

№ 4913

51

М 34

# **МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

**Методические указания**

**НОВОСИБИРСК**

**2021**

# МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания  
к расчетно-графическому заданию студентов III курса  
направлений 01.03.02 «Прикладная математика и информатика»  
и 02.03.03 «Математическое обеспечение и администрирование  
информационных систем» ФПМИ

УДК 519.2(07)  
М 34

Составители:

д-р техн. наук, доцент *С.Н. Постовалов*  
д-р техн. наук, доцент *Е.В. Чимитова*

Рецензент

канд. техн. наук *В.С. Карманов*

Методические указания предназначены для студентов, выполняющих расчетно-графическое задание по курсу «Математическая статистика» для студентов III курса направлений 01.03.02 «Прикладная математика и информатика» 02.03.03 «Математическое обеспечение и администрирование информационных систем» ФПМИ. 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.

Работа подготовлена на кафедре теоретической  
и прикладной информатики

© Новосибирский государственный  
технический университет, 2021

## ВВЕДЕНИЕ

Математическая статистика изучает способы построения 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, принадлежащему некоторому известному классу  $F = \{F(x; \theta), \theta \in \Omega\}$ . Цель статистического анализа состоит в том, чтобы указать правдоподобное значение параметра  $\theta$  либо определить подмножество  $\Omega$ , 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.

## 1. ОЦЕНИВАНИЕ ПАРАМЕТРОВ

Пусть имеется выборка  $\mathbf{X}_n = \{X_1, \dots, X_n\}$  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 статистический вывод об истинном значении неизвестного параметра  $\theta$ .

*Точечной* оценкой неизвестного параметра  $\theta$  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  $\theta$ .

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

### 1.1. МЕТОДЫ ОЦЕНИВАНИЯ ПАРАМЕТРОВ

Существует множество различных методов построения оценок для 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.

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

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

#### Пример 1.1

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

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

Требуется найти оценку по методу моментов векторного параметра  $(\alpha, \beta)$ .

*Решение*

Найдем первый и второй теоретические моменты:

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

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

Приравнявая теоретические и выборочные моменты, получим

$$\left\{ \begin{array}{l} \alpha\beta = \frac{1}{n} \sum_{i=1}^n X_i = A_1, \\ \alpha^2\beta(\beta+1) = \frac{1}{n} \sum_{i=1}^n X_i^2 = A_2, \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \hat{\alpha} = \frac{A_2 - A_1^2}{A_1}, \\ \hat{\beta} = \frac{A_1^2}{A_2 - A_1^2}. \end{array} \right.$$

### 1.1.2. МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

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

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

#### Пример 1.2

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

*Решение*

Логарифмическая функция правдоподобия равна

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

Уравнение правдоподобия и ОМП:

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

где  $\bar{X}$  – среднее выборочное значение. Вторая производная

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

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

$$\begin{aligned} \left. \frac{\partial^2 \ln L}{\partial p^2} \right|_{p=\bar{X}} &= -\frac{n\bar{X}}{p^2} - \frac{1}{(1-p)^2} (n - n\bar{X}) = -\frac{n\bar{X}}{\bar{X}^2} - \frac{n}{(1-\bar{X})^2} (1 - \bar{X}) = \\ &= -\frac{n}{\bar{X}} - \frac{n}{(1-\bar{X})} = -\frac{n(1-\bar{X} + \bar{X})}{\bar{X}(1-\bar{X})} = -\frac{n}{\bar{X}(1-\bar{X})} < 0. \end{aligned}$$

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

### Пример 1.3

Построить ОМП параметра  $\theta > 0$  равномерного распределения на отрезке  $[0, \theta]$ .

*Решение*

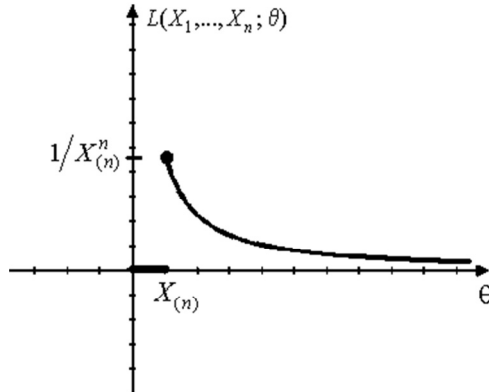
Функция правдоподобия выборки равна

$$\begin{aligned} L(X_1, \dots, X_n; \theta) &= \begin{cases} \theta^{-n}, & \text{если все } X_j \in [0, \theta], \\ 0, & \text{если хотя бы одно } X_j \notin [0, \theta] \end{cases} = \\ &= \begin{cases} \theta^{-n}, & \text{если } X_{(n)} \leq \theta, \\ 0, & \text{если } X_{(n)} > \theta, \end{cases} \end{aligned}$$

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



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



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  $\theta$  и известна ее функция распределения  $F_T(t, \theta)$ , *непрерывная и монотонная* по  $\theta$ , то, опираясь на эту функцию, доверительный интервал можно построить следующим образом.

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

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

3. Определяем границы доверительного интервала:

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

### 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  $\theta$  и при любом фиксированном  $\theta$  статистика  $G(\mathbf{X}_n, \theta)$  непрерывна и строго монотонна по  $\theta$ .

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

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

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

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

3. Определяем границы доверительного интервала:

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

При построении 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  $\theta$ . Можно положить в качестве центральной статистики функцию  $G(\mathbf{X}_n, \theta) = -\sum_{i=1}^n \ln F(X_i, \theta)$ , которая подчиняется гамма-распределению с параметром формы  $n$ .

#### Пример 1.4

Построить точный  $\gamma$ -доверительный интервал по выборке  $X_1, \dots, X_n$  для параметра  $\theta$  экспоненциального распределения

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

*Решение*

Функция распределения  $F(x; \theta) = 1 - \exp\left\{-\frac{x}{\theta}\right\}$  является монотон-

ной (возрастающей) по параметру  $\theta$ , так как  $F'_\theta(x; \theta) = \frac{x}{\theta^2} e^{-\frac{x}{\theta}} > 0$ .

Следовательно, в качестве центральной статистики можно взять

$G(\mathbf{X}_n, \theta) = -\sum_{i=1}^n \ln F(X_i, \theta)$ , которая подчиняется гамма-распределению

с функцией плотности  $f(x; \alpha, n) = \frac{x^{n-1} e^{-x/\alpha}}{\Gamma(n)\alpha^n}$ ,  $x > 0$ ,  $\alpha > 0$ , где  $n$  –

объем выборки.

Тогда границы  $\gamma$ -доверительного интервала  $(T_1, T_2)$  определяются при численном решении уравнений:  $G(\mathbf{X}_n, T_1) = g_1$ ,  $G(\mathbf{X}_n, T_2) = g_2$ ,

где  $g_1$  и  $g_2$  выбираются такими, что  $P\{g_1 < G(\mathbf{X}_n, \theta) < g_2\} =$

$$= \int_{g_1}^{g_2} \frac{x^{n-1} e^{-x/\alpha}}{\Gamma(n)\alpha^n} dx = \gamma.$$

### 1.1.5. ПОСТРОЕНИЕ АСИМПТОТИЧЕСКОГО ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА

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

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

где  $\Phi(x)$  – функция распределения стандартного нормального закона,

$i(\theta) = \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x, \theta)}{\partial \theta}\right)^2 f(x, \theta) dx$  – информационное количество

Фишера,  $\hat{\theta}_n$  – ОМП. Отсюда  $c_\gamma = \Phi^{-1}\left(\frac{\gamma+1}{2}\right)$ , тогда  $\left(\hat{\theta}_n - \frac{c_\gamma}{\sqrt{ni(\hat{\theta}_n)}}; \hat{\theta}_n + \frac{c_\gamma}{\sqrt{ni(\hat{\theta}_n)}}\right)$  – асимптотически кратчайший  $\gamma$ -доверительный интервал для  $\theta$ .

### Пример 1.5

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

*Решение*

Пусть  $\gamma = 0.99$ .

ОМП параметра  $\lambda$  при известном параметре формы  $\alpha$  имеет вид  $\hat{\lambda}_n = \frac{\bar{X}}{\alpha}$ . ОМП параметров гамма-распределения являются асимптотически нормальными, поэтому  $|\hat{\lambda}_n - \lambda| \sqrt{ni(\hat{\lambda}_n)}$  сходится к стандартному нормальному распределению. По таблице из приложения 2 находим:  $c_\gamma = \Phi^{-1}\left(\frac{\gamma+1}{2}\right) = \Phi^{-1}(0,995) = 2,57$ .

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

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

В случае, если модель не является регулярной и нельзя вычислить информационное количество Фишера, то асимптотический довери-

тельный 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

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

или

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

где  $S$  – стандартное отклонение.

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

## 1.2. СВОЙСТВА ОЦЕНОК ПАРАМЕТРОВ

### 1.2.1. НЕСМЕЩЕННОСТЬ

Статистика  $T(\mathbf{X}_n)$  называется *несмещенной* оценкой параметра  $\theta$ , 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  $D[T^*(\mathbf{X}_n)] \leq D[T(\mathbf{X}_n)]: \forall \theta \in \Theta, \forall T(\mathbf{X}_n): M[T(\mathbf{X}_n)] = \theta$ .

### 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  $n \rightarrow \infty$ . То есть  $\forall \varepsilon > 0: P\{|T(\mathbf{X}_n) - \tau(\theta)| > \varepsilon\} \rightarrow 0$  при  $n \rightarrow \infty$ .

Свойство состоятельности обязательно для любого правила оценивания, однако оно является асимптотическим и не связано со свой-

ствами 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  $\varepsilon_n = \varepsilon_n(\theta) \rightarrow 0$ ,  $\delta_n = \delta_n(\theta) \rightarrow 0$  при  $n \rightarrow \infty$ . 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  $\theta, \theta \in \Omega$ , плотность  $f(x; \theta)$  дифференцируема по  $\theta$ , т. е. существует  $\frac{\partial}{\partial \theta} f(x, \theta)$ ;

2) множество  $\{x : f(x, \theta) = 0\}$  не зависит от  $\theta$ .

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

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

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

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

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

#### Пример 1.6

Пусть  $X_1, \dots, X_n$  – выборка из распределения Максвелла с функцией

плотности  $f(x; \theta) = \frac{2x^2}{\theta^3 \sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\theta^2}\right\}$ ,  $x > 0$ ,  $\theta > 0$ .

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

*Решение*

1. *Несмещенность.*

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

оценка  $\hat{\theta} = \frac{\sqrt{2\pi}}{4} \bar{X}$  является несмещенной оценкой параметра  $\theta$ .

2. *Состоятельность.*

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

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

по критерию состоятельности оценка  $\hat{\theta} = \frac{\sqrt{2\pi}}{4} \bar{X}$  является состоятельной.

3. *Эффективность.*

Проверим, достигается ли нижняя граница в неравенстве Рао-Крамера.

Найдем информационное количество Фишера:

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

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

### Пример 1.7

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

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

*Решение*

Вероятностная модель является регулярной, так как область определения случайной величины не зависит от параметров закона и функция плотности дифференцируема по  $\theta$ . Поэтому можно воспользоваться критерием эффективности (1.3). Логарифмическая функция правдоподобия и ее производная имеют вид

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

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

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  $\mathbf{X}_n$  (при условии  $T(\mathbf{X}_n) = t$ ) не зависит от параметра  $\theta$ .



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

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

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

### Пример 1.8

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

*Решение*

Воспользуемся критерием факторизации (1.4):

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

Тогда

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

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

## 1.4. ВАРИАНТЫ ЗАДАНИЙ

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

1. Найти числовые характеристики заданной модели:

- а) математическое ожидание (1 балл);
- б) дисперсию (1 балл).

2. Найти точечную оценку неизвестного параметра  $\theta$ :
- а) по методу моментов (2 балла);
  - б) 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  $\theta$  из п. 1. Проверить свойства  $\hat{\tau}(\theta)$ :
- а) несмещенность (2 балла),
  - б) состоятельность (2 балла),
  - в) 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  $\theta$  (2 балла).

№ варианта	Закон	Неизвестные параметры	Известные параметры
1	Биномиальное	$p$	–
2	Пуассона	$p$	–
3	Двустороннее экспоненциальное	$\lambda$	$a = 0, \alpha = 1$
4	Двустороннее экспоненциальное	$\lambda$	$a = 0, \alpha = 2$
5	Двустороннее экспоненциальное	$\lambda$	$a = 0, \alpha = 3$
6	Двустороннее экспоненциальное	$\lambda$	$a = 0, \alpha = 4$
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	$\lambda$	$\alpha = 1, \beta = 1$
12	Бета-распределение	$\lambda$	$\alpha = 2, \beta = 1$

## Окончание таблицы

№ варианта	Закон	Неизвестные параметры	Известные параметры
13	Бета-распределение	$\lambda$	$\alpha = 3, \beta = 1$
14	Бета-распределение	$\lambda$	$\alpha = 1, \beta = 2$
15	Бета-распределение	$\lambda$	$\alpha = 1, \beta = 3$
16	Бета-распределение	$\alpha$	$\lambda = 1, \beta = 1$
17	Бета-распределение	$\beta$	$\lambda = 1, \alpha = 1$
18	Гамма-распределение	$\lambda$	$\alpha = 1$
19	Гамма-распределение	$\lambda$	$\alpha = 2$
20	Гамма-распределение	$\lambda$	$\alpha = 3$
21	Гамма-распределение	$\lambda$	$\alpha = 4$
22	Вейбулла–Гнеденко	$\lambda$	$\alpha = 2$
23	Вейбулла–Гнеденко	$\lambda$	$\alpha = 3$
24	Вейбулла–Гнеденко	$\lambda$	$\alpha = 4$
25	Полунормальное	$\lambda$	–

## 2. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

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

Для проверки статистической гипотезы необходимо 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  $\mathbf{X}_n = \{X_1, \dots, X_n\}$  наблюдаемой случайной величины с функцией распределения  $F_\xi(x)$ .

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

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

Для проверки гипотезы о виде распределения используются критерии: Колмогорова, Смирнова,  $\omega^2$  Крамера–Мизеса–Смирнова и  $\Omega^2$  Андерсона–Дарлинга (при негруппированных наблюдениях),  $\chi^2$  Пирсона, отношения правдоподобия (при группированных наблюдениях).

### Пример 2.1

Дана выборка объемом  $n = 30$  :

$X_i$	1	2	3	4
$n_i$	15	8	4	3

Требуется проверить гипотезу о согласии данной выборки с законом Пуассона.

*Решение*

Зададимся уровнем значимости  $\alpha = 0,05$ .

Поскольку распределение случайных величин является дискретным, для проверки гипотезы о согласии воспользуемся критерием  $\chi^2$  Пирсона:

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

ОМП для параметра  $\lambda$  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:

$$\begin{aligned} X_n^2 &= \sum_{i=0}^5 \frac{(n_i - nP_i)^2}{nP_i} = \frac{(0 - 30 \cdot 0,16)^2}{30 \cdot 0,16} + \frac{(15 - 30 \cdot 0,29)^2}{30 \cdot 0,29} + \frac{(8 - 30 \cdot 0,27)^2}{30 \cdot 0,27} + \\ &+ \frac{(4 - 30 \cdot 0,16)^2}{30 \cdot 0,16} + \frac{(3 - 30 \cdot 0,08)^2}{30 \cdot 0,08} + \frac{(0 - 30 \cdot 0,04)^2}{30 \cdot 0,04} = 10,85. \end{aligned}$$

В случае оценивания по анализируемой выборке  $m$  параметров закона статистика  $X_n^2$  Пирсона подчиняется  $\chi^2$ -распределению со степенью свободы  $K - m - 1$ , 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  $S_K = \frac{6nD_n + 1}{6\sqrt{n}}$ , где  $D_n =$

$$= \max(D_n^+, D_n^-), \quad D_n^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F(x_i, \theta) \right\}, \quad D_n^- = \max_{1 \leq i \leq n} \left\{ F(x_i, \theta) - \frac{i-1}{n} \right\}.$$

Объем выборки  $n = 50$ ,  $x_1, x_2, \dots, x_n$  – упорядоченные по возрастанию выборочные значения,  $F(x, \theta)$  – функция распределения Лапласа.

Для нахождения ОМП параметров распределения воспользуемся программной системой ISW 5.4 [10]:  $\hat{a}_n = 9,87$ ,  $\hat{\theta}_n = 0,94$ .

Вычисляем значение статистики Колмогорова  $S_K = 0,55$ . Находим по таблице из приложения 6 критическое значение статистики Колмогорова при  $\alpha = 0,05$ :  $S_\alpha = 0,94$ . 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  $\mathbf{X}_{n_1} = \{X_1^1, \dots, X_{n_1}^1\}$ ,  $\mathbf{X}_{n_2} = \{X_1^2, \dots, X_{n_2}^2\}$ , ...,  $\mathbf{X}_{n_k} = \{X_1^k, \dots, X_{n_k}^k\}$  и пусть  $F_i(x)$  – функция распределения  $i$ -й серии. Чтобы проверить, менялось ли распределение от серии к серии, можно сформулировать гипотезу однородности:

$$H_0: F_1(x) = F_2(x) = \dots = F_k(x),$$

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

Для проверки гипотезы однородности используется критерий Смирнова (если выборки негруппированы) и  $\chi^2$  Пирсона (если выборки сгруппированы).

### Пример 2.3

Проверить гипотезу об однородности двух выборок:

$X_i:$	3,49	3,5	3,52	3,62	3,79	3,8	3,81	3,99	4,01	4,05
$Y_i:$	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  $S_{nm} = D_{nm} \sqrt{\frac{nm}{n+m}}$ ,

где  $D_{nm} = \sup_x |F_n^1(x) - F_m^2(x)|$ , 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 приложе-

ния 5 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  $(X, Y)$ :

$$\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}.$$

Гипотеза независимости  $X$  и  $Y$ :

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

Для проверки гипотезы независимости используется критерий  $\chi^2$  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  $t_0 < t_1 < \dots < t_{k-1} < t_k$ , 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  $C_1, \dots, C_s$ , а  $Y$  – в интервалы  $B_1, \dots, B_k$ . Обозначим  $v_{ij}$  количество наблюдений:

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

Таблица сопряженности  $X$  и  $Y$  имеет вид

$X$	$Y$			Всего
	$B_1$	...	$B_k$	
$C_1$	$v_{11}$	...	$v_{1k}$	$v_{1\bullet}$
...	...		...	...
$C_s$	$v_{s1}$	...	$v_{sk}$	$v_{s\bullet}$
Всего	$v_{\bullet 1}$	...	$v_{\bullet k}$	$n$



Статистика критерия независимости  $\chi^2$  Пирсона

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

имеет распределение  $\chi_{(s-1)(k-1)}^2$  при  $n \rightarrow \infty$ .

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

### Пример 2.4

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  $\chi^2$ . Зададимся уровнем значимости  $\alpha = 0,05$ . Составим таблицу сопряженности двух признаков:  $i = \overline{1, s}$ ,  $j = \overline{1, k}$ :

$X$	$Y$				$v_{i \cdot}$
	(151, 161]	(161, 171]	(171, 181]	(181, 191]	
(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
$v_{\cdot j}$	2	3	3	4	12

Статистика критерия независимости  $\chi^2$ :  $X_n^2 = n \left( \sum_{i,j} \frac{v_{ij}^2}{v_{i \cdot} \cdot v_{\cdot j}} - 1 \right)$

имеет  $\chi^2$ -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  $S_{XY} = \overline{XY} - \bar{X} \cdot \bar{Y}$  – выборочный коэффициент ковариации;

$$\overline{XY} = \frac{1}{n} \sum_{i=1}^n X_i Y_i, \quad S_x^2 = \overline{X^2} - (\bar{X})^2, \quad S_y^2 = \overline{Y^2} - (\bar{Y})^2,$$

$$\overline{X^2} = \frac{1}{n} \sum_{i=1}^n (X_i)^2, \quad \overline{Y^2} = \frac{1}{n} \sum_{i=1}^n (Y_i)^2,$$

$S_X = \sqrt{S_X^2}$  и  $S_Y$  – выборочное стандартное отклонение;  $\bar{Y}$ ,  $\bar{X}$  – 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  $\alpha$ , если

$$|T_{\text{набл}}| > t_{cr}(\alpha, n-2) .$$

## 2.5. ЛИНЕЙНАЯ РЕГРЕССИЯ

Пусть  $Y$  – зависимая,  $X$  – объясняющая переменные.

Функция регрессии

$$y = f_r(x) \stackrel{\text{def}}{=} E(Y | X = x) .$$

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

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

где  $\beta = (\beta_0, \beta_1, \dots, \beta_m)$  – вектор параметров.

Линейная регрессионная модель

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

Тогда

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

где  $\varepsilon_i$  – случайная ошибка,  $i = 1, \dots, n$ .

Оценки параметров по методу наименьших квадратов:

$$\hat{\beta}_0 = \boxed{\bar{Y} - \beta_1 \bar{X}}, \quad \hat{\beta}_1 = \frac{\sum_i Y_i X_i - n \bar{Y} \bar{X}}{\sum_i (X_i)^2 - n (\bar{X})^2}$$

### Анализ качества модели

Пусть  $e_i = Y_i - \hat{Y}_i$  –  $i$ -й остаток, где

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

– 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

$$TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Вариация, объясненная регрессией:

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

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 единице, тем лучше). Для парной линейной модели  $R^2 = r_{xy}^2$  (квадрат коэффициента корреляции).

Гипотеза о незначимости 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  $\alpha$ , если

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

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

Если  $H_0$  отвергается, то модель линейной регрессии является значимой.

## 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 выполнить следующее.

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 (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 чашелистика	Iris versicolor, Iris virginica, ширина чашелистика	Iris Setosa, длина лепестка, ширина 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

## Окончание таблицы

Номер варианта	Задача 1 Класс, переменная	Задача 2 Классы, переменная	Задача 3 Класс, переменные
18	<i>Iris versicolor</i> , ширина чашелистика	<i>Iris versicolor</i> , <i>Iris virginica</i> , ширина чашелистика	<i>Iris virginica</i> , длина лепестка, ширина лепестка
19	<i>Iris versicolor</i> , длина лепестка	<i>Iris Setosa</i> , <i>Iris versicolor</i> , длина лепестка	<i>Iris</i> (все виды), длина чашелистика, ширина чашелистика
20	<i>Iris versicolor</i> , ширина лепестка	<i>Iris Setosa</i> , <i>Iris virginica</i> , ширина лепестка	<i>Iris virginica</i> , длина чашелистика, длина лепестка
21	<i>Iris virginica</i> , длина чашелистика	<i>Iris versicolor</i> , <i>Iris virginica</i> , длина чашелистика	<i>Iris</i> (все виды), длина чашелистика, ширина лепестка
22	<i>Iris virginica</i> , ширина чашелистика	<i>Iris Setosa</i> , <i>Iris versicolor</i> , ширина чашелистика	<i>Iris</i> (все виды), ширина чашелистика, длина лепестка
23	<i>Iris virginica</i> , длина лепестка	<i>Iris Setosa</i> , <i>Iris virginica</i> , длина лепестка	<i>Iris</i> (все виды), ширина чашелистика, ширина лепестка
24	<i>Iris virginica</i> , ширина лепестка	<i>Iris versicolor</i> , <i>Iris virginica</i> , ширина лепестка	<i>Iris</i> (все виды), длина лепестка, ширина лепестка

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Большев Л. Н.* Таблицы математической статистики / Л. Н. Большев, Н. В. Смирнов. – Москва : Наука, 1983. – 416 с.
2. *Губарев В. В.* Вероятностные модели: Справочник. В 2 ч. / В. В. Губарев. – Новосибирск, 1992. – Ч. 1. – 198 с. – Ч. 2. – 188 с.
3. *Гланц С.* Медико-биологическая статистика / С. Гланц. – Москва : Практика, 1998. – 459 с.
4. *Гмурман В. Е.* Руководство к решению задач по теории вероятностей и математической статистике / В. Е. Гмурман. – Москва : Высшая школа, 1979. – 400 с.
5. *Ивченко Г. И.* Математическая статистика: учеб. пособие для вузов / Г. И. Ивченко, Ю. И. Медведев. – Москва : Высшая школа, 1994. – 248 с.
6. *Ивченко Г. И.* Сборник задач по математической статистике / Г. И. Ивченко, Ю. И. Медведев, А. В. Чистяков. – Москва : Высшая школа, 1989. – 255 с.
7. *Коршунов Д. А.* Сборник задач и упражнений по математической статистике: учеб. пособие / Д. А. Коршунов, Н. И. Чернова. – Новосибирск : Изд-во Института математики, 2001. – 120 с.
8. *Лемешко Б. Ю.* Компьютерные технологии анализа данных и исследования статистических закономерностей: учеб. пособие / Б. Ю. Лемешко, С. Н. Постовалов. – Новосибирск : Изд-во НГТУ, 2004. – 120 с.
9. 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»).
10. ISW – Программная система статистического анализа одномерных наблюдений. – URL: <https://ami.nstu.ru/~headrd/ISW.htm> (дата обращения: 08.01.2021)
11. *Лемешко Б. Ю.* Непараметрические критерии согласия: Руководство по применению: монография / Б. Ю. Лемешко. – Москва : ИНФРА-М, 2014. – 163 с. DOI: 10.12737/11873.



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

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

14. *Лемешко Б. Ю.* Критерии проверки гипотез об однородности. Руководство по применению: монография / Б. Ю. Лемешко. – Москва : ИНФРА-М, 2017. – 208 с. – (Научная мысль). DOI: 10.12737/22368.

15. *Лемешко Б. Ю.* Критерии проверки гипотез об однородности. Руководство по применению: монография / Б. Ю. Лемешко. – 2-е изд., перераб. и доп. – Москва : ИНФРА-М, 2021. – 248 с. – (Научная мысль). DOI: 10.12737/986695.

16. *Лемешко Б. Ю.* Критерии проверки отклонения от экспоненциального закона. Руководство по применению: монография / Б. Ю. Лемешко, П. Ю. Блинов. – Москва : ИНФРА-М, 2021. – 352 с. – (Научная мысль). DOI: 10.12737/1097477.

# ПРИЛОЖЕНИЯ

## ПРИЛОЖЕНИЕ 1

### Основные законы распределения случайных величин

№ п/п	Распределение, параметры	Плотность (вероятность) распределения, область значений случайной величины
<b>Дискретные</b>		
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, $p \in [0, 1]$	$P\{\xi = k\} = \frac{p^k}{k!} e^{-p}, k = 0, 1, \dots$
5	Паскаля, $p \in [0, 1]$	$P\{\xi = k\} = \frac{p^k}{(1+p)^{k+1}}, k = 0, 1, \dots$
<b>Непрерывные</b>		
6	Равномерное, $\lambda > 0$	$f(x) = \frac{1}{\lambda}, x \in [0, \lambda]$
7	Бета-распределение, $\lambda > 0, \alpha > 0, \beta > 0$	$f(x) = \frac{\Gamma(\alpha+\beta)}{\lambda^{\alpha+\beta-1} \Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (\lambda-x)^{\beta-1}, x \in [0, \lambda]$
8	Нормальное, $a \in R, \lambda > 0$	$f(x) = \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{(x-a)^2}{2\lambda^2}}, x \in R$
9	Лапласа, $a \in R, \lambda > 0$	$f(x) = \frac{1}{2\lambda} e^{-\frac{ x-a }{\lambda}}, x \in R$
10	Двустороннее экспоненциальное, $a \in R, \lambda > 0, \alpha > 0$	$f(x) = \frac{\alpha}{2\lambda\Gamma(1/\alpha)} e^{-\left \frac{x-a}{\lambda}\right ^\alpha}, x \in R$
11	Экспоненциальное, $a \in R, \lambda > 0$	$f(x) = \frac{1}{\lambda} e^{-(x-a)/\lambda}, x \geq a$

Окончание таблицы

№ п/п	Распределение, параметры	Плотность (вероятность) распределения, область значений случайной величины
12	Полунормальное, $\lambda > 0$	$f(x) = \frac{2}{\lambda\sqrt{2\pi}} e^{-\frac{x^2}{2\lambda^2}}, x \geq 0$
13	Рэлея, $\lambda > 0$	$f(x) = \frac{x}{\lambda^2} e^{-\frac{x^2}{2\lambda^2}}, x \geq 0$
14	Максвелла, $\lambda > 0$	$f(x) = \frac{2x^2}{\lambda^3\sqrt{2\pi}} e^{-\frac{x^2}{2\lambda^2}}, x \geq 0$
15	Гамма, $\lambda > 0, \alpha > 0$	$f(x) = \frac{1}{\lambda^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\lambda}, x \geq 0$
16	Вейбулла–Гнеденко, $\lambda > 0, \alpha > 0$	$f(x, \theta) = \frac{\alpha x^{\alpha-1}}{\lambda^\alpha} \exp\left\{-\left(\frac{x}{\lambda}\right)^\alpha\right\}, x \geq 0$
17	Логнормальное, $\alpha \in R, \beta > 0$	$f(x) = \frac{1}{x\beta\sqrt{2\pi}} e^{-(\ln x - \alpha)^2 / 2\beta^2}, x \geq 0$
18	Парето $\beta > 0, \theta > 0$	$f(x) = \beta\theta^\beta x^{-(\beta+1)}, x \geq \theta$

## 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.

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

<b>t</b>	<b>0,00</b>	<b>0,01</b>	<b>0,02</b>	<b>0,03</b>	<b>0,04</b>	<b>0,05</b>	<b>0,06</b>	<b>0,07</b>	<b>0,08</b>	<b>0,09</b>
<b>0,0</b>	0,504	0,504	0,508	0,512	0,516	0,5199	0,5239	0,5279	0,5319	0,5359
<b>0,1</b>	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
<b>0,2</b>	0,5793	0,5832	0,5871	0,591	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
<b>0,3</b>	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,648	0,6517
<b>0,4</b>	0,6554	0,6591	0,6628	0,6664	0,67	0,6736	0,6772	0,6808	0,6844	0,6879
<b>0,5</b>	0,6915	0,695	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,719	0,7224
<b>0,6</b>	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
<b>0,7</b>	0,758	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
<b>0,8</b>	0,7881	0,791	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
<b>0,9</b>	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,834	0,8365	0,8389
<b>1</b>	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
<b>1,1</b>	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,877	0,879	0,881	0,883
<b>1,2</b>	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,898	0,8997	0,9015
<b>1,3</b>	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
<b>1,4</b>	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
<b>1,5</b>	0,9332	0,9345	0,9357	0,937	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
<b>1,6</b>	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545

## Окончание таблицы

<i>t</i>	<b>0,00</b>	<b>0,01</b>	<b>0,02</b>	<b>0,03</b>	<b>0,04</b>	<b>0,05</b>	<b>0,06</b>	<b>0,07</b>	<b>0,08</b>	<b>0,09</b>
<b>1,7</b>	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
<b>1,8</b>	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
<b>1,9</b>	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,975	0,9756	0,9761	0,9767
<b>2</b>	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
<b>2,1</b>	0,9821	0,9826	0,983	0,9834	0,9838	0,9842	0,9846	0,985	0,9854	0,9857
<b>2,2</b>	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,989
<b>2,3</b>	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
<b>2,4</b>	0,9918	0,992	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
<b>2,5</b>	0,9938	0,994	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
<b>2,6</b>	0,9953	0,9955	0,9956	0,9957	0,9959	0,996	0,9961	0,9962	0,9963	0,9964
<b>2,7</b>	0,9965	0,9966	0,9967	0,9968	0,9969	0,997	0,9971	0,9972	0,9973	0,9974
<b>2,8</b>	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,998	0,9981
<b>2,9</b>	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
<b>3</b>	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,999	0,999
<b>3,1</b>	0,999	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
<b>3,2</b>	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
<b>3,3</b>	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  $n$ :

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

$n$	$p = 0,95$	$p = 0,975$	$p = 0,99$	$p = 0,995$
	$\gamma = 0,9$	$\gamma = 0,95$	$\gamma = 0,98$	$\gamma = 0,99$
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

## Окончание таблицы

<i>n</i>	<i>p</i> = 0,95	<i>p</i> = 0,975	<i>p</i> = 0,99	<i>p</i> = 0,995
	$\gamma = 0,9$	$\gamma = 0,95$	$\gamma = 0,98$	$\gamma = 0,99$
<b>19</b>	1,7291	2,0930	2,5395	2,8609
<b>20</b>	1,7247	2,0860	2,5280	2,8453
<b>21</b>	1,7207	2,0796	2,5177	2,8314
<b>22</b>	1,7171	2,0739	2,5083	2,8188
<b>23</b>	1,7139	2,0687	2,4999	2,8073
<b>24</b>	1,7109	2,0639	2,4922	2,7969
<b>25</b>	1,7081	2,0595	2,4851	2,7874
<b>26</b>	1,7056	2,0555	2,4786	2,7787
<b>27</b>	1,7033	2,0518	2,4727	2,7707
<b>28</b>	1,7011	2,0484	2,4671	2,7633
<b>29</b>	1,6991	2,0452	2,4620	2,7564
<b>30</b>	1,6973	2,0423	2,4573	2,7500
$\infty$	1,6449	1,9600	2,3264	2,5758

## ПРИЛОЖЕНИЕ 4

### Таблица распределения $\chi^2$

В таблице показаны квантили  $t_{1-\alpha}$  функции распределения  $\chi^2$  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 квантилей распределения Колмогорова

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

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

## ПРИЛОЖЕНИЕ 6

**Таблица критических значений 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 случайной величины	Оцениваемый параметр	Верхние процентные точки		
		0,9	0,95	0,99
Экспоненциальное	Масштабный	0,995	1,094	1,292
Полунормальное	Масштабный	1,051	1,160	1,381
Рэля	Масштабный	0,995	1,094	1,292
Максвелла	Масштабный	0,969	1,062	1,251
Лапласа	Масштабный	1,177	1,313	1,587
	Сдвиг	0,957	1,045	1,223
	Два параметра	0,863	0,940	1,095
Нормальное	Масштабный	1,190	1,327	1,600
	Сдвиг	0,888	0,963	1,114
	Два параметра	0,835	0,909	1,057
Логнормальное	Масштабный	1,190	1,327	1,600
	Сдвиг	0,888	0,963	1,114
	Два параметра	0,835	0,909	1,057
Коши	Масштабный	1,137	1,275	1,550
	Сдвиг	0,975	1,070	1,260
	Два параметра	0,815	0,893	1,048
Логистическое	Масштабный	1,180	1,316	1,589
	Сдвиг	0,837	0,907	1,046
	Два параметра	0,746	0,805	0,923
Наибольшего значения	Масштабный	1,182	1,316	1,583
	Сдвиг	0,995	1,093	1,292
	Два параметра	0,824	0,895	1,037
Наименьшего значения	Масштабный	1,182	1,316	1,583
	Сдвиг	0,995	1,093	1,292
	Два параметра	0,824	0,895	1,037
Вейбулла	Формы	1,182	1,316	1,583
	Масштабный	0,995	1,093	1,292
	Два параметра	0,824	0,895	1,037

## ПРИЛОЖЕНИЕ 7

### Таблица данных для части II

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 трех 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
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

Продолжение таблицы

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Вид ириса
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
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

Продолжение таблицы

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Вид ириса
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

Продолжение таблицы

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

## Окончание таблицы

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Вид ириса
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
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

## СОДЕРЖАНИЕ

Введение .....	3
1. Оценивание параметров .....	4
1.1. Методы оценивания параметров.....	4
1.1.1. Метод моментов .....	5
1.1.2. Метод максимального правдоподобия .....	6
1.1.3. Построение доверительного интервала с использованием распределения точечной оценки параметров .....	8
1.1.4. Построение доверительного интервала с использованием центральной статистики .....	9
1.1.5. Построение асимптотического доверительного интервала .....	10
1.2. Свойства оценок параметров .....	12
1.2.1. Несмещенность .....	12
1.2.2. Состоятельность .....	12
1.2.3. Эффективность .....	13
1.3. Достаточные статистики.....	15
1.4. Варианты заданий .....	16
2. Проверка статистических гипотез.....	19
2.1. Гипотеза о виде распределения.....	19
2.2. Гипотеза однородности .....	22
2.3. Гипотеза независимости .....	23
2.4. Коэффициент корреляции .....	25
2.5. Линейная регрессия.....	26
2.6. Варианты заданий .....	28
Библиографический список .....	31
Приложения.....	33
Приложение 1. Основные законы распределения случайных величин .....	33
Приложение 2. Таблица стандартного нормального распределения .....	35
Приложение 3. Таблица распределения Стьюдента .....	37
Приложение 4. Таблица распределения $\chi^2$ .....	39
Приложение 5. Таблица квантилей распределения Колмогорова.....	40
Приложение 6. Таблица критических значений статистики Колмогорова при проверке сложных гипотез .....	41
Приложение 7. Таблица данных для части II .....	42



# МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

## Методические указания

Выпускающий редактор *И.П. Брованова*  
Корректор *Л.Н. Кишит*  
Компьютерная верстка *Н.В. Гаврилова*

Налоговая льгота – Общероссийский классификатор продукции  
Издание соответствует коду 95 3000 ОК 005-93 (ОКП)

---

Подписано в печать 22.03.2021 . Формат 60 × 84 1/16. Бумага офсетная  
Тираж 50 экз. Уч.-изд. л. 2,79. Печ. л. 3,0. Изд. № 273/20. Заказ № 337  
Цена договорная

---

Отпечатано в типографии  
Новосибирского государственного технического университета  
630073, г. Новосибирск, пр. К. Маркса, 20