ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Основные задачи математической статистики:

1. Разработка методологии сбора и группировки статистического материала, полученного в результате наблюдений за случайными процессами.

2. Разработка методов анализа полученных статистических данных. Этот анализ включает оценку вероятностей события, функции распределения вероятностей или плотности вероятности, оценку параметров известного распределения, а также оценку связей между случайными величинами.

1. Определение на основе данных опыта (неизвестного) закона распределения случайной величины.
2. Определение по данным опыта неизвестных параметров распределения.
3. Проверка статистических гипотез.

ВЫБОРОЧНЫЙ МЕТОД

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ И ВЫБОРОЧНАЯ

Статистические данные представляют собой данные, полученные в результате обследования большого числа объектов или явлений.

Экспериментальные данные - это результаты измерения некоторых признаков объектов, выбранных из большой совокупности объектов.

Часть объектов исследования, определенным образом выбранная из более обширной совокупности, называется *выборкой*, а *вся* исходная совокупность, из которой взята выборка,- *генеральной (основной) совокупностью*.

Исследования, в которых участвуют все без исключения объекты, составляющие генеральную совокупность, называются *сплошными исследованиями*. Может использоваться *выборочный метод*, суть которого в том, что для обследования привлекается часть генеральной совокупности (*выборка*), но по результатам этого обследования судят о свойствах всей генеральной совокупности.

Предметом изучения в статистике являются варьирующиеся признаки (называемые *статистическими*). Они делятся на качественные и количественные.

Качественными признаками объект обладает либо не обладает. Они не поддаются непосредственному измерению (спортивная специализация, квалификация, национальность, территориальная принадлежность и т. п.).

Количественные признаки представляют собой результаты подсчета или измерения. В соответствии с этим они делятся на *дискретные и непрерывные*.

Например, измеряемая температура воздуха в некотором пункте – непрерывная случайная величина (может меняться на сколь угодно малую величину), и соответствующая генеральная совокупность представляет собой бесконечное множество значений.

Повторной называют выборку, при которой объект перед отбором следующего возвращается в генеральную совокупность. *Бесповторной* называют выборку, при которой отобранный объект в генеральную совокупность не возвращается.

Если выборка правильно отражает соотношения в генеральной совокупности, то ее называют *репрезентативной* (представительной). Например, результаты социологического опроса населения будут зависеть от того, в каком месте он проводится, среди каких групп.

ВАРИАЦИОННЫЙ РЯД. ПОЛИГОН ЧАСТОТ И ГИСТОГРАММА ЭМПИРИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

Пусть X — некоторый признак изучаемого объекта или явления (срок службы электролампы, вес студента, диаметр шарика для подшипника и т.п.). Генеральной совокупностью является множество всех возможных значений этого признака, а результаты n наблюдений над признаком X дадут нам выборку объема n — первоначальные статистические данные, значения x_1, x_2, ..., x_n (простая выборка, не сгруппированные данные)

При этом значение x_1 получено при первом наблюдении случайной величины X, x_2 – при втором наблюдении той же случайной величины и т.д.

Выборку преобразуют в *вариационный ряд*, располагая результаты наблюдений в порядке возрастания: $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$. Каждый член $x_{(i)}$ вариационного ряда называется *вариантой*.

Пример 1.

1. Измерена масса тела 10-ти детей 6-ти лет. Полученные данные образуют простой статистический ряд: 24 22 23 28 24 23 25 27 25 25.

2. Из 10000 выпущенных на конвейере электрических лампочек отобрано 300 штук для проверки качества всей партии. Здесь $N = 10000$, а $n = 300$.

Отдельные значения статистического ряда называются *вариантами*. Если варианта x_i появилась m раз, то число m называют *частотой*, а ее отношение к объему выборки m/n – *относительной частотой*.

Последовательность вариант, записанная в возрастающем (убывающем) порядке, называется *ранжированным рядом*.

Пример 2. Для ранжированного ряда: 23 23 24 24 25 25 25 27 28 в нижеприведенной таблице в первой строке записаны все значения величины (варианты), во второй – соответствующие им частоты (безынтервальный вариационный ряд), в третьей – накопленные частоты, в четвертой – относительные частоты (табл.4.1).

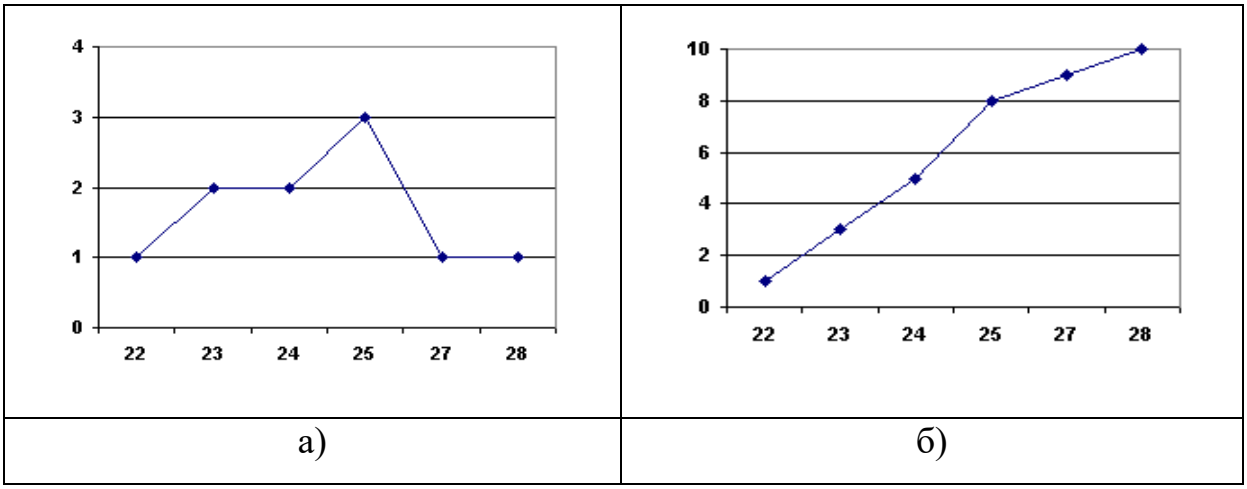
Таблица 4.1. Значения вариант и их частот

X	22	23	24	25	27	28
n_i	1	2	2	3	1	1
$n_н$	1	3	5	8	9	10
$\frac{n_i}{n}$	0.1	0.2	0.2	0.3	0.1	0.1

Полигоном частот называют ломаную линию, отрезки которой соединяют точки с координатами (x_i; n_i) (рис. 4.1).

Отметим, что сумма частот статистического ряда равна объему выборки. Часто статистический ряд составляют, используя относительные частоты вариант: $h_i = \frac{n_i}{n}$, $i = 1, 2, ..., m$ (m — количество различных вариант). Сумма относительных частот равна единице.

Полигоном относительных частот называют ломаную линию, отрезки которой соединяют точки с координатами (x_i; h_i).



а) б)

Рисунок 4.1. Полигон частот а), кумулятивная кривая б)

Эмпирическим аналогом графика интегральной функции распределения является *кумулятивная кривая* (*кумулята*). Для ее построения на оси *OX* откладывают значения вариант, на оси *OY* – накопленные частоты или относительные частоты. Полученная плавная кривая называется кумулятой.

В том случае, если выборка представлена большим количеством различных значений непрерывной случайной величины, то группировку данных проводят в виде интервального вариационного ряда (ИВР). Для этого диапазон варьирования признака разбивают на несколько (5–10) равных

интервалов и указывают количество вариант, попавших в каждый интервал.

Алгоритм построения интервального вариационного ряда.

1. Исходя из объема выборки (n), определить количество интервалов (k) (см. табл. 4.2).

Таблица.Рекомендуемое соотношение объем выборки–число интервалов

n	25–40	40–60	60–100	100–200	>200
k	5–6	6–8	7–10	8–12	10–15

2. Вычислить размах ряда: $R=X_{max} - X_{min}$

3. Определить ширину интервала: $h=R/(k–1)$

4. Найти начало первого интервала $X_0 = X_{min} – h/2$

5. Составить интервальный вариационный ряд.

Графическим изображением ИВР является *гистограмма*. Для ее построения на оси OX откладывают интервалы шириной h, на каждом интервале строят прямоугольник высотой m/h. Величина m/h называется *плотностью частоты*. **Гистограмма** является эмпирическим аналогом графика дифференциальной функции распределения.

Пример 3. Измерена масса тела 100 женщин 30 лет, получены значения от 60 до 90 кг. Построить интервальный вариационный ряд (табл. 4.3) и гистограмму.

Таблица 4.3. Интервальный вариационный ряд

Интервал	Середина интервала	m	m/h
60–65	62.5	14	2.8
65–70	67.5	32	6.4
70–75	72.5	28	5.6
75–80	77.5	14	2.8

80–85	82.5	7	1.4
85–90	87.5	2	0.4

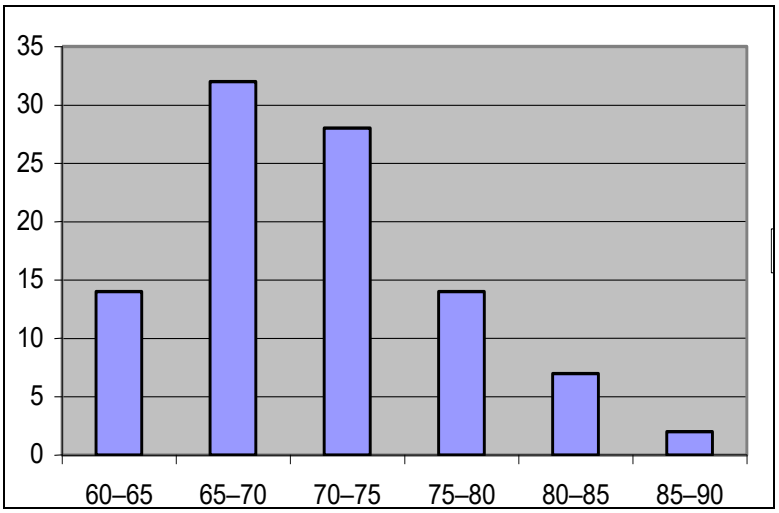


Рисунок 4.2. Гистограмма

Эмпирическая функция распределения находится по следующей формуле (отношение накопленных частот к объему выборки):

$$F^*(x) = \begin{cases} 0, & x \le x_1, \\ \dfrac{n_1}{n}, & x_1 < x \le x_2, \\ \dfrac{n_1 + n_2}{n}, & x_2 < x \le x_3, \\ \dfrac{n_1 + n_2 + n_3}{n}, & x_3 < x \le x_4, \\ \ldots & \ldots \\ n = \sum\limits_{i=1}^{m-1} \dfrac{n_i}{n}, & x_{m-1} < x \le x_m, \\ 1, & x > x_m. \end{cases} \qquad (4.1)$$

СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ

ОЦЕНКА ПАРАМЕТРОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ. ТОЧЕЧНАЯ ОЦЕНКА И ЕЕ СВОЙСТВА

Числовые характеристики генеральной совокупности называются *параметрами* генеральной совокупности.

Например, для нормального распределения это математическое ожидание и среднее квадратическое отклонение, для равномерного распределения – это границы интервала, в котором наблюдаются значения этой случайной величины

Оценка параметра – соответствующая числовая характеристика, рассчитанная по выборке. Если оценка определяется одним числом, она называется *точечной оценкой*.

Например, среднее арифметическое выборочных значений служит оценкой математического ожидания. Выборочные значения случайны, поэтому оценки можно рассматривать как случайные величины. Построим точечную оценку параметра θ по выборке $x_1, x_2, ..., x_n$ как значение некоторой функции и перечислим «желаемые» свойства оценки θ^*.

Определение. Оценка θ^* называется *несмещенной*, если ее математическое ожидание равно истинному значению оцениваемого параметра: $M\theta^* = \theta$.

Данное свойство характеризует отсутствие *систематической ошибки*, т.е. при многократном использовании вместо параметра θ его оценки θ^* среднее значение ошибки приближения $|\theta - \theta^*|$ равно нулю.

Так, выборочное среднее арифметическое $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ является *несмещенной оценкой* математического ожидания, а выборочная дисперсия $\bar{D} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ – *смещенная оценка* генеральной дисперсии *D*. Несмещенной оценкой генеральной дисперсии является оценка («исправленная дисперсия») $S^2 = \frac{1}{n-1}\cdot\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Определение. Оценка $\theta^*(X_1, X_2, ..., X_n)$ называется *состоятельной*, если она сходится по вероятности к оцениваемому параметру θ при $n \to \infty$: $\theta^* \xrightarrow[n\to\infty]{P} \theta$.

Данное свойство характеризует улучшение оценки с увеличением объема выборки.

Сходимость по вероятности означает, что при большом объеме выборки вероятность больших отклонений оценки от истинного значения мала.

Определение . Несмещенная оценка является *эффективной*, если она имеет наименьшую среди всех несмещенных оценок дисперсию.

Пример 4.

1. Вычислить среднее значение массы тела детей 6 лет.

$$\bar{X} = \frac{24+22+23+28+24+23+25+27+25+25}{10} = 24.6$$

2. Если выборочное среднее вычисляется по вариационному ряду, то находят сумму произведений вариант на соответствующие частоты, и делят на количество элементов в выборке: $\bar{X} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i$.

$$\bar{X} = \frac{22+23\cdot 2+24\cdot 2+25\cdot 3+27+28}{10} = 24.6$$

3. В том случае, когда статистические данные представлены в виде интервального вариационного ряда, при вычислении выборочного среднего значениями вариант считают середины интервалов. Так, для вычисления среднего значения массы тела женщин 30 лет из примера 4.3. используют формулу:

$$\bar{X} = \frac{62.5\cdot 14+67.5\cdot 32+72.5\cdot 14+77.5\cdot 14+82.5\cdot 7+87.5\cdot 2}{100} = 71.5\ \hat{e}\tilde{a}.$$

Другими характеристиками являются **мода** и **медиана**.

В теории вероятностей ***модой*** M_o дискретной случайной величины называется ее значение, которое имеет максимальную вероятность. Модой

непрерывной случайной величины называется такое ее значение, при котором достигается максимум плотности распределения $f(x)$. Закон распределения называется унимодальным, если мода единственна. В математической статистике мода M_o определяется по выборке, как *варианта с наибольшей частотой*.

Медианой называется варианта, расположенная в центре *ранжированного ряда*. Если ряд состоит из четного числа вариант, то медианой считают среднее арифметическое двух вариант, расположенных в центре ранжированного ряда.

Пример 5. Найти моду и медиану выборочной совокупности по массе тела детей 6 лет.

Ответ: $M_o = 24$; $M_e = 24$.

Основные числовые характеристики выборочной совокупности:

1) *размах вариационного ряда* $R=X_{max} - X_{min.}$ Этот показатель является наиболее простой характеристикой рассеяния и показывает диапазон варьирования величины. Этой характеристикой пользуются при работе с малыми выборками;

2) *выборочное среднее* находится как взвешенное среднее арифметическое $\bar{x}_в = \frac{\sum_{i=1}^{m} x_i \cdot n_i}{n}$, которое характеризует среднее значение признака X в пределах рассматриваемой выборки;

3) *выборочная дисперсия* определяется по формуле: $D_в = \frac{\sum_{i=1}^{m} \left(x_i - \bar{x}_в\right)^2 \cdot n_i}{n}$, которая является мерой рассеяния возможных значений показателя X вокруг своего среднего значения, и ее размерность совпадает с квадратом размерности варианты;

4) *выборочное среднее квадратическое отклонение* $\sigma_в(X) = \sqrt{D_в}$ описывает абсолютный разброс значений показателя X. Его размерность

совпадает с размерностью варианты;

5) «*исправленная*» *дисперсия* (вычисляют при малых n, n<30) $S^2(X)=\frac{n}{n-1}D_{в}=\frac{\sum_{i=1}^{m}\left(x_i-\bar{x}_{в}\right)^2\cdot n_i}{n-1}$ и «*исправленное*» *стандартное отклонение* $S(X)=\sqrt{S^2(X)}$;

6) *коэффициент вариации* $V=\frac{\sigma_{в}(X)}{\bar{x}_{в}}\cdot 100\%$ характеризует относительную изменчивость показателя X, то есть относительный разброс вокруг его среднего значения $\bar{x}_{в}$. Коэффициент вариации является безразмерной величиной, поэтому он пригоден для сравнения рассеяния вариационных рядов, варианты которых имеют различную размерность.

Пример 6.: Измерена длина (X) и масса тела (Y) девочек 10-ти лет. Получены следующие показатели: X=130 см, σ_X = 5 см, Y = 32 кг, σ_Y = 4 кг. Какая величина имеет большую вариативность?

Так как длина и масса тела измеряются в разных единицах, то вариативность нельзя сравнить при помощи СКО. Необходимо вычислить относительный показатель вариации.

$$V_X=\frac{5}{130}\cdot 100\%=3.8\%;$$

$$V_Y=\frac{4}{32}\cdot 100\%=12.5\%.$$

Таким образом, масса тела имеет большую вариативность, чем длина тела.

ОЦЕНКА С ПОМОЩЬЮ ИНТЕРВАЛОВ

Оценка параметров с помощью интервалов заключается в нахождении интервалов, называемых *доверительными*, между границами которых с определенными вероятностями (доверительными) находятся истинные значения оцениваемых параметров. *Интервальная оценка* определяется двумя числами – концами интервала.

Пусть найденная по данным выборки величина θ^* служит оценкой неизвестного параметра θ. Оценка θ^* определяется тем точнее, чем меньше $|\theta - \theta^*|$, т. е. чем меньше δ в неравенстве $|\theta - \theta^*| < \delta$, $\delta > 0$.

Доверительной вероятностью (*надежностью*) оценки θ^* параметра θ называется вероятность γ, с которой оценивается неравенство $|\theta - \theta^*| < \delta$.

Число $\alpha = 1 - \gamma$ называется *уровнем значимости*, определяющим вероятность того, что оцениваемый параметр не попадет в доверительный интервал.

Обычно задается надежность γ и определяется δ. Чаще всего вероятность γ задается значениями от 0.95 и выше. Неравенство $|\theta - \theta^*| < \delta$ можно записать в виде

$$-\delta < \theta - \theta^* < \delta \text{ или } \theta^* - \delta < \theta < \theta^* + \delta.$$

Доверительным интервалом называется интервал $(\theta^* - \delta,\ \theta^* + \delta)$, который покрывает неизвестный параметр θ с заданной надежностью.

Определение доверительного интервала для среднего значения нормально распределенной измеряемой случайной величины X при известной дисперсии σ_X^2.

Нам уже известно, что $\bar{m} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Можно показать [1-5], что $M(\bar{m}) = m,\ \ \sigma^2[\bar{m}] = \frac{\sigma_X^2}{n}$ (сумма $\bar{m}$ нормально распределенных случайных величин X_i сама является нормальной).

Зададим доверительную вероятность γ и найдем доверительный интервал $(\bar{m} - \delta,\ \bar{m} + \delta)$, который покрывал бы неизвестный параметр $\bar{m}$ с заданной надежностью γ.

Согласно формуле **В** (свойства нормального распределения, раздел 3)

$$\mathbf{P}\left(|\bar{m} - m| < \delta\right) = 2 \cdot \Phi_{0,1}\left(\frac{\delta\sqrt{n}}{\sigma_X}\right) = 2 \cdot \Phi_{0,1}\left(t_\gamma\right) = \gamma. \qquad (4.1)$$

Таким образом, для отыскания величины *доверительной границы* случайного отклонения результатов наблюдений по *доверительной вероятности* γ имеем уравнение:

$$2 \cdot \Phi_{0,1}\left(t_\gamma\right) = \gamma \text{, где } t_\gamma = \frac{\delta\sqrt{n}}{\sigma_X},$$

где значение t_γ находим по таблице Лапласа (приложение 1), $t_\gamma : \Phi_{0,1}\left(t_\gamma\right) = \gamma / 2$.

Пример 7. По результатам наблюдений была найдена оценка неизвестного математического ожидания m случайной величины $\xi \sim N\left(m, \sigma^2\right)$, если точечная оценка $\bar{m}$=10.2, а дисперсия оценки σ_X=4. Требуется оценить доверительный интервал для оценки математического ожидания по 36-ти наблюдениям с заданной надежностью γ=0.99.

Решение. Из (4.1) следует, что $\Phi_{0,1}\left(t_\gamma\right) = \frac{0.99}{2} = 0.495$. Отсюда получаем, что $t_\gamma = \frac{\delta\sqrt{n}}{\sigma_X}$=2.58 и половина искомого интервала $\delta = \frac{2.58 \cdot 4}{\sqrt{36}} = 1.89$. Так как $\bar{m} - \delta < m < \bar{m} + \delta$, то с вероятностью 0.99 доверительный интервал для оценки математического ожидания: $8.31 < m < 12.09$.

Со случаем, когда распределение результатов наблюдений нормально, но их дисперсия неизвестна, можно ознакомится в [3, 4, 6].

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Статистическая гипотеза — это предположение

о виде закона распределения («данная генеральная совокупность нормально распределена»);

о значениях его параметров («генеральное среднее равно нулю»);

об однородности данных («эти две выборки извлечены из одной генеральной совокупности»).

Статистическая проверка гипотезы состоит в выяснении того, согласуются ли результаты наблюдений (выборочные данные) с нашим предположением.

Результатом проверки может быть отрицательный ответ: выборочные данные противоречат высказанной гипотезе, поэтому от нее следует отказаться.

В случае ответа неотрицательного (выборочные данные не противоречат гипотезе) гипотезу принимают в качестве *одного из допустимых решений* (не единственно верного).

Различают *основную* (нулевую) гипотезу (гипотеза, которая проверяется, H_0) и альтернативную (конкурирующую, противопоставленную основной, H_1).

Например, если нулевая гипотеза H_0: MX= 10 (т. е. математическое ожидание нормально распределенной величины равно 10), тогда гипотеза H_1, может иметь вид H_1: $MX \neq 10$.

Цель статистической проверки гипотез: на основании выборочных данных принять решение о справедливости основной гипотезы или отклонить в ее пользу альтернативной.

Так как проверка осуществляется на основании выборки, а не всей генеральной совокупности, то существует вероятность, возможно, очень малая, ошибочного заключения.

Так, нулевая гипотеза может быть отвергнута, в то время как в действительности в генеральной совокупности она является справедливой.

Такую ошибку называют *ошибкой первого рода*, а её вероятность — *уровнем значимости* и обозначают α (стандартные значения α: 0.1, 0.05, 0.01, 0.001).

Возможно, что нулевая гипотеза принимается, в то время как в генеральной совокупности справедлива альтернативная гипотеза.

Такую ошибку называют *ошибкой второго рода*, а её вероятность обозначают β.

Замечание. Ошибка первого рода состоит в том, что будет отвергнута правильная гипотеза. Ошибка второго рода состоит в том, что будет принята неправильная гипотеза.

Проверка статистических гипотез осуществляется с *помощью статистического критерия K* — правила (функции от результатов наблюдений), определяющего меру расхождения результатов наблюдений с нулевой гипотезой.

Вероятность $(1-\beta)$ называют *мощностью* критерия.

Например, основная гипотеза состоит в том, что предприятие получает прибыль. Если это правильная гипотеза, то ошибка первого рода состоит в том, что данная гипотеза отвергается.

Если принимается решение о том, что прибыль предприятие не получает, то это ошибка второго рода.

Иногда ошибку первого рода называют «альфа-риск» (α-риск) а ошибку второго рода «бета-риск» (β-риск).

Из двух критериев, характеризующихся одной и той же вероятностью α, выбирают тот, которому соответствует меньшая ошибка 2-го рода, т.е. большая мощность.

Уменьшить вероятности обеих ошибок α и β одновременно можно, увеличив объем выборки.

Значения критерия *K* разделяются на две части: область ***допустимых значений*** (область принятия гипотезы H_0) и ***критическую область*** (область принятия гипотезы H_1).

Критическая область состоит из тех же значений критерия *K*, которые маловероятны при справедливости гипотезы H_0.

Если значение $K_{\text{набл}}$ критерия *K*, рассчитанное по выборочным данным, попадает в критическую область, то гипотеза H_0 отвергается в пользу альтернативной H_1. В противном случае мы утверждаем, что нет оснований отклонять гипотезу H_0.

Пример. Для подготовки к зачету преподаватель сформулировал 100 вопросов (генеральная совокупность) и считает, что студенту можно поставить «зачтено», если тот знает как ответить на 60 % вопросов (критерий).

Преподаватель задает студенту 5 вопросов (выборка из генеральной совокупности) и ставит «зачтено», если правильных ответов не меньше трех.

Гипотеза H_0: «студент курс усвоил», а множество {3, 4, 5}— область принятия этой гипотезы.

Критической областью является множество {0, 1, 2}— правильных ответов меньше трех, в этом случае основная гипотеза отвергается в пользу альтернативной H_1 «студент курс не усвоил, знает меньше 60 % вопросов».

Студент *A* выучил 70 вопросов из 100, но ответил правильно только на два из пяти, предложенных преподавателем.

Зачет не сдан.

В этом случае преподаватель совершает ошибку первого рода.

Студент *Б* выучил 50 вопросов из 100, но ему повезло, и он ответил правильно на 3 вопроса. Зачет сдан, но совершена ошибка второго рода.

Преподаватель может уменьшить вероятность этих ошибок, увеличив количество задаваемых на зачете вопросов.

Алгоритм проверки статистических гипотез сводится к следующему:

1) сформулировать основную H_0 и альтернативную H_1 гипотезы;

2) выбрать уровень значимости α;

3) в соответствии с видом гипотезы H_0 выбрать статистический критерий для ее проверки, т.е. случайную величину *K*, распределение которой известно;

4) по таблицам распределения случайной величины *K* найти границу критической области $K_{кр}$ (вид критической области определить по виду альтернативной гипотезы H_1);

5) по выборочным данным вычислить наблюдаемое значение критерия $K_{набл}$;

6) принять статистическое решение: если $K_{набл}$ попадает в критическую область — отклонить гипотезу H_0 в пользу альтернативной H_1; если $K_{набл}$ попадает в область допустимых значений, то *нет оснований отклонять основную гипотезу*.

http://www.aup.ru/books/m155/

А.И. Орлов

Математика случая

Вероятность и статистика – основные факты

Учебное пособие. М.: МЗ-Пресс, 2004.

Распределения Пирсона (хи – квадрат), Стьюдента и Фишера

С помощью нормального распределения определяются три распределения, которые в настоящее время часто используются при статистической обработке данных. В дальнейших разделах книги много раз встречаются эти распределения.

Распределение Пирсона χ^2 (хи - квадрат) – распределение случайной величины

$$X = X_1^2 + X_2^2 + ... + X_n^2,$$

где случайные величины $X_1, X_2, ..., X_n$ независимы и имеют одно и тоже распределение $N(0,1)$. При этом число слагаемых, т.е. n, называется «числом степеней свободы» распределения хи – квадрат.

Распределение хи-квадрат используют при оценивании дисперсии (с помощью доверительного интервала), при проверке гипотез согласия, однородности, независимости, прежде всего для качественных (категоризованных) переменных, принимающих конечное число значений, и во многих других задачах статистического анализа данных [8, 9, 11, 16].

Распределение t Стьюдента – это распределение случайной величины

$$T = \frac{U\sqrt{n}}{\sqrt{X}},$$

где случайные величины U и X независимы, U имеет распределение стандартное нормальное распределение $N(0,1)$, а X – распределение хи – квадрат с n степенями свободы. При этом nназывается «числом степеней свободы» распределения Стьюдента.

Распределение Стьюдента было введено в 1908 г. английским статистиком В. Госсетом, работавшем на фабрике, выпускающей пиво. Вероятностно-статистические методы использовались для принятия экономических и технических решений на этой фабрике, поэтому ее руководство запрещало В. Госсету публиковать научные статьи под своим именем. Таким способом охранялась коммерческая тайна, «ноу-хау» в виде вероятностно-статистических методов, разработанных В. Госсетом. Однако он имел возможность публиковаться под псевдонимом «Стьюдент». История Госсета - Стьюдента показывает, что еще сто лет назад менеджерам Великобритании была очевидна большая экономическая эффективность вероятностно-статистических методов.

В настоящее время распределение Стьюдента – одно из наиболее известных распределений среди используемых при анализе реальных данных. Его применяют при оценивании математического ожидания, прогнозного значения и других характеристик с помощью доверительных интервалов, по проверке гипотез о значениях математических ожиданий, коэффициентов регрессионной зависимости, гипотез однородности выборок и т.д. [8, 9, 11, 16].

Распределение Фишера – это распределение случайной величины

$$F = \frac{\frac{1}{k_1} X_1}{\frac{1}{k_2} X_2},$$

где случайные величины X_1 и X_2 независимы и имеют распределения хи – квадрат с числом степеней свободы k_1 и k_2 соответственно. При этом пара (k_1, k_2) – пара «чисел степеней свободы» распределения Фишера, а именно, k_1 – число степеней свободы числителя, а k_2 – число степеней свободы знаменателя. Распределение случайной величины F названо в честь великого английского статистика Р.Фишера (1890-1962), активно использовавшего его в своих работах.

Распределение Фишера используют при проверке гипотез об адекватности модели в регрессионном анализе, о равенстве дисперсий и в других задачах прикладной статистики [8, 9, 11, 16].

Выражения для функций распределения хи - квадрат, Стьюдента и Фишера, их плотностей и характеристик, а также таблицы, необходимые для их практического использования, можно найти в специальной литературе (см., например, [8]).

http://www.aup.ru/books/m155/

Хи-квадрат критерий

Chi-square test

Синонимы на русском:

Критерий согласия Пирсона

Критерий согласия для проверки гипотезы о законе распределения исследуемой случайной величины. Во многих практических задачах точный закон распределения неизвестен. Поэтому выдвигается гипотеза о соответствии имеющегося эмпирического закона, построенного по наблюдениям, некоторому теоретическому. Данная гипотеза требует статистической проверки, по результатам которой будет либо подтверждена, либо опровергнута.

Пусть X – исследуемая случайная величина. Требуется проверить гипотезу H_0 о том, что данная случайная величина подчиняется закону распределения $F(x)$. Для этого необходимо произвести выборку из n независимых наблюдений и по ней построить эмпирический закон распределения $F'(x)$. Для сравнения эмпирического и гипотетического законов используется правило, называемое критерием согласия. Одним из популярных является критерий согласия хи-квадрат К. Пирсона.

В нем вычисляется статистика хи-квадрат:

$$x2=N\sum i=1N(pti−pei)2pti,$$

где N – число интервалов, по которому строился эмпирический закон распределения (число столбцов соответствующей гистограммы), i – номер интервала, p_{ti} - вероятность попадания значения случайной величины в i-й интервал для теоретического закона распределения, p_{ei} – вероятность попадания значения случайной величины в i-й интервал для эмпирического закона распределения. Она и должна подчинятьсяраспределению хи-квадрат.

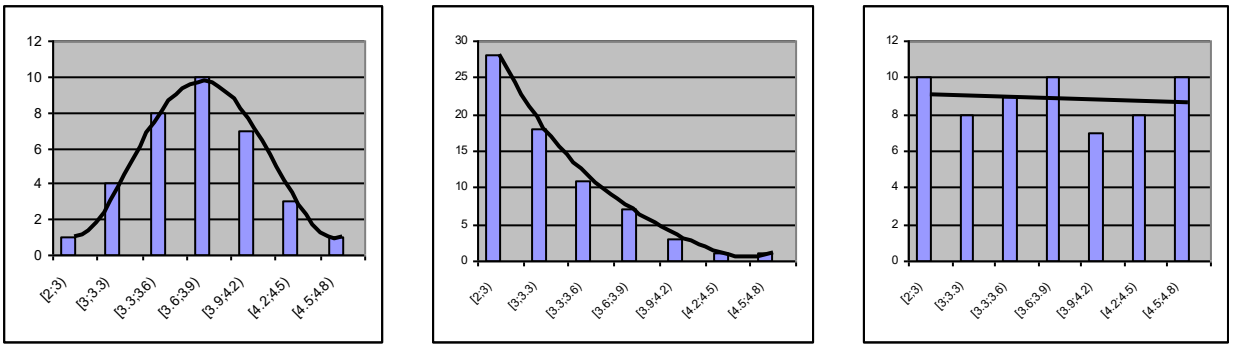
Если вычисленное значение статистики превосходит квантиль распределения хи-квадрат с $k−p−1$ степенями свободы для заданного уровня значимости, то гипотеза H_0 отвергается. В противном случае она принимается на заданном уровне значимости. Здесь k – число наблюдений, p – число оцениваемых параметров закона распределения.

https://basegroup.ru/community/glossary/chi-square-test

ПРОВЕРКА ГИПОТЕЗ О ВИДЕ РАСПРЕДЕЛЕНИЯ. КРИТЕРИЙ СОГЛАСИЯ ПИРСОНА

Одной из важных задач математической статистики является установление теоретического закона распределения случайной величины, характеризующей изучаемый признак по эмпирическому распределению, представляющему вариационный ряд.

Предположение о виде закона распределения можно сделать по гистограмме или полигону (Рис. 4.3)



а) б) в)

Рис. 4.3. Возможные виды гистограмм:

а) нормального, б) показательного, в) равномерного распределений

Например, по гистограмме (рис. 4.3, а)) можно сделать предположение о том, что генеральная совокупность распределена по нормальному закону.

Для проверки гипотез о виде распределения служат специальные критерии — *критерии согласия*. Они отвечают на вопрос: согласуются ли результаты экспериментов с предположением о том, что генеральная совокупность имеет заданное распределение.

Проверим это предположение с помощью ***критерия согласия Пирсона***.

В этом критерии мерой расхождения между гипотетическим (предполагаемым) и эмпирическим распределением служит статистика

$$K = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j},$$

где n — объем выборки;

k — количество интервалов (групп наблюдений);

n_j — количество наблюдений, попавших в j-й интервал;

p_j — вероятность попадания в j-й интервал случайной величины, распределенной по гипотетическому закону.

Если предположение о виде закона распределения справедливо, то статистика Пирсона распределена по закону «хи-квадрат» с числом степеней свободы $k-r-1$ (r — число параметров распределения, оцениваемых по выборке): $K \sim \chi^2_{(k-r-1)}$.

Оцениваются неизвестные параметры с использованием теории точечных оценок (см. источник [Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике.], гл.16 и раздел 3.8. настоящего пособия), некоторые оценки приведены в табл. 4.4.

Таблица 4.4. Оцениваемые параметры и их точечные оценки

Вид распределения	Оцениваемые параметры	Точечные оценки параметров
$f_X(x)=\begin{cases}0 & ,x<0\\ \lambda e^{-\lambda x} & ,x\ge 0\end{cases}$	$\lambda>0$	$\bar{\lambda}=\frac{1}{\bar{x}_B}$
$f_X(x)=\begin{cases}0, & x<\alpha\\ 1/(\beta-\alpha), & x\in[\alpha,\beta]\\ 0, & x>\beta\end{cases}$	$\alpha,\quad \beta$	$\alpha=\bar{x}_B-\sqrt{3}\cdot\sqrt{D_B}$ $\beta=\bar{x}_B+\sqrt{3}\cdot\sqrt{D_B}$
$f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-m)^2}{2\sigma^2}}$	$m,\quad \sigma^2$	$m=\bar{x}_B,\quad \sigma^2=\sqrt{D_B}$

Здесь $\bar{x}_B=\dfrac{\sum_{i=1}^{m}x_i\cdot n_i}{n}$ $\quad D_B=\dfrac{\sum_{i=1}^{m}\left(x_i-\bar{x}_B\right)^2\cdot n_i}{n}$.

Количество интервалов k рекомендуется рассчитывать по формуле Старджеса $k=1+3.3\cdot\lg n,$ где n — объем выборки. Длину i-го интервала принимают равной $h=\dfrac{x_{(n)}-x_{(1)}}{k}$, где $x_{(n)}$—наибольшее, а $x_{(1)}$ — наименьшее значение в вариационном ряду.

Пример.

Для среднего балла среди 30-ти групп (с точностью до сотых долей балла) получили выборку x_i:

3.7, 3.85, 3.7, 3.78, 3.6, 4.45, 4.2, 3.87, 3.33, 3.76, 3.75, 4.03, 3.8, 4.75, 3.25, 4.1, 3.55, 3.35, 3.38, 3.05, 3.56, 4.05, 3.24, 4.08, 3.58, 3.98, 3.4, 3.8, 3.06, 4.38.

Проверить гипотезу о нормальном распределении среднего балла на уровне значимости $\alpha = 0.025$.

Решение. Сгруппируем эту выборку.

Наименьший средний балл равен 3.05, наибольший — 4.75.

Интервал [3; 4.8] разобьем на 6 частей длиной $h = 0.3$, применяя формулу Старджеса ($k = 5.875 \approx 6$). Подсчитаем частоту n_i (относительную частоту $\frac{n_i}{n}$) для каждого интервала и получим сгруппированный статистический ряд (табл. 4.5).

Таблица 4.5. Статистический ряд

Интервалы	[3;3.3)	[3.3;3.6)	[3.6;3.9)	[3.9;4.2)	[4.2;4.5)	[4.5;4.8)
Частоты n_i	4	7	10	5	3	1
Относительные частоты $\frac{n_i}{n}$	0.133	0.233	0.3	0.167	0.1	0.033

Правило Стёрджеса — эмпирическое правило определения оптимального количества интервалов, на которые разбивается наблюдаемый диапазон изменения случайной величины при построении гистограммы плотности её распределения. Названо по имени американского статистика Герберта Стёрджеса (*Herbert Arthur Sturges*, 1882—1958).

Количество интервалов n определяется как:

$$n = 1 + \lfloor \log_2 N \rfloor,$$

где N — общее число наблюдений величины, $\log_2$ — логарифм по основанию 2, $\lfloor x \rfloor$ — обозначает целую часть числа x.

Часто встречается записанным через десятичный логарифм:

$$n = 1 + \lfloor 3.322 \lg N \rfloor,$$

Основанием для него служит оценка количества событий с разными вероятностями в схеме испытаний Бернулли длительностью в $n-1$ этап. Если имеются серии испытаний с 2 альтернативными исходами с постоянной вероятностью каждого, то число видов серий, где в составе имеется k исходов, принимающих первое из альтернативных значений, и, соответственно, $n-k-1$ — принимающих второе, равно: n (от $k=0$ до $k=n-1$), а общее число серий $N=2^{n-1}$.

Если аппроксимировать значения наблюдаемой случайной величины результатами сложения случайно выпадающих в серии испытаний значений двух чисел a и b (например 0 и 1), соответствующих исходам схемы Бернулли, то каждой серии испытаний содержащей k исходов с результатом a и $n-k-1$ исходов с результатом b будет соответствовать сумма $ka+(n-k+1)b$. Количество различных значений (в рассматриваемом случае: $a(n-1), a(n-2)+b, ..a+b(n-2), b(n-1)$, для пары $0,1$ — $0,1,2,..n-1$) будет равно количеству последовательностей с различным числом исходов n. Т.о., если ставить задачу, чтобы на каждый интервал между a и b приходилось в среднем не меньше одного значения суммы, а значит и не меньше одной серии испытаний, моделирующей получение случайной величины, то число этапов в серии, равное числу интервалов, на которые разбивается диапазон изменения наблюдаемых значений, должно быть не больше, чем $n=1+\lfloor \log_2 N \rfloor$

https://ru.wikipedia.org

Таблица 4.5. Статистический ряд

Интервалы	[3;3.3)	[3.3;3.6)	[3.6;3.9)	[3.9;4.2)	[4.2;4.5)	[4.5;4.8)
Частоты n_i	4	7	10	5	3	1
Относительные частоты $\frac{n_i}{n}$	0.133	0.233	0.3	0.167	0.1	0.033

Сформулируем основную и альтернативную гипотезы.

$H_0: X \sim N(\bar{a}, \bar{\sigma})$ — случайная величина X (средний балл) подчиняется нормальному закону с параметрами $\bar{a}$, $\bar{\sigma}$. Так как истинных значений параметров a, σ мы не знаем, возьмем их оценки, рассчитанные по выборке: $\bar{a}=3.746$, $\bar{\sigma}=0.399$.

H_1: случайная величина X не подчиняется нормальному закону с данными параметрами.

Рассчитаем наблюдаемое значение $K_{\text{набл}}$ статистики Пирсона. Эмпирические частоты n_j уже известны (табл. 4.5), а для вычисления вероятностей p_j (в предположении, что гипотеза H_0 справедлива) применим уже известную формулу (свойство **В**):

$$p_j = P(a_j < X < a_{j+1}) = \Phi\left(\frac{a_{j+1} - \bar{a}}{\bar{\sigma}}\right) - \Phi\left(\frac{a_j - \bar{a}}{\bar{\sigma}}\right), \quad j = 1,\ 2,\ ...,\ k$$

и таблицу функции Лапласа (приложение 1). Полученные результаты сведем в таблицу (табл. 4.6). Наблюдаемое значение статистики Пирсона равно $K_{\text{набл}} = 0.978$.

Определим границу критической области. Так как статистика Пирсона измеряет разницу между эмпирическим и теоретическим распределениями, то чем больше ее наблюдаемое значение $K_{\text{набл}}$, тем сильнее довод против основной гипотезы.

Поэтому критическая область для этой статистики всегда правосторонняя: $[K_{\text{кр}};\ +\infty)$. Её границу $K_{\text{кр}} = \chi^2_{(k-r-1;\ \alpha)}$ находим по таблицам распределения «хи-квадрат» (приложение 3) и заданным значениям $\alpha = 0.025,\ k = 6$ (число интервалов), $r = 2$ (параметры a и σ оценены по выборке): $K_{\text{кр}} = \chi^2(6-2-1;\ 0.025) = \chi^2(3;\ 0.025) = 9.4$.

Наблюдаемое значение статистики Пирсона не попадает в критическую область: $K_{\text{набл}} < K_{\text{кр}}$, поэтому *нет оснований отвергать основную гипотезу*.

Вывод: на уровне значимости 0.025 справедливо предположение о том, что средний балл имеет нормальное распределение.

Таблица 4.6. Сравнение наблюдаемых и ожидаемых частот

№ п/п	Интервалы группировки $[a_{j;}a_{j+1})$	Наблюдаемая частота n_j	Вероятность p_j попадания в j-й интервал	Ожидаемая частота $n \cdot p_j$	Слагаемые статистики Пирсона $\frac{(n_j - np_j)^2}{np_j}$
1.	[3; 3.3)	4	0.101	3.032	0.309
2.	[3.3; 3.6)	7	0.225	6.761	0.008
3.	[3.6; 3.9)	10	0.295	8.79	0.166
4.	[3.9; 4.2)	5	0.222	6.665	0.416
5.	[4.2; 4.5)	3	0.098	2.946	0.001
6.	[4.5; 4.8)	1	0.025	0.758	0.077
$\sum$	—	30	0.965	28.95	$K_{\text{набл}} = 0.978.$

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

Основная литература.

1. Вентцель Е.С. Теория вероятностей. – М.: Высшая школа. 2004.–576с.

2. Вентцель Е.С. Задачи и упражнения по теории вероятностей. – М.: Высш. шк. 2004. – 166 с.

3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. Высшая школа, 2005.

4. Гмурман В.Е. Теория вероятностей и математическая статистика. Высшая школа, 2004, 480 с.

5. Данко П.Е., Попов А.Г., Кожевникова Т.Я. Высшая математика в упражнениях и задачах. – М.: Высшая школа. 1999. – 415 с.

6. Магазинников Л.И. Теория вероятностей. – Томск.: ТУСУР, 2000.–150 с

Дополнительнаялитература.

1. Р.Курант, Г.Робинс. Что такое математика? Элементарный очерк идей и методов. М.: Просвещение, 1967. 560 с.

2. Мостеллер Ф., Рурке Р., Томас Дж. Вероятность.-М.: Мир, 1969.
3. Скворцов В.В. Теория вероятностей? – Это интересно!– М.: Мир. 1992.– 118 с.

ПРИЛОЖЕНИЕ 1

Таблица значений функции $$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-z^2/2}\, dz$$

	0	1	2	3	4	5	6	7	8	9
0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997
4	0,49997	0,49997	0,49997	0,49997	0,49997	0,49997	0,49998	0,49998	0,49998	0,49998
4,1	0,49998	0,49998	0,49998	0,49998	0,49998	0,49998	0,49998	0,49998	0,49999	0,49999
4,2	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999
4,3	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999
4,4	0,49999	0,49999	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000

ПРИЛОЖЕНИЕ 2

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

	0	1	2	3	4	5	6	7	8	9
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	0,2420	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
2	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

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

(продолжение)

	0	1	2	3	4	5	6	7	8	9
3	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	0,0008	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	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
4,1	0,00009	0,00009	0,00008	0,00008	0,00008	0,00007	0,00007	0,00007	0,00006	0,00006
4,2	0,00006	0,00006	0,00005	0,00005	0,00005	0,00005	0,00005	0,00004	0,00004	0,00004
4,3	0,00004	0,00004	0,00004	0,00003	0,00003	0,00003	0,00003	0,00003	0,00003	0,00003
4,4	0,00002	0,00002	0,00002	0,00002	0,00002	0,00002	0,00002	0,00002	0,00002	0,00002
4,5	0,00002	0,00002	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001
4,6	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001	0,00001
4,7	0,000006									
4,8	0,000004									
4,9	0,000002									
5	0,00000									

ПРИЛОЖЕНИЕ 3

Критические точки распределения χ^2

Число степеней свободы k	Уровень значимости α					
	0.01	0.025	0.05	0.95	0.975	0.99
1	6.6	5.0	3.8	0.0039	0.00098	0.00016
2	9.2	7.4	6.0	0.103	0.051	0.020
3	11.3	9.4	7.8	0.352	0.216	0.115
4	133	11.1	9.5	0.711	0.484	0.297
5	151	12.8	11.1	1.15	0.831	0.554
6	16.8	14.4	12.6	1.64	1.24	0.872
7	18.5	16.0	14.1	2.17	1.69	1.24
8	20.1	17.5	15.5	2.73	2.18	1.65
9	21.7	19.0	16.9	3.33	2.70	2.09
10	23.2	20.5	18.3	3.94	3.25	2.56
11	24.7	21.9	19.7	4.57	3.82	3.05
12	26.2	23.3	21 .0	5.23	4.40	3.57
13	27.7	24.7	22.4	5.89	5.01	4.11
14	29.1	26.1	23.7	6.57	5.63	4.66
15	30.6	27.5	25.0	7.26	6.26	5.23
16	32.0	28.8	26.3	7.96	6.91	5.81
17	33.4	30.2	27.6	8.67	7.56	6.41
18	34.8	31.5	28.9	9.39	8.23	7.01
19	36.2	32.9	30.1	10.1	8.91	7.63
20	37.6	34.2	31.4	10.9	9.59	8.26
21	38.9	35.5	32.7	11.6	10.3	8.90
22	40.3	36.8	33.9	12.3	11.0	9.54
23	41.6	38.1	35.2	13.1	11.7	10.2
24	43.0	39.4	36.4	13.8	12.4	10.9
25	44.3	40.6	37.7	14.6	13.1	11.5
26	45.6	41.9	38.9	15.4	13.8	12.2
27	47.0	43.2	40.1	16.2	14.6	12.9
28	48.3	44.5	41.3	16.9	15.3	13.6
29	49.6	45.7	42.6	17.7	16.0	14.3
30	50.9	47.0	43.8	18.5	16.8	15.0

ПРИЛОЖЕНИЕ 4

Критические точки распределения Стьюдента

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0.10	0.05	0.02	0.01	0.002	0.001
1	6.31	12.7	31.82	63.7	318.3	637.0
2	2.92	4.30	6.97	9.92	22.33	31.6
3	2.35	3.18	4.54	5.84	10.22	12.9
4	2.13	2.78	3.75	4.60	7.17	8.61
5	2.01	2.57	3.37	4.03	5.89	6.86
6	1.94	2.45	3.14	3.71	5.21	5.96
7	1.89	2.36	3.00	3.50	4.79	5.40
8	1.86	2.31	2.90	3.36	4.50	5.04
9	1.83	2.26	2.82	3.25	4.30	4.78
10	1.81	2.23	2.76	3.17	4.14	4.59
11	1.80	2.20	2.72	3.11	4.03	4.44
12	1.78	2.18	2.68	3.05	3.93	4.32
13	1.77	2.16	2.65	3.01	3.85	4.22
14	1.76	2.14	2.62	2.98	3.79	4.14
15	1.75	2.13	2.60	2.95	3.73	4.07
16	1.75	2.12	2.58	2.92	3.69	4.01
17	1.74	2.11	2.57	2.90	3.65	3.95
18	1.73	2.10	2.55	2.88	3.61	3.92
19	1.73	2.09	2.54	2.86	3.58	3.88
20	1.73	2.09	2.53	2.85	3.55	3.85
21	1.72	2.08	2.52	2.83	3.53	3.82
22	1.72	2.07	2.51	2.82	3.51	3.79
23	1.71	2.07	2.50	2.81	3.59	3.77
24	1.71	2.06	2.49	2.80	3.47	3.74
25	1.71	2.06	2.49	2.79	3.45	3.72
26	1.71	2.06	2.48	2.78	3.44	3.71
27	1.71	2.05	2.47	2.77	3.42	3.69
28	1.70	2.05	2.46	2.76	3.40	3.66
29	1.70	2.05	2.46	2.76	3.40	3.66
30	1.70	2.04	2.46	2.75	3.39	3.65
40	1.68	2.02	2.42	2.70	3.31	3.55
60	1.67	2.00	2.39	2.66	3.23	3.46
120	1.66	1.98	2.36	2.62	3.17	3.37
∞	1.64	1.96	2.33	2.58	3.09	3.29