Министерство образования Российской Федерации

Новосибирский государственный технический университет

51
????

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания
к расчетно-графическому заданию для студентов III курса ФПМИ

Новосибирск,
2020

УДК ?????

Методические указания предназначены для студентов, выполняющих расчетно-графическое задание по курсу «Математическая статистика». Указания содержат теоретические сведения, необходимые для решения задач по оцениванию параметров и проверке статистических гипотез. В работе разобраны примеры решения задач и приведены варианты заданий.

Составители: доктор техн. наук, доцент С.Н. Постовалов,
доктор техн. наук Е.В. Чимитова

Рецензент:

Работа подготовлена на кафедре теоретической и прикладной информатики

СОДЕРЖАНИЕ

Введение

Математическая статистика изучает способы получения статистических закономерностей на основании *наблюдений* случайных величин. В классической теории статистических выводов постулируется, что наблюдения являются значениями, принимаемыми случайными величинами, которые подчиняются совместному распределению вероятностей, принадлежащего некоторому известному классу $F = \{F(x;\theta), \theta \in \Omega\}$. Цель статистического анализа состоит в том, чтобы указать правдоподобное значение параметра θ, либо определить подмножество Ω, о котором мы можем утверждать, что оно содержит истинное значение параметра с заданной вероятностью.

Расчетно-графическое задание состоит из двух частей. В первой части необходимо найти оценки параметров распределений и исследовать их свойства: несмещенность, состоятельность и эффективность, а также построить доверительные интервалы. Во второй части требуется проверить гипотезу о виде распределения, гипотезу независимости и гипотезу однородности, гипотезу о незначимости коэффициента корреляции, гипотезу о значимости модели линейной регрессии.

При решении задач необходимо сначала выполнить постановку задачи, а именно – определить тип наблюдаемой случайной величины (непрерывная или дискретная, регулярная или нерегулярная модель), а также способ представления наблюдений случайной величины (группированный или негруппированный). В соответствии с этим необходимо выбрать подходящий метод оценивания или критерий для проверки гипотезы.

При выполнении расчетно-графического задания допускается использование специального программного обеспечения. Вычисление оценок и статистик критериев можно выполнять численно на компьютере.

Отчет по расчетно-графическому заданию должен содержать титульный лист, лист задания, текст решения задач с необходимой степенью детализации, ссылки на соответствующие теоремы, свойства, статистические таблицы, использованные при получении решения.

Часть 1. Оценивание параметров

Пусть имеется выборка $\mathbf{X}_n = \{X_1,...,X_n\}$ из распределения случайной величины $\xi \in \{F(x;\theta), \theta \in \Theta\}$. В общем случае задача оценивания заключается в том, чтобы, используя статистическую информацию, доставляемую выборкой, сделать статистические выводы об истинном значении неизвестного параметра θ.

Точечной оценкой неизвестного параметра θ по выборке $\mathbf{X}_n$ называется значение некоторой статистики $T_n = T(\mathbf{X}_n)$, которое приближенно равно значению параметра $\theta : \hat{\theta} = T(\mathbf{X}_n)$.

Так как любая статистика является случайной величиной (имеющей некоторое распределение $G_{T_n}(x)$), то для каждой новой реализации выборки $\mathbf{X}_n$ будет получаться другое значение оценки, в общем случае отличное от истинного значения параметра θ.

Интервальной оценкой (или *доверительным интервалом*) параметра θ называют интервал $[T_1(\mathbf{X}_n), T_2(\mathbf{X}_n)]$, содержащий истинное значение параметра θ с вероятностью γ.

1.1. Методы оценивания параметров

Существует множество различных методов построения оценок неизвестных параметров закона распределения случайной величины по выборке $\mathbf{X}_n$. Рассмотрим наиболее простые методы.

1.1.1. Метод моментов

Приравнивая теоретические и выборочные моменты можно найти точечные оценки неизвестных параметров.

$$M\theta^k = \frac{1}{n}\sum_{i=1}^{n} X_i^k,\ k = 1,2,\dots. \qquad (1.1)$$

Такой метод называется *методом моментов*. Если в векторе θ содержится r неизвестных параметров, то необходимо взять столько уравнений (1.1), чтобы можно было выразить неизвестные параметры.

Пример 1.1

Пусть $X_1,...,X_n$ – выборка из гамма-распределения с функцией плотности:

$$f(x;\alpha,\beta) = \frac{x^{\beta-1} e^{-x/\alpha}}{\Gamma(\beta)\alpha^{\beta}},\ x > 0,\ \alpha,\beta > 0$$

Требуется найти оценку по методу моментов векторного параметра (α,β).

<u>Решение:</u>

Найдем первый и второй теоретические моменты:

$$MX_1=\int_0^\infty x\frac{x^{\beta-1}e^{-x/\alpha}}{\Gamma(\beta)\alpha^\beta}dx=\frac{\alpha}{\Gamma(\beta)}\int_0^\infty y^\beta e^{-y}dy=\frac{\alpha}{\Gamma(\beta)}\Gamma(\beta+1)=\alpha\cdot\beta,$$

$$MX_1^2=\int_0^\infty x^2\frac{x^{\beta-1}e^{-x/\alpha}}{\Gamma(\beta)\alpha^\beta}dx=\frac{\alpha^2}{\Gamma(\beta)}\int_0^\infty y^{\beta+1}e^{-y}dy=\frac{\alpha^2}{\Gamma(\beta)}\Gamma(\beta+2)=\alpha^2\cdot\beta\cdot(\beta+1).$$

Приравнивая теоретические и выборочные моменты, получим:

$$\begin{cases}\alpha\beta=\dfrac{1}{n}\sum_{i=1}^n X_i=A_1\\ \alpha^2\beta(\beta+1)=\dfrac{1}{n}\sum_{i=1}^n X_i^2=A_2\end{cases}\quad\Rightarrow\quad \begin{array}{l}\hat\alpha=\dfrac{A_2-A_1^2}{A_1}\\ \hat\beta=\dfrac{A_1^2}{A_2-A_1^2}\end{array}$$

1.1.2. Метод максимального правдоподобия

Оценкой максимального правдоподобия (ОМП) параметра θ называется точка параметрического множества Θ, в которой функция максимального правдоподобия $L(\mathbf{X}_n,\theta)=\prod_{i=1}^n f(X_i,\theta)$ достигает наибольшего значения: $L(\mathbf{X}_n,\hat\theta)=\sup\limits_{\theta\in\Theta}L(\mathbf{X}_n,\theta)$.

Если для любой выборки $\mathbf{X}_n$ из выборочного пространства максимум $L(\mathbf{X}_n,\theta)$ достигается во внутренней точке θ, и $L(\mathbf{X}_n,\theta)$ дифференцируема по θ, то ОМП $\hat\theta$ удовлетворяет уравнению $\dfrac{\partial\ln L(\mathbf{X}_n,\theta)}{\partial\theta_i}=0,\ i=1,...,r$, которое называется *уравнением правдоподобия*.

Пример 1.2

Построить оценку максимального правдоподобия параметра p распределения Бернулли: $P\{\xi=k\}=p^k(1-p)^{1-k}$, $k=0,1$.

<u>Решение:</u>

Логарифмическая функция правдоподобия равна

$$\ln L(X_1,...,X_n;p)=\sum_{i=1}^n\left(X_i\ln p+(1-X_i)\ln(1-p)\right)=$$

$$=\sum_{i=1}^n X_i\ln p+\left(n-\sum_{i=1}^n X_i\right)\ln(1-p);$$

$$\frac{\partial \ln L}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}\left(n - \sum_{i=1}^{n} X_i\right) = 0 \quad => \quad \hat{p} = \overline{X},$$

где $\overline{X}$ – среднее выборочное значение.

$$\frac{\partial^2 \ln L}{\partial p^2} = -\frac{1}{p^2}\sum_{i=1}^{n} X_i - \frac{1}{(1-p)^2}\left(n - \sum_{i=1}^{n} X_i\right) =$$

$$= \left(\frac{1}{(1-p)^2} - \frac{1}{p^2}\right)\sum_{i=1}^{n} X_i - \frac{n}{(1-p)^2} = \frac{2p-1}{p^2(1-p)^2}\sum_{i=1}^{n} X_i - \frac{n}{(1-p)^2}.$$

Проверим знак второй производной при $p = \overline{X}$:

$$\left.\frac{\partial^2 \ln L}{\partial p^2}\right|_{p=\overline{X}} = -\frac{n\overline{X}}{p^2} - \frac{1}{(1-p)^2}\left(n - n\overline{X}\right) = -\frac{n\overline{X}}{\overline{X}^2} - \frac{n}{(1-\overline{X})^2}\left(1-\overline{X}\right) =$$

$$= -\frac{n}{\overline{X}} - \frac{n}{(1-\overline{X})} = -\frac{n(1-\overline{X}+\overline{X})}{\overline{X}(1-\overline{X})} = -\frac{n}{\overline{X}(1-\overline{X})} < 0.$$

Таким образом, при $p = \overline{X}$ функция правдоподобия достигает максимума.

Пример 1.3

Построить оценку максимального правдоподобия параметра $\theta > 0$ равномерного распределения на отрезке $[0, \theta]$.

Решение:

Функция правдоподобия выборки равна

$$L(X_1,...,X_n;\theta) = \begin{cases} \theta^{-n}, \textit{ если все } X_j \in [0, \theta] \\ 0, \textit{ если хотя бы одно } X_j \notin [0, \theta] \end{cases} =$$

$$= \begin{cases} \theta^{-n}, \textit{ если } X_{(n)} \le \theta \\ 0, \textit{ если } X_{(n)} > \theta \end{cases},$$

где $X_{(n)}$ – максимальная порядковая статистика.

При фиксированных значениях выборки (и, следовательно, при фиксированном значении $X_{(n)}$) зависимость $L(X_1,...,X_n;\theta)$ от θ показана на рисунке 1.1. Максимум функции правдоподобия достигается в точке $\theta = X_{(n)}$. Поэтому искомая оценка максимального правдоподобия есть $\hat{\theta} = X_{(n)}$.

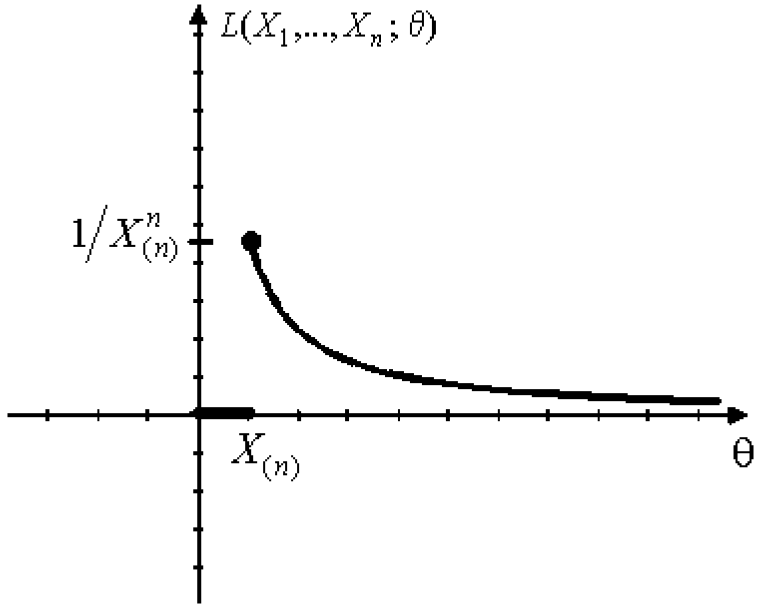


Рис. 1.1. Функция правдоподобия

1.1.3. Построение доверительного интервала с использованием распределения точечной оценки параметров

Если имеется некоторая точечная оценка $T_n = T(\mathbf{X}_n)$ для параметра θ и известна ее функция распределения $F_T(t,\theta)$, *непрерывная и монотонная* по θ, то доверительный интервал можно построить, основываясь на этой функции:

1. Вычисляем точечную оценку $T_n = T(\mathbf{X}_n)$.
2. Решаем относительно $\tilde{\theta}_1, \tilde{\theta}_2$ уравнения

$$F_T\left(T_n, \tilde{\theta}_1\right) = \frac{1-\gamma}{2},\ F_T\left(T_n, \tilde{\theta}_2\right) = \frac{1+\gamma}{2}.$$

3. Определяем границы доверительного интервала:

$$T_1 = \min\left(\tilde{\theta}_1, \tilde{\theta}_2\right),\ T_2 = \max\left(\tilde{\theta}_1, \tilde{\theta}_2\right).$$

1.1.4. Построение доверительного интервала с использованием центральной статистики

Статистика $G(\mathbf{X}_n,\theta)$ называется *центральной статистикой*, если распределение $G(\mathbf{X}_n,\theta)$ не зависит от θ, и при любом фиксированном θ статистика $G(\mathbf{X}_n,\theta)$ непрерывна и строго монотонна по θ.

С помощью центральной статистики можно построить доверительный интервал. Пусть $f_G(g)$ плотность распределения статистики $G(\mathbf{X}_n,\theta)$.

1. Найдем такие значения g_1, g_2, что

$$P\left\{g_1 < G\left(X_n, \theta\right) < g_2\right\} = \int\limits_{g_1}^{g_2} f_G\left(g\right) dg = \gamma.$$

2. Решим относительно $\tilde{T}_1, \tilde{T}_2$ уравнения

$$G\left(X_n, \tilde{T}_1\right) = g_1,\; G\left(X_n, \tilde{T}_2\right) = g_2.$$

3. Определяем границы доверительного интервала:

$$T_1 = \min\left\{\tilde{T}_1, \tilde{T}_2\right\}, T_2 = \max\left\{\tilde{T}_1, \tilde{T}_2\right\},$$

Для построения доверительного интервала с помощью центральной статистики основная проблема заключается в нахождении этой центральной статистики. Можно выделить класс моделей, для которых центральная статистика существует и имеет простой вид.

Пусть $F(x, \theta)$ – функция распределения наблюдаемой случайной величины, *монотонная* по параметру θ. Можно положить в качестве центральной статистики функцию $G(\mathbf{X}_n, \theta) = -\sum\limits_{i=1}^{n} \ln F(X_i, \theta)$, которая подчинена гамма-распределению с параметром формы n.

Пример 1.4

Построить точный γ-доверительный интервал по выборке $X_1, \ldots, X_n$ для параметра θ экспоненциального распределения $f(x;\theta) = \dfrac{1}{\theta}\exp\left\{-\dfrac{x}{\theta}\right\}$, $x \geq 0$.

Решение:

Функция распределения $F(x;\theta) = 1 - \exp\left\{-\dfrac{x}{\theta}\right\}$ является монотонной (возрастающей) по параметру θ ($F_\theta^{'}(x;\theta) = \dfrac{x}{\theta^2} e^{-\frac{x}{\theta}} > 0$), следовательно, в качестве центральной статистики можно взять $G(\mathbf{X}_n, \theta) = -\sum\limits_{i=1}^{n} \ln F(X_i, \theta)$, которая подчинена гамма-распределению с функцией плотности $f(x;\alpha, n) = \dfrac{x^{n-1} e^{-x/\alpha}}{\Gamma(n)\alpha^n}$, $x > 0$, $\alpha > 0$, n – объем выборки.

Тогда границы γ-доверительного интервала (T_1, T_2) определяются при численном решении уравнений: $G\left(\mathbf{X}_n, T_1\right) = g_1$, $G\left(\mathbf{X}_n, T_2\right) = g_2$, где g_1 и g_2 выбираются такими, что $P\left\{g_1 < G\left(\mathbf{X}_n, \theta\right) < g_2\right\} = \int\limits_{g_1}^{g_2} \dfrac{x^{n-1} e^{-x/\alpha}}{\Gamma(n)\alpha^n} dx = \gamma$.

1.1.5. Построение асимптотического доверительного интервала

Оценки максимального правдоподобия при достаточно общих условиях являются асимптотически эффективными и асимптотически нормальными, следовательно

$$P\left\{\left|\hat{\theta}_n-\theta\right|\sqrt{ni\left(\hat{\theta}_n\right)}\le c_\gamma\right\}\to\Phi\left(c_\gamma\right)-\Phi\left(-c_\gamma\right)=2\Phi\left(c_\gamma\right)-1=\gamma,$$

где $\Phi\left(x\right)$ – функция распределения стандартного нормального закона, $i(\theta)=\int\limits_{-\infty}^{\infty}\left(\frac{\partial\ln f(x,\theta)}{\partial\theta}\right)^2 f(x,\theta)dx$ – информационное количество Фишера, $\hat{\theta}_n$ – ОМП. Отсюда $c_\gamma=\Phi^{-1}\left(\frac{\gamma+1}{2}\right)$, тогда $\left(\hat{\theta}_n-\frac{c_\gamma}{\sqrt{ni\left(\hat{\theta}_n\right)}};\hat{\theta}_n+\frac{c_\gamma}{\sqrt{ni\left(\hat{\theta}_n\right)}}\right)$ – асимптотически кратчайший γ-доверительный интервал для θ.

Пример 1.5

Пусть $X_1,...,X_n$ – выборка из гамма-распределения $\Gamma(\lambda,\alpha)$ с функцией плотности $f(x)=\frac{1}{\lambda^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\lambda}$, $x\ge0$. Построить асимптотический γ-доверительный интервал для параметра масштаба λ, считая, что $\alpha>0$ – известно.

Решение:

Пусть $\gamma=0.99$.

Оценкой максимального правдоподобия параметра λ при известном параметре формы α имеет вид: $\hat{\lambda}_n=\frac{\overline{X}}{\alpha}$. ОМП параметров гамма-распределения являются асимптотически нормальными, поэтому $\left|\hat{\lambda}_n-\lambda\right|\sqrt{ni\left(\hat{\lambda}_n\right)}$ сходится к стандартному нормальному распределению. По таблице из приложения 2: $c_\gamma=\Phi^{-1}\left(\frac{\gamma+1}{2}\right)=\Phi^{-1}(0.995)=2.57$.

Информационное количество Фишера: $i\left(\hat{\lambda}_n\right)=\frac{\alpha}{\hat{\lambda}_n^2}\cdot$.

Следовательно, случайный интервал $\left[\frac{\overline{X}}{\alpha}\left(1-\frac{2,57}{\sqrt{n\alpha}}\right);\frac{\overline{X}}{\alpha}\left(1+\frac{2,57}{\sqrt{n\alpha}}\right)\right]$ является асимптотическим 99% - доверительным интервалом.

В случае, если модель не является регулярной, и нельзя вычислить информационное количество Фишера, то асимптотический доверительный интервал можно построить на основании центральной предельной теоремы

(ЦПТ). При выполнении условий ЦПТ в качестве *асимптотической* центральной статистики (см. п.1.1.4) может рассматриваться статистика

$$G(\mathbf{X}_n,\theta)=\sqrt{n}\,\frac{\bar{X}_n-M_\theta\xi}{\sqrt{D_\theta\xi}}\succ N(0,1) \text{ , при } n\to\infty$$

или

$$G(\mathbf{X}_n,\theta)=\sqrt{n}\,\frac{\bar{X}_n-M_\theta\xi}{S}\succ N(0,1) \text{ , при } n\to\infty,$$

где S – это стандартное отклонение.

Для того чтобы эти статистики были центральными, необходимо, чтобы выполнялось условие монотонности и непрерывности по параметру θ.

1.2. Свойства оценок параметров

1.2.1. Несмещенность

Статистика $T(\mathbf{X}_n)$ называется *несмещенной* оценкой параметра θ, если выполняется условие: $M[T(\mathbf{X}_n)]=\theta, \forall\theta\in\Theta$.

Несмещенной оценкой с равномерно минимальной дисперсией (НОРМД) называется такая оценка $T^*(\mathbf{X}_n)$, что $D[T^*(\mathbf{X}_n)]\le D[T(\mathbf{X}_n)]$: $\forall\theta\in\Theta$, $\forall T(\mathbf{X}_n): M[T(\mathbf{X}_n)]=\theta$.

1.2.2. Состоятельность

Оценка $T(\mathbf{X}_n)$ некоторой функции $\tau(\theta)$ называется *состоятельной*, если $T(\mathbf{X}_n)\xrightarrow{P}\tau(\theta)$, $\forall\theta\in\Theta$, при $n\to\infty$. То есть $\forall\varepsilon>0$: $P\{|T(\mathbf{X}_n)-\tau(\theta)|>\varepsilon\}\to 0$, при $n\to\infty$.

Свойство состоятельности обязательно для любого правила оценивания, однако оно является асимптотическим и не связано со свойствами оценки при фиксированном объеме выборки (в отличие от свойств несмещенности и минимальной дисперсии).

Критерий состоятельности. Пусть $M_\theta T_n=\tau(\theta)+\varepsilon_n, D_\theta T_n=\delta_n$ и $\varepsilon_n=\varepsilon_n(\theta)\to 0, \delta_n=\delta_n(\theta)\to 0$ при $n\to\infty$. Тогда $T(\mathbf{X}_n)$ – состоятельная оценка функции $\tau=\tau(\theta)$.

1.2.3. Эффективность

Семейство $\{F(x;\theta), \theta\in\Omega\}$ является *регулярным*, если выполняются следующие условия:

1) для любого θ, $\theta\in\Omega$, плотность $f(x;\theta)$ дифференцируема по θ, то есть существует $\frac{\partial}{\partial\theta}f(x,\theta)$;
2) множество $\{x: f(x,\theta)=0\}$ не зависит от θ.

Неравенство Рао-Крамера. Если выполняются условия регулярности, то для любой несмещенной оценки $T(\mathbf{X}_n)$ параметрической функции $\tau(\theta)$ справедливо неравенство:

$$D[T(\mathbf{X}_n)] \geq \frac{[\tau'(\theta)]^2}{ni(\theta)}, \tag{1.2}$$

где $i(\theta)$ – информационное количество Фишера. Оценка, при которой достигается нижняя граница неравенства (1.2), называется *эффективной*.

Критерий эффективности. $T(\mathbf{X}_n)$ – эффективная оценка $\tau(\theta)$, если

$$T(\mathbf{X}_n) - \tau(\theta) = a(\theta)U(\mathbf{X}_n, \theta), \tag{1.3}$$

где $a(\theta)$ – некоторая функция от θ, $U(\mathbf{X}_n, \theta) = \dfrac{\partial \ln L(\mathbf{X}_n, \theta)}{\partial \theta}$.

Пример 1.6

Пусть $X_1, ..., X_n$ – выборка из распределения Максвелла с функцией плотности $f(x;\theta) = \dfrac{2x^2}{\theta^3\sqrt{2\pi}} \exp\left\{-\dfrac{x^2}{2\theta^2}\right\}$, $x > 0, \theta > 0$.

Требуется проверить оценку $\hat{\theta} = \dfrac{\sqrt{2\pi}}{4}\overline{X}$ на несмещенность, состоятельность и эффективность.

<u>Решение:</u>

1. *Несмещенность.*

$$M\left(\frac{\sqrt{2\pi}}{4}\overline{X}\right) = \frac{\sqrt{2\pi}}{4} M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{\sqrt{2\pi}}{4n}\sum_{i=1}^{n} MX_i = \frac{\sqrt{2\pi}}{4n}\sum_{i=1}^{n}\frac{4\theta}{\sqrt{2\pi}} = \theta,$$

$\Rightarrow$ оценка $\hat{\theta} = \dfrac{\sqrt{2\pi}}{4}\overline{X}$ является несмещенной оценкой параметра θ.

2. *Состоятельность.*

Поскольку $\hat{\theta}$ является несмещенной, то нам достаточно исследовать дисперсию оценки $D(\hat{\theta})$.

$$D(\hat{\theta}) = \frac{2\pi}{16n^2}\sum_{i=1}^{n} DX_i = \frac{\pi}{8n}\cdot\frac{3\pi - 8}{\pi}\theta^2 \to 0,\ n \to \infty,$$

$\Rightarrow$ по критерию состоятельности, оценка $\hat{\theta} = \dfrac{\sqrt{2\pi}}{4}\overline{X}$ является состоятельной.

3. *Эффективность.*

Проверим, достигается ли нижняя граница в неравенстве Рао-Крамера. Найдем информационное количество Фишера:

$$i(\theta) = -M\left(\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right) = -M\left(\frac{3}{\theta^2} - \frac{3x^2}{\theta^4}\right) = \frac{3}{\theta^2}\left(\frac{MX^2}{\theta^2} - 1\right) = \frac{3}{\theta^2}\left(\frac{3\theta^2}{\theta^2} - 1\right) = \frac{6}{\theta^2};$$

$$\frac{1}{ni(\theta)} = \frac{\theta^2}{6n} \neq \frac{3\pi - 8}{8n}\theta^2 = D(\hat{\theta}), \Rightarrow \hat{\theta}$$ не является эффективной оценкой θ.

Пример 1.7

Найти функцию $\tau(\theta)$, допускающую эффективную оценку для параметра масштаба распределения Вейбулла:

$$f(x,\theta) = \frac{\alpha x^{\alpha-1}}{\theta^\alpha}\exp\left\{-\left(\frac{x}{\theta}\right)^\alpha\right\},\ x \geq 0,\ \theta > 0,\ \alpha > 0.$$

Решение:

Вероятностная модель является регулярной, так как область определения случайно величины не зависит от параметров и функция плотности дифференцируема по θ. Поэтому можно воспользоваться критерием эффективности (1.3). Логарифмическая функция правдоподобия и её производная имеют вид:

$$\ln L(\mathbf{X}_n,\theta) = \sum_{i=1}^{n}\left[\ln\alpha \cdot X_i^{\alpha-1} - \alpha\ln\theta - \frac{X_i^\alpha}{\theta^\alpha}\right],$$

$$U(\mathbf{X}_n,\theta) = \frac{\partial \ln L(\mathbf{X}_n,\theta)}{\partial\theta} = -\frac{n\alpha}{\theta} + \frac{\alpha}{\theta^{\alpha+1}}\sum_{i=1}^{n} X_i^\alpha .$$

Отсюда

$$\frac{\theta^{\alpha+1}}{n\alpha}U(\mathbf{X}_n,\theta) = \frac{1}{n}\sum_{i=1}^{n} X_i^\alpha - \theta^\alpha .$$

Таким образом, оценка $T(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^{n} X_i^\alpha$ является эффективной оценкой функции $\tau(\theta) = \theta^\alpha$.

1.3. Достаточные статистики

Статистика $T = T(\mathbf{X}_n)$ называется *достаточной* для модели $F = \{F(x,\theta), \theta \in \Theta\}$, если условная плотность (или условная вероятность в дискретном случае) $L(\mathbf{X}_n / t;\theta)$ случайного вектора X_n при условии $T(\mathbf{X}_n) = t$ не зависит от параметра θ.

Критерий факторизации. Для того чтобы статистика $T(\mathbf{X}_n)$ была достаточной для θ, необходимо и достаточно, чтобы функция правдоподобия имела вид

$$L(\mathbf{X}_n,\theta) = g(T(\mathbf{X}_n);\theta)h(\mathbf{X}_n), \tag{1.4}$$

где функция $g(t,\theta)$ зависит от выборки только через $T(\mathbf{X}_n) = t$, а функция $h(\mathbf{X}_n)$ не зависит от θ.

Пример 1.8

Пусть $X_1,...,X_n$ – выборка из распределения Рэлея с функцией плотности $f(x;\theta)=\frac{x}{\theta}\exp\left\{-\frac{x^2}{2\theta^2}\right\}$, $x\ge 0$, $\theta>0$. Найти достаточную статистику для параметра θ.

Решение:

Воспользуемся критерием факторизации (1.4).

$$L(\mathbf{X}_n,\theta)=\prod_{i=1}^{n}\frac{X_i}{\theta^2}\exp\left\{-\frac{X_i^2}{2\theta^2}\right\}=\frac{1}{\left(\theta^2\right)^n}\cdot\prod_{i=1}^{n}X_i\cdot\exp\left\{-\frac{1}{2\theta^2}\sum_{i=1}^{n}X_i^2\right\}.$$

Тогда

$$g(T;\theta)=\frac{1}{\left(\theta^2\right)^n}\cdot\exp\left\{-\frac{T}{2\theta^2}\right\},\quad h(\mathbf{X}_n)=\prod_{i=1}^{n}X_i,$$

где $T=\sum_{i=1}^{n}X_i^2$ – достаточная статистика.

1.4. Варианты заданий

Пусть $X_1, X_2,..., X_n$ – выборка из заданного в соответствии с вариантом закона распределения.

1. Найти числовые характеристики заданной модели:
   a. математическое ожидание (1 балл);
   b. дисперсию (1 балл).
2. Найти точечную оценку неизвестного параметра θ
   a. по методу моментов (2 балла);
   b. по методу максимального правдоподобия (2 балла).
3. Проверить условия регулярности модели. В случае регулярности модели, вычислить информационное количество Фишера $i(\theta)$ (2 балла).
4. Подобрать удобную параметрическую функцию $\tau(\theta)$ для исследования свойств оценок. Записать оценку $\hat{\tau}(\theta)$ на основании любой оценки θ из п.1. Проверить свойства $\hat{\tau}(\theta)$:
   a. несмещенность (2 балла),
   b. состоятельность (2 балла),
   c. эффективность (не используя критерий эффективности) (2 балла).
5. Найти достаточную статистику для заданной модели (2 балла).
6. Найти функцию $\tau(\theta)$, допускающую эффективную оценку (с помощью критерия эффективности) (2 балла).
7. Построить асимптотический доверительный интервал для θ (2 балла).

№ варианта	Закон	Неизвестные параметры	Известные параметры
1.	Биномиальное	p	–
2.	Пуассона	p	–
3.	Двустороннее экспоненциальное	λ	$a=0,\alpha=1$
4.	Двустороннее экспоненциальное	λ	$a=0,\alpha=2$
5.	Двустороннее экспоненциальное	λ	$a=0,\alpha=3$
6.	Двустороннее экспоненциальное	λ	$a=0,\alpha=4$
7.	Двустороннее экспоненциальное	a	$\lambda=1,\alpha=1$
8.	Двустороннее экспоненциальное	a	$\lambda=1,\alpha=2$
9.	Двустороннее экспоненциальное	a	$\lambda=1,\alpha=3$
10.	Двустороннее экспоненциальное	a	$\lambda=1,\alpha=4$
11.	Бета-распределение	λ	$\alpha=1,\beta=1$
12.	Бета-распределение	λ	$\alpha=2,\beta=1$
13.	Бета-распределение	λ	$\alpha=3,\beta=1$
14.	Бета-распределение	λ	$\alpha=1,\beta=2$
15.	Бета-распределение	λ	$\alpha=1,\beta=3$
16.	Бета-распределение	α	$\lambda=1,\beta=1$
17.	Бета-распределение	β	$\lambda=1,\alpha=1$
18.	Гамма-распределение	λ	$\alpha=1$
19.	Гамма-распределение	λ	$\alpha=2$
20.	Гамма-распределение	λ	$\alpha=3$
21.	Гамма-распределение	λ	$\alpha=4$
22.	Вейбулла-Гнеденко	λ	$\alpha=2$
23.	Вейбулла-Гнеденко	λ	$\alpha=3$
24.	Вейбулла-Гнеденко	λ	$\alpha=4$
25.	Полунормальное	λ	–

Часть 2. Проверка статистических гипотез

Статистической гипотезой называется любое утверждение о виде или свойствах распределения наблюдаемых в эксперименте случайных величин (обычно она обозначается H_0 и называется основной).

Проверка статистической гипотезы состоит в том, чтобы сформулировать такое правило, которое позволило бы по результатам соответствующих наблюдений принять или отклонить гипотезу.

Правило, согласно которому гипотеза принимается или отвергается, называется *статистическим критерием* проверки гипотезы.

2.1. Гипотеза о виде распределения

Пусть имеется выборка $X_n = \{x_1,...,x_n\}$ наблюдаемой случайной величины с функцией распределения $F_\xi(x)$.

а) Простой гипотезой является утверждение H_0: $F_\xi(x)=F(x)$, где $F(x)$ полностью задана.

б) Сложной гипотезой является утверждение H_0: $F_\xi(x)\in\{F(x,\theta),\theta\in\Theta\}$.

Для проверки гипотезы о виде распределения используются критерии: Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса (при негруппированных наблюдениях), χ^2 Пирсона, отношения правдоподобия (при группированных наблюдениях).

Пример 2.1

Дана выборка объема $n=30$:

X	1	2	3	4
n_i	15	8	4	3

Требуется проверить гипотезу о согласии данной выборки с законом Пуассона.

<u>Решение:</u>

Зададимся уровнем значимости $\alpha = 0.05$.

Поскольку распределение случайных величин является дискретным, для проверки гипотезы о согласии воспользуемся критерием χ^2 Пирсона.

$$P\{\xi = k\} = \frac{\lambda^k}{k!}e^{-\lambda},\ k = 0,\ 1,\ 2,\ ...$$

ОМП для параметра λ является $\hat{\lambda}_n = \bar{X}$. Для данной выборки $\hat{\lambda}_n = 1.83$. Тогда $P_0 = P\{\xi = 0\} = 0.16$, $P_1 = P\{\xi = 1\} = 0.29$, $P_2 = P\{\xi = 2\} = 0.27$, $P_3 = P\{\xi = 3\} = 0.16$, $P_4 = P\{\xi = 4\} = 0.08$, $P_5 = \sum_{k=5}^{\infty} P\{\xi = k\} = 0.04$, $\sum_{i=0}^{5} P_i = 1$.

Статистика Пирсона:

$$X_n^2 = \sum_{i=0}^{5} \frac{(n_i - nP_i)^2}{nP_i} = \frac{(0-30\cdot 0.16)^2}{30\cdot 0.16} + \frac{(15-30\cdot 0.29)^2}{30\cdot 0.29} + \frac{(8-30\cdot 0.27)^2}{30\cdot 0.27} +$$

$$+ \frac{(4-30\cdot 0.16)^2}{30\cdot 0.16} + \frac{(3-30\cdot 0.08)^2}{30\cdot 0.08} + \frac{(0-30\cdot 0.04)^2}{30\cdot 0.04} = 10.85$$

В случае оценивания по данной выборке m параметров распределения, статистика X_n^2 Пирсона подчиняется χ^2-распределению с $K-m-1$ степенью свободы, где K – число групп. В данном случае число степеней свободы равно $6-1-1=4$. Находим по таблице из приложения 3 критическое значение статистики Пирсона при $\alpha = 0.05$: $S_\alpha = 9.49$. Поскольку $X_n^2 > S_\alpha$, то гипотеза о согласии данной выборки с распределением Пуассона отвергается. Отметим, что если $\alpha = 0.01$, то гипотеза о согласии не отвергается.

Пример 2.2

В следующей таблице представлены результаты измерений длин чайных ложечек в сантиметрах.

9.65	9.05	9.20	9.79	6.69	9.14	9.93	11.95	10.20	10.21
8.58	9.82	11.75	9.05	12.31	10.47	10.10	8.40	10.77	10.19
8.78	10.36	7.30	11.03	12.47	11.06	10.31	7.43	9.87	10.29
9.41	10.37	9.52	10.15	5.36	11.02	8.52	8.34	10.94	9.33
10.01	9.87	9.43	8.27	10.34	9.48	9.61	10.95	10.01	9.86

Требуется проверить гипотезу о согласии данной выборки с распределением Лапласа.

Решение:

Зададимся уровнем значимости $\alpha = 0.05$.

Поскольку мы имеем непрерывную случайную величину, то для проверки гипотезы о согласии воспользуемся критерием типа Колмогорова, статистика которого имеет вид: $S_K = \dfrac{6nD_n + 1}{6\sqrt{n}}$, где $D_n = \max(D_n^+, D_n^-)$, $D_n^+ = \max\limits_{1\le i\le n}\left\{\dfrac{i}{n} - F(x_i,\theta)\right\}$, $D_n^- = \max\limits_{1\le i\le n}\left\{F(x_i,\theta) - \dfrac{i-1}{n}\right\}$. Объем выборки $n=50$, $x_1, x_2, \ldots, x_n$ – упорядоченные по возрастанию выборочные значения, $F(x,\theta)$ – функция распределения Лапласа.

Для нахождения ОМП параметров распределения воспользуемся программной системой ISW 4.0 [10]: $\hat{a}_n = 9.87$, $\hat{\theta}_n = 0.94$.

Вычисляем значение статистики Колмогорова $S_K = 0.55$. Находим по таблице из приложения 5 критическое значение статистики Колмогорова при $\alpha = 0.05$: $S_\alpha = 0.95$. Поскольку $S_K < S_\alpha$, то гипотеза о согласии данной выборки с распределением Лапласа не отвергается.

2.2. Гипотеза однородности

Пусть произведено k серий независимых наблюдений $X_{n_1}=\{x_1^1,...,x_{n_1}^1\}$, $X_{n_2}=\{x_1^2,...,x_{n_2}^2\}$,..., $X_{n_k}=\{x_1^k,...,x_{n_k}^k\}$ и пусть $F_i(x)$ – функция распределения *i*-й серии. Чтобы проверить менялось ли распределение от серии к серии, можно сформулировать гипотезу однородности:

$$H_0 : F_1(x) = F_2(x) = ... = F_k(x),$$

при этом само распределение $F(x)$ может быть неизвестным.

Для проверки гипотезы однородности используется критерий Смирнова (**если выборки негруппированы**) и χ^2 Пирсона (**если выборки группированы**).

Пример

Проверить гипотезу об однородности двух выборок:

X:	3.49	3.5	3.52	3.62	3.79	3.8	3.81	3.99	4.01	4.05
Y:	3.8	3.81	3.83	3.85	3.86	3.9	4.1	4.38	4.66	4.96

Решение:

Так как выборка является негруппированной, то для проверки гипотезы однородности выборок X и Y можно воспользоваться критерием однородности Смирнова. Зададимся уровнем значимости $\alpha = 0.05$.

Статистика критерия однородности Смирнова: $S_{nm} = D_{nm}\sqrt{\dfrac{nm}{n+m}}$, где $D_{nm} = \sup\limits_x \left|F_n^1(x) - F_m^2(x)\right|$ подчиняется распределению Колмогорова $K(S)$. $F_n^1(x)$ – эмпирическая функция распределения по первой выборке, $F_m^2(x)$ – по второй. Проводя вычисления, получаем: $D_{nm} = 0.5$, $n = m = 10$, $S_{nm} = 1.118$. Находим по таблице из приложения 4 критическое значение статистики Смирнова при $\alpha = 0.05$: $S_\alpha = 1.36$. Поскольку $S_{nm} < S_\alpha$, то нет оснований для отклонения гипотезы об однородности выборок X и Y.

2.3. Гипотеза независимости

Пусть дана выборка двумерной случайной величины (X,Y):

$$\left\{(x_1,y_1),(x_2,y_2),...,(x_n,y_n)\right\}$$

Гипотеза независимости X и Y:

$$H_0 : F_{X,Y}(x,y) = F_X(x)F_Y(y).$$

Для проверки гипотезы независимости используется критерий χ^2 Пирсона.

Если исходные данные негруппированы, то предварительно производится группировка наблюдений. Группирование — это такое преобразование выборки, когда область значений случайной величины разбивается на k непересекающихся интервалов граничными точками $t_0 < t_1 < ... < t_{k-1} < t_k$, где t_0 – нижняя граница области определения, t_k – верхняя грань. В соответствии с заданным разбиением подсчитывается число n_i выборочных значений, попавших в i-й интервал. Для дискретной случайной величины вместо интервалов могут быть взяты ее значения.

Пусть случайная величина попадает в интервалы $C_1,...,C_s$, а Y в интервалы $B_1,...,B_k$. Обозначим ν_{ij} количество наблюдений

$$\{(x,y) : x \in C_i, y \in B_j\},\ \sum_{i=1}^{s}\sum_{j=1}^{k}\nu_{ij} = n.$$

Таблица сопряженности X и Y имеет вид:

X	Y			Всего
	B_1	$\cdots$	B_k	
C_1	ν_{11}	$\cdots$	ν_{1k}	$\nu_{1\bullet}$
$\cdots$	$\cdots$		$\cdots$	$\cdots$
C_s	ν_{s1}	$\cdots$	ν_{sk}	$\nu_{s\bullet}$
Всего	$\nu_{\bullet 1}$	$\cdots$	$\nu_{\bullet k}$	n

Статистика критерия независимости χ^2 Пирсона

$$\chi_n^2 = n\sum_{i=1}^{s}\sum_{j=1}^{k}\frac{\left(\nu_{ij} - \nu_{i\bullet}\nu_{\bullet j} / n\right)^2}{\nu_{i\bullet}\nu_{\bullet j}} = n\left(\sum_{i=1}^{s}\sum_{j=1}^{k}\frac{\nu_{ij}^{\ 2}}{\nu_{i\bullet}\nu_{\bullet j}} - 1\right)$$

имеет распределение $\chi^2_{(s-1)(k-1)}$ при $n \to \infty$.

Гипотеза о независимости X и Y отвергается, если $\chi_n^2 > F^{-1}_{\chi^2_{(s-1)(k-1)}}(1-\alpha)$, где α - вероятность ошибки первого рода.

Пример. В следующей таблице представлены значения показателя Y и значения показателя X в течение 12 лет.

Год	Y	X	Год	Y	X
1986	152	170	1992	177	200
1987	159	179	1993	179	207
1988	162	187	1994	184	215
1989	165	189	1995	186	216
1990	170	193	1996	190	220
1991	172	199	1997	191	225

Проверить гипотезу о независимости величин X и Y.

Решение. Для проверки гипотезы независимости воспользуемся критерием независимости χ^2. Зададимся уровнем значимости $\alpha = 0.05$. Составим таблицу сопряженности двух признаков: $i = \overline{1,s}$, $j = \overline{1,k}$:

$X \backslash Y$	(151,161]	(161,171]	(171,181]	(181,191]	$\nu_{i\bullet}$
(165,180]	2	0	0	0	2
(180,195]	0	3	0	0	3
(195,210]	0	0	3	0	3
(210,225]	0	0	0	4	4
$\nu_{\bullet j}$	2	3	3	4	12

Статистика критерия независимости χ^2: $X_n^2 = n\left(\sum_{i,j} \frac{\nu_{ij}^2}{\nu_{i\bullet}\nu_{\bullet j}} - 1\right)$ имеет χ^2-распределение с числом степеней свободы $(s-1)(k-1)$. Вычислим значение статистики: $X_n^2 = 36$, число степеней свободы $(s-1)(k-1) = 9$. Находим по таблице из приложения 6 критическое значение статистики Пирсона при $\alpha = 0.05$: $S_\alpha = F_{\chi_9^2}^{-1}(0,95) = 16.9$. Поскольку $X_n^2 > S_\alpha$, то гипотеза о независимости признаков X и Y отвергается.

2.4. Коэффициент корреляции

Корреляционный анализ служит для выявления зависимостей между статистическими показателями. Выборочный коэффициент линейной корреляции:

$$r_{xy} = \frac{s_{xy}}{s_x s_y},$$

где $s_{xy} = \overline{xy} - \overline{x} \cdot \overline{y}$ - выборочный коэффициент ковариации;

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i\,,\ s_x^2 = \overline{x^2} - (\overline{x})^2,\ s_y^2 = \overline{y^2} - (\overline{y})^2$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i)^2\,,\ \overline{y^2} = \frac{1}{n}\sum_{i=1}^{n}(y_i)^2\,,$$

$s_x = \sqrt{s_x^2}$ и s_y - выборочное стандартное отклонение,

$\overline{y}, \overline{x}$ - выборочные средние.

Гипотеза о незначимости коэффициента линейной корреляции

Пусть

$$H_0 : \rho_{X,Y} = 0\,.$$

Распределение статистики

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

при нормально распределенных X, Y и достаточно больших n описывается распределением Стьюдента t_{n-2}.

Таким образом, H_0 отвергается на уровне значимости α, если

$$|T_{набл}| > \mathrm{t}_{cr}(\alpha, n-2).$$

2.5. Линейная регрессия

Пусть Y - зависимая, X - объясняющая переменные.
Функция регрессии

$$y = f_r(x) \overset{\text{def}}{=} E\left(Y \mid X = x\right).$$

Регрессионная модель – предполагаемый вид регрессионной зависимости, с точностью до неизвестных параметров:

$$y = f_r(x;\beta),$$

где $\beta = \left(\beta_0, \beta_1, ..., \beta_m\right)$ - вектор параметров.

Линейная регрессионная модель:

$$y = \beta_0 + \beta_1 \cdot x.$$

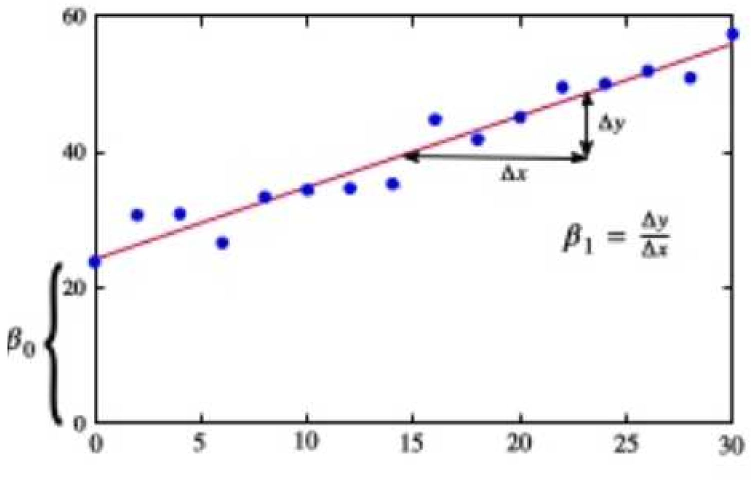
Тогда

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

где ε_i - случайная ошибка, $i = 1, ..., n$.

Оценки параметров по методу наименьших квадратов:

$$\hat{\beta}_0 = \boxed{\overline{y} - \beta_1\,\overline{x}},\ \hat{\beta}_1 = \boxed{\frac{\sum_i y_i x_i - n\,\overline{y}\ \overline{x}}{\sum_i (x_i)^2 - n\,(\overline{x})^2}}$$



β_0 коэффициент пересечения (с осью X=0)

β_1 коэффициент наклона

Анализ качества модели

Пусть $e_i = y_i - \hat{y}_i$ i-й остаток, где

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

прогноз для i-го наблюдения.

Остаточная вариация (residual sum of squares)

$$RSS = \sum_{i=1}^{n} (e_i)^2 ;$$

Стандартная ошибка (несмещенная оценка дисперсии ошибки):

$$s^2 = RSS \,/\, (n - m - 1),$$

где $m = 1$ - число независимых переменных.

Общая вариация

$$TSS = \sum_{i=1}^{n} (y_i - \overline{y})^2 ;$$

Вариация, объясненная регрессией

$$ESS = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 .$$

Коэффициент детерминации

$$R^2 = 1 - RSS/TSS = ESS/TSS ; \quad R^2 \in [0,1]$$

показывает степень подгонки модели к наблюдаемым значениям Y (чем ближе к 1, тем лучше). Для парной линейной модели $R^2 = r_{xy}^2$ (квадрат коэффициента корреляции).

Гипотеза о значимости регрессии:

$$H_0 : \beta_1 \;=\; 0$$

Статистика *F*-критерия:

$$F = \frac{R^2}{1 - R^2} \frac{n - m - 1}{m},$$

где R^2 - коэффициент детерминации, $m = 1$ - число независимых переменных.

H_0 отвергается на уровне значимости α, если

$$F \;>\; F(\alpha; m, n - m - 1),$$

где $F(\alpha; m, n - m - 1)$ определяется из таблицы *F*-распределения.

2.6. Варианты заданий

1. По заданному набору данных одной переменной проверить гипотезу о нормальном законе распределения по критерию Колмогорова (4 балла)

2. Проверить гипотезу об однородности по критерию (3 балла)

3. По заданному набору данных двух переменных:
3.1. Проверить гипотезу о независимости переменных по критерию Хи-квадрат (3 балла)
3.2. Вычислить оценку ковариации, коэффициента корреляции (1 балл). Проверить гипотезу о равенстве коэффициента корреляции нулю (3 балла).
3.3. Оценить параметры линейной регрессии (2 балла), вычислить коэффициент детерминации (1 балл), проверить значимость модели по критерию Фишера (3 балла).

Наборы данных содержит результаты измерений трех видов ирисов (см. приложение 7). Варианты заданий по задачам приведены в таблице.

Номер варианта	Задача 1 Класс, переменная	Задача 2 Классы, переменная	Задача 3 Класс, переменные
1.	Iris Setosa, Длина чашелистика	Iris Setosa, Iris versicolor Длина чашелистика	Iris Setosa, Длина чашелистика, Ширина чашелистика
2.	Iris Setosa, Ширина чашелистика	Iris Setosa, Iris virginica Ширина чашелистика	Iris Setosa, Длина чашелистика, Длина лепестка
3.	Iris Setosa, Длина лепестка	Iris versicolor, Iris virginica Длина лепестка	Iris Setosa, Длина чашелистика, Ширина лепестка
4.	Iris Setosa, Ширина лепестка	Iris Setosa, Iris versicolor Ширина лепестка	Iris Setosa, Ширина чашелистика, Длина лепестка
5.	Iris versicolor, Длина чашелистика	Iris Setosa, Iris virginica Длина чашелистика	Iris Setosa, Ширина чашелистика, Ширина лепестка
6.	Iris versicolor, Ширина	Iris versicolor, Iris virginica	Iris Setosa, Длина лепестка, Ширина

	чашелистика	Ширина чашелистика	лепестка
7.	Iris versicolor, Длина лепестка	Iris Setosa, Iris versicolor Длина лепестка	Iris versicolor, Длина чашелистика, Ширина чашелистика
8.	Iris versicolor, Ширина лепестка	Iris Setosa, Iris virginica Ширина лепестка	Iris versicolor, Длина чашелистика, Длина лепестка
9.	Iris virginica, Длина чашелистика	Iris versicolor, Iris virginica Длина чашелистика	Iris versicolor, Длина чашелистика, Ширина лепестка
10.	Iris virginica, Ширина чашелистика	Iris Setosa, Iris versicolor Ширина чашелистика	Iris versicolor, Ширина чашелистика, Длина лепестка
11.	Iris virginica, Длина лепестка	Iris Setosa, Iris virginica Длина лепестка	Iris versicolor, Ширина чашелистика, Ширина лепестка
12.	Iris virginica, Ширина лепестка	Iris versicolor, Iris virginica Ширина лепестка	Iris versicolor, Длина лепестка, Ширина лепестка
13.	Iris Setosa, Длина чашелистика	Iris Setosa, Iris versicolor Длина чашелистика	Iris virginica, Длина чашелистика, Ширина чашелистика
14.	Iris Setosa, Ширина чашелистика	Iris Setosa, Iris virginica Ширина чашелистика	Iris virginica, Длина чашелистика, Длина лепестка
15.	Iris Setosa, Длина лепестка	Iris versicolor, Iris virginica Длина лепестка	Iris virginica, Длина чашелистика, Ширина лепестка
16.	Iris Setosa, Ширина лепестка	Iris Setosa, Iris versicolor Ширина лепестка	Iris virginica, Ширина чашелистика, Длина лепестка
17.	Iris versicolor, Длина чашелистика	Iris Setosa, Iris virginica Длина чашелистика	Iris virginica, Ширина чашелистика, Ширина лепестка
18.	Iris versicolor, Ширина чашелистика	Iris versicolor, Iris virginica Ширина чашелистика	Iris virginica, Длина лепестка, Ширина лепестка
19.	Iris versicolor, Длина лепестка	Iris Setosa, Iris versicolor Длина лепестка	Iris (все виды), Длина чашелистика, Ширина чашелистика
20.	Iris versicolor, Ширина лепестка	Iris Setosa, Iris virginica Ширина лепестка	Iris virginica, Длина чашелистика, Длина лепестка
21.	Iris virginica, Длина чашелистика	Iris versicolor, Iris virginica Длина чашелистика	Iris (все виды), Длина чашелистика, Ширина лепестка
22.	Iris virginica, Ширина чашелистика	Iris Setosa, Iris versicolor Ширина чашелистика	Iris (все виды), Ширина чашелистика, Длина лепестка
23.	Iris virginica, Длина лепестка	Iris Setosa, Iris virginica Длина лепестка	Iris (все виды), Ширина чашелистика, Ширина лепестка
24.	Iris virginica, Ширина лепестка	Iris versicolor, Iris virginica Ширина лепестка	Iris (все виды), Длина лепестка, Ширина лепестка

Литература

1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
2. Губарев В.В. Вероятностные модели: Справочник. В 2-х ч. /Новосиб.электротехн. ин-т. – Новосибирск, 1992. – Ч.1 – 198 с. Ч.2 – 188 с.
3. Гланц С. Медико-биологическая статистика. – М.: Практика, 1998. – 459 с.
4. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. М.: Высшая школа, 1979– 400 с.
5. Ивченко Г.И., Медведев Ю.А. Математическая статистика: Учеб. пособие для втузов. – М.: Высшая школа, 1994. – 248 с.
6. Ивченко Г.И., Медведев Ю.А., Чистяков А.В. Сборник задач по математической статистике. – М.: Высшая школа, 1989. – 255 с.
7. Коршунов Д.А., Чернова Н.И. Сборник задач и упражнений по математической статистике: учебное пособие. – Новосибирск: Изд-во Института математики, 2001. – 120 с.
8. Лемешко Б.Ю., Постовалов С.Н. Компьютерные технологии анализа данных и исследования статистических закономерностей: Учеб. пособие. – Новосибирск: изд-во НГТУ, 2004. – 120 с.

9. Лемешко Б.Ю. Критерии проверки гипотез об однородности. Руководство по применению : монография / Б.Ю. Лемешко. – М. : ИНФРА-М, 2017. – 208 с. – (Научная мысль). – ISBN: 978-5-16-012557-2 (print), 978-5-16-105463-5(online) DOI: 10.12737/22368

10. Лемешко Б.Ю., Блинов П.Ю. Критерии проверки отклонения распределения от равномерного закона. Руководство по применению: Монография / Б.Ю. Лемешко, П.Ю. Блинов. – М.: ИНФРА-М, 2015. – 183 с. – (Научная мысль). DOI: 10.12737/11304

11. Лемешко Б.Ю. Критерии проверки отклонения распределения от нормального закона. Руководство по применению: Монография / Б.Ю. Лемешко. – М.: ИНФРА-М, 2015. – 160 с. – (Научная мысль). DOI: 10.12737/6086

12. Лемешко Б.Ю. Непараметрические критерии согласия: Руководство по применению: Монография / Б.Ю. Лемешко.– М.: ИНФРА-М, 2014. – 163 с. DOI: 10.12737/11873

13. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: Монография / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов, Е.В. Чимитова. – Новосибирск : Изд-во НГТУ, 2011. – 888 с. (серия «Монографии НГТУ»).

Приложение 1

Основные законы распределения случайных величин

№	Распределение, параметры	Плотность (вероятность) распределения, область значений случайной величины
		Дискретные
1.	Биномиальное, $p \in [0,1]$	$P\{\xi = k\} = C_m^k p^k (1-p)^{m-k}$, $k = 0, \dots, m$
2.	Отрицательное биномиальное, $p \in [0,1]$	$P\{\xi = k\} = C_{m+k-1}^k (1-p)^k p^m$, $k = 0, 1, \dots$
3.	Геометрическое, $p \in [0,1]$	$P\{\xi = k\} = (1-p)^k p$, $k = 0, 1, \dots$
4.	Пуассона, $p \in [0,1]$	$P\{\xi = k\} = \dfrac{p^k}{k!} e^{-p}, k = 0, 1, \dots$
5.	Паскаля, $p \in [0,1]$	$P\{\xi = k\} = \dfrac{p^k}{(1+p)^{k+1}}, k = 0, 1, \dots$
		Непрерывные
6.	Равномерное, $\lambda > 0$	$f(x) = \dfrac{1}{\lambda}$, $x \in [0, \lambda]$
7.	Бета-распределение, $\lambda > 0$, $\alpha > 0$, $\beta > 0$	$f(x) = \dfrac{\Gamma(\alpha+\beta)}{\lambda^{\alpha+\beta-1}\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(\lambda - x)^{\beta-1}$, $x \in [0, \lambda]$
8.	Нормальное, $a \in R$, $\lambda > 0$	$f(x) = \dfrac{1}{\sqrt{2\pi}\lambda} e^{-\frac{(x-a)^2}{2\lambda^2}}, x \in R$
9.	Лапласа, $a \in R$, $\lambda > 0$	$f(x) = \dfrac{1}{2\lambda} e^{-\frac{\lvert x-a \rvert}{\lambda}}, x \in R$
10.	Двустороннее экспоненциальное, $a \in R$, $\lambda > 0$, $\alpha > 0$	$f(x) = \dfrac{\alpha}{2\lambda\Gamma(1/\alpha)} e^{-\left\lvert \frac{x-a}{\lambda} \right\rvert^{\alpha}}, x \in R$
11.	Экспоненциальное, $a \in R$, $\lambda > 0$	$f(x) = \dfrac{1}{\lambda} e^{-(x-a)/\lambda}$, $x \ge a$
12.	Полунормальное, $\lambda > 0$	$f(x) = \dfrac{2}{\lambda\sqrt{2\pi}} e^{-\frac{x^2}{2\lambda^2}}$, $x \ge 0$
13.	Рэлея, $\lambda > 0$	$f(x) = \dfrac{x}{\lambda^2} e^{-\frac{x^2}{2\lambda^2}}$, $x \ge 0$
14.	Максвелла, $\lambda > 0$	$f(x) = \dfrac{2x^2}{\lambda^3\sqrt{2\pi}} e^{-\frac{x^2}{2\lambda^2}}$, $x \ge 0$
15.	Гамма, $\lambda > 0, \alpha > 0$	$f(x) = \dfrac{1}{\lambda^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\lambda}$, $x \ge 0$
16.	Вейбулла-Гнеденко, $\lambda > 0$, $\alpha > 0$	$f(x, \theta) = \dfrac{\alpha x^{\alpha-1}}{\lambda^{\alpha}} \exp\left\{-\left(\dfrac{x}{\lambda}\right)^{\alpha}\right\}$, $x \ge 0$
17.	Логнормальное, $\alpha \in R$, $\beta > 0$	$f(x) = \dfrac{1}{x\beta\sqrt{2\pi}} e^{-(\ln x - \alpha)^2/2\beta^2}$, $x \ge 0$
18.	Парето $\beta > 0$, $\theta > 0$	$f(x) = \beta\theta^{\beta} x^{-(\beta+1)}$, $x \ge \theta$

Приложение 2

Таблица стандартного нормального распределения

В таблице показаны значения функции распределения стандартного нормального закона, например, значение функции в точке 2,57 находится в строке "2,5" и колонке "0,07" и равно 0,9949.

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

t	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,504	0,504	0,508	0,512	0,516	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	0,591	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	0,648	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,67	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,695	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,719	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	0,758	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,791	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	0,834	0,8365	0,8389
1	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	0,877	0,879	0,881	0,883
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,898	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	0,937	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
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	0,975	0,9756	0,9761	0,9767
2	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	0,983	0,9834	0,9838	0,9842	0,9846	0,985	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,989
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	0,992	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,994	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	0,996	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,997	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	0,998	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	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,999	0,999
3,1	0,999	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

Приложение 3

Таблица распределения Стьюдента

В таблице показаны квантили $t_{p,\gamma}$ функции распределения Стьюдента при разных степенях свободы *n*.

$$p = \frac{\gamma+1}{2} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma(n/2)} \int_{-\infty}^{t_{p,n}} \frac{dx}{(x^2/n+1)^{(n+1)/2}}$$

n	$p = 0{,}95$	$p = 0{,}975$	$p = 0{,}99$	$p = 0{,}995$
	$\gamma = 0{,}9$	$\gamma = 0{,}95$	$\gamma = 0{,}98$	$\gamma = 0{,}99$
1	6,3138	12,7062	31,8205	63,6567
2	2,9200	4,3027	6,9646	9,9248
3	2,3534	3,1825	4,5407	5,8409
4	2,1318	2,7765	3,7470	4,6041
5	2,0150	2,5706	3,3649	4,0321
6	1,9432	2,4469	3,1427	3,7074
7	1,8946	2,3646	2,9980	3,4995
8	1,8595	2,3060	2,8965	3,3554
9	1,8331	2,2622	2,8214	3,2498
10	1,8125	2,2281	2,7638	3,1693
11	1,7959	2,2010	2,7181	3,1058
12	1,7823	2,1788	2,6810	3,0545
13	1,7709	2,1604	2,6503	3,0123
14	1,7613	2,1448	2,6245	2,9768
15	1,7531	2,1315	2,6025	2,9467
16	1,7459	2,1199	2,5835	2,9208
17	1,7396	2,1098	2,5669	2,8982
18	1,7341	2,1009	2,5524	2,8784
19	1,7291	2,0930	2,5395	2,8609
20	1,7247	2,0860	2,5280	2,8453
21	1,7207	2,0796	2,5177	2,8314
22	1,7171	2,0739	2,5083	2,8188
23	1,7139	2,0687	2,4999	2,8073
24	1,7109	2,0639	2,4922	2,7969
25	1,7081	2,0595	2,4851	2,7874
26	1,7056	2,0555	2,4786	2,7787
27	1,7033	2,0518	2,4727	2,7707
28	1,7011	2,0484	2,4671	2,7633
29	1,6991	2,0452	2,4620	2,7564
30	1,6973	2,0423	2,4573	2,7500
∞	1,6449	1,9600	2,3264	2,5758

Приложение 4

Таблица распределения χ^2

В таблице показаны квантили $t_{1-\alpha}$ функции распределения χ^2 при разных степенях свободы *n*.

$$1-\alpha = \frac{1}{2^{n/2}\Gamma(n/2)}\int_0^{t_{1-\alpha}} x^{n/2-1}e^{-x/2}dx$$

n	$\alpha = 0,1$	$\alpha = 0,05$	$\alpha = 0,01$
1	2,71	3,84	6,64
2	4,61	5,99	9,21
3	6,25	7,81	11,3
4	7,78	9,49	13,3
5	9,24	11,1	15,1
6	10,6	12,6	16,8
7	12,0	14,1	18,5
8	13,4	15,5	20,1
9	14,7	16,9	21,7
10	16,0	18,3	23,2
11	17,3	19,7	24,7
12	18,5	21,0	26,2
13	19,8	22,4	27,7
14	21,1	23,7	29,1
15	22,3	25,0	30,6
16	23,5	26,3	32,0
17	24,8	27,6	33,4
18	26,0	28,9	34,8
19	27,2	30,1	36,2
20	28,4	31,4	37,6

Приложение 5

Таблица распределения Колмогорова

В таблице показаны квантили $t_{1-\alpha} = K^{-1}(1-\alpha)$ функции распределения Колмогорова.

	$\alpha = 0,1$	$\alpha = 0,05$	$\alpha = 0,01$
$t_{1-\alpha}$	1,2238	1,3581	1,6276

Приложение 6

Таблица распределения статистики Колмогорова при проверке сложных гипотез

В таблице показаны квантили $t_{1-\alpha} = G^{-1}(1-\alpha)$ функции распределения статистики критерия Колмогорова при проверке сложной гипотезы, когда параметры распределения при верной гипотезе H_0 оцениваются по методу максимального правдоподобия.

№	Распределение	$\alpha = 0,1$	$\alpha = 0,05$	$\alpha = 0,01$
1	Экспоненциальное	0,9841	1,0794	1,2838
2	Лапласа	0,8710	0,9497	1,1206
3	Нормальное	0,8333	0,9042	1,0599
4	Логистическое	0,7451	0,8036	0,9261

Приложение 7

Таблица данных для II части

Ирисы Фишера (Fisher, R.A. The Use of Multiple Measurements in Taxonomic Problems // Annals of Eugenics: journal. — 1936. — Vol. 7. — P. 179—188.) состоят из данных о 150 экземплярах ириса, по 50 экземпляров из трёх видов — Ирис щетинистый (Iris setosa), Ирис виргинский (Iris virginica) и Ирис разноцветный (Iris versicolor). Для каждого экземпляра измерялись четыре характеристики (в сантиметрах):

- Длина чашелистика (sepal length);
- Ширина чашелистика (sepal width);
- Длина лепестка (petal length);
- Ширина лепестка (petal width).

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Вид ириса
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Вид ириса
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa
5.4	3.4	1.7	0.2	setosa
5.1	3.7	1.5	0.4	setosa
4.6	3.6	1.0	0.2	setosa
5.1	3.3	1.7	0.5	setosa
4.8	3.4	1.9	0.2	setosa
5.0	3.0	1.6	0.2	setosa
5.0	3.4	1.6	0.4	setosa
5.2	3.5	1.5	0.2	setosa
5.2	3.4	1.4	0.2	setosa
4.7	3.2	1.6	0.2	setosa
4.8	3.1	1.6	0.2	setosa
5.4	3.4	1.5	0.4	setosa
5.2	4.1	1.5	0.1	setosa
5.5	4.2	1.4	0.2	setosa
4.9	3.1	1.5	0.2	setosa
5.0	3.2	1.2	0.2	setosa
5.5	3.5	1.3	0.2	setosa
4.9	3.6	1.4	0.1	setosa
4.4	3.0	1.3	0.2	setosa
5.1	3.4	1.5	0.2	setosa
5.0	3.5	1.3	0.3	setosa
4.5	2.3	1.3	0.3	setosa
4.4	3.2	1.3	0.2	setosa
5.0	3.5	1.6	0.6	setosa
5.1	3.8	1.9	0.4	setosa

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Вид ириса
4.8	3.0	1.4	0.3	setosa
5.1	3.8	1.6	0.2	setosa
4.6	3.2	1.4	0.2	setosa
5.3	3.7	1.5	0.2	setosa
5.0	3.3	1.4	0.2	setosa
7.0	3.2	4.7	1.4	versicolor
6.4	3.2	4.5	1.5	versicolor
6.9	3.1	4.9	1.5	versicolor
5.5	2.3	4.0	1.3	versicolor
6.5	2.8	4.6	1.5	versicolor
5.7	2.8	4.5	1.3	versicolor
6.3	3.3	4.7	1.6	versicolor
4.9	2.4	3.3	1.0	versicolor
6.6	2.9	4.6	1.3	versicolor
5.2	2.7	3.9	1.4	versicolor
5.0	2.0	3.5	1.0	versicolor
5.9	3.0	4.2	1.5	versicolor
6.0	2.2	4.0	1.0	versicolor
6.1	2.9	4.7	1.4	versicolor
5.6	2.9	3.6	1.3	versicolor
6.7	3.1	4.4	1.4	versicolor
5.6	3.0	4.5	1.5	versicolor
5.8	2.7	4.1	1.0	versicolor
6.2	2.2	4.5	1.5	versicolor
5.6	2.5	3.9	1.1	versicolor
5.9	3.2	4.8	1.8	versicolor
6.1	2.8	4.0	1.3	versicolor
6.3	2.5	4.9	1.5	versicolor
6.1	2.8	4.7	1.2	versicolor
6.4	2.9	4.3	1.3	versicolor
6.6	3.0	4.4	1.4	versicolor
6.8	2.8	4.8	1.4	versicolor
6.7	3.0	5.0	1.7	versicolor
6.0	2.9	4.5	1.5	versicolor
5.7	2.6	3.5	1.0	versicolor
5.5	2.4	3.8	1.1	versicolor
5.5	2.4	3.7	1.0	versicolor
5.8	2.7	3.9	1.2	versicolor
6.0	2.7	5.1	1.6	versicolor
5.4	3.0	4.5	1.5	versicolor
6.0	3.4	4.5	1.6	versicolor

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Вид ириса
6.7	3.1	4.7	1.5	versicolor
6.3	2.3	4.4	1.3	versicolor
5.6	3.0	4.1	1.3	versicolor
5.5	2.5	4.0	1.3	versicolor
5.5	2.6	4.4	1.2	versicolor
6.1	3.0	4.6	1.4	versicolor
5.8	2.6	4.0	1.2	versicolor
5.0	2.3	3.3	1.0	versicolor
5.6	2.7	4.2	1.3	versicolor
5.7	3.0	4.2	1.2	versicolor
5.7	2.9	4.2	1.3	versicolor
6.2	2.9	4.3	1.3	versicolor
5.1	2.5	3.0	1.1	versicolor
5.7	2.8	4.1	1.3	versicolor
6.3	3.3	6.0	2.5	virginica
5.8	2.7	5.1	1.9	virginica
7.1	3.0	5.9	2.1	virginica
6.3	2.9	5.6	1.8	virginica
6.5	3.0	5.8	2.2	virginica
7.6	3.0	6.6	2.1	virginica
4.9	2.5	4.5	1.7	virginica
7.3	2.9	6.3	1.8	virginica
6.7	2.5	5.8	1.8	virginica
7.2	3.6	6.1	2.5	virginica
6.5	3.2	5.1	2.0	virginica
6.4	2.7	5.3	1.9	virginica
6.8	3.0	5.5	2.1	virginica
5.7	2.5	5.0	2.0	virginica
5.8	2.8	5.1	2.4	virginica
6.4	3.2	5.3	2.3	virginica
6.5	3.0	5.5	1.8	virginica
7.7	3.8	6.7	2.2	virginica
7.7	2.6	6.9	2.3	virginica
6.0	2.2	5.0	1.5	virginica
6.9	3.2	5.7	2.3	virginica
5.6	2.8	4.9	2.0	virginica
7.7	2.8	6.7	2.0	virginica
6.3	2.7	4.9	1.8	virginica
6.7	3.3	5.7	2.1	virginica
7.2	3.2	6.0	1.8	virginica
6.2	2.8	4.8	1.8	virginica

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Вид ириса
6.1	3.0	4.9	1.8	virginica
6.4	2.8	5.6	2.1	virginica
7.2	3.0	5.8	1.6	virginica
7.4	2.8	6.1	1.9	virginica
7.9	3.8	6.4	2.0	virginica
6.4	2.8	5.6	2.2	virginica
6.3	2.8	5.1	1.5	virginica
6.1	2.6	5.6	1.4	virginica
7.7	3.0	6.1	2.3	virginica
6.3	3.4	5.6	2.4	virginica
6.4	3.1	5.5	1.8	virginica
6.0	3.0	4.8	1.8	virginica
6.9	3.1	5.4	2.1	virginica
6.7	3.1	5.6	2.4	virginica
6.9	3.1	5.1	2.3	virginica
5.8	2.7	5.1	1.9	virginica
6.8	3.2	5.9	2.3	virginica
6.7	3.3	5.7	2.5	virginica
6.7	3.0	5.2	2.3	virginica
6.3	2.5	5.0	1.9	virginica
6.5	3.0	5.2	2.0	virginica
6.2	3.4	5.4	2.3	virginica
5.9	3.0	5.1	1.8	virginica