и $\chi_2^2 = 155.529$. Аналогично для α_2 находим $u_{\frac{1+\alpha_2}{2}} = u_{0.975} = 1.96$, и $\chi_1^2 = 97.5177$, $\chi_2^2 = 160.4823$. Подставляя эти значения в выражения для границ доверительных интервалов

$$\frac{(n-1)\hat{\sigma}^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)\hat{\sigma}^2}{\chi_1^2} \text{ и } \frac{(n-1)^{1/2}\hat{\sigma}}{\chi_2} < \sigma < \frac{(n-1)^{1/2}\hat{\sigma}}{\chi_1},$$

находим, что интервалы (0.4889; 0.7419) и (0.6992; 0.8613) содержат значения дисперсии σ^2 и среднего квадратического отклонения σ соответственно с вероятностью $\alpha_1 = 0.90$, а интервалы (0.4738; 0.7798), (0.6883; 0.8831) – значения этих же параметров с вероятностью $\alpha_2 = 0.95$.

Выполнение пп. *г*, *д*, и *ж* ничем принципиально не отличается от выполнения соответствующих пунктов в примере 1, поэтому мы приводим здесь только результаты вычислений, сведенные в таблицу.

Границы интервалов	Частоты n_i	z_i	$\dfrac{z_i - \hat{a}}{\hat{\sigma}}$	$\hat{F}_\xi(z_i)$	$\hat{f}(z_i)$	$\hat{F}_n(z_i)$	$g(z_i)$
3.0 - 3.6	2	3.3	-2.405	0.0080	0.029	0.0077	0.015
3.6 - 4.2	8	3.9	-1.623	0.0526	0.139	0.0462	0.061
4.2 - 4.8	35	4.5	-0.842	0.2010	0.365	0.2115	0.269
4.8 - 5.4	43	5.1	-0.060	0.4963	0.519	0.5115	0.331
5.4 - 6.0	22	5.7	0.721	0.7640	0.401	0.7615	0.169
6.0 - 6.6	15	6.3	1.503	0.9333	0.168	0.9038	0.115
6.6 - 7.2	5	6.9	2.284	0.9891	0.038	0.9808	0.038

Варианты заданий к лабораторной работе № 1

Номер варианта		Нижняя граница Xmin	Число интер-валов r	Шаг h	Частоты							
1	а)	0.1	6	0.3	2	3	8	12	4	1	-	-
	б)	1.2	8.	0.4	10	25	30	42	40	35	15	8
2	а)	2.2	6	0.5	1	4	7	11	3	2	-	-
	б)	4.4	8	0.6	8	16	26	38	40	35	18	10
3	а)	-2.4	6	0.5	2	4	9	10	3	2	-	-
	б)	10.0	8	0.5	9	26	32	40	38	34	20	10
4	а)	0.0	6	1.0	3	3	9	14	9	5	-	-
	б)	-5.0	8	0.4	10	22	33	42	38	35	17	8
5	а)	-15.2	6	0.6	3	7	12	10	3	2	-	-
	б)	2.7	8	0.3	12	20	34	41	36	18	12	4
6	а)	20.0	6	07	1	3	6	14	8	3	2	
	б)	1.3	8	2.4	12	24	28	42	38	24	12	
7	а)	-12.4	6	2.1	1	2	10	14	5	1	-	-
	б)	18.0	8	3.0	8	26	30	43	42	32	18	6
8	а)	-5.0	6	1.4	4	6	10	8	5	3	-	-
	б)	2.2	8	0.3	14	22	34	48	40	38	24	10
9	а)	34.2	6	7.8	2	3	8	12	4	1	-	-
	б)	-12.4	8	0.6	10	25	30	42	40	35	15	8
10	а)	16.2	6	1.8	1	5	10	10	4	2	-	-
	б)	-8.2	8	2.2	8	18	28	42	40	32	24	8
11	а)	3.0	6	0.6	3	5	10	12	5	1	-	-
	б)	0.2	8	0.3	11	27	33	53	42	20	16	12
12	а)	-4.2	6	1.2	1	4	12	8	3	2	-	-
	б)	12.4	8	0.6	18	24	36	45	38	30	20	15
13	а)	2.8	6	0.2	3	7	16	8	4	2	-	-
	б)	-10.2	8	1.2	12	28	39	50	43	32	20	16
14	а)	0.8	6	0.3	2	4	7	12	3	2	-	-
	б)	18.2	8	0.8	13	26	33	43	39	30	16	4
15	а)	8.0	6	1	1	3	8	12	4	2	-	-
	б)	2.2	8	0.8	8	15	35	40	42	30	25	10
16	а)	-4.2	6	1.2	2	4	6	14	5	1	-	-
	б)	0.2	8	0.4	8	27	28	44	38	37	13	10
17	а)	6.0	6	1	2	2	10	5	2	1	-	-
	б)	-8.4	8	1.4	12	23	32	40	42	33	17	6
18	а)	5.2	6	0.3	1	3	9	6	3	2	-	-
	б)	-5.0	8	0.6	8	26	29	50	37	24	16	8
19	а)	2.2	6	0.8	2	4	6	12	2	2	-	-
	б)	24.4	8	1.6	20	50	60	84	70	30	16	
20	а)	-1.7	6	0.7	4	6	8	10	8	2	-	-
	б)	20.0	8	0.5	16	54	56	88	76	70	26	18

ЛАБОРАТОРНАЯ РАБОТА № 2

Проверка статистических гипотез

Предлагаемая лабораторная работа состоит из трех частей. В первой части требуется проверить гипотезу о параметрах нормального закона в случае малых выборок, используя точные распределения статистик критерия. Во второй части требуется проверить гипотезу о параметрах биномиального, пуассоновского или показательного распределения для больших выборок, используя асимптотическое распределение статистик критерия. В третьей части требуется проверить гипотезу о типе распределения с помощью критерия χ^2.

Задание к первой части лабораторной работы

В предположении, что данные эксперимента есть выборка из генеральной совокупности случайных величин $\xi \sim N(a,\sigma^2)$, требуется проверить гипотезу о неизвестном параметре распределения. При решении следует выполнить следующие пункты:

1) выбрать нулевую гипотезу и альтернативу;

2) указать, в чем состоят ошибки первого и второго рода;

3) предложить критерий для проверки нулевой гипотезы, указав распределение статистики и критическое множество;

4) построить график функции мощности предложенного критерия, вычислив значение функции мощности в 5-10 равноотстоящих узлах;

5) проверить нулевую гипотезу.

Замечание. Пункты 4 и 5 предлагается выполнить двумя способами - используя статистические таблицы, приведенные в приложении, а также методами статистического моделирования с помощью системы MatLAB (или MATHCAD).

Пример выполнения задания. Для стальной проволоки, используемой при изготовлении стальных тросов, среднее предельное значение растягивающего усилия равно 6720 кг/см^2 (обычно при таком усилии происходит разрыв проволоки). Стандартное уклонение предельного усилия от среднего равно 220 кг/см^2. С целью контроля из каждой партии проволоки, поступающей на завод, подвергается испытанию на разрыв 20 экземпляров. Результаты одного из таких испытаний следующие: 6300, 6780, 7130, 6930, 6870, 6720, 6690,

6720, 6720, 6630, 6750, 6950, 6980, 6660, 6560, 6960, 6780, 6900, 6700, 6780. Следует ли на основании этих данных забраковать партию? ($\alpha = 0.05$).

1) В качестве нулевой гипотезы рассмотрим гипотезу H_0: $a = a_0 = 6720$, а в качестве альтернативы H_1: $a < a_0$.

2) Ошибка первого рода состоит в том, что проволоку с растягивающим усилием на разрыв 6720 кг/см^2 мы будем браковать. Ошибка второго рода состоит в том, что проволоку с растягивающим усилием на разрыв, равным $a < 6720$, мы будем классифицировать как проволоку, выдерживающую усилие на разрыв в 6720 кг/см^2.

3) В качестве статистики критерия рассмотрим статистику

$$Z = \frac{\overline{X} - a_0}{\sigma/\sqrt{n}},$$

имеющую, в случае истинности H_0, стандартное нормальное распределение. Пусть z_α - квантиль порядка $1 - \alpha = 0.95$ для стандартного нормального распределения. В качестве критического множества рассмотрим множество $S = \{s | s \leqslant -z_\alpha\}$.

4) Если истинное значение среднего предельного усилия равно a, то статистика Z имеет нормальное распределение с параметрами $EZ = \frac{a - a_0}{\sigma/\sqrt{n}}$, $DZ = 1$, следовательно, статистика

$$Z_a = Z - \frac{a - a_0}{\sigma/\sqrt{n}}.$$

имеет стандартное нормальное распределение. Поэтому $P\{Z \leqslant -z_\alpha\} =$

$$= P\left\{Z_a + \frac{a - a_0}{\sigma/\sqrt{n}} \leqslant -z_\alpha\right\} = P\left\{Z_a \leqslant \frac{a - a_0}{\sigma/\sqrt{n}} - z_\alpha\right\} =$$

$= \Phi\left[\frac{a - a_0}{\sigma/\sqrt{n}} - z_\alpha\right]$, где Φ - функция распределения стандартного нормального закона. Таким образом,

$$\beta(a) = \Phi\left[\frac{a - a_0}{\sigma/\sqrt{n}} - z_\alpha\right].$$

Для построения графика функции мощности по точкам воспользуемся таблицами стандартного нормального распределения и его квантилей. По таблице квантилей находим $z_\alpha = 1.65$. Учитывая, что $n = 20$, а $\sigma = 220$ кг/см^2, по таблицам стандартного нормального распределения для точек $a_i = 6500, 6550, 6600, 6650, 6700$ получим

$\beta(6500) = 0.99$, $\beta(6550) = 0.96$, $\beta(6600) = 0.79$, $\beta(6650) = 0.41$, $\beta(6700) = 0.11$.

5) Фактическое значение статистики $Z = \frac{6775 - 6720}{220/\sqrt{20}} = 1.12$.

Поскольку $1.12 > -1.65 = -z_\alpha$, то гипотеза H_0 не отклоняется.

Замечание. Для проверки гипотезы H_0 и построения графика функции мощности при помощи компьютера можно воспользоваться методами статистического моделирования, оценивания вероятности событий частотами их наступления в достаточно длинной серии испытаний. В нашем случае статистика критерия Z при справедливости нулевой гипотезы имеет стандартное нормальное распределение и приняла фактическое значение 1.12, что, очевидно, больше отрицательного числа $-z_\alpha$, так что гипотеза H_0 принимается очевидным образом (без использования таблиц). Допустим, что для другой партии проволоки статистика Z приняла значение -1.12. Пусть p_N - частота события $\{Z_i < -1.12\}$ в достаточно длинной серии N испытаний, представляющих собой независимые реализации стандартной нормальной случайной величины. Если окажется, что $p_N > \alpha$, то гипотеза H_0 принимается, а если $p_N \leqslant \alpha$, то H_0 отклоняется. Например, для $N = 1000$ моделирование в системе MatLAB дает следующий результат: $p_{1000} = 0.1480$. Поскольку $0.1480 > 0.05$, то в этом случае гипотеза H_0 также будет принята.

Задание ко второй части лабораторной работы

В предположении, что данные эксперимента представляют собой выборку из генеральной совокупности случайных величин, имеющих заданный тип распределения (бернуллиевское ($B(p)$), пуассоновское ($\Pi(\lambda)$), показательное ($Exp(\lambda)$)), проверить указанную гипотезу H_0 против альтернативы H_1, если известны объем выборки n и фактическое значение статистики $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Пример выполнения задания. Фактическая частота наступления некоторого события в $n = 100$ независимых испытаниях равна 0.1. Требуется проверить гипотезу $H_0: p = p_0 = 0.15$ против альтернативы $H_1: p \neq p_0$ ($\alpha = 0.05$).

Известно, что статистика $\hat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ является

Варианты заданий к лабораторной работе №2

Номер варианта	Тип распределения	Основная гипотеза H_0	Альтернативная гипотеза H_1	Значение статистики $\bar{X}$	Объем выборки
1	B(p)	$p = 0.5$	$p \neq 0.5$	0.46	100
2	П(λ)	$\lambda = 1$	$\lambda > 1$	1.23	100
3	Exp(λ)	$\lambda = \frac{3}{2}$	$\lambda < 1.5$	0.75	100
4	B(p)	$p = 0.2$	$p > 0.2$	0.25	100
5	П(λ)	$\lambda = 3$	$\lambda < 3$	2.53	100
6	Exp(λ)	$\lambda = 1$	$\lambda \neq 1$	0.95	100
7	Exp(λ)	$\lambda = 0.5$	$\lambda > 0.5$	1.75	100
8	B(p)	$p = 0.8$	$p < 0.8$	0.76	100
9	П(λ)	$\lambda = 0.5$	$\lambda \neq 0.5$	0.64	100
10	П(λ)	$\lambda = 10$	$\lambda < 10$	10.2	100
11	Exp(λ)	$\lambda = 3$	$\lambda < 3$	0.31	100
12	B(p)	$p = 0.3$	$p \neq 0.3$	0.22	100
13	B(p)	$p = 0.7$	$p < 0.7$	0.73	100
14	Exp(λ)	$\lambda = 2$	$\lambda < 2$	0.8	100
15	Exp(λ)	$\lambda = 5$	$\lambda > 5$	0.24	100
16	П(λ)	$\lambda = 7$	$\lambda < 7$	0.72	100
17	П(λ)	$\lambda = 5$	$\lambda \neq 5$	4.32	100
18	B(p)	$p = 0.5$	$p < 0.5$	0.38	100
19	B(p)	$p = 0.5$	$p > 0.5$	0.56	100
20	Exp(λ)	$\lambda = 4$	$\lambda > 4$	0.21	100

состоятельной, несмещенной, асимптотически нормальной оценкой параметра p распределения Бернулли. В частности, при справедливости основной гипотезы статистика

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

имеет при больших n распределение, близкое к стандартному нормальному распределению. Пусть z_α – квантиль порядка $1 - \frac{\alpha}{2}$ для стандартного нормального распределения, т.е.

$$\int_{-z_\alpha}^{z_\alpha} \varphi(t)\, dt = 1 - \alpha \text{ , где } \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} .$$

При больших n $P\{|Z| \geqslant z_\alpha\} \approx 1 - \alpha$. Поэтому гипотезу H_0 будем отклонять, если $|Z| \geqslant z_\alpha$, и не отклонять в противном случае. Имеем

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.1 - 0.15}{\sqrt{\frac{0.15 * 0.85}{100}}} = -1.4 .$$

По таблице квантилей стандартного нормального распределения находим $z_\alpha = 1.96$. Поскольку $|Z| < z_\alpha$, то гипотеза H_0 не отклоняется.

Задание к третьей части лабораторной работы

В третьей части лабораторной работы требуется проверить гипотезу H_0 о том, что распределение генеральной совокупности является нормальным по данным к лабораторной работе №1 (см.стр.44). Для проверки предлагается использовать критерий согласия χ^2.

Пример выполнения задания. В таблице приведены значения времени опозданий 130 пассажирских поездов.

Границы интервалов, мин	3.0–3.6	3.6–4.2	4.2–4.8	4.8–5.4	5.4–6.0	6.0–6.6	6.6–7.2
Частота	2	8	35	43	22	15	5

Имеем (см. пример 2 к ЛР № 1) $\hat{a} = \bar{X} = 5.146$, $\hat{\sigma}^2 = 0.5895$, $\hat{\sigma} = 0.7678$. Для того, чтобы вычислить значение статистики χ^2_n, найдем гипотетические (при условии справедливости основной гипотезы H_0: $\xi \sim N(\hat{a},\hat{\sigma})$) вероятности попадания с.в. ξ в интервалы группировки:

$$p_i = \hat{F}_\xi(a_i) - \hat{F}_\xi(a_{i-1}) = \Phi\left(\frac{a_i - \bar{X}}{\hat{\sigma}}\right) - \Phi\left(\frac{a_{i-1} - \bar{X}}{\hat{\sigma}}\right),$$

где a_i – границы интервалов. При этом ввиду того, что частоты попадания значений с.в. ξ в крайние интервалы малы (2 и 5), объединим крайние соседние интервалы (см. рекомендации по применению критерия χ^2 [6],[8]). Таким образом, мы будем вычислять статистику χ^2_n при пяти интервалах: $(-\infty , 4.2)$; $(4.2 , 4.8)$; $(4.8 , 5.4)$; $(5.4 , 6.0)$; $(6 , +\infty)$, частоты n_i попадания в которые равны соответственно: 10; 35; 43; 22; 20. Найдем значения гипотетической функции распределения $\hat{F}_\xi(x)$ в граничных точках интервалов a_i, $i=0, 1, 2, 3, 4, 5$, (где $a_0=-\infty$, $a_5=+\infty$). Имеем: $\hat{F}_\xi(a_0) = 0$;

$$\hat{F}_\xi(a_1) = \Phi\left(\frac{4.2 - 5.146}{0.7678}\right) = \Phi(-1.2321) = 1 - \Phi(1.2321) = 0.1093;$$

$$\hat{F}_\xi(a_2) = \Phi\left(\frac{4.8 - 5.146}{0.7678}\right) = \Phi(-0.4507) = 1 - \Phi(0.4507) = 0.3264;$$

$$\hat{F}_\xi(a_3) = \Phi\left(\frac{5.4 - 5.146}{0.7678}\right) = \Phi(0.3308) = 0.6293;$$

$$\hat{F}_\xi(a_4) = \Phi\left(\frac{6.0 - 5.146}{0.7678}\right) = \Phi(1.1123) = 0.8665; \quad \hat{F}_\xi(a_5) = 1.$$

Теперь найдем гипотетические вероятности p_i:

$p_1 = \hat{F}_\xi(a_1) = 0.1093$; $p_2 = \hat{F}_\xi(a_2) - \hat{F}_\xi(a_1) = 0.2171$;

$p_3 = \hat{F}_\xi(a_3) - \hat{F}_\xi(a_2) = 0.3029$; $p_4 = \hat{F}_\xi(a_4) - \hat{F}_\xi(a_3) = 0.2372$;

$p_5 = \hat{F}_\xi(a_5) - \hat{F}_\xi(a_4) = 1 - 0.8665 = 0.1335$.

Вычисляя значение статистики χ^2_n (сумму Пирсона), получаем

$$\chi^2_n = \sum_{i=1}^{5} \frac{(n_i - np_i)^2}{np_i} = 6.1425.$$

Статистика χ^2_n в случае справедливости основной гипотезы имеет распределение хи-квадрат с числом степеней свободы, равным $r = k - 1 - m$, где k – число интервалов, m – число параметров гипотетического распределения, которые оценивались по результатам

наблюдений. В нашем случае $r = 5 - 1 - 2 = 2$. По таблицам квантилей χ^2 – распределения (см. приложение, табл.5) с двумя степенями свободы находим t_α так, чтобы $P\{\chi^2_2 \geqslant t_\alpha\} = 0.05$. Получим $t_\alpha = 5.99$. Поскольку фактическое значение $\chi^2_n = 6.1425 > 5.99$, то гипотеза H_0 при уровне значимости $\alpha = 0.05$ противоречит опытным данным и отклоняется.

Задачи к лабораторной работе № 2

1. Партия шин проверяется на износ протекторов. Обычной процедурой является сравнение с эталоном (контрольным образцом), имеющим характеристику износа 100 усл.ед. Для рассматриваемой партии (из десяти шин) измерения этой характеристики дали следующие результаты: 113, 100, 100, 94, 102, 92, 97, 104, 103, 99. Можно ли по этим данным считать, что характеристика на износ у исследуемой партии не ниже эталонной? ($\alpha = 0.05$).

2. В лаборатории, изучающей воздействие окружающей среды на человека, с целью определения комнатной температуры, при которой наиболее комфортабельно чувствуют себя мужчины и женщины, были обследованы 10 мужчин и 10 женщин. Были получены следующие температуры (в $^\circ C$) наибольшей комфортабельности.

Мужчины: 22, 25, 24, 24, 21, 23, 24, 22, 26, 22, 23.
Женщины: 23, 26, 25, 25, 22, 22, 27, 26, 23, 25.

Можно ли на основании этих данных считать, что температура наибольшего комфорта для мужчин и женщин одинакова? ($\alpha = 0.05$).

3. Для исследования на антикоррозийную стойкость нержавеющей стали 18Cr-10Ni-2Mo отобраны 12 образцов. Предварительно образцы подвергают исследованию на процентное содержание входящих в них элементов. Для данной партии измерение процентного содержания хрома дало следующие результаты: 17.4, 17.9, 17.6, 18.1, 17.6, 18.4, 16.9, 17.5, 17.8, 17.4, 24.6, 26.0. Можно ли считать, что содержание хрома в стали равно 18%? ($\alpha = 0.05$).

4. В полиметаллическом руднике жилы залегают в кварцевых кератофилах и известняках. Было взято по 7 проб каждой породы. Измерение процентного содержания руды дало следующие результаты.

В кератофилах: 2.50, 2.55, 2.60, 2.75, 2.80, 2.80, 2.95.
В известняках: 2.50, 2.80, 2.85, 2.94, 2.95, 2.95, 3.40.
Можно ли считать, что содержание руды в обоих типах пород одинаково? ($\alpha = 0.05$).

5. Двумя лабораториями определялась концентрация серы в дизельном топливе в стандартном образце, для которого она равна 0.870 усл.ед. Восемь равноточечных измерений в первой лаборатории дали результаты: 0.869, 0.874, 0.867, 0.875, 0.870, 0.869, 0.864, 0.872; десять независимых равноточечных измерений во второй лаборатории следующие: 0.866, 0.871, 0.868, 0.870, 0.871, 0.870, 0.869, 0.874. Можно ли считать точность измерений этими лабораториями одинаковой? ($\alpha = 0.05$).

6. С десяти участков цинкового месторождения взято по одной пробе породы. Измерение процентного содержания цинка дало следующие результаты: 0.6, 2.4, 2.1, 1.4, 1.2, 4.8, 0.9, 3.1, 3.5, 3.0. Можно ли на основании этих данных считать, что процентное содержание цинка на этом месторождении равно 2.7%? ($\alpha = 0.05$).

7. Детям одной из групп детского сада, состоящей из 10 человек, ежедневно выдавали молоко, а детям из другой группы, состоящей из 12 человек, - апельсиновый сок. Через некоторое время было зафиксировано следующее увеличение веса детей (в килограммах).
I группа: 1.6, 1.0, 1.4, 1.6, 0.6, 0.4, 1.4, 1.2, 1.0.
II группа: 0.6, 1.4, 1.0, 1.2, 1.0, 0.8, 0.8, 1.0, 0.6, 0.8, 1.2.
Можно ли считать увеличение веса в обеих группах одинаковым? (α 0.05).

8. Средняя разность в длине правого и левого бедер, вычисленная по 36 скелетам, составила 2.0234 мм, а сумма квадратов отклонений разностей от средней составила 418.687. Можно ли считать, что правое и левое бедро имеют одинаковую длину? ($\alpha = 0.05$).

9. На железнорудном месторождении проводилось опробование двумя методами - пунктирной бороздой и сплошной бороздой. Содержание железа в пробах, взятых первым методом, %: 45, 48, 53, 59, 44, 60, 41, 43, 57; вторым: 51, 50, 42, 44, 39, 40, 48, 38, 59, 55, 51. Можно ли считать, что эти методы дают одинаковое стандартное отклонение? ($\alpha = 0.05$).

10. При исследовании влияния двух типов покрытия на удельную проводимость телевизионных трубок, получены следующие результаты (в усл.ед.).
I тип 6, 5, 12, 9, 10.
II тип 14, 11, 0, 6, 8, 5.
Можно ли на основании этих данных считать, что тип покрытия не влияет на удельную проводимость трубок? ($\alpha = 0.10$).

11. Чтобы определить, какое влияние оказывает температура окружающей среды на систематическую ошибку угломерного инструмента, проведены измерения угла объекта утром ($t = 10^{o}C$) и днем ($t = 26^{o}C$). Измерения дали следующие результаты (в угловых секундах).
Утром: 38.2, 36.4, 37.7, 36.1, 37.9, 37.8.
Днем: 39.5, 38.7, 37.8, 38.6, 39.2, 39.1, 38.9, 39.2.
Можно ли считать, что температура окружающей среды не влияет на систематическую ошибку угломерного инструмента? ($\alpha = 0.05$).

12. Определение концентрации SiO_2 в мартеновском шлаке проводилось в 5 пробах весом и в 6 пробах фотоколориметрическими методами. Были получены следующие данные (в усл.ед.).
Весовой метод: 20.57, 20.45, 20.42, 20.43, 20.94.
Фотоколориметрический метод: 21.15, 21.18, 20.91, 21.11, 21.88, 20.92.
Свидетельствуют ли эти данные о систематическом расхождении между результатами применения первого и второго метода? ($\alpha = 0.05$).

13. Одна из гипотез объясняет массовое вымирание динозавров на границе между меловым и третичным периодами ($\cong$ 65 млн.лет назад) столкновением Земли с кометным ядром или крупным метеоритом, содержащим большое количество иридия. Для проверки этой гипотезы измерялось содержание иридия в нерастворимом остатке карбонатных пород отложений пограничной зоны (65 ± 0.1 млн.лет) и в прилегающих к ней зонах. Измерения, проведенные в различных местах земного шара, дали следующие результаты (в частях на миллиард).
В пограничной зоне: 1.2, 1.4, 2.3, 3.5, 0.9, 0.6, 0.9, 4.0, 6.0, 7.3.
В прилегающих зонах: 1.8, 1.7, 0.8, 2.1, 2.4, 0.7, 1.6, 2.4, 0.9, 2.2, 2.5.
Можно ли на основании этих данных утверждать, что в пограничной зоне имеет место иридиевая аномалия? ($\alpha = 0.05$).

14. Для исследования влияния на потомство употребления родителями алкоголя самцам крыс в течение двух недель добавляли в пищу спирт, затем две недели кормили обычной пищей, после чего их спаривали с самками (питавшимися обычной пищей). Тестирование умственных способностей потомства представляло собой измерение времени, необходимого крысам для нахождения пищи в лабиринте. Измерения времени (в минутах) в подопытной и контрольной (не подвергавшейся влиянию алкоголя) группах, дали следующие результаты.
Контрольная группа: 2.4, 1.8, 1.6, 1.6, 1.7, 1.5, 0.8, 1.1, 2.7, 2.0.
Подопытная группа: 2.2, 2.4, 1.6, 2.25, 4.57, 2.9, 1.03, 2.35, 2.14.
Можно ли на основании этих данных утверждать, что употребление алкоголя родителями влияет на умственные способности потомства? ($\alpha = 0.05$).

15. С целью уменьшения дисперсии отражательной способности краски были внесены изменения в технологию ее приготовления. Затем измерялась отражательная способность (в усл.ед.) у образцов, изготовленных по старой технологии и по новой. Были получены следующие данные.
Старая технология: 45, 40, 195, 65, 145, 137.
Новая технология: 110, 55, 120, 50, 80.
Свидетельствуют ли эти данные об уменьшении дисперсии при изготовлении краски по новой технологии? ($\alpha = 0.05$).

16. Дисперсия предела прочности на разрыв волокна составляет 35.63 $кг^2$. Ожидается, что внесение в технологический процесс определенных изменений снизило указанную дисперсию. В выборке объема 15 были зарегистрированы следующие значения предела прочности (в килограммах): 151, 156, 147, 153, 155, 148, 160, 156, 149, 160, 161, 154, 162, 163, 149. Привело ли изменение технологического процесса к снижению дисперсии? ($\alpha = 0.05$).

17. Для того чтобы проверить, влияет ли на прочность бетона особый способ его приготовления, измерялось предельное значение нагрузки для трех образцов бетона, изготовленного обычным способом (*бетон I*), и трех образцов бетона, изготовленного особым способом (*бетон II*). Измерения дали следующие результаты.

бетон I: 290, 311, 284.
бетон II: 309, 318, 318.
Свидетельствуют ли эти данные о наличии эффекта специальной обработки? ($\alpha = 0.05$).

18. В процессе производства электроэнергии электрические счетчики с вращающимся диском отрегулированы так, что их работа синхронизирована с работой стандартного счетчика. Проверка 10 счетчиков заключалась в определении постоянной счетчика, причем постоянная, характеризующая стандартный счетчик, равна 1.000. Измерения дали следующие результаты: 0.983, 1.002, 0.998, 0.996, 1.002, 0.983, 0.994, 0.991, 1.005, 0.986. Можно ли отклонения от стандарта рассматривать как случайные или, напротив, результаты указывают на систематическое отклонение от стандарта? ($\alpha = 0.05$).

19. При обработке втулок на станке с ЧПУ (числовым программным управлением) было отобрано две пробы по 10 штук деталей в каждой (изготовленных до обеда и после обеда). Измерения диаметров этих втулок (в сантиметрах) дали следующие результаты.

Проба 1: 2.0666, 2.063, 2.068, 2.060, 2.067, 2.063, 2.059, 2,062, 2,062, 2.068.

Проба 2: 2.063 , 2.060, 2.057, 2.056, 2.059, 2.058, 2.062, 2,059, 2,059, 2.057.
Диаметр образца равен 2.06 см. Свидетельствуют ли эти данные о том, что точность обработки в обеих пробах одинакова? ($\alpha = 0.05$).

20. Тринадцать крупнейших землетрясений, произошедших в первой половине *XX* века, имели магнитуды: 7.5, 6.5, 8.0, 8.2, 8.0, 8.2, 7.8, 8.2, 8.0, 6.5, 7.5, 7.9, 7.5. Двадцать крупнейших землетрясений, произошедших во второй половине *XX* века, имели магнитуды: 8.2, 7.5, 8.7, 6.0, 7.8, 5.3, 7.0, 6.5, 6.8, 7.3, 8.1, 7.8, 8.0, 6.0, 8.2, 8.6, 8.4, 8.8, 8.0, 7.3. Свидетельствуют ли эти данные о том, что во второй половине *XX* века землетрясения стали сильнее? ($\alpha = 0.05$).

ДОПОЛНЕНИЯ

1. Стандартные распределения, связанные с нормальным законом

χ^2_n-распределение. Пусть $\xi_1,\ldots,\xi_n$ – независимые стандартные нормальные с.в. (т.е. $\xi_i \sim N(0,1)$). Распределение с.в.

$$\chi^2_n = \sum_{i=1}^{n} \xi_i^2 \tag{1}$$

называется *хи - квадрат распределением с n степенями свободы*.

Распределение χ^2_n – это, очевидно, распределение квадрата длины нормального сферически симметричного случайного вектора $(\xi_1,\ldots,\xi_n)$ с нулевым вектором математических ожиданий и единичной ковариационной матрицей I_n. Функция плотности распределения с.в. χ^2_n имеет вид:

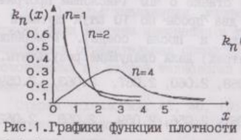


Рис.1.Графики функции плотности $k_n(x)$ при различных значениях n

$$k_n(x) = \begin{cases} \dfrac{1}{2^{n/2}\,\Gamma(n/2)}\, x^{n/2-1}e^{-x/2} & \text{при } x > 0, \\ 0 & \text{при } x \leqslant 0, \end{cases}$$

где $\Gamma(n/2)$ – гамма-функция, определяемая равенством $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$. Из определения следует: $\mathrm{E}\chi^2_n = \sum_{i=1}^{n} \xi_i^2 = n$.

Легко доказать, что $\mathrm{D}\chi^2_n = 2n$. Так как с.в. χ^2_n представляет собой сумму независимых одинаково распределенных с.в. ξ_i^2, имеющих конечный второй момент, для нее справедлива центральная предельная теорема : $\dfrac{\chi^2_n - n}{\sqrt{2n}} \xrightarrow{d} N(0,1)$ при $n \to \infty$. Квантили χ^2_n-распределения для малых значений n находятся по таблицам квантилей распределения хи-квадрат (см. приложение,табл.5). При больших n квантили можно вычислить, пользуясь нормальной аппроксимацией:

$$F_{\chi^2_n}(x) = \mathrm{P}\{\chi^2_n < x\} = \mathrm{P}\left\{\frac{\chi^2_n - n}{\sqrt{2n}} < \frac{x-n}{\sqrt{2n}}\right\} \approx \Phi\left(\frac{x-n}{\sqrt{2n}}\right),$$

где $\Phi(x)$ – стандартная нормальная функция распределения. Из последнего соотношения вытекает, что если нужно определить значение квантили χ^2_α уровня α распределения с.в. χ^2_n, то есть корень уравнения $F_{\chi^2_n}(\chi^2_\alpha) = \mathrm{P}\{\chi^2_n < \chi^2_\alpha\} = \alpha$, то $\Phi\left(\dfrac{\chi^2_\alpha - n}{\sqrt{2n}}\right) \approx \alpha$, и для искомой квантили имеем: $\chi^2_\alpha = n + u_\alpha\sqrt{2n}$, где u_α – квантиль уровня α стандартного нормального закона, т.е. корень уравнения $\Phi(x) = \alpha$.

t_n-распределение Стьюдента. Пусть с.в. ξ имеет стандартное нормальное распределение $N(0,1)$, рассмотрим χ_n – независимую от ξ случайную величину, распределенную как $\sqrt{\chi^2_n}$, где χ^2_n – случайная величина с распределением хи-квадрат с n степенями свободы. Распределение с.в.

$$t_n = \frac{\sqrt{n}\,\xi}{\chi_n} \tag{2}$$

называется *t_n-распределением Стьюдента с n степенями свободы*. Функция плотности t_n – распределения имеет вид

$$s_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)} \left[\frac{x^2}{n} + 1\right]^{-\frac{n+1}{2}},$$

$-\infty < x < \infty$.

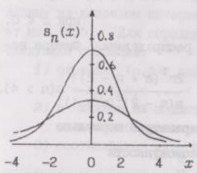


Рис.2.Графики плотности $s_n(x)$ при различных значениях n

Распределение Стьюдента симметрично. $\mathrm{E}(t_n^k) = 0$ для всех нечетных k. В частности, $\mathrm{E}\,t_n = 0$ (если $n > 1$), $\mathrm{D}(t_n) = \sigma^2_{t_n} = \dfrac{n}{n-2}$ (если $n > 2$). При $n=1$ плотность распределения Стьюдента совпадает с плотностью распределения Коши $\dfrac{1}{\pi(x^2+1)}$, у которого не существуют конечные математическое ожидание и дисперсия. При $n \to \infty$ распределение Стьюдента сходится к стандартному нормальному:

$$S_n(x) \xrightarrow[n\to\infty]{} \Phi(x),$$

где $S_n(x)$ – функция распределения (ф.р.) Стьюдента с n степенями свободы. Значения ф.р. $S_n(x)$ при малых n находят по таблицам распределения Стьюдента (см. приложение, табл.6). При больших n (на практике при $n > 30$) применяют приближенное равенство $S(x) \approx \Phi(x)$, где $\Phi(x)$ – стандартная нормальная ф.р. Распределение с.в. t_n инвариантно относительно одновременного умножения с.в. ξ и с.в. ξ_i (на основании которых образована с.в. $\chi_n = \sqrt{\xi_1^2+\ldots+\xi_n^2}$ в знаменателе в соотношении (2)) на одну и ту же величину $\sigma > 0$, поэтому распределение t_n не изменится при переходе от стандартных нормальных с.в. к нормальным с.в. с параметрами 0 и $\sigma > 0$.

$F_{m,n}$-распределение Фишера. Пусть χ^2_m и χ^2_n – две независимые с.в., имеющие хи-квадрат распределение с m и n степенями свободы соответственно. Распределение с.в.

$$F_{m,n} = \frac{n\chi^2_m}{m\chi^2_n} \tag{3}$$

называется *$F_{m,n}$-распределением Фишера с (m,n) степенями свободы*, где m – число степеней свободы числителя, n – число степеней свободы знаменателя. Функция плотности $F_{m,n}$-распределения имеет вид:

$$f_{m,n}(x) = \begin{cases} \dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left[\dfrac{m}{n}\right]^{m/2} x^{\frac{m}{2}-1} \left[1 + \dfrac{mx}{n}\right]^{-\frac{m+n}{2}} & \text{при } x > 0, \\ 0 & \text{при } x \leqslant 0. \end{cases}$$

Математическое ожидание и дисперсия распределения Фишера находятся по формулам:

$$\mathrm{E}(F_{m,n}) = \frac{n}{n-2} \quad (n > 2); \quad \mathrm{D}(F_{m,n}) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad (n > 4).$$

2. Свойства стандартных оценок параметров нормально распределенной генеральной совокупности

Пусть $X = (X_1,\ldots,X_n)$ – выборка из генеральной совокупности с.в. $\xi \sim N(a,\sigma)$ с неизвестными параметрами. Для оценки параметров используются стандартные статистики (см.разд.4):

$$\hat{a} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{– эмпирическое среднее;}$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad \text{– эмпирическая дисперсия;}$$

$$\hat{\sigma}^2 = \frac{n}{n-1}S^2 \quad \text{– несмещенная оценка дисперсии.}$$

В случае нормального распределения генеральной совокупности эти статистики обладают следующими свойствами:

1) с.в. $\overline{X}$ и S^2 взаимно независимы;

2) с.в. $\overline{X} \sim N\left(a, \dfrac{\sigma}{\sqrt{n}}\right)$;

3) с.в. $\dfrac{n\,S^2}{\sigma^2} = \dfrac{(n-1)\hat{\sigma}^2}{\sigma^2} = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2$ имеет χ^2_{n-1}-распределение с $(n-1)$ степенями свободы;

4) с.в. $t_{n-1} = \sqrt{n-1}\,\dfrac{\overline{X} - a}{S} = \dfrac{\sqrt{n}\,(\overline{X} - a)}{\hat{\sigma}}$, где $S = (S^2)^{1/2}$, $\hat{\sigma} = (\hat{\sigma}^2)^{1/2}$, имеет распределение Стьюдента с $(n-1)$ степенями свободы.

3. Интеграл Стилтьеса с монотонной интегрирующей функцией

Пусть $f(x)$ и $F(x)$ – две кусочно непрерывные функции, определенные на конечном интервале $[a,b]$, причем функция $F(x)$ не убывает на $[a,b]$. Для определения интеграла Стилтьеса производятся следующие операции:

1) отрезок $[a,b]$ разбивается на n частей точками $a = x_0 < x_1 < \ldots < x_n = b$;
2) в каждом из подынтервалов $[x_{i-1},x_i]$, $i=1,\ldots,n$, произвольно выбирается точка ξ_i;
3) составляется интегральная сумма:

$$S_n = \sum_{i=1}^{n} f(\xi_i)\left[F(x_i) - F(x_{i-1})\right]. \tag{4}$$

Если интегральная сумма S_n сходится к пределу, когда ранг разбиения $r_n = \max\limits_{1 \leq i \leq n} |x_i - x_{i-1}|$ стремится к нулю, и значение этого предела не зависит от способа разбиения интервала $[a,b]$ и выбора точек ξ_i, то этот предел называется интегралом Стилтьеса и обозначается

$$\int\limits_a^b f(x)dF(x) = \lim\limits_{r_n \to 0} S_n . \qquad (5)$$

Имеются два важных частных случая:

I) функция $F(x)$ дифференцируема: $p(x) = \frac{d}{dx}(F(x))$.
В этом случае интеграл Стилтьеса совпадает с обычным интегралом Римана от функции $f(x)p(x)$ по промежутку $[a,b]$:

$$\int\limits_a^b f(x)\,dF(x) = \int\limits_a^b f(x)p(x)dx ;$$

II) функция $F(x)$ кусочно постоянна и имеет скачки в точках $a_1, \dots, a_m \in [a,b]$. В этом случае интеграл Стилтьеса превращается в конечную сумму:

$$\int\limits_a^b f(x)\,dF(x) = \sum_{j=1}^{m} f(a_j)\,[F(a_j + 0) - F(a_j - 0)].$$

Несобственный интеграл Стилтьеса по бесконечному промежутку $[a,+\infty)$ определяется как предел:

$$\int\limits_a^{+\infty} f(x)\,dF(x) = \lim\limits_{A \to +\infty} \int\limits_a^A f(x)\,dF(x). \qquad (6)$$

Если $F(x)$ – ф.р.некоторой с.в. ξ, то интеграл $\int\limits_{-\infty}^{+\infty} f(x)\,dF(x)$ представляет собой математическое ожидание $E(f(\xi))$ с.в. $f(\xi)$. В частности, если $f(x) = x$, то

$$E(\xi) = \int\limits_{-\infty}^{+\infty} x\,dF(x), \qquad (7)$$

причем в соответствии со свойствами I и II интеграла Стилтьеса формула (7) в случае дискретного или абсолютно непрерывного распределений приводит к обычным формулам вычисления математического ожидания.

4. Асимптотически наиболее мощные критерии для сближающихся альтернатив

Пусть функции $p_0(x)$ и $p_1(x)$ лежат в одномерном параметрическом семействе: $p_0(x) = p(x,\theta_0)$ и $p_1(x) = p(x,\theta_{1,n})$; $\theta_{1,n} = \theta_0 + \delta/\sqrt{n}$, $\delta > 0$ (т.е. рассматриваются сближающиеся со скоростью $\delta/\sqrt{n}$ правосторонние альтернативы). Тогда при достаточно общих предположениях "регулярности" асимптотически наиболее мощный критерий проверки H_0 против H_1 основан на производной логарифма отношения правдоподобия

$$l_n = n^{-1/2} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log p(X_i,\theta_0) .$$

При гипотезе H_0 статистика l_n асимптотически нормальна с нулевым средним и дисперсией $I(\theta_0)$, где $I(\theta_0)$ есть информация Фишера в точке θ_0:

$$I(\theta_0) = E_{\theta_0}\left(\frac{\partial}{\partial\theta} \log p(X_i,\theta_0)\right)^2 ,$$

через E_θ мы обозначаем математическое ожидание, вычисленное для распределения с функцией плотности или вероятностей $p(x,\theta)$. Поэтому при больших n порог критерия $\phi_{n,\alpha}$ можно выбрать в виде $T_n(\alpha) = t_\alpha\sqrt{I(\theta_0)}$, где t_α есть $(1 - \alpha)$ – я квантиль стандартного нормального распределения: $\Phi(t_\alpha) = 1 - \alpha$ и определяется из таблиц функции $\Phi(t)$. При альтернативе H_1 статистика l_n асимптотически нормальна со средним $\delta I(\theta_0)$ и той же дисперсией $I(\theta_0)$. Отсюда вытекают предельные выражения для вероятностей ошибок I, II рода и мощности:

$$\alpha(\phi,\theta_n) \to \alpha ; \qquad \beta(\phi,\theta_{s,n}) \to \Phi(t_\alpha - \rho) ;$$

$$\gamma(\phi,\theta_{1,n}) \to \Phi(\rho - t_\alpha) ; \quad \rho = \delta\sqrt{I(\theta_0)} .$$

Критерий $\phi_{n,\alpha}$ не зависит от величины $\delta > 0$ и поэтому является асимптотически равномерно наиболее мощным для всех сближающихся правосторонних альтернатив. Такой критерий называют *локально наиболее мощным*.

5. Вычисление статистики ω_n^2 по вариационному ряду выборки

Пусть $X_{(1)},\dots,X_{(n)}$ – вариационный ряд выборки. Разбивая область интегрирования в формуле (7.6) на интервалы

$$(-\infty , X_{(1)}),\ (X_{(1)} , X_{(2)}),\dots,(X_{(n)} , +\infty),$$

на которых $F_n(x)$ постоянна, и делая в интегралах замену переменной $y = F_0(x)$, преобразуем сумму интегралов по интервалам к виду (7.7):

$$\omega_n^2 = \int\limits_{-\infty}^{X_{(1)}} (F_0(x) - 0)^2\,dF_0(x) + \sum_{i=1}^{n-1} \int\limits_{X_{(i)}}^{X_{(i+1)}} (F_0(x) - \frac{i}{n})^2\,dF_0(x) +$$

$$+ \int\limits_{X_{(n)}}^{+\infty} (F_0(x) - 1)^2\,dF_0(x) = \frac{\left[F_0(X_{(1)})\right]^3}{3} +$$

$$+ \frac{1}{3}\sum_{i=1}^{n-1}\left[\left[F_0(X_{(i+1)} - \frac{i}{n})\right]^3 - \left[F_0(X_{(i)} - \frac{i}{n})\right]^3\right] + \frac{\left[F_0(X_{(n)}) - 1\right]^3}{3} =$$

$$= \frac{1}{12n^2} + \frac{1}{n}\sum_{i=1}^{n}\left[F_0(X_{(i)}) - \frac{2i - 1}{2n}\right]^2 .$$

СПИСОК ЛИТЕРАТУРЫ

1. *Боровков А.А.* Математическая статистика.- М.: Наука, 1984. - 476 с.

2. *Бородин А.Н.* Элементарный курс теории вероятностей и математической статистики. - СПб.: Изд.-во "Лань", 1998.- 224 с.

3. *Бочаров П.П.*, Печинкин А.В. Теория вероятностей. Математическая статистика.- М: Гардарика, 1998. - 328 с.

4. *Дьяконов В.П.* Справочник по применению системы PC MatLAB.- М.: Физматлит, 1993.- 112 с.

5. *Дьяконов В.П.* Руководство по применению системы MathCAD. - Смоленск: ГНЦ "КИТ", 1992.- 114 с.

6. *Ивченко Г.И.*, Медведев Ю.И. Математическая статистика. - М.: Высшая школа, 1992.- 379 с.

7. *Ингстер Ю.И.* Асимптотические методы статистики: Учебное пособие.- СПб: ПГУПС, 1999 . - 130 с.

8. *Колемаев В.А.* и др. Теория вероятностей и математическая статистика / В.А.Колемаев, О.В.Староверов, В.Б.Турундаевский. - М.: Высшая школа, 1991. - 400 с.

9. *Леман Э.* Проверка статистических гипотез/ Пер. с англ.- М.: Наука, 1979. - 408 с.

10. *Очков В.Ф.* Mathcad PLUS 6.0 для студентов и инженеров . - М.: ТОО фирма "Компьютер Пресс", 1996. - 238 с.

11. *Потемкин В.Г.* Система MATLAB 5 для студентов.- М.: ДИАЛОГ-МИФИ, 1998.- 314 с.

12. *Потемкин В.Г.* Система инженерных и научных расчетов MATLAB 5.x. - М: ДИАЛОГ-МИФИ, 1999. - Т.1.- 367 с.- Т.2.- 304 с.

13. *Тюрин Ю.И.*, Макаров А.А. Статистический анализ данных на компьютере. - М: Инфра-М, 1998. - 114 с.

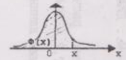
14. *Хампель Ф.Р.* и др. Робастность в статистике. Подход на основе функций влияния/ Пер. с англ. - М.:, 1989. - 685 с.

15. *Хьюбер Дж.П.* Робастность в статистике/ Пер. с англ. - М.: Мир, 1984. - 304 с.

ПРИЛОЖЕНИЕ

Таблица 1

Функция распределения Φ(x)



нормального закона N (0, 1); Φ(-x) =1- Φ(x), $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

Продолжение табл.1

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Квантили U_p нормального распределения N (0,1)

p	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
U_p	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Таблица 2

Значения функции плотности нормального распределения N(0,1)

$f(x)=\frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.3989	0.3989	0.3989	0.3988	0.3986	0.3984	0.3982	0.3980	0.3977	0.3973
0.1	0.3970	0.3965	0.3961	0.3956	0.3951	0.3945	0.3939	0.3932	0.3925	0.3918
0.2	0.3910	0.3902	0.3894	0.3885	0.3876	0.3867	0.3857	0.3847	0.3836	0.3825
0.3	0.3814	0.3802	0.3790	0.3778	0.3765	0.3752	0.3739	0.3725	0.3712	0.3697
0.4	0.3683	0.3668	0.3653	0.3637	0.3621	0.3605	0.3589	0.3572	0.3555	0.3538
0.5	0.3521	0.3503	0.3485	0.3467	0.3448	0.3429	0.3410	0.3391	0.3372	0.3352
0.6	0.3332	0.3312	0.3292	0.3271	0.3251	0.3230	0.3209	0.3187	0.3166	0.3144
0.7	0.3123	0.3101	0.3079	0.3056	0.3034	0.3011	0.2989	0.2966	0.2943	0.2920
0.8	0.2897	0.2874	0.2850	0.2827	0.2803	0.2780	0.2756	0.2732	0.2709	0.2685
0.9	0.2661	0.2637	0.2613	0.2589	0.2565	0.2541	0.2516	0.2492	0.2468	0.2444
1.0	0.2420	0.2396	0.2371	0.2347	0.2323	0.2299	0.2275	0.2251	0.2227	0.2203
1.1	0.2179	0.2155	0.2131	0.2107	0.2083	0.2059	0.2036	0.2012	0.1989	0.1965
1.2	0.1942	0.1919	0.1895	0.1872	0.1849	0.1826	0.1804	0.1781	0.1758	0.1736
1.3	0.1714	0.1691	0.1669	0.1647	0.1626	0.1604	0.1582	0.1561	0.1539	0.1518
1.4	0.1497	0.1476	0.1456	0.1435	0.1415	0.1394	0.1374	0.1354	0.1334	0.1315
1.5	0.1295	0.1276	0.1257	0.1238	0.1219	0.1200	0.1182	0.1163	0.1145	0.1127
1.6	0.1109	0.1092	0.1074	0.1057	0.1040	0.1023	0.1006	0.0989	0.0973	0.0957
1.7	0.0940	0.0925	0.0909	0.0893	0.0878	0.0863	0.0848	0.0833	0.0818	0.0804
1.8	0.0790	0.0775	0.0761	0.0748	0.0734	0.0721	0.0707	0.0694	0.0681	0.0669
1.9	0.0656	0.0644	0.0632	0.0620	0.0608	0.0596	0.0584	0.0573	0.0562	0.0551
x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09

Продолжение табл.2

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.0	0.0540	0.0529	0.0519	0.0508	0.0498	0.0488	0.0478	0.0468	0.0459	0.0449
2.1	0.0440	0.0431	0.0422	0.0413	0.0404	0.0396	0.0387	0.0379	0.0371	0.0363
2.2	0.0355	0.0347	0.0339	0.0332	0.0325	0.0317	0.0310	0.0303	0.0297	0.0290
2.3	0.0283	0.0277	0.0270	0.0264	0.0258	0.0252	0.0246	0.0241	0.0235	0.0229
2.4	0.0224	0.0219	0.0213	0.0208	0.0203	0.0198	0.0194	0.0189	0.0184	0.0180
2.5	0.0175	0.0171	0.0167	0.0163	0.0158	0.0154	0.0151	0.0147	0.0143	0.0139
2.6	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110	0.0107
2.7	0.0104	0.0101	0.0099	0.0096	0.0093	0.0091	0.0088	0.0086	0.0084	0.0081
2.8	0.0079	0.0077	0.0075	0.0073	0.0071	0.0069	0.0067	0.0065	0.0063	0.0061
2.9	0.0060	0.0058	0.0056	0.0055	0.0053	0.0051	0.0050	0.0048	0.0047	0.0046
3.0	0.0044	0.0043	0.0042	0.0040	0.0039	0.0038	0.0037	0.0036	0.0035	0.0034
3.1	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026	0.0025	0.0025
3.2	0.0024	0.0023	0.0022	0.0022	0.0021	0.0020	0.0020	0.0019	0.0018	0.0018
3.3	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014	0.0013	0.0013
3.4	0.0012	0.0012	0.0012	0.0011	0.0011	0.0010	0.0010	0.0010	0.0009	0.0009
3.5	0.0009	0.0008	0.0009	0.0008	0.0008	0.0007	0.0007	0.0007	0.0007	0.0006
3.6	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0004
3.7	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003	0.0003	0.0003	0.0003
3.8	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002	0.0002	0.0002	0.0002	0.0002
3.9	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001
4.0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09

Таблица 3

Распределение Пуассона $P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$

k \ λ	0.1	0.2	0.3	0.4	0.5
0	0.90484	0.81873	0.74082	0.67032	0.60653
1	0.09048	0.16375	0.22223	0.26813	0.30327
2	0.00452	0.01638	0.03334	0.05363	0.07582
3	0.00015	0.00109	0.00333	0.00715	0.01204
4		0.00006	0.00025	0.00072	0.00158
5			0.00002	0.00006	0.00016
					0.00001

k \ λ	0.6	0.7	0.8	0.9
0	0.54881	0.49659	0.44933	0.40657
1	0.32929	0.34761	0.35946	0.36591
2	0.09879	0.12166	0.14379	0.16466
3	0.01976	0.02839	0.03834	0.04940
4	0.00296	0.00497	0.00767	0.01112
5	0.00036	0.00070	0.00123	0.00200
6	0.00004	0.00008	0.00016	0.00030
7		0.00001	0.00002	0.00004

k \ λ	1.0	2.0	3.0	4.0	5.0
0	0.36788	0.13534	0.04979	0.01832	0.00674
1	0.36788	0.27067	0.14936	0.07326	0.03369
2	0.18394	0.27067	0.22404	0.14653	0.08422
3	0.06131	0.18045	0.22404	0.19537	0.14037
4	0.01533	0.09022	0.16803	0.19537	0.17547
5	0.00307	0.03609	0.10082	0.15629	0.17547
6	0.00051	0.01203	0.05041	0.10419	0.14622
7	0.00007	0.00344	0.02160	0.05954	0.10445
8	0.00001	0.00086	0.00810	0.02977	0.06528
9		0.00019	0.00270	0.01323	0.03627
10		0.00004	0.00081	0.00529	0.01813
11		0.00001	0.00022	0.00193	0.00824
12			0.00006	0.00064	0.00343
13			0.00001	0.00020	0.00132
14				0.00006	0.00047
15				0.00002	0.00016
16					0.00005
17					0.00001

Таблица 4

Суммарные вероятности для распределения Пуассона

$$P\{X \geqslant x\} = \sum_{k=x}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda}$$

x \ λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	0.09516	0.18127	0.25918	0.32968	0.39347	0.45119	0.50341
2	00468	01752	03694	06155	09020	12190	15580
3	00016	00115	00360	00793	01439	02312	03414
4	0	00006	00027	00078	00175	00336	00575
5		0	00002	00006	00017	00039	00079
6					00001	00004	00009
7							00001

x \ λ	0.8	0.9	1.0	2.0	4.0	6.0	8.0
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
1	0.55067	0.59343	0.63212	0.86466	0.98168	0.99752	0.99968
2	19121	22752	26424	59399	90842	98265	99698
3	04742	06286	08030	32332	76190	84880	98625
4	00908	01346	01899	14288	56653	71494	95762
5	00141	00234	00366	05265	37116	55432	90037
6	00018	00034	00059	01656	21487	39379	80876
7	00002	00004	00008	00453	11067	25602	68663
8		00001	00002	00110	05113	15276	54704
9				00024	02136	08392	40745
10				00005	00823	04262	28338
11				00001	00284	02009	18411
12					00092	00883	11192
13					00027	00363	06380
14					00008	00140	03418
15					00002	00051	01726
16					00001	00017	00823
17						00006	00372
18						00002	00159
19						00001	00065
20							00025
21							00009
22							00003
23							00001

Квантили хи-квадрат

k \ p	0.005	0.010	0.025	0.05	0.10	0.20	0.30
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	0.0642	0.148
2	0.0100	0.0201	0.0506	0.0103	0.211	0.446	0.713
3	0.0717	0.115	0.216	0.352	0.584	1.00	1.42
4	0.207	0.297	0.484	0.711	1.06	1.65	2.19
5	0.412	0.554	0.831	1.152	1.61	2.34	3.00
6	0.676	0.872	1.24	1.64	2.20	3.07	3.83
7	0.989	1.24	1.69	2.17	2.83	3.82	4.67
8	1.34	1.65	2.18	2.73	3.49	4.59	5.53
9	1.73	2.09	2.70	3.33	4.17	5.38	6.39
10	2.16	2.56	3.25	3.94	4.87	6.18	7.27
11	2.60	3.05	3.82	4.57	5.58	6.99	8.15
12	3.07	3.57	4.40	5.23	6.30	7.81	9.03
13	3.57	4.11	5.01	5.89	7.04	8.63	9.93
14	4.07	4.66	5.63	6.57	7.79	9.47	10.8
15	4.60	5.23	6.26	7.26	8.55	10.3	11.7
16	5.14	5.81	6.91	7.96	9.31	11.2	12.6
17	5.70	6.41	7.56	8.67	10.1	12.0	13.5
18	6.26	7.01	8.23	9.39	10.9	12.9	14.4
19	6.84	7.63	8.91	10.1	11.7	13.7	15.4
20	7.43	8.26	9.59	10.9	12.4	14.6	16.3
21	8.03	8.90	10.3	11.6	13.2	15.4	17.2
22	8.64	9.54	11.0	12.3	14.0	16.3	18.1
23	9.26	10.2	11.7	13.1	14.8	17.2	19.0
24	9.89	10.9	12.4	13.8	15.7	18.1	19.9
25	10.5	11.5	13.1	14.6	16.5	18.9	20.9
26	11.2	12.2	13.8	15.4	17.3	19.8	21.8
27	11.8	12.9	14.6	16.2	18.1	20.7	22.7
28	12.5	13.6	15.3	16.9	18.9	21.6	23.6
29	13.1	14.3	16.0	17.7	19.8	22.5	24.6
30	13.8	15.0	16.8	18.5	20.6	23.4	25.5
35	17.2	18.5	20.6	22.5	24.8	27.8	30.2
40	20.7	22.2	24.4	26.5	29.1	32.3	34.9
45	24.3	25.9	28.4	30.6	33.4	36.9	39.6
50	28.0	29.7	32.4	34.8	37.7	41.4	44.3
75	47.2	49.5	52.9	56.1	59.8	64.5	68.1
100	67.3	70.1	74.2	77.9	82.4	87.9	92.1

распределения $\chi_p^2(k)$ Таблица 5

0.70	0.80	0.90	0.95	0.975	0.990	0.995	0.999	p / k
1.07	1.64	2.71	3.84	5.02	6.63	7.88	10.8	1
2.41	3.22	4.61	5.99	7.38	9.21	10.6	13.8	2
3.67	4.64	6.25	7.81	9.35	11.3	12.8	16.3	3
4.88	5.99	7.78	9.49	11.1	13.3	14.9	18.5	4
6.06	7.29	9.24	11.1	12.8	15.1	16.7	20.5	5
7.23	8.56	10.6	12.6	14.4	16.8	18.5	22.5	6
8.38	9.80	12.0	14.1	16.0	18.5	20.3	24.3	7
9.52	11.0	13.4	15.5	17.5	20.1	22.0	26.1	8
10.7	12.2	14.7	16.9	19.0	21.7	23.6	27.9	9
11.8	13.4	16.0	18.3	20.5	23.2	25.2	29.6	10
12.9	14.6	17.3	19.7	21.9	24.7	26.8	31.3	11
14.0	15.8	18.5	21.0	23.3	26.2	28.3	32.9	12
15.1	17.0	19.8	22.4	24.7	27.7	29.8	34.5	13
16.2	18.2	21.1	23.7	26.1	29.1	31.3	36.1	14
17.3	19.3	22.3	25.0	27.5	30.6	32.8	37.7	15
18.4	20.5	23.5	26.3	28.8	32.0	34.3	39.3	16
19.5	21.6	24.8	27.6	30.2	33.4	35.7	40.8	17
20.6	22.8	26.0	28.9	31.5	34.8	37.2	42.3	18
21.7	23.9	27.2	30.1	32.9	36.2	38.6	43.8	19
22.8	25.0	28.4	31.4	34.2	37.6	40.0	45.3	20
23.9	26.9	29.6	32.7	35.5	38.9	41.4	46.8	21
24.9	27.3	30.8	33.9	36.8	40.3	42.8	48.3	22
26.0	28.4	32.0	35.2	38.1	41.6	44.2	49.7	23
27.1	29.6	33.2	36.4	39.4	43.0	45.6	51.2	24
28.2	30.7	34.4	37.7	40.6	44.3	46.9	52.6	25
29.2	31.8	35.6	38.9	41.9	45.6	48.3	54.1	26
30.3	32.9	36.7	40.1	43.2	47.0	49.6	55.5	27
31.4	34.0	37.9	41.3	44.5	48.3	51.0	56.9	28
32.5	35.1	39.1	42.6	45.7	49.6	52.3	58.3	29
33.5	36.3	40.3	43.8	47.0	50.9	53.7	59.7	30
38.9	41.8	46.1	49.8	53.2	57.3	60.3	66.6	35
44.2	47.3	51.8	55.8	59.3	63.7	66.8	73.4	40
49.5	52.7	57.5	61.7	65.4	70.0	73.2	80.1	45
54.7	58.2	63.2	67.5	71.4	76.2	79.5	86.7	50
80.9	85.1	91.1	96.2	100.8	106.4	110.3	118.6	75
106.9	111.7	118.5	124.3	129.6	135.6	140.2	149.4	100

Таблица 6

Квантили распределения Стьюдента $t_p(k)$

$k \backslash p$	0.750	0.900	0.950	0.975	0.990	0.995	0.999
1	1.000	3.078	6.314	12.706	31.821	63.657	318
2	0.816	1.886	2.920	4.303	6.965	9.925	22.3
3	0.765	1.638	2.353	3.182	4.541	5.841	10.2
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090

Таблица 7

Вероятности $P(z)=2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2z^2}$ распределения Колмогорова

z	0	1	2	3	4	5	6	7	8	9
0.3				9999	9998	9997	9995	9992	9987	9981
0.4	0.9972	9960	9945	9926	9903	9874	9840	9800	9753	9700
0.5	9639	9572	9497	9415	9325	9228	9124	9013	8896	8772
0.6	8643	8508	8368	8222	8073	7920	7764	7604	7442	7278
0.7	7112	6945	6777	6609	6440	6272	6104	5936	5770	5605
0.8	5441	5280	5120	4962	4806	4653	4503	4355	4209	4067
0.9	3927	3791	3657	3527	3399	3275	3154	3036	2921	2809
1.0	2700	2594	2492	2392	2296	2202	2111	2024	1939	1857
1.1	1777	1700	1626	1555	1486	1420	1356	1294	1235	1177
1.2	1122	1070	1019	0970	0924	0879	0836	0794	0755	0717
1.3	0681	0646	0613	0582	0551	0522	0495	0469	0444	0420
1.4	0397	0375	0354	0335	0316	0298	0282	0266	0250	0236
1.5	0222	0209	0197	0185	0174	0164	0154	0145	0136	0127
1.6	0120	0112	0105	0098	0092	0086	0081	0076	0071	0066
1.7	0062	0058	0054	0050	0047	0044	0041	0038	0035	0033
1.8	0031	0029	0027	0025	0023	0021	0020	0019	0017	0016
1.9	0015	0014	0013	0012	0011	0010	0009	0009	0008	0007
2.0	0007	0006	0006	0005	0005	0004	0004	0004	0003	0003
2.1	0003	0003	0002	0002	0002	0002	0002	0002	0001	0001
2.2	0001	0001	0001	0001	0001	0001	0001	0001	0001	0001
2.3	0001	0000								

Квантили распределения Фишера $F_p(k_1,k_2)$

Таблица 8

$p = 0.9$

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	k_2
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	1
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	2
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	3
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	4
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	5
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	6
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	7
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	8
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	9
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	10
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	11
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	12
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	13
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	14
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	15
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	16
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	17
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	18
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	19
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	20
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	21
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	22
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	23
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	24
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	25
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	26
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	27
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	28
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	29
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	30
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	40
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	60
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	120
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	∞

k_2 \ k_1	1	2	3	4	5	6	7	8	9
								$p = 0.95$	
1	161.8	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Продолжение табл. 8

10	12	15	20	24	30	40	60	120	k_1 / k_2
241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	1
19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	2
8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	3
5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	4
4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	5
4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	6
3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	7
3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	8
3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	9
2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	10
2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	11
2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	12
2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	13
2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	14
2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	15
2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	16
2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	17
2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	18
2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	19
2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	20
2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	21
2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	22
2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	23
2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	24
2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	25
2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	26
2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	27
2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	28
2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	29
2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	30
2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	40
1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	60
1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	120
1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	∞

k_2 \ k_1	1	2	3	4	5	6	7	8	9
								$p = 0.975$	
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	1.19	2.11

Продолжение табл. 8

10	12	15	20	24	30	40	60	120	k_1 / k_2
968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1
39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	2
14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	3
8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	4
6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	5
5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	6
4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	7
4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	8
3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	9
3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	10
3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	11
3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	12
3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	13
3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	14
3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	15
2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	16
2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	17
2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	18
2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	19
2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	20
2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	21
2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	22
2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	23
2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	24
2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	25
2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	26
2.57	2.47	2.36	2.28	2.22	2.16	2.09	2.03	1.95	27
2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	28
2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	29
2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	30
2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	40
2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	60
2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	120
2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	∞

k_2 \ k_1	1	2	3	4	5	6	7	8	9
								$p = 0.99$	
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.68
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

10	12	15	20	24	30	40	60	120	k_1 / k_2
6056	6106	6157	6209	6235	6261	6287	6313	6339	1
99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	2
27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	3
14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	4
10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	5
7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6
6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	7
5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	8
5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	9
4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	10
4.54	4.40	4.25	4.10	4.02	3.94	3.86	4.78	3.69	11
4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	12
4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	13
3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	14
3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	15
3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	16
3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	17
3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	18
3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	19
3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	20
3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	21
3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	22
3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	23
3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	24
3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	25
3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	26
3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	27
3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	28
3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	29
2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	30
2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	40
2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	60
2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	120
2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	∞

k_2 \ k_1	1	2	3	4	5	6	7	8	9
								$p = 0.995$	
1	16211	20000	21615	22500	23056	23437	23715	23925	24091
2	198.5	199.00	199.2	199.2	199.30	199.3	199.4	199.4	199.4
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62

10	12	15	20	24	30	40	60	120	k_1 / k_2
24224	24426	24630	24836	24940	25044	25148	25253	25359	1
199.40	199.4	199.4	199.4	199.5	199.5	199.5	199.5	199.5	2
43.69	43.39	43.08	42.78	42.62	42.47	42.31	42.15	41.99	3
20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	4
13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	5
10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	6
8.38	8.18	7.97	7.75	7.65	7.53	7.42	7.31	7.19	7
7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	8
6.42	6.23	6.03	5.83	5.73	5.62	5.52	5.41	5.30	9
5.85	5.66	5.47	5.27	5.17	5.07	4.97	4.86	4.75	10
5.42	5.24	5.05	4.86	4.76	4.65	4.55	4.44	4.34	11
5.09	4.91	4.72	4.53	4.43	4.33	4.23	4.12	4.01	12
4.82	4.64	4.46	4.27	4.17	4.07	3.97	3.87	3.76	13
4.60	4.43	4.25	4.06	3.96	3.86	3.76	3.66	3.55	14
4.42	4.25	4.07	3.88	3.79	3.69	3.58	3.48	3.37	15
4.27	4.10	3.92	3.73	3.64	3.54	3.44	3.33	3.22	16
4.14	3.97	3.79	3.61	3.51	3.41	3.31	3.21	3.10	17
4.03	3.86	3.68	3.50	3.40	3.30	3.20	3.10	2.99	18
3.93	3.76	3.59	3.40	3.31	3.21	3.11	3.00	2.89	19
3.85	3.68	3.50	3.32	3.22	3.12	3.02	2.92	2.81	20
3.77	3.60	3.43	3.24	3.15	3.05	2.95	2.84	2.73	21
3.70	3.54	3.36	3.18	3.08	2.98	2.88	2.77	2.66	22
3.64	3.47	3.30	3.12	3.02	2.92	2.82	2.71	2.60	23
3.59	3.42	3.25	3.06	2.97	2.87	2.77	2.66	2.55	24
3.54	3.37	3.20	3.01	2.92	2.82	2.72	2.61	2.50	25
3.49	3.33	3.15	2.97	2.87	2.77	2.67	2.56	2.45	26
3.45	3.28	3.11	2.93	2.83	2.73	2.63	2.52	2.41	27
3.41	3.25	3.07	2.89	2.79	2.69	2.59	2.48	2.37	28
3.38	3.21	3.04	2.86	2.76	2.66	2.56	2.45	2.33	29
3.34	3.18	3.01	2.82	2.73	2.63	2.52	2.42	2.30	30
3.12	2.95	2.78	2.60	2.50	2.40	2.30	2.18	2.06	40
2.90	2.74	2.57	2.39	2.29	2.19	2.08	1.96	1.83	60
2.71	2.54	2.37	2.19	2.09	1.98	1.87	1.75	1.61	120
2.52	2.36	2.19	2.00	1.90	1.79	1.67	1.53	1.36	∞

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9
								$p = 0.999$	
1	4053+	5000+	5404+	5625+	5764+	5859+	5929+	5981+	6023+
2	998.5	999.0	999.2	999.2	999.3	999.3	999.4	999.4	999.4
3	167.0	148.5	141.1	137.1	134.6	132.8	131.6	130.6	129.9
4	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
5	47.18	37.12	33.20	31.09	29.75	28.84	28.16	27.64	27.24
6	35.51	27.00	23.70	21.92	20.81	20.03	19.46	19.03	18.69
7	29.25	21.69	18.77	17.19	16.21	15.52	15.02	14.63	14.33
8	25.42	18.49	15.83	14.39	13.49	12.86	12.40	12.04	11.77
9	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11
10	21.04	14.91	12.55	11.28	10.48	9.92	9.52	9.20	8.96
11	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12
12	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48
13	17.81	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98
14	17.14	11.78	9.73	8.62	7.92	7.43	7.08	6.80	6.58
15	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26
16	16.12	10.97	9.00	7.94	7.27	6.81	6.46	6.19	5.98
17	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75
18	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56
19	15.08	10.16	8.28	7.26	6.62	6.18	5.85	5.59	5.39
20	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24
21	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11
22	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99
23	14.19	9.47	7.67	6.69	6.08	5.65	5.33	5.09	4.89
24	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80
25	13.88	9.22	7.45	6.49	5.88	5.46	5.15	4.91	4.71
26	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64
27	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57
28	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50
29	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45
30	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39
40	12.61	8.25	6.60	5.70	5.13	4.73	4.44	4.21	4.02
60	11.97	7.76	6.17	5.31	4.76	4.37	4.09	3.87	3.69
120	11.38	7.32	5.79	4.95	4.42	4.04	3.77	3.55	3.38
∞	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10

4053+ – означает $4053 \cdot 10^2$

Продолжение табл. 8

10	12	15	20	24	30	40	60	120	k_1 / k_2
6056+	6107+	6158+	6209+	6235+	6261+	6287+	6313+	6340+	1
999.4	999.4	999.4	999.4	999.5	999.5	999.5	999.5	999.5	2
129.2	128.3	127.4	126.4	125.9	125.4	125.0	124.5	124.0	3
48.05	47.41	46.76	46.10	45.77	45.43	45.09	44.75	44.40	4
26.92	26.42	25.91	25.39	25.14	24.87	24.60	24.33	24.06	5
18.41	17.99	17.56	17.12	16.89	16.67	16.44	16.21	15.99	6
14.08	13.71	13.32	12.98	12.73	12.53	12.33	12.12	11.91	7
11.54	11.19	10.84	10.48	10.30	10.11	9.92	9.73	9.53	8
9.89	9.57	9.24	8.90	8.72	8.55	8.37	8.19	8.00	9
8.75	8.45	8.13	7.80	7.64	7.47	7.30	7.12	6.94	10
7.92	7.63	7.32	7.01	6.85	6.68	6.52	6.35	6.17	11
7.29	7.00	6.71	6.40	6.25	6.09	5.93	5.76	5.59	12
6.80	6.52	6.23	5.93	5.78	5.63	5.47	5.30	5.14	13
6.40	6.13	5.85	5.56	5.41	5.25	5.10	4.94	4.77	14
6.08	5.81	5.54	5.25	5.10	4.95	4.80	4.64	4.47	15
5.81	5.55	5.27	4.99	4.85	4.70	4.54	4.39	4.23	16
5.58	5.32	5.05	4.78	4.63	4.48	4.33	4.18	4.02	17
5.39	5.13	4.87	4.59	4.45	4.30	4.15	4.00	3.84	18
5.22	4.97	4.70	4.43	4.29	4.14	3.99	3.84	3.68	19
5.08	4.82	4.56	4.29	4.15	4.00	3.86	3.70	3.54	20
4.95	4.70	4.44	4.17	4.03	3.88	3.74	3.58	3.42	21
4.83	4.58	4.33	4.06	3.92	3.78	3.63	3.48	3.32	22
4.73	4.48	4.23	3.96	3.82	3.68	3.53	3.38	3.22	23
4.64	4.39	4.14	3.87	3.74	3.59	3.45	3.29	3.14	24
4.56	4.31	4.06	3.79	3.66	3.52	3.37	3.22	3.06	25
4.48	4.24	3.99	3.72	3.59	3.44	3.30	3.15	2.99	26
4.41	4.17	3.92	3.66	3.52	3.38	3.23	3.08	2.92	27
4.35	4.11	3.86	3.60	3.46	3.32	3.18	3.02	2.86	28
4.29	4.05	3.80	3.54	3.41	3.27	3.12	2.97	2.81	29
4.24	4.00	3.75	3.49	3.36	3.22	3.07	2.92	2.76	30
3.87	3.64	3.40	3.15	3.01	2.87	2.73	2.57	2.41	40
3.54	3.31	3.08	2.83	2.69	2.55	2.41	2.25	2.08	60
3.24	3.02	2.78	2.53	2.40	2.26	2.11	1.95	1.76	120
2.96	2.74	2.51	2.27	2.13	1.99	1.84	1.66	1.45	∞

СОДЕРЖАНИЕ

МОДЕЛИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Методические указания

Составители:

доц. БАУШЕВ Алексей Николаевич,

доц. ГРИБКОВА Надежда Викторовна,

проф. ИНГСТЕР Юрий Измайлович

Редактор *Н.В. Фролова*

Корректор *Н.Г. Берзина*

План 1997 г., № 147

Подписано в печать с оригинал-макета 09.10.00.
Формат 60x84 1/16. Бумага для множ. апп. Печать офсетная.
Усл. печ. л. 5,5. Уч.-изд.л. 5,5. Тираж 500.
Заказ 1138. Цена 27 р.50 к.
Петербургский государственный университет путей сообщения.
190031, СПб., Московский пр., 9.

Типография ПГУПС. 190031, СПб., Московский пр., 9.