В.Л. БОРЩ, Ю.П. СОВИТ

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В БИОЛОГИИ В ПРИМЕРАХ И ЗАДАЧАХ

(для студентов-иностранцев биолого-медицинского института)

1998

Міністерство освіти Україны
Дніпропетровський державний університет

В.Л. Борщ, Ю.П. Совіт

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В БИОЛОГИИ В ПРИМЕРАХ И ЗАДАЧАХ

(для студентов-иностранцев биолого-медицинского института)

Затверджено редакційно-видавничою радою університету як навчальний посібник

Дніпропетровськ
ДДУ

1998
УДК 519.22/.25 (075.8)
Б82
Рецензенты: д-р фіз.-мат. наук, проф. М.В. Поляков
д-р фіз.-мат. наук, проф. С.Б. Вакарчук

Для студентов-иностранцев биолого-медицинского института университета, изучающих курс “Математические методы в биологии” и выполняющих курсовое задание.

Борщ В.Л., Совіт Ю.П.
Б82
Математичні методи в біології в прикладах і задачах (для студентів-іноземців біолого-медичного інституту). − Дніпропетровськ: ДДУ, 1998. − 80 с. − Рос. мовою.

Для студентів-іноземців біолого-медичного інституту університету, які вивчають курс “Математичні методи в біології” і виконують курсове завдання.

© Борщ В.Л., Совит Ю.П., 1998

ПРЕДИСЛОВИЕ

Учебное пособие предназначено для студентов-иностранцев, обучающихся в биолого-медицинском институте ДГУ и изучающих курс “Математические методы в биологии”.

Пособие состоит из трех частей. В первой части (разделы 1-5) приведены необходимые теоретические сведения (теоретический минимум), алгоритмы и примеры решения типовых задач, входящих в курсовое задание. Подобранные задачи соответствует основным темам изучаемого курса (оценка параметров непрерывного распределения, сравнение параметров независимых и зависимых распределений, корреляция и регрессия).

Во второй части (раздел 6) приведены 13 вариантов условий задач по каждой теме. При необходимости число вариантов может быть увеличено за счет [2].

В третьей части (раздел 7) дан необходимый справочный материал, оформленный в виде таблиц.

Решение каждой задачи должно сопровождаться соответствующими таблицами и графическими иллюстрациями (гистограммы, графики, корреляционные поля), а также медико-биологическим обоснованием полученных результатов.

Учебное пособие ориентировано на самостоятельную работу студентов, поэтому может использоваться и студентами-заочниками.

1. ТОЧЕЧНАЯ И ИНТЕРВАЛЬНАЯ ОЦЕНКА ЧИСЛОВЫХ ХАРАКТЕРИСТИК РАСПРЕДЕЛЕНИЯ, ЭМПИРИЧЕСКИЕ ФУНКЦИЯ И ЗАКОН РАСПРЕДЕЛЕНИЯ

1.1. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

1.1.1. Точечная оценка параметров. Для любого параметра γ изучаемого распределения $f(x;\gamma)$ случайной величины X по выборке может быть вычислена точечная оценка − статистика g. Так как состав выборки случаен, то можно сказать лишь, что g примет заранее неизвестное числовое значение, изменяющееся от выборки к выборке; то есть статистика является случайной величиной, определенной на выборке объемом n:

$$\tilde{g} = G[\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_i, \ldots, \tilde{x}_n] \qquad (1.1)$$

со своим законом распределения и параметрами (например, математическим ожиданием $\mu_{\tilde{g}} = M[\tilde{g}]$ и дисперсией $\sigma^2_{\tilde{g}} = D[\tilde{g}]$).

Выбор оценки, позволяющей получить хорошее приближение оцениваемого параметра, и составляет задачу точечного оценивания. ***Основными свойствами оценок*** являются ***несмещенность***, ***эффективность*** и ***состоятельность***.

Оценка g ***параметра*** γ называется ***несмещенной***, если ее математическое ожидание равно оцениваемому параметру γ, то есть $M[\tilde{g}] = \gamma$. Если это равенство не выполняется, то оценка g может либо завышать значение γ (то есть $M[\tilde{g}] > \gamma$), либо занижать его ($M[\tilde{g}] < \gamma$), что приводит к систематическим (одного знака) ошибкам в оценке параметра. Требование несмещенности гарантирует отсутствие систематических ошибок при оценке параметра.

Так как $\tilde{g}$ − случайная величина, то меру ее рассеивания около математического ожидания $M[\tilde{g}]$ характеризует дисперсия $D[\tilde{g}]$. Пусть $\tilde{g}_1$ и $\tilde{g}_2$ − две несмещенные оценки параметра γ, то есть $M[\tilde{g}_1] = \gamma$ и $M[\tilde{g}_2] = \gamma$, дисперсии которых равны соответственно $D[\tilde{g}_1]$ и $D[\tilde{g}_2]$. Из двух оценок $\tilde{g}_1$ и $\tilde{g}_2$ следует отдать предпочтение той, которая обладает меньшим рассеиванием около оцениваемого параметра, так как в этом случае значение оценки, вычисленное по конкретной выборке, меньше отличается от истинного значения оцениваемого параметра. Если $D[\tilde{g}_1] < D[\tilde{g}_2]$, то в качестве оценки γ принимают $\tilde{g}_1$. Несмещенная оценка $\tilde{g}$, которая имеет наименьшую дисперсию среди всех возможных несмещенных оценок па-

раметра, вычисленных по выборкам одного и того же объема, называется ***эффективной оценкой***.

Оценка $\tilde{g}$ параметра γ называется ***состоятельной***, если при достаточно большом числе независимых наблюдений n с вероятностью, близкой к единице, разность между $\tilde{g}$ и γ по абсолютной величине окажется меньше сколь угодно малого положительного числа ε, или

$$P\left(\left\{\left|\tilde{g}-\gamma\right|<\varepsilon\right\}\right)>1-\eta, \qquad (1.2)$$

где η − положительное число, близкое к нулю. Следовательно, в случае использования состоятельных оценок оправдывается увеличение объема выборки, так как при этом уменьшается вероятность возникновения значительной ошибки в оценке неизвестного параметра.

Для двух наиболее часто используемых параметров − математического ожидания μ и дисперсии σ^2 величины X − формулы для подсчета состоятельных и несмещенных точечных оценок m, $\tilde{s}^2$ по выборке $x_1, x_2, \ldots, x_n$ имеют следующий вид:

$$m \equiv m_X = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad \tilde{s}^2 \equiv \tilde{s}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_X)^2. \qquad (1.3)$$

1.1.2. Интервальная оценка параметров. ***Доверительным интервалом*** $I_\beta(a, b)$ неизвестного параметра γ называется такой интервал, который с вероятностью β включает неизвестное значение параметра, то есть

$$P\left(\left\{\gamma \in I_\beta(a, b)\right\}\right) = P\left(\left\{a < \gamma < b\right\}\right) = \beta. \qquad (1.4)$$

Вероятность противоположного события $\left\{\gamma \notin I_\beta\right\}$ обозначается α=1− β и называется ***уровнем значимости***.

Обычно выбираются доверительные вероятности β=0,9; 0,95; 0,99 и т.д. Например, если β=0,99, то $I_{0.99}(a, b)$ − такой интервал, вероятность попадания в который значения неизвестного параметра равна 0,99.

Вычисление доверительного интервала математического ожидания μ ***нормального распределения***. Если среднее квадратическое отклонение σ нормального распределения ***известно***, то для уровня значимости α доверительный интервал параметра μ вычисляется по формуле:

$$m - \frac{\sigma}{\sqrt{n}}\, z_{1-\alpha/2} < \mu < m + \frac{\sigma}{\sqrt{n}}\, z_{1-\alpha/2}\,, \qquad (1.5)$$

где $z_{1-\alpha/2}$ – квантиль стандартизованного нормального распределения $\tilde{z} \sim N(0,1)$ (см. 7.1); по определению $P\left(\left\{\tilde{z} < z_{1-\alpha/2}\right\}\right) = 1 - \alpha/2$.

Если среднее квадратическое отклонение σ нормального распределения ***не известно***, то для уровня значимости α доверительный интервал параметра μ вычисляется так:

$$m - \frac{\hat{s}}{\sqrt{n}} t_{n-1,1-\alpha/2} < \mu < m + \frac{\hat{s}}{\sqrt{n}} t_{n-1,1-\alpha/2}\,, \qquad (1.6)$$

где $\hat{s}/\sqrt{n} \equiv S_m$ – ***стандартная ошибка выборочного среднего***, $t_{n-1,\,1-\alpha/2}$ – квантиль распределения Стьюдента $\tilde{t}_\nu$ с числом степеней свободы ν=n-1 (см. 7.4); по определению $P\left(\left\{\tilde{t}_\nu < t_{\nu,1-\alpha/2}\right\}\right) = 1 - \alpha/2$.

Приведенный способ интервального оценивания математического ожидания μ является ***робастным***, то есть он ***не чувствителен*** к умеренным отклонениям изучаемого распределения от нормального закона.

Доверительный интервал σ^2 ***дисперсии нормального распределения*** вычисляется по формуле

$$\frac{\hat{s}^2(n-1)}{\chi^2_{n-1,\,1-\alpha/2}} < \sigma^2 < \frac{\hat{s}^2(n-1)}{\chi^2_{n-1,\,\alpha/2}}, \qquad (1.7)$$

где – $\chi^2_{n-1,\,\alpha/2}$, $\chi^2_{n-1,\,1-\alpha/2}$ – квантили распределения Пирсона $\tilde{\chi}^2_\nu$ с числом степеней свободы ν=n-1 (см.7.2); по определению $P\left(\left\{\tilde{\chi}^2_\nu < \chi^2_{\nu,\,\alpha/2}\right\}\right) = \alpha/2$, $P\left(\left\{\tilde{\chi}^2_\nu < \chi^2_{\nu,\,1-\alpha/2}\right\}\right) = 1 - \alpha/2$.

Приведенный способ интервального оценивания дисперсии σ^2 ***очень чувствителен*** к умеренным отклонениям изучаемого распределения от нормального закона.

1.1.3. Проверка нормальности закона распределения случайной величины. Построение доверительного интервала (а также другие статистические выводы, например о коэффициенте корреляции) основывается на том, что изучаемая случайная величина распределена по нормальному закону. Поэтому проверка гипотезы H_0 о нормальности распределения предваряет статистический анализ. Наводящие соображения

о законе распределения может дать эмпирическая функция распределения и (или) функция плотности, а также гистограммы частот и (или) накопленных частот. Подсчет числа интервалов при построении эмпирических функций проводится по одной из следующих формул (см. 7.5):

$$K_1 = 1 + 3{,}32 \lg n = 1 + 1{,}44 \ln n, \tag{1.8}$$

$$K_2 = 5 \lg n, \tag{1.9}$$

$$K_3 = \sqrt{n}. \tag{1.10}$$

Ширина интервала группировки, обозначаемая Δa, вычисляется по формуле

$$\Delta a \approx \Delta X / K, \tag{1.11}$$

где $\Delta X = x_{max} - x_{min}$ – размах выборки. После этого строятся интервалы, то есть записываются левая и правая границы $[a_k, a_{k+1})$ для каждого из них, начиная с первого (k=1):

$$a_{k+1} = a_k + \Delta a, \quad k = \overline{1, K}, \quad a_1 \le x_{min}, \quad a_{K+1} > x_{max}. \tag{1.12}$$

Ширина интервала Δa обычно округляется с той точностью, с которой получены выборочные данные $x_1, x_2, \ldots, x_n$.

После определения границ интервалов по ним распределяются выборочные значения и формируются таблицы частот, относительных частот или выборочной (эмпирической) плотности (табл.1.1), что позволяет построить гистограммы трех типов:

а) гистограмма частот;

б) гистограмма относительных частот;

в) гистограмма выборочной плотности.

Таблица 1.1

Интервал	$[a_1, a_2)$	$[a_2, a_3)$	...	$[a_k, a_{k+1})$	...	$[a_K, a_{K+1})$
Частота	n_1	n_2	...	n_k	...	n_K
Относительная частота	$p_1^* = \frac{n_1}{n}$	$p_2^* = \frac{n_2}{n}$	...	$p_k^* = \frac{n_k}{n}$	...	$p_K^* = \frac{n_K}{n}$
Выборочная плотность	$\frac{p_1^*}{\Delta a}$	$\frac{p_2^*}{\Delta a}$	...	$\frac{p_k^*}{\Delta a}$	...	$\frac{p_K^*}{\Delta a}$

По полученным данным находится ***выборочная*** (***эмпирическая***) ***функция распределения*** $F^*(x)$

$$F^*(x) = \begin{cases} F_0^* = 0, & x < a_1, \\ F_1^* = p_1^*, & a_1 \le x < a_2, \\ F_2^* = F_1^* + p_2^* = p_1^* + p_2^*, & a_2 \le x < a_3, \\ \dots & \dots \\ F_k^* = F_{k-1}^* + p_k^* = \sum_{j=1}^{k} p_j^*, & a_k \le x < a_{k+1}, \\ \dots & \dots \\ F_K^* = F_{K-1}^* + p_K^* = \sum_{j=1}^{K} p_j^* = 1, & a_K \le x < a_{K+1}, \\ F_{K+1}^* \equiv 1, & a_{K+1} \le x, \end{cases} \tag{1.13}$$

которая позволяет сформулировать гипотезу $H_0 : \{X \approx F(x)\}$ или

H_0: { случайная величина X имеет функцию распределения $F(x)$},

а отклонения от функции распределения $F(x)$ в сторону эмпирической функции распределения $F^*(x)$ обусловлены случайными причинами, другими словами, методологией выборочного анализа.

Для решения вопроса о состоятельности гипотезы H_0 используются различные критерии проверки гипотезы, из которых наиболее часто применяется ***критерий Пирсона*** χ^2.

Для построения критерия проверки гипотезы необходимо подсчитать теоретические вероятности p_k попадания значений случайной величины в интервалы группировки $[a_k, a_{k+1})$. Они могут быть вычислены точно

$$p_k = \int_{a_k}^{a_{k+1}} f(x)\,dx = F(a_{k+1}) - F(a_k) \tag{1.14}$$

или приближенно (см.рис. 1.1)

$$\begin{aligned} p_k &\approx f[(a_k + a_{k+1})/2] \times (a_{k+1} - a_k), \\ p_k &\approx 1/2[f(a_k) + f(a_{k+1})] \times (a_{k+1} - a_k). \end{aligned} \tag{1.15}$$

По полученным данным построим табл.1.2, во второй строке которой поместим относительные частоты (выборочные данные), а в третьей − теоретические вероятности, вычисленные в предположении о справедливости гипотезы H_0.

Таблица 1.2

Интервал	$[a_1, a_2)$	$[a_2, a_3)$	...	$[a_k, a_{k+1})$	...	$[a_K, a_{K+1})$
Относительная частота	p_1^*	p_2^*	...	p_k^*	...	p_K^*
Теоретическая вероятность	p_1	p_2	...	p_k	...	p_K

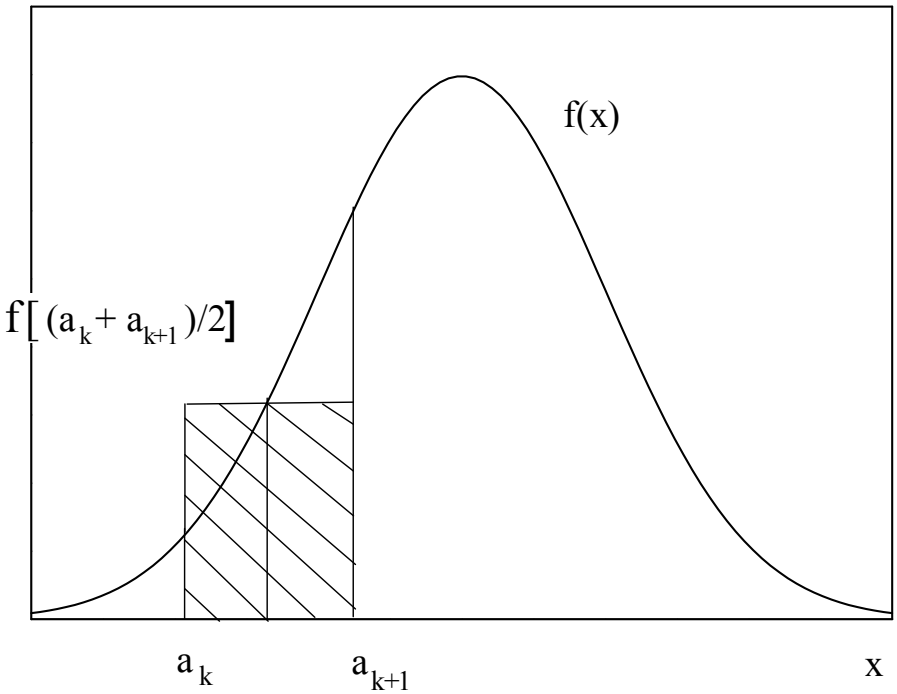

Рис.1.1 Интервалы группировки для вычисления p_k

Если плотность вероятности $f(x)$ и функция распределения $F(x)$ определены на всей числовой оси (например, в случае гипотезы H_0: {случайная величина X распределена по нормальному закону}), необходимо ввести нулевые интервалы $(-\infty, a_1)$ и $[a_{K+1}, +\infty)$, для которых вероятности p_0 и p_{K+1} вычисляются следующим образом:

$$p_0 = \int_{-\infty}^{a_1} f(x)\,dx = F(a_1) - F(-\infty) = F(a_1), \qquad (1.16)$$

$$p_{K+1} = \int_{a_{K+1}}^{+\infty} f(x)dx = F(+\infty) - F(a_{K+1}) = 1 - F(a_{K+1}). \qquad (1.17)$$

Используя данные, построим табл.1.3.

Таблица 1.3

Интервал	$(-\infty, a_1)$	$[a_1, a_2)$	$[a_2, a_3)$	...	$[a_k, a_{k+1})$	...	$[a_K, a_{K+1})$	$[a_{K+1}, +\infty)$
Относительная частота	0	p_1^*	p_2^*	...	p_k^*	...	p_K^*	0
Теоретическая вероятность	p_0	p_1	p_2	...	p_k	...	p_K	p_{K+1}

В критерии Пирсона χ^2 статистикой служит величина

$$\chi^2 = n\sum_{k=0}^{K+1}\frac{(p_k - p_k^*)^2}{p_k} = \sum_{k=0}^{K+1}\frac{(np_k - n_k)^2}{np_k}, \qquad (1.18)$$

распределенная по закону χ_ν^2 с числом степеней свободы ν=K+2-q, где q-число независимых условий налагаемых на параметры теоретического распределения. Если выборочное значение статистики χ_*^2 попадает в ***одностороннюю критическую область*** S={$\chi_*^2 \ge \chi_{\nu,1-\alpha}^2$}, то ***гипотеза*** H_0 ***отвергается***, в противном случае ***гипотеза*** H_0 **принимается**. Так как слишком малые значения χ_*^2 не обязательно возникают вследствие хорошего совпадения с теоретическим распределением, то ***критическая область*** может выбираться ***двусторонней*** S={$\chi_\nu^2 \le \chi_{\nu,\alpha/2}^2$}+ {$\chi_\nu^2 \ge \chi_{\nu,1-\alpha/2}^2$}.

В аналитической записи формул почти всех законов распределения участвуют числовые параметры. Так, в случае нормального закона это математическое ожидание μ и дисперсия σ^2. Поскольку они не известны для генеральной совокупности, то при вычислении величин p_k используются их выборочные оценки, а "исходное" число степеней свободы

ν=K+2 следует уменьшить на q=3, т.е. ν=K+2-q=K-1 (две единицы − за счет согласования выборочных данных со значениями m и $\hat{s}^2$ и еще одна единица − за счет выполнения условия по частотам $n_1 + n_2 + \ldots + n_K = n$).

1.2. АЛГОРИТМ РЕШЕНИЯ

1.2.1. Построение гистограммы и вычисление точечных оценок параметров распределения. Для этого необходимо выполнить следующие действия:

1) определить лимиты x_{min}, x_{max} и размах вариационного ряда ΔX;

2) по одной из зависимостей (1.8) − (1.10) вычислить число интервалов группировки K;

3) определить по (1.11) ширину интервала Δa и по (1.12) границы интервалов $[a_k, a_{k+1})$, k=1,...,K;

4) добавить два нулевых интервала (-∞, a_1), $[a_K, \infty)$;

5) составить таблицу частот n_k, относительных частот p_k^*, значений эмпирической функции плотности f_k^* и функции распределения F_k^* (см.(1.13)) на интервалах;

6) построить гистограмму частот или эмпирических функций плотности $f^*(x)$ и распределения $F^*(x)$;

7) по (1.3) вычислить математическое ожидание m, выборочное среднеквадратическое отклонение $\hat{s}$ и выборочную дисперсию $\hat{s}^2$.

1.2.2. Проверка гипотезы H_0: $F(x) = F_0(x;\mu,\sigma)$ о нормальности закона распределения с помощью критерия Пирсона χ^2. Выполняется так:

1) задать уровень значимости α;

2) по найденным границам интервалов a_k вычислить теоретические вероятности попадания значений X в каждый интервал

$$p_k = P(\{a_k \le x < a_{k+1}\}) = \Phi((a_{k+1} - m)/\hat{s}) - \Phi((a_k - m)/\hat{s}),$$

k=0, ..., K+1, где $\Phi(z)$ − функция Лапласа (см. табл.7.1);

3) определить теоретические частоты, соответствующие вероятностям p_k: $n_k^0 = n\,p_k$, k=0, ..., K+1;

4) вычислить по (1.18) выборочное значение статистики критерия Пирсона χ_*^2;

5) по табл. 7.2 для заданного уровня значимости α и числа степеней свободы $\nu = K - 1$ найти квантиль $\chi^2_{\nu,1-\alpha}$;

6) если выполняется условие $\chi_*^2 < \chi_{\nu,1-\alpha}^2$, то принять гипотезу $H_0: F(x) = F_0(x;\mu,\sigma)$;

7) в случае принятия гипотезы H_0 на график эмпирической функции плотности $f^*(x)$ нанести значения функции распределения $f(x) = \frac{1}{\sqrt{2\pi}\hat{s}}\exp\left[-\frac{(x-m)^2}{2\hat{s}^2}\right]$, вычисленные в серединах интервалов $[a_k, a_{k+1})$, k=1, ..., K, и соединить их отрезками прямых.

1.2.3. Интервальная оценка параметров распределения. Выполняется следующим способом:

1) для заданного значения уровня значимости α и числа степеней свободы $\nu = n-1$ по табл. 7.4 определить квантиль распределения Стьюдента $t_{\nu,1-\alpha/2}$, после чего по формуле (1.6) выполнить доверительное оценивание математического ожидания μ генеральной совокупности;

2) для заданного значения уровня значимости α и числа степеней свободы $\nu = n-1$ по табл. 7.2 определить квантили распределения Пирсона $\chi_{\nu,\alpha/2}^2, \chi_{\nu,1-\alpha/2}^2$, после чего по формуле (1.7) выполнить доверительное оценивание дисперсии σ^2 генеральной совокупности.

1.3. ПРИМЕР РАСЧЕТА

Выборка объемом n=115 представляет рост взрослых мужчин, см:
162 151 161 170 167 164 166 164 173 172 165 153 164 169 170 154 163
159 161 167 168 164 170 166 176 157 159 158 160 161 167 155 166 167
173 165 175 165 174 167 170 169 159 159 160 156 161 162 161 181 159
169 160 169 161 161 166 164 170 180 158 167 169 165 166 172 168 171
178 178 171 165 161 162 182 164 171 169 176 177 170 169 171 160 165
165 179 161 178 173 168 171 163 165 166 166 166 169 167 166 167 172
169 171 168 162 165 168 171 174 165 168 167 170 170

1.3.1. Построение гистограммы и вычисление точечных оценок параметров распределения. По выборочным данным находим лимиты x_{min}=151,0, x_{max}= 182,0 и размах Δx=31. Для n=115 вычисляем количество классов K=$\left[\sqrt{115}\right]$≈10 и ширину класса Δa =3,1. Интервалы, частоты n_k, относительные частоты p_k^* и значения эмпирических функций плотности f_k^* и распределения F_k^* представлены в табл. 1.4.

По полученным данным строим гистограммы частот n_k (рис.1.2) и накопленных частот (рис. 1.3).

Точечные оценки параметров выборки − m = 166,58, $\hat{s}^2$ = 5,98.

Таблица 1.4

k	$[a_k, a_{k+1})$	n_k	p_k^*	f_k^*	F_k^*
0	$-\infty$ − 151,0	0	0,000	0,000	0,000
1	151,0 − 154,1	3	0,026	0,008	0,026
2	154,1 − 157,2	3	0,026	0,008	0,052
3	157,2 − 160,3	11	0,096	0,031	0,148
4	160,3 − 163,4	15	0,130	0,042	0,278
5	163,4 − 166,5	25	0,217	0,070	0,495
6	166,5 − 169,6	24	0,209	0,067	0,704
7	169,6 − 172,7	18	0,157	0,051	0,861
8	172,7 − 175,8	6	0,052	0,017	0,913
9	175,8 − 178,9	7	0,061	0,020	0,974
10	178,9 − 182,0	3	0,026	0,008	1,000
11	182,0 − $+\infty$	0	0,000	0,000	1,000

1.3.2. Проверка гипотезы $H_0: F(x) = F_0(x;\mu,\sigma)$ о нормальности закона распределения с помощью критерия Пирсона χ^2. Поскольку гистограмма частот является куполообразной, одномодальной и более или менее симметричной, то формулируем гипотезу $H_0: F(x) = F_0(x;\mu,\sigma)$ − выборка произведена из генеральной совокупности, распределенной по нормальному закону с параметрами $\mu = m$ =166,58 и $\sigma = \hat{s}$=5,98. Задаем уровень значимости α=0,05 и вычисляем вероятности попадания значений X в k-й интервал p_k, теоретические частоты $n_k^0 = n\,p_k$, k=0,...,K+1. Полученные данные приведены в табл. 1.5.

В результате вычислений получаем значение статистики критерия Пирсона χ_*^2=7,544. Для ν = (K+2)-3=K-1 и α=0,05 по табл. 7.2 находим $\chi_{\nu,1-\alpha}^2 \approx$ 16,9. Поскольку выполняется условие $\chi_*^2 < \chi_{\nu,1-\alpha}^2$, то гипотеза $H_0: F(x) = F_0(x;\mu,\sigma)$ принимается и с доверительной вероятностью $\beta = 1-\alpha$=0,95 можно полагать, что рост взрослых мужчин в генеральной совокупности распределен по нормальному закону с функцией плотности

$$f(x) = \frac{1}{5{,}98\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{(x-166{,}58)}{5{,}98}\right)^2\right].$$

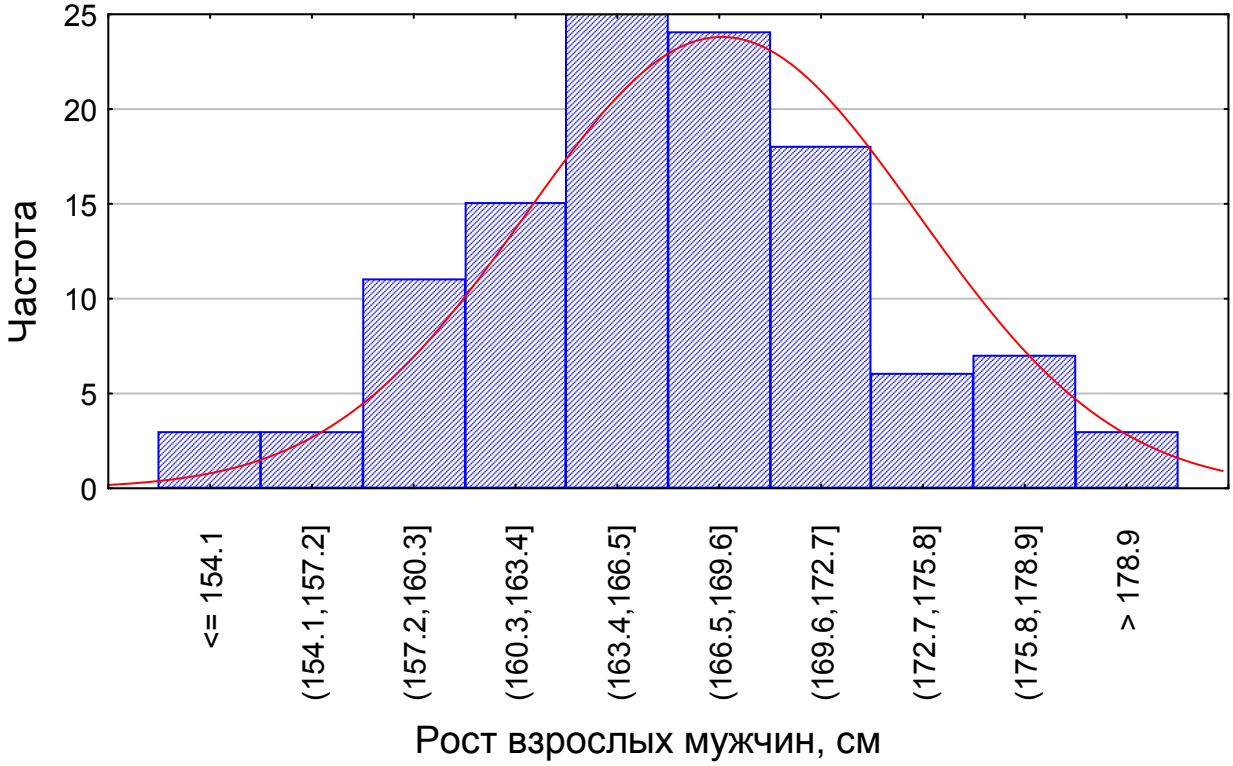

Рис. 1.2. Гистограмма частот

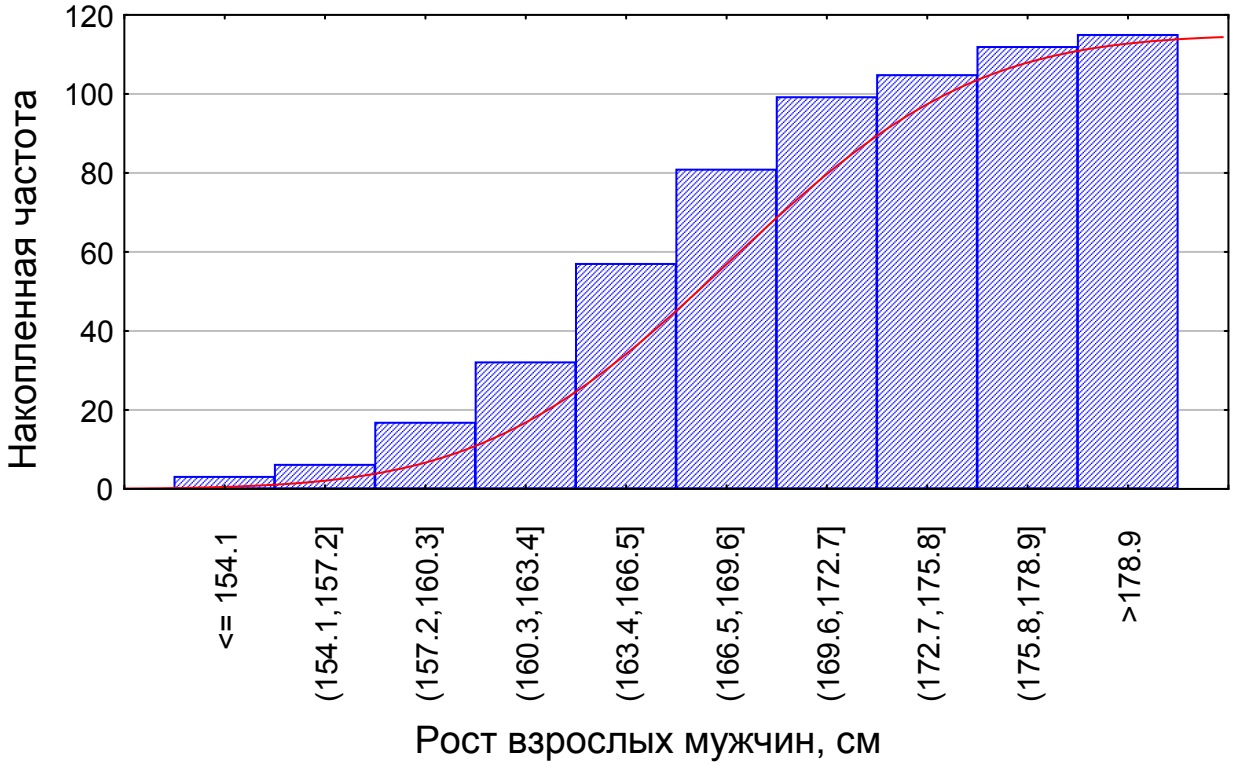

Рис. 1.3. Гистограмма накопленных частот

Таблица 1.5

k	n_k	p_k	n_k^0	$\left(n_k - n_k^0\right)^2$	$\frac{\left(n_k - n_k^0\right)^2}{n_k^0}$
0	0	0,0046	0,529	0,280	0,529
1	3	0,0140	1,61	1,93	1,20
2	3	0,0400	4,60	2,56	0,56
3	11	0,0900	10,35	0,42	0,04
4	15	0,1510	17,39	5,71	0,33
5	25	0,1980	22,71	5,29	0,23
6	24	0,2000	22,94	1,21	0,05
7	18	0,1510	17,39	0,36	0,02
8	6	0,0900	10,35	18,92	1,83
9	7	0,0400	4,60	5,76	1,25
10	3	0,0150	1,73	1,61	0,93
11	0	0,0050	0,575	0,331	0,575
					χ_*^2=7,544

1.3.3. Интервальная оценка параметров распределения. При выполнении интервального оценивания параметров выборки по табл. 7.4 и 7.2 для заданного значения α=0,05 и числа степеней свободы $\nu = n - 1$=114 находим квантиль распределения Стьюдента $t_{\nu,1-\alpha/2}$≈1,98 и квантили распределения Пирсона $\chi^2_{\nu,\alpha/2}$≈ 86,34 и $\chi^2_{\nu,1-\alpha/2}$≈ 145,40, откуда согласно формулам (1.6) и (1.7)

$$166{,}58 - \frac{5{,}98}{\sqrt{115}} \cdot 1{,}98 < \mu < 166{,}58 + \frac{5{,}98}{\sqrt{115}} \cdot 1{,}98 \quad \text{или} \quad 165{,}48 < \mu < 167{,}68,$$

$$\frac{35{,}70 \cdot 114}{145{,}40} < \sigma^2 < \frac{35{,}70 \cdot 114}{86{,}34} \quad \text{или} \quad 27{,}99 < \sigma^2 < 47{,}14.$$

2. СРАВНЕНИЕ ПАРАМЕТРОВ ДВУХ НЕЗАВИСИМЫХ РАСПРЕДЕЛЕНИЙ

2.1. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

2.1.1. Постановка задачи о сравнении параметров двух независимых распределений. Пусть некоторая случайная величина определена на двух непересекающихся подпопуляциях или слоях из более широкой популяции. Тогда можно считать две реализации этой случайной величины на непересекающихся подпопуляциях двумя независимыми случайными величинами.

Пусть из первой популяции извлекается выборка объемом n_X, результаты наблюдений (выборочные значения) соответственно $x_1, x_2, \ldots, x_{n_X}$. Из второй популяции извлекается выборка объемом n_Y, результаты наблюдений соответственно $y_1, y_2, \ldots, y_{n_Y}$. По каждой группе наблюдений можно получить точечные и интервальные оценки m_X, $\hat{s}_X^2$, m_Y, $\hat{s}_Y^2$ параметров $\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2$ распределений случайных величин X и Y, однако практически более важными являются задачи сравнения параметров распределений X и Y. Результаты подобного сравнения позволяют судить о влиянии на параметры распределений того фактора, по которому производится разделение на подпопуляции.

2.1.2. Проверка гипотезы о равенстве дисперсий H_0: $\sigma_X^2 = \sigma_Y^2$. Построение ***критерия проверки гипотезы*** основывается на том, что:

1) для ***нормально распределенной случайной величины*** с дисперсией σ^2 и выборочной дисперсией $\hat{s}^2$, рассчитанной по выборке объемом n, ***величина*** $(n-1)\tilde{\hat{s}}^2/\sigma^2$ имеет ***распределение Пирсона*** $\tilde{\chi}_\nu^2$ с числом степеней свободы ν=n-1, и $\dfrac{(n_X-1)\,\tilde{\hat{s}}_X^2}{\sigma_X^2} \sim \tilde{\chi}_{n_x-1}^2$, $\dfrac{(n_Y-1)\,\tilde{\hat{s}}_Y^2}{\sigma_Y^2} \sim \tilde{\chi}_{n_y-1}^2$;

2) в предположении о справедливости гипотезы H_0 ***отношение выборочных дисперсий*** $\hat{s}_X^2, \hat{s}_Y^2$ ***независимых случайных величин*** X и Y

$$\tilde{F} = \frac{\tilde{\hat{s}}_X^2}{\tilde{\hat{s}}_Y^2} = \frac{\tilde{\chi}_{n_X-1}^2 \sigma_X^2 / (n_X-1)}{\tilde{\chi}_{n_Y-1}^2 \sigma_Y^2 / (n_Y-1)} = \frac{\tilde{\chi}_{n_X-1}^2 / (n_X-1)}{\tilde{\chi}_{n_Y-1}^2 / (n_Y-1)} \qquad (2.1)$$

имеет ***распределение Фишера*** $\tilde{F}_{\nu_1,\nu_2}$ с числом степеней свободы $\nu_X = n_X - 1$ числителя и числом степеней свободы $\nu_Y = n_Y - 1$ знаменателя (см. 7.3).

Поэтому можно построить β-доверительный интервал $\tilde{F}$

$$P\left(\left\{F_{\nu_1,\nu_2,\alpha/2} < \tilde{F} < F_{\nu_1,\nu_2,1-\alpha/2}\right\}\right) = \beta, \qquad (2.2)$$

границы которого − квантили $F_{\nu_1,\nu_2,\alpha/2}$ и $F_{\nu_1,\nu_2,1-\alpha/2}$ − определены таким образом: $P\left(\left\{\tilde{F} < F_{\nu_1,\nu_2,\alpha/2}\right\}\right) = \alpha/2$, $P\left(\left\{\tilde{F} < F_{\nu_1,\nu_2,1-\alpha/2}\right\}\right) = 1-\alpha/2$.

Если выборочное значение F^*, вычисленное по (2.1), попадает в доверительный интервал (2.2), то ***гипотеза*** H_0: $\sigma_X^2 = \sigma_Y^2$ ***принимается***, в противном случае ***гипотеза отвергается***.

Для сокращения вычислений обычно пользуются свойством F-распределения, которое состоит в том, что если при вычислении выборочного значения F^* в числитель помещать бóльшее значение выборочной дисперсии, а в знаменатель мéньшее, то для принятия гипотезы H_0 достаточно выполнения одностороннего неравенства

$$F^* < F_{\nu_1,\, \nu_2,\, 1-\alpha/2}. \tag{2.3}$$

Критерий F очень ***чувствителен*** к отклонениям распределений исследуемых величин от нормального. Поэтому его нельзя использовать, если есть сомнения относительно нормальности распределений.

2.1.3. Проверка гипотезы о равенстве математических ожиданий $H_0\colon\ \mu_X = \mu_Y$**.** Построение критерия проверки гипотезы основывается на том, что:

1) независимые величины X и Y распределены нормально;

2) известен результат проверки гипотезы $H_0\colon\ \sigma_X^2 = \sigma_Y^2$ (см.2.1.2).

Если ***гипотеза*** $H_0\colon\ \sigma_X^2 = \sigma_Y^2$ ***принята***, то есть $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, то $\hat{s}_X^2$ и $\hat{s}_Y^2$ являются точечными оценками одного и того же параметра σ^2 (о точечных оценках см. 1.1.1). Средневзвешенной (обобщенной) точечной оценкой параметра σ^2 по двум выборкам будет величина

$$\hat{s}^2 = \frac{\nu_X \hat{s}_X^2 + \nu_y \hat{s}_Y^2}{\nu_X + \nu_Y} = \frac{\sum\limits_{i=1}^{n_X}(x_i - m_X)^2 + \sum\limits_{i=1}^{n_Y}(y_i - m_Y)^2}{\nu_X + \nu_Y}. \tag{2.4}$$

С ее помощью строится критерий проверки гипотезы $H_0\colon\ \mu_X = \mu_Y$ (***двухвыборочный t-критерий***), основанный на статистике

$$\tilde{t} = \frac{\tilde{m}_X - \tilde{m}_Y}{\hat{s}\sqrt{1/n_X + 1/n_Y}}, \tag{2.5}$$

имеющей ***распределение Стьюдента*** $\tilde{t}_\nu$ с числом степеней свободы $\nu = n_X + n_Y - 2 = \nu_X + \nu_Y$. Поэтому можно построить β-доверительный интервал $\tilde{t}$

$$P\left(\left\{t_{\nu,\, \alpha/2} < \tilde{t} < t_{\nu,\, 1-\alpha/2}\right\}\right) = \beta, \tag{2.6}$$

границами которого будут квантили распределения $t_{\nu,\,\alpha/2}$ и $t_{\nu,\,1-\alpha/2}$, определенные так: $P\left(\left\{\tilde{t} < t_{\nu,\alpha/2}\right\}\right) = \alpha/2$, $P\left(\left\{\tilde{t} < t_{\nu,1-\alpha/2}\right\}\right) = 1-\alpha/2$.

Если выборочное значение t^*, вычисленное по (2.5), попадает в доверительный интервал (2.6), то ***гипотеза*** H_0: $\mu_X = \mu_Y$ ***принимается***, в противном случае - ***отвергается***.

Если ***гипотеза*** H_0: $\sigma_X^2 = \sigma_Y^2$ ***отвергнута***, то применение двухвыборочного t-критерия становится не надежным, и более подходящим оказывается ***критерий Уэлча***. В критерии Уэлча статистикой является величина

$$\tilde{t} = \frac{\tilde{m}_X - \tilde{m}_Y}{\sqrt{\hat{s}_X^2/n_X + \hat{s}_Y^2/n_Y}}, \tag{2.7}$$

имеющая ***распределение Стьюдента*** $\tilde{t}_\nu$ с числом степеней свободы

$$\nu = \left(\frac{\hat{s}_X^2}{n_X} + \frac{\hat{s}_Y^2}{n_Y}\right)^2 \Bigg/ \left(\frac{\hat{s}_X^4}{n_X^2(n_X-1)} + \frac{\hat{s}_Y^2}{n_Y^2(n_Y-1)}\right). \tag{2.8}$$

Если выборочное значение t^*, вычисленное по (2.7) и (2.8), попадает в доверительный интервал (2.6), то ***гипотеза*** H_0: $\mu_X = \mu_Y$ ***принимается***, в противном случае - ***отвергается***.

Так как число степеней свободы ν в критерии Уэлча не обязательно целое, то для вычисления квантилей $t_{\nu,\,\alpha/2}$ и $t_{\nu,\,1-\alpha/2}$ следует использовать интерполяцию по их значениям для ближайших целочисленных значений ν_a, ν_b ($\nu_a < \nu < \nu_b$), то есть $t_\nu = t_{\nu_a} + \left(t_{\nu_b} - t_{\nu_a}\right)\dfrac{\nu - \nu_a}{\nu_b - \nu_a}$.

t-критерии (двухвыборочный и Уэлча) являются ***робастными***, то есть они ***не чувствительны*** к умеренным отклонениям исследуемых распределений от нормального.

2.1.4. Проверка гипотезы о различии математических ожиданий H_0: $\mu_X - \mu_Y = \delta$**.** Построение критерия проверки гипотезы основывается на тех же предположениях, что и в случае проверки гипотезы H_0: $\mu_X = \mu_Y$ (см. 2.1.3). Первоначально проверяется гипотеза о равенстве дисперсий H_0: $\sigma_X^2 = \sigma_Y^2$ (см. 2.1.2).

Если ***гипотеза*** H_0: $\sigma_X^2 = \sigma_Y^2$ ***принята***, то далее используется ***двухвыборочный t-критерий***, основанный на статистике

$$\tilde{t} = \frac{\tilde{m}_X - \tilde{m}_Y - \delta}{\hat{s}\sqrt{1/n_X + 1/n_Y}} \tag{2.9}$$

имеющей распределение Стьюдента $\tilde{t}_\nu$ с числом степеней свободы $\nu = n_X + n_Y - 2 = \nu_X + \nu_Y$, и в которой обобщенная выборочная дисперсия по-прежнему вычисляется по формуле (2.4). Если выборочное значение t^*, вычисленное по (2.9), попадает в доверительный интервал (2.6), то ***гипотеза*** H_0: $\mu_X - \mu_Y = \delta$ ***принимается***, в противном случае ***гипотеза отвергается***.

Если ***гипотеза*** H_0: $\sigma_X^2 = \sigma_Y^2$ ***отвергнута***, то применяется ***критерий Уэлча***, но не со статистикой (2.7), а со статистикой

$$\tilde{t} = \frac{\tilde{m}_X - \tilde{m}_Y - \delta}{\sqrt{\hat{s}_X^2/n_X + \hat{s}_Y^2/n_Y}}, \tag{2.10}$$

имеющей распределение Стьюдента $\tilde{t}_\nu$ с числом степеней свободы (2.8).

Если выборочное значение t^*, вычисленное по (2.8) и (2.10), попадает в доверительный интервал (2.6), то ***гипотеза*** H_0: $\mu_X = \mu_Y - \delta$ ***принимается***, в противном случае ***гипотеза отвергается***.

t-критерии (двухвыборочный и Уэлча) являются ***робастными***, то есть они ***не чувствительны*** к умеренным отклонениям исследуемых распределений от нормального; как правило, достаточно, чтобы X и Y имели одномодальные и приблизительно симметричные эмпирические функции плотности.

2.2. АЛГОРИТМ РЕШЕНИЯ

2.2.1. Проверка гипотезы о равенстве дисперсий H_0: $\sigma_X^2 = \sigma_Y^2$. Выполняется по алгоритму:

1) по выборочным данным вычислить точечные оценки параметров согласно формулам (1.3):

$$m_X = \frac{1}{n_X}\sum_{i=1}^{n_X} x_i, \quad m_Y = \frac{1}{n_Y}\sum_{i=1}^{n_Y} y_i,$$

$$\hat{s}_X^2 = \frac{1}{n_X - 1}\sum_{i=1}^{n_X}(x_i - m_X)^2, \quad \hat{s}_Y^2 = \frac{1}{n_Y - 1}\sum_{i=1}^{n_Y}(y_i - m_Y)^2.$$

2) вычислить статистику критерия Фишера

$$F^* = \frac{\hat{s}_X^2}{\hat{s}_Y^2}, \text{ если } \hat{s}_X^2 \ge \hat{s}_Y^2 \text{ , или } F^* = \frac{\hat{s}_Y^2}{\hat{s}_X^2}, \text{ если } \hat{s}_Y^2 \ge \hat{s}_X^2;$$

3) для заданного уровня значимости α=0,05 и чисел степеней свободы $\nu_X = n_X - 1$ и $\nu_Y = n_Y - 1$ по табл. 7.3 найти квантиль статистики Фишера $F_{\nu_x,\nu_y,1-\alpha/2}$;

4) если выполняется неравенство $F^* > F_{\nu_X,\nu_Y,1-\alpha/2}$, то отклонить гипотезу H_0: $\sigma_X^2 = \sigma_Y^2$, в противном случае принять нуль-гипотезу и сделать вывод о совпадении дисперсий генеральных совокупностей.

2.2.2. Проверка гипотезы о равенстве математических ожиданий H_0: $\mu_X = \mu_Y$**.** Для ее выполнения необходимо:

1) если гипотеза H_0: $\sigma_X^2 = \sigma_Y^2$ принята, то вычислить обобщенную точечную оценку $\hat{s}^2$ по (2.4), статистику критерия Стьюдента t^* по (2.5) и число степеней свободы $\nu = \nu_X + \nu_Y = n_X + n_Y - 2$;

2) если гипотеза H_0: $\sigma_X^2 = \sigma_Y^2$ отвергнута, то вычислить по (2.7) статистику критерия Уэлча t^* и число степеней свободы ν (2.8);

3) для определенного в 2.2.1 уровня значимости α и вычисленного в 1) или 2) числа степеней свободы ν по табл. 7.4 найти критическое значение статистики $t_{\nu,1-\alpha/2}$;

4) если $t^* > t_{\nu,1-\alpha/2}$, то отвергнуть гипотезу H_0: $\mu_X = \mu_Y$, в противном случае принять гипотезу и сделать вывод о совпадении математических ожиданий.

2.3. ПРИМЕР РАСЧЕТА

Данные о времени, мин, до наступления паралича у комнатной мухи под влиянием двух доз (X и Y) яда представлены следующими результатами опытов:

X	0,48	0,70	0,70	0,85	0,95	0,95	1,00	1,08	1,30	1,38	1,38
	1,53	1,63	1,66	1,76	2,15						
Y	0,30	0,70	0,70	0,85	0,90	0,95	1,15	1,26	1,38	1,42	1,42
	1,53	1,57	1,62	1,95							

2.3.1. Проверка гипотезы о равенстве дисперсий H_0: $\sigma_X^2 = \sigma_Y^2$. По исходным данным вычисляем: m_X=1,219, m_Y=1,180, $\hat{s}_X$=0,455, $\hat{s}_Y$=0,439.

Учитывая, что $\hat{s}_X > \hat{s}_Y$, вычисляем статистику критерия Фишера $F^* = \hat{s}_X^2 / \hat{s}_Y^2 \approx 0{,}455^2 / 0{,}439^2 \approx 1{,}075$. Для заданного уровня значимости α=0,05 и $\nu_X = n_X - 1 = 15$, $\nu_Y = n_Y - 1 = 14$ по табл.7.3 определяем критическое значение $F_{\nu_x, \nu_y, 1-\alpha/2} \approx 2{,}95$. Поскольку $F^* < F_{\nu_x, \nu_y, 1-\alpha/2}$ (1,075<2,95), то гипотезу H_0: $\sigma_X^2 = \sigma_Y^2$ принимаем.

2.3.2. Проверка гипотезы о равенстве математических ожиданий H_0: $\mu_X = \mu_Y$. Вычисляем обобщенную выборочную дисперсию $\hat{s}^2$ по (2.4):

$$\hat{s}^2 = \frac{\nu_X \hat{s}_X^2 + \nu_Y \hat{s}_Y^2}{\nu_X + \nu_Y} = \frac{15 \cdot 0{,}455^2 + 14 \cdot 0{,}439^2}{29} \approx \frac{5{,}804}{29} \approx 0{,}2001,$$

$$\hat{s} = \sqrt{\hat{s}^2} \approx 0{,}447.$$

Так как гипотеза H_0: $\sigma_X^2 = \sigma_Y^2$ принята, то по обобщенной выборочной дисперсии $\hat{s}^2$ вычисляем статистику t^* (2.5):

$$t^* = \frac{m_X - m_Y}{\hat{s}\sqrt{1/n_X + 1/n_Y}} \approx \frac{1{,}219 - 1{,}180}{0{,}447\sqrt{1/16 + 1/15}} \approx \frac{0{,}039}{0{,}1607} \approx 0{,}243.$$

Для α=0,05 и $\nu = n_X + n_Y - 2$=29 по табл. 7.4 находим критическое значение статистики Стьюдента $t_{\nu, 1-\alpha/2} \approx 2{,}05$. Поскольку выполняется неравенство $t^* < t_{\nu, 1-\alpha/2}$, то принимаем гипотезу H_0: $\mu_X = \mu_Y$.

На основании полученных данных можно сделать вывод о том, что время до наступления паралича у мухи не зависит от используемых в эксперименте доз яда.

3. СРАВНЕНИЕ ПАРАМЕТРОВ ДВУХ ЗАВИСИМЫХ РАСПРЕДЕЛЕНИЙ

3.1. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

3.1.1. Постановка задачи о сравнении параметров двух зависимых распределений. Пусть некоторая случайная величина определена на популяции, подвергающейся действию определенного фактора. Тогда можно считать две реализации этой случайной величины, соответствен-

но до и после действия фактора, зависимыми случайными величинами (так называемые ***парные наблюдения "до-после"***).

Пусть результаты измерений в одной группе наблюдений ("до") соответственно $x_1, x_2, \ldots, x_n$, а во второй группе ("после") - $y_1, y_2, \ldots, y_n$. По каждой группе наблюдений можно получить точечные и интервальные оценки $m_X, \hat{s}_X^2, m_Y, \hat{s}_Y^2$ параметров $\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2$ распределений случайных величин X и Y, однако практически более важными являются задачи сравнения параметров распределений X и Y. Результаты подобного сравнения позволяют судить о влиянии на некоторый процесс изучаемого фактора.

3.1.2. Проверка гипотезы о равенстве математических ожиданий H_0: $\mu_X = \mu_Y$**.** Построение критерия проверки гипотезы основывается на предположении о том, что зависимые величины X и Y ***распределены нормально***.

Исходная гипотеза переформулируется в виде H_0: $\mu_Z = \mu_Y - \mu_X = 0$ для случайной величины $Z=Y-X$. ***Статистикой критерия проверки*** гипотезы H_0 служит величина

$$\tilde{t} = \frac{\tilde{m}_Z}{\tilde{\hat{s}}_Z / \sqrt{n}}, \tag{3.1}$$

которая ***распределена по закону Стьюдента*** $\tilde{t}_\nu$ с числом степеней свободы $\nu=n-1$, где

$$m_Z = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i), \tag{3.2}$$

$$\hat{s}_Z^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[(y_i - x_i) - m_Z\right]^2 = \hat{s}_X^2 + \hat{s}_Y^2 - 2\hat{s}_X\hat{s}_Y r_{X,Y}, \tag{3.3}$$

$$r_{X,Y} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_X)\times(y_i - m_Y)}{\sqrt{\hat{s}_X^2}\sqrt{\hat{s}_Y^2}}, \tag{3.4}$$

где $r_{X,Y}$ - выборочный коэффициент корреляции (точечная оценка коэффициента корреляции $\rho_{X,Y}$, см. 4.1.2).

Так как распределение случайной величины $\tilde{t}$ известно, то можно построить β-доверительный интервал $\tilde{t}$

$$P\left(\left\{t_{\nu,\alpha/2} < \tilde{t} < t_{\nu,1-\alpha/2}\right\}\right) = \beta, \tag{3.5}$$

границами которого будут квантили распределения $t_{\nu,\,\alpha/2}$ и $t_{\nu,\,1-\alpha/2}$, определенные по формулам: $P\left(\left\{\tilde{t}_\nu < t_{\nu,\alpha/2}\right\}\right) = \alpha/2$, $P\left(\left\{\tilde{t}_\nu < t_{\nu,1-\alpha/2}\right\}\right) = 1-\alpha/2$.

Если выборочное значение t^*, вычисленное по (3.1), попадает в доверительный интервал (3.5), то ***гипотеза*** H_0: $\mu_X = \mu_Y$ ***принимается***, в противном случае ***гипотеза отвергается***.

Для сокращения вычислений обычно пользуются свойством симметрии распределения $\tilde{t}_\nu$. Для принятия гипотезы H_0 достаточно выполнения одностороннего неравенства :

$$\left|t^*\right| < t_{\nu,\,1-\alpha/2}. \tag{3.6}$$

t-критерий проверки гипотезы является ***робастным***, то есть он ***не чувствителен*** к умеренным отклонениям исследуемых распределений от нормального закона.

3.1.3. Проверка гипотезы о различии математических ожиданий H_0: $\mu_X - \mu_Y = \delta$**.** Построение критерия проверки гипотезы основывается на тех же предположениях, что и в случае проверки гипотезы H_0: $\mu_X = \mu_Y$ (см. 3.1.2), однако используется не статистика (3.1), а статистика

$$\tilde{t} = \frac{\tilde{m}_Z - \delta}{\tilde{s}_Z / \sqrt{n}}, \tag{3.7}$$

которая также ***распределена по закону Стьюдента*** $\tilde{t}_\nu$ с числом степеней свободы ν=n-1. Если выборочное значение t^*, вычисленное по (3.7), попадает в доверительный интервал (3.5), или выполняется одностороннее неравенство (3.6), то ***гипотеза*** H_0: $\mu_X - \mu_Y = \delta$ ***принимается***, в противном случае ***гипотеза отвергается***.

t-критерий проверки гипотезы является ***робастным***, то есть он ***не чувствителен*** к умеренным отклонениям исследуемых распределений от нормального закона.

3.2. АЛГОРИТМ РЕШЕНИЯ

3.2.1. Проверка гипотезы о равенстве математических ожиданий $H_0: \mu_X = \mu_Y$. Для этого необходимо:

1) по исходным данным x_i, y_i, $i = \overline{1,n}$ вычислить значения $z_i = x_i - y_i$, $i = \overline{1,n}$;

2) вычислить точечные оценки параметров распределения $Z = X - Y$ по (1.3):

$$m_Z = \frac{1}{n}\sum_{i=1}^{n} z_i, \quad \hat{s}_Z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - m_z)^2;$$

3) для проверки гипотезы $H_0: \mu_Z = \mu_Y - \mu_X = 0$ вычислить по (3.1) статистику t^* и число степеней свободы ν=n-1;

4) по табл. 7.4 для заданного уровня значимости α и числа степеней свободы ν=n-1 найти квантиль распределения Стьюдента $t_{\nu,\,1-\alpha/2}$;

5) если $t^* < t_{\nu,\,1-\alpha/2}$, то принять гипотезу $H_0: \mu_Z = 0$, в противном случае гипотезу о равенстве математических ожиданий генеральных совокупностей отвергнуть и сделать вывод о влиянии исследуемого фактора на параметры распределений.

3.3. ПРИМЕР РАСЧЕТА

Данные о скорости кровотока, мл/с, у здоровых людей в норме (X) и после физической нагрузки (20 приседаний, Y) представлены результатами 13 измерений:

X	3,5	4,0	4,2	4,2	4,0	4,7	4,3	3,0	3,0	3,7	3,8	3,1	4,3
Y	4,0	5,0	5,0	5,0	5,1	5,2	6,0	6,8	5,8	5,0	5,7	4,5	5,9

3.3.1. Проверка гипотезы о равенстве математических ожиданий $H_0: \mu_X = \mu_Y$. По выборочным данным x_i, y_i вычисляем значения $z_i = y_i - x_i$, $i = \overline{1,13}$:

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

Вычисляем точечные оценки $m_Z \approx 1{,}477$, $\hat{s}_Z \approx 0{,}941$ и значение статистики $t^* \approx \frac{1{,}477\sqrt{13}}{0{,}941} \approx 5{,}660$. Для заданного уровня значимости $\alpha = 0{,}05$ и числа степеней свободы $\nu = n-1 = 12$ по табл. 7.4 находим квантиль распределения Стьюдента $t_{\nu,\,1-\alpha/2} \approx 2{,}180$. Поскольку $t^* > t_{\nu,\,1-\alpha/2}$, то отклоняем гипотезу H_0: $\mu_Z = 0$ о равенстве математических ожиданий X и Y. Таким образом, физическая нагрузка влияет на скорость кровотока у здоровых людей.

4. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ЗАВИСИМЫХ РАСПРЕДЕЛЕНИЙ

4.1. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

4.1.1. Понятие корреляции двух зависимых распределений. Изученные в первых трех разделах методы статистического анализа дают возможность исследовать изменчивость биологических объектов по отдельному признаку. Однако в ряде случаев важно знать, изменяются ли два признака самостоятельно, независимо друг от друга, или изменчивость одного признака в какой-то степени связана с изменчивостью другого. Если зависимость между признаками не является функциональной, то есть когда определенному значению одного признака соответствует не одно значение, а целое распределение значений другого признака, то такого рода ***взаимная связь*** или взаимная зависимость ***различных признаков при их изменчивости называется корреляционной или стохастической***.

При ***положительной корреляции*** зависимость между признаками ***прямая***: при увеличении значений одного увеличиваются, как правило, и значения другого. При ***отрицательной корреляции*** зависимость между признаками ***обратная***: увеличение значений одного признака соответственно связано с уменьшением значений другого.

4.1.2. Диаграмма рассеяния и коэффициент корреляции. Первым шагом для определения по выборке существования и качественного вида зависимости между случайными величинами (признаками) X и Y является графическое отображение точек (x_i, y_i), $\overline{i = 1, n}$, на плоскость XY. Такой график называется ***диаграммой рассеяния*** (***корреляционным полем***).

Второй шаг - вычисление меры корреляции между величинами X и Y, обоснованно указывающей на степень зависимости между ними. Положение одного объекта относительно остальных в выборке по X и Y,

определяемое выборочными средними m_X и m_Y, проявляется в величинах и знаках отклонений $(x_i - m_X)$ и $(y_i - m_Y)$. Если величины X и Y подчинены прямой связи, то большинству объектов с положительными значениями $(x_i - m_X)$ будут соответствовать положительные значения $(y_i - m_Y)$, а с отрицательными значениями $(x_i - m_X)$ - отрицательные $(y_i - m_Y)$, т.е. большинство произведений $(x_i - m_X)\times(y_i - m_Y)$ будут положительными. Следовательно, сумма этих произведений для всех объектов будет положительной.

Если величины X и Y подчинены обратной связи, то большинству объектов с положительными значениями $(x_i - m_X)$ будут соответствовать отрицательные значения $(y_i - m_Y)$, а с отрицательными значениями $(x_i - m_X)$ - положительные $(y_i - m_Y)$. В этом случае произведения $(x_i - m_X)\times(y_i - m_Y)$ будут, как правило, отрицательными. Следовательно, и сумма этих произведений отрицательна.

Если величины X и Y не имеют систематической связи (большие значения X сочетаются с малыми значениями Y столь же часто, как и с большими значениями Y, и то же самое справедливо для сочетания малых значений X со значениями Y), то среди объектов с большими положительными значениями $(x_i - m_X)$ у некоторых из них значения $(y_i - m_Y)$ будут положительные, а у других - отрицательные. В сумме произведений $(x_i - m_X)\times(y_i - m_Y)$ для всех объектов выборки положительные и отрицательные члены приблизительно компенсируют друг друга, поэтому при большом числе слагаемых сумма, как правило, близка к нулю.

Поскольку сумма произведений отклонений зависит от числа пар значений, полученных в выборке, то с помощью осреднения переходят к величине

$$\hat{s}_{X,Y} = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_i - m_X)\times(y_i - m_Y), \qquad (4.1)$$

называемой ***выборочной ковариацией***. Выборочная ковариация признака X с самим собой - это просто выборочная дисперсия X: $\hat{s}_{X,X} = \hat{s}_X^2$.

Поскольку выборочная ковариация является размерной величиной, которая может быть как сколь угодно малой, так и сколь угодно большой, удобно перейти к безразмерной величине

$$r_{X,Y} = \frac{\hat{s}_{X,Y}}{\hat{s}_X\,\hat{s}_Y}, \qquad (4.2)$$

называемой ***выборочным коэффициентом корреляции***.

Выборочный коэффициент корреляции является точечной оценкой ***коэффициента корреляции***

$$\rho_{X,Y} = \frac{M[(X - M[X])(Y - M[Y])]}{\sqrt{D[X]}\ \sqrt{D[Y]}} = \frac{\sigma_{X,Y}}{\sigma_X\ \sigma_Y}, \tag{4.3}$$

где $\sigma_{X,Y}$ - ковариация величин X и Y.

Множество значений коэффициента корреляции $-1 \le \rho_{X,Y} \le +1$, причем значения $\rho_{X,Y}$ не меняются при линейном преобразовании переменных (вида $aX+b$, $cY+d$). Равенство $|\rho_{X,Y}| = 1$ имеет место только тогда, когда ***связь*** между величинами X и Y строго ***функциональная*** и ***линейная*** (X и Y зависят между собой без посредства каких-либо других величин). Если $\rho_{X,Y}$=0, то величины ***некоррелированы***. Понятие некоррелированности не следует смешивать с понятием независимости, поскольку независимые величины всегда некоррелированы. Однако обратное утверждение неверно: ***некоррелированные величины могут быть зависимы*** и даже функционально, однако эта связь не линейная. Неизменность $\rho_{X,Y}$ при линейном преобразовании величин означает, что значение коэффициента корреляции будет одним и тем же, если измерения проводить в различных единицах, метрических или неметрических (в метрах или дюймах, футах; в килограммах или унциях, фунтах и т.д.). ***Свойства выборочного коэффициента корреляции*** $r_{X,Y}$ совпадают с перечисленными свойствами коэффициента корреляции $\rho_{X,Y}$.

Пусть величины X и Y имеют двумерное нормальное распределение. При некотором значении X=x распределение величины Y, называемое условным распределением Y при X=x, есть одномерное нормальное распределение с ***условным математическим ожиданием***

$$\mu_{Y|X=x} = \mu_Y + \frac{\sigma_{X,Y}}{\sigma_X^2}(x - \mu_X) \tag{4.4}$$

и ***условной дисперсией*** $\sigma^2_{Y|X=x} = \sigma^2_Y(1 - \rho^2_{X,Y})$, откуда

$$\rho^2_{X,Y} = \frac{\sigma^2_Y - \sigma^2_{Y|X=x}}{\sigma^2_Y}. \tag{4.5}$$

То есть ***квадрат коэффициента корреляции равен относительному уменьшению дисперсии величины Y от значения σ_Y^2 до значения $\sigma_{Y|X=x}^2$, связанному с принятием величиной X фиксированного значения x***. В связи с данной интерпретацией может быть введена классификация связи между двумя величинами, в зависимости от значения коэффициента корреляции. В табл.4.1 перечислены значения ρ с указанием наименований для различных видов корреляционной связи.

Таблица 4.1

$\rho_{X,Y}$	Вид корреляционной связи
$\lvert\rho_{X,Y}\rvert = 1$	строгая (имеется функциональная зависимость между переменными)
$0{,}7 \le \lvert\rho_{X,Y}\rvert < 1$	сильная (тесная)
$0{,}5 \le \lvert\rho_{X,Y}\rvert < 0{,}7$	средняя
$0{,}3 \le \lvert\rho_{X,Y}\rvert < 0{,}5$	умеренная
$0{,}2 \le \lvert\rho_{X,Y}\rvert < 0{,}3$	слабая
$0 \le \lvert\rho_{X,Y}\rvert < 0{,}2$	очень слабая
$\lvert\rho_{X,Y}\rvert = 0$	отсутствует

Например, при $\rho_{X,Y}$=0,5 25% ($0{,}5^2 \cdot 100\%$) изменчивости одного признака объясняется изменчивостью другого признака, в остальном же связь между признаками на 75% случайная. При $\rho_{X,Y}$=0,3 менее 10% ($0{,}3^2 \cdot 100\%$), а при $\rho_{X,Y}$=0,7 около 50% ($0{,}7^2 \cdot 100\%$) изменчивости одного признака объясняется таким же образом. При $\rho_{X,Y}$=0,9 уже более 80% ($0{,}9^2 \cdot 100\%$) изменчивости одного признака закономерно связано с изменчивостью другого признака, в остальных же примерно 20% совпадение или несовпадение их изменчивостей случайно.

4.1.3. Проверка значимости коэффициента корреляции. Поскольку выборочный коэффициент корреляции $r_{X,Y}$ является точечной оценкой коэффициента корреляции $\rho_{X,Y}$, то равенство нулю $r_{X,Y}$ еще не свидетельствует о равенстве нулю $\rho_{X,Y}$, а следовательно, о некоррелированности величин X и Y. Чтобы выяснить, находятся ли величины X и Y в корреляционной зависимости, нужно проверить, достаточна ли величина $r_{X,Y}$ для обоснованного вывода о наличии корреляционной связи. Для этого проверяется нулевая ***гипотеза о равенстве коэффициента***

корреляции ρ нулю H_0: $\rho_{X,Y} = 0$. Двусторонней альтернативой является гипотеза $H_1: \rho_{X,Y} \neq 0$. ***Построение критерия проверки основывается на предположении о нормальности распределений X и Y. Статистикой критерия*** является величина

$$\tilde{t} = \sqrt{n-2}\,\frac{\tilde{r}_{X,Y}}{\sqrt{1-\tilde{r}_{X,Y}^2}}, \tag{4.6}$$

имеющая распределение Стьюдента $\tilde{t}_\nu$ с числом степеней свободы ν=n-2. Следовательно, можно построить β-доверительный интервал

$$P\left(\left\{t_{\nu,\alpha/2} < \tilde{t}_\nu < t_{\nu,1-\alpha/2}\right\}\right) = \beta, \tag{4.7}$$

границами которого будут квантили распределения, определенные так:

$$P\left(\left\{\tilde{t}_\nu < t_{\nu,\alpha/2}\right\}\right) = \alpha/2, \quad P\left(\left\{\tilde{t}_\nu > t_{\nu,1-\alpha/2}\right\}\right) = 1-\alpha/2. \tag{4.8}$$

Если выборочное значение t^* статистики (4.5) попадает в доверительный интервал, то ***гипотеза*** H_0: $\rho_{X,Y} = 0$ ***принимается***, в противном случае ***гипотеза*** H_0 ***отвергается***, и коэффициент корреляции статистически значимо отличается от нуля (***коэффициент корреляции значим)***. Фактически проверяется выполнение неравенства

$$\left|t^*\right| < t_{\nu,1-\alpha/2}. \tag{4.9}$$

Проверка гипотезы H_0: $\rho_{X,Y} = 0$ может быть осуществлена и непосредственно по выборочному значению $r_{X,Y}$. ***Гипотеза*** H_0: $\rho_{X,Y} = 0$ ***принимается***, если

$$\left|r_{X,Y}\right| < \sqrt{\frac{t_{\nu,1-\alpha/2}}{n-2+t_{\nu,1-\alpha/2}}}, \tag{4.10}$$

в противном случае ***гипотеза отвергается***.

Приведенный t-критерий проверки гипотезы H_0 является ***робастным***, то есть ***не чувствительным*** к умеренным отклонениям исследуемых распределений от нормального закона.

4.1.4. Интервальная оценка коэффициента корреляции. Выборочный коэффициент корреляции является очень неустойчивой оценкой, и при небольшом объеме выборки даже заметно отличающееся от нуля значение выборочного коэффициента корреляции может оказаться не-

значимым. Для хороших оценок коэффициента корреляции, как правило, необходимо несколько десятков или даже сотен наблюдений.

В случае значимого выборочного коэффициента корреляции может быть построен доверительный интервал для коэффициента корреляции. Однако для этого нужно знать закон распределения выборочного коэффициента корреляции как случайной величины $\tilde{r}_{X,Y}$.

Распределение $\tilde{r}_{X,Y}$ для нормально распределенных величин X и Y существенно зависит от значения коэффициента корреляции $\rho_{X,Y}$ (оно неизвестно) и объема выборки n (задается исследователем) и имеет сложный вид, поэтому прибегают к специально подобранным функциям (преобразованиям) выборочного коэффициента корреляции, которые сводятся к хорошо изученным распределениям. Чаще всего применяется преобразование Фишера

$$\tilde{y} = \frac{1}{2}\ln\frac{1+\tilde{r}_{X,Y}}{1-\tilde{r}_{X,Y}}, \tag{4.11}$$

переводящее случайную величину $\tilde{r}_{X,Y}$ в нормально распределенную величину $\tilde{y} \sim N(\mu_{\tilde{y}}, \sigma_{\tilde{y}})$ с математическим ожиданием

$$\mu_{\tilde{y}} = \frac{1}{2}\ln\frac{1+\rho_{X,Y}}{1-\rho_{X,Y}} + \frac{\rho_{X,Y}}{2n-2} \tag{4.12}$$

и дисперсией $\sigma^2_{\tilde{y}} = 1/(n-3)$, зависящей только от объема выборки. Поправочный член $\rho_{X,Y}/(2n-2)$ в выражении (4.12) для $\mu_{\tilde{y}}$ мал по сравнению со среднеквадратичным отклонением $\sigma_{\tilde{y}}$, и им, как правило, пренебрегают. График преобразования $y(r)$ показан на рис.7.10. В табл. 7.5 представлены значения $y(r)$ только для $r \geq 0$, так как $y(r)$ - нечетная функция, то есть $y(-r) = -y(r)$.

От нормально распределенной величины $\tilde{y}$ можно перейти к нормально распределенной стандартизованной величине $\tilde{z} = (\tilde{y} - \sigma_{\tilde{y}})/\mu_{\tilde{y}}$ и выполнить интервальную оценку коэффициента корреляции, построив β-доверительный интервал

$$P\left(\left\{z_{\alpha/2} < \tilde{z} < z_{1-\alpha/2}\right\}\right) = \beta, \tag{4.13}$$

границами которого будут квантили стандартизованного нормального распределения, определенные по формуле:

$$P(\{\tilde{z} < z_{\alpha/2}\}) = \alpha/2,\ P(\{\tilde{z} < z_{1-\alpha/2}\}) = 1-\alpha/2 \Leftrightarrow P(\{\tilde{z} > z_{1-\alpha/2}\}) = \alpha/2, \quad (4.14)$$

причем $z_{\alpha/2} = -z_{1-\alpha/2}$.

С помощью обратного преобразования $\tilde{z} \mapsto \tilde{y} \mapsto \tilde{r}$ получается интервальная оценка для $\rho_{X,Y}$

$$\operatorname{th} y_{\alpha/2} < \rho_{X,Y} < \operatorname{th} y_{1-\alpha/2},$$

$$y_{\alpha/2} = \frac{1}{2}\ln\frac{1+r_{X,Y}}{1-r_{X,Y}} + \frac{z_{\alpha/2}}{\sqrt{n-3}} = \frac{1}{2}\ln\frac{1+r_{X,Y}}{1-r_{X,Y}} - \frac{z_{1-\alpha/2}}{\sqrt{n-3}},$$

$$y_{1-\alpha/2} = \frac{1}{2}\ln\frac{1+r_{X,Y}}{1-r_{X,Y}} + \frac{z_{1-\alpha/2}}{\sqrt{n-3}}, \quad (4.15)$$

$$\operatorname{th} y = \frac{e^{y} - e^{-y}}{e^{y} + e^{-y}}.$$

Функция “гиперболический тангенс” $\operatorname{th}(y)$ представлена графиком на рис. 7.11 и значениями в табл.7.6.

Стоит обратить внимание на то, что:

1) из-за нелинейности функции $\operatorname{th}(\cdot)$ доверительный интервал для коэффициента корреляции $\rho_{X,Y}$ получается несимметричным относительно значения выборочного коэффициента корреляции $r_{X,Y}$ (ср. с доверительным интервалом для математического ожидания (1.5), (1.6)) ;

2) при построенном доверительном интервале для коэффициента корреляции может быть проверена гипотеза $H_0 : \rho_{X,Y} = \rho_0$ − достаточно убедиться в том, что значение ρ_0 попадает в доверительный интервал;

3) интервальная оценка коэффициента корреляции $\rho_{X,Y}$ проводится для тех переменных X и Y, которые распределены нормально, даже если они не совпадают с исходными.

4.2. АЛГОРИТМ РЕШЕНИЯ

4.2.1. Построение диаграммы рассеяния и вычисление коэффициента корреляции. Выполняется в следующем порядке:

1) по выборочным значениям x_i, y_i построить диаграмму рассеяния;

2) вычислить точечные оценки параметров m_X, m_Y, $\hat{s}_X$, $\hat{s}_Y$ по (1.3);

3) вычислить значение выборочной ковариации $\hat{s}_{XY}$ по (4.1);

4) вычислить значение выборочного коэффициента корреляции r_{XY} по (4.2).

4.2.2. Проверка гипотезы $H_0: \rho_{X,Y} = 0$ о равенстве нулю коэффициента корреляции. Для ее выполнения необходимо:

1) вычислить по (4.6) статистику t^*;

2) для заданного значения уровня значимости α и числа степеней свободы ν=n-2 по табл. 7.4 вычислить значение квантиля распределения Стьюдента $t_{\nu,\,1-\alpha/2}$;

3) если $t^* > t_{\nu,\,1-\alpha/2}$, то отвергнуть гипотезу $H_0: \rho_{X,Y} = 0$ и сделать вывод о значимости коэффициента корреляции $\rho_{X,Y}$;

4) если коэффициент корреляции статистически значимо отличается от нуля, то выполнить интервальную оценку коэффициента корреляции по (4.13).

4.3. ПРИМЕР РАСЧЕТА

При сравнении двух методов (X и Y) определения содержания крахмала в картофеле, %, получено 16 результатов:

X	21.7	18.7	18.3	17.5	18.5	15.6	17.0	16.6	14.0	17.2	21.7	18.6	17.9	17.7	18.3	15.6
Y	21.5	18.7	18.3	17.4	18.3	15.4	16.7	16.9	13.9	17.0	21.4	18.6	18.0	17.6	18.5	15.5

4.3.1. Построение диаграммы рассеяния и вычисление коэффициента корреляции. По исходным данным вычисляем точечные оценки параметров выборок : m_X=17,806, m_Y=17,731, $\hat{s}_X$= 1,992, $\hat{s}_Y$= 1,969.

Вычисляем значение выборочной ковариации по (4.1)

$$\hat{s}_{X,Y} \approx \frac{58{,}655}{15} \approx 3{,}910.$$

Результаты промежуточных вычислений представлены в табл. 4.1. Значение выборочного коэффициента корреляции $r_{X,Y}$ равно

$$r_{X,Y} \approx \frac{3{,}910}{1{,}992 \cdot 1{,}969} \approx 0{,}996.$$

4.3.2. Проверка гипотезы $H_0: \rho_{X,Y} = 0$ о равенстве нулю коэффициента корреляции. Для этого вычисляется выборочное значение статистики t^* (4.6)

$$t^* \approx \sqrt{14}\,\frac{0{,}996}{\sqrt{1-0{,}996^2}} \approx 44{,}038.$$

Таблица 4.1

I	x_i	$(x_i - m_X)$	y_i	$(y_i - m_Y)$	$(x_i - m_X)\times(y_i - m_Y)$
1	21,7	3,899	21,5	3,769	14,695
2	18,7	0,894	18,7	0,969	0,866
3	18,3	0,494	18,3	0,569	0,281
4	17,5	-0,306	17,4	-0,331	0,101
5	18,5	0,694	18,3	0,569	0,395
6	15,6	-2,206	15,4	-2,331	5,142
7	17,0	-0,806	16,7	-1,031	0,831
8	16,6	-1,206	16,9	-0,831	1,002
9	14,0	-3,806	13,9	-3,831	14,581
10	17,2	-0,606	17,0	-0,731	0,443
11	21,7	3,894	21,4	3,669	14,287
12	18,6	0,794	18,6	0,869	0,690
13	17,9	0,094	18,0	0,269	0,025
14	17,7	-0,106	17,6	-0,131	0,014
15	18,3	0,494	18,5	0,769	0,380
16	15,6	-2,206	15,5	-2,231	4,922
					58,655

Найденное по табл.7.4 значение квантиля распределения Стьюдента $t_{14;0,975} = 2{,}145$. Поскольку $\left|t_{14}^*\right| > t_{14;\,0,975}$, то гипотеза $H_0 : \rho_{X,Y} = 0$ отвергается, вследствие чего можно говорить о значимости выборочного коэффициента корреляции. Для построения доверительного интервала для $\rho_{X,Y}$ вычисляются

$$y_{\alpha/2} = \frac{1}{2}\ln\frac{1+r_{xy}}{1-r_{xy}} - \frac{z_{1-\alpha/2}}{\sqrt{n-3}} = 2{,}563, \quad y_{1-\alpha/2} = \frac{1}{2}\ln\frac{1+r_{xy}}{1-r_{xy}} + \frac{z_{1-\alpha/2}}{\sqrt{n-3}} = 3{,}650,$$

$$\mathrm{th}\,y_{\alpha/2} = 0{,}988, \quad \mathrm{th}\,y_{1-\alpha/2} = 0{,}999.$$

Тогда 0,988 < $\rho_{X,Y}$ < 0,999, то есть между двумя методам определения содержания крахмала в картофеле существует сильная линейная корреляционная связь.

Диаграмма рассеяния представлена на рис. 4.1.

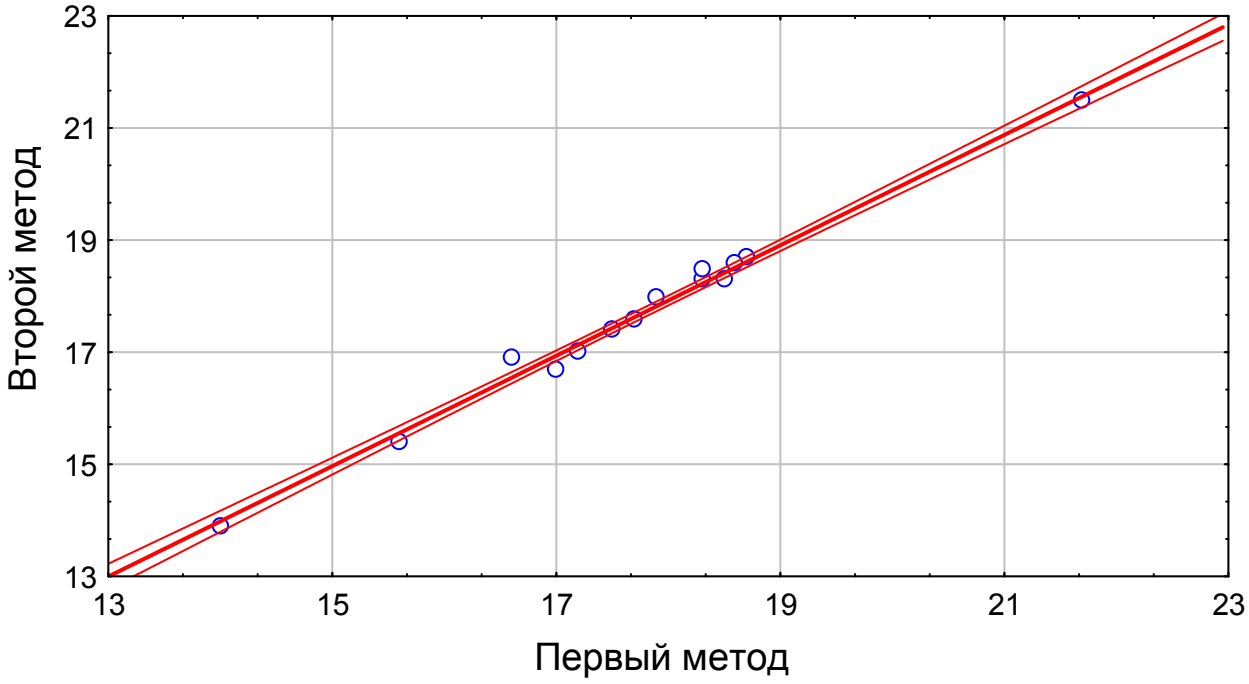

Рис. 4.1. Диаграмма рассеяния

5. РЕГРЕССИОННЫЙ АНАЛИЗ ЗАВИСИМЫХ РАСПРЕДЕЛЕНИЙ

5.1. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

5.1.1. Понятие регрессии двух зависимых распределений. С помощью корреляционного анализа вычисляется ***степень линейной связи*** двух зависимых случайных величин (признаков), однако установление вида связи (в том числе и нелинейной) возможно только с помощью ***регрессионного анализа***.

Кривой регрессии Y по X (или Y ***на*** X) называется условное математическое ожидание случайной величины Y, рассматриваемое как функция определенного класса, параметры которой оценивают по выборочным данным $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$:

$$M\,[Y \mid X=x]=\varphi(x;\ \text{параметры}). \qquad (5.1)$$

Аналогично определяется ***кривая регрессии X по Y*** (X ***на*** Y):

$$M\,[X \mid Y=y]=\psi(y;\ \text{параметры}). \qquad (5.2)$$

Если есть основания предполагать, что изучаемые случайные величины X и Y связаны некоторой зависимостью, тогда задача ее описания заключается в:

1) установлении общего вида зависимости;

2) точечном и интервальном оценивании параметров зависимости.

В статистике отсутствуют стандартные методы указания общего вида кривой регрессии. Поэтому на практике визуальный анализ диаграммы рассеяния (см. 4.1.2), как правило, предшествует проведению регрессионного анализа. Иногда точки на корреляционном поле ложатся так, что без априорного знания природы изучаемых величин трудно выбрать конкретный вид кривой регрессии.

Основными допущениями регрессионного анализа являются:

1) величины $x_1, x_2, \ldots, x_n$, которые выбираются экспериментатором;

2) результаты наблюдений $y_1, y_2, \ldots, y_n$ − выборка нормально распределенной случайной величины Y, то есть $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n$ − независимые нормально распределенные случайные величины;

3) дисперсии $\sigma^2_{\tilde{y}_i} = D[Y|X = x_i] = \sigma^2$, $\overline{i = 1, n}$, то есть равны.

5.1.2. Линейная регрессия. **Точечная оценка параметров линейной регрессии**. Наиболее часто используется модель линейной регрессии, позволяющая оценить линейную корреляционную зависимость между X и Y. Это означает, что для произвольного фиксированного значения X величина Y распределена в генеральной совокупности нормально с математическим ожиданием

$$M\,[Y \mid X=x]=\beta_0 + \beta_1 x \qquad (5.3)$$

и дисперсией σ^2, независимой от X (третье допущение о регрессии).

Для получения точечных оценок коэффициентов β_0 и β_1, соответственно b_0 и b_1, определяющих наиболее близкую к экспериментальным точкам регрессионную прямую, используется ***метод наименьших квадратов***. В этом методе "близость" измеряется суммой квадратов разностей между наблюдаемыми значениями y_i и значениями $\bar{y}_i = b_0 + b_1 x_i$, которые вычисляются по выборочному уравнению регрессии. Разница $d_i = y_i - \bar{y}_i$ между наблюдаемым y_i и оцениваемым по выборочному уравнению регрессии $\bar{y}_i$ значениями при $X = x_i$ называется ***отклонением***, или ***остатком***. В методе наименьших квадратов величины b_0 и b_1 находятся из условия минимума ***оценки*** − суммы квадратов отклонений

$$G(b_0, b_1) = \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_i)]^2 = \sum_{i=1}^{n} [y_i - \bar{y}_i]^2 = \sum_{i=1}^{n} d_i^2. \qquad (5.4)$$

Варьированием значений b_0 и b_1 можно добиться минимально возможного значения этой суммы квадратов отклонения: получается обычная задача на нахождение экстремума функции двух переменных. Поэтому для минимизации оценки G приравниваются к нулю частные производные G по b_0 и b_1

$$\begin{aligned} \frac{\partial G}{\partial b_0} &= -\sum_{i=1}^{n} [y_i - (b_0 + b_1 x_i)] = 0, \\ \frac{\partial G}{\partial b_1} &= -2\sum_{i=1}^{n} [y_i - (b_0 + b_1 x_i)] x_i = 0. \end{aligned} \qquad (5.5)$$

Решение системы (5.5), дающее выражения для коэффициентов b_0, b_1, имеет вид

$$b_0 = m_Y - r_{X,Y} \frac{\hat{s}_Y}{\hat{s}_X} m_X, \quad b_1 = r_{X,Y} \frac{\hat{s}_Y}{\hat{s}_X}, \qquad (5.6)$$

то есть получается ***выборочное уравнение регрессии***

$$\bar{y} = b_0 + b_1 x = m_Y - r_{X,Y} \frac{\hat{s}_Y}{\hat{s}_X} m_X + \left(r_{X,Y} \frac{\hat{s}_Y}{\hat{s}_X} \right) x \qquad (5.7)$$

или в симметричной форме, удобной для запоминания:

$$\bar{y} - m_Y = r_{X,Y} \frac{\hat{s}_Y}{\hat{s}_X} (x - m_X). \qquad (5.8)$$

Это будет наилучшая, в смысле метода наименьших квадратов, прямая среди всех прямых на плоскости, которая приближенно описывает закон изменения условного среднего величины Y в зависимости от изменения значений величины X.

5.1.3. Проверка гипотезы о равенстве коэффициента регрессии нулю. Выборочный коэффициент b_1 является точечной оценкой ***коэффициента регрессии*** β_1, задающего наклон прямой (5.3) к оси x. Часто оказывается необходимым выяснить, достаточна ли точечная оценка b_1

для статистически обоснованного вывода о том, что коэффициент регрессии β_1 отличен от нуля (статистически значимо отличается от нуля). Для этого проверяется гипотеза H_0: $\beta_1 = 0$.

Основанием для построения ***критерия проверки гипотезы*** служит ***предположение о нормальности распределения величины*** Y, откуда следует, что величина $n\tilde{\hat{s}}_d^2 / \sigma^2$ имеет ***распределение Пирсона*** $\tilde{\chi}_\nu^2$ с числом степеней свободы ν=n-2, где $\hat{s}_d^2$ − остаточная дисперсия

$$\hat{s}_d^2 = \frac{1}{n-2}\sum_{i=1}^{n} d_i^2, \qquad d_i = y_i - \bar{y}_i. \tag{5.9}$$

Статистикой критерия служит величина

$$\tilde{F} = \frac{\sum_{i=1}^{n} (\tilde{\bar{y}}_i - \tilde{m}_Y)^2}{\tilde{s}_d^2}, \tag{5.10}$$

распределенная по закону Фишера $\tilde{F}_{\nu_1, \nu_2}$ с числом степеней свободы числителя $\nu_1 = 1$ и знаменателя $\nu_2 = n - 2$.

Следовательно, можно построить β-доверительный интервал $\tilde{F}$

$$P\left(\left\{F_{\nu_1, \nu_2, \alpha/2} < \tilde{F} < F_{\nu_1, \nu_2, 1-\alpha/2}\right\}\right) = \beta, \tag{5.11}$$

границы которого − квантили $F_{\nu_1, \nu_2, \alpha/2}$ и $F_{\nu_1, \nu_2, 1-\alpha/2}$ − определены так: $P\left(\left\{\tilde{F} < F_{\nu_1, \nu_2, \alpha/2}\right\}\right) = \alpha/2$, $P\left(\left\{\tilde{F} < F_{\nu_1, \nu_2, 1-\alpha/2}\right\}\right) = 1 - \alpha/2$.

Если выборочное значение F^*, вычисленное по (5.10), попадает в доверительный интервал (5.11), то ***гипотеза*** H_0: $\beta_1 = 0$ ***принимается***, в противном случае ***гипотеза отвергается***.

F-критерий очень ***чувствителен*** к отклонениям исследуемых распределений от нормального. Поэтому его нельзя использовать, если есть сомнения относительно нормальности распределений.

В случае отклонения гипотезы H_0: $\beta_1 = 0$ коэффициент регрессии статистически значимо отличается от нуля (***коэффициент регрессии значим***).

5.1.4. Интервальная оценка параметров линейной регрессии. Если коэффициент регрессии оказывается значимым (между величинами

имеется линейная связь), то можно построить β-доверительные интервалы для параметров линейной регрессии. Построение доверительных интервалов основывается на статистиках

$$\tilde{t} = \frac{\tilde{b}_0 - \beta_0}{\tilde{S}_m(\tilde{b}_0)}, \quad \tilde{t} = \frac{\tilde{b}_1 - \beta_1}{\tilde{S}_m(\tilde{b}_1)}, \tag{5.12}$$

распределенных по закону Стьюдента с числом степеней свободы ν=n-2; $S_m(b_0)$, $S_m(b_1)$ **−** ***стандартные ошибки*** b_0 и b_1 (ср. с (1.6)):

$$S_m(b_0) = \frac{\hat{s}_d^2 - \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - m_X)^2} = \frac{\hat{s}_d^2\left[(n-1)\hat{s}_X^2 + nm_X^2\right]}{n(n-1)\hat{s}_X^2},$$

$$S_m(b_1) = \frac{\hat{s}_d^2}{\sum_{i=1}^{n}(x_i - m_X)^2} = \frac{\hat{s}_d^2}{(n-1)\hat{s}_X^2}. \tag{5.13}$$

Следовательно, для уровня значимости α интервальные оценки параметров линейной регрессии имеют вид

$$\begin{aligned} b_0 - S_m(b_0)\cdot t_{n-2,1-\alpha/2} &< \beta_0 < b_0 + S_m(b_0)\cdot t_{n-2,1-\alpha/2}, \\ b_1 - S_m(b_1)\cdot t_{n-2,1-\alpha/2} &< \beta_1 < b_1 + S_m(b_1)\cdot t_{n-2,1-\alpha/2}. \end{aligned} \tag{5.14}$$

5.1.5. Проверка гипотез о значениях параметров линейной регрессии. Если есть основания предполагать, что значения параметров линейной регрессии равны некоторым значениям, соответственно $\beta_0^{(0)}$ и $\beta_1^{(0)}$, то следует проверить гипотезы H_0: $\beta_0 = \beta_0^{(0)}$ и H_0: $\beta_1 = \beta_1^{(0)}$. Для заданного уровня значимости ***гипотезы принимаются*** , если значения $\beta_0^{(0)}$ и $\beta_1^{(0)}$ попадают в доверительные интервалы (5.14), в противном случае ***гипотезы отвергаются***.

5.2. АЛГОРИТМ РЕШЕНИЯ

5.2.1. Выборочное уравнение регрессии. Для его нахождения необходимо:

1) построить корреляционное поле;

2) по выборочным данным x_i, y_i, $i=\overline{1,n}$ вычислить точечные оценки m_X, m_Y, $\hat{s}_X$, $\hat{s}_Y$;

3) вычислить значение выборочного коэффициента корреляции $r_{X,Y}$;

4) вычислить коэффициенты уравнения линейной регрессии b_0, b_1 по (5.6) и получить выборочное уравнение регрессии

$$\bar{y}=b_0+b_1x;$$

5) на корреляционное поле нанести линию регрессии.

5.2.2. Проверка гипотезы о равенстве коэффициента регрессии нулю $H_0:\beta_1=0$**.** Осуществляется следующим образом:

1) по выборочному уравнению регрессии для значений x_i, $i=\overline{1,n}$ вычислить теоретические значения $\bar{y}_i=b_0+b_1x_i$;

2) вычислить по (5.9) остаточную дисперсию отклонений экспериментальных данных y_i от линии регрессии;

3) вычислить по (5.10) значение статистики F^*;

4) для заданного уровня значимости α и степеней свободы ν_1=1 ν_2=n-2 по табл. 7.3 вычислить значение квантиля распределения Фишера $F_{\nu_1,\nu_2,1-\alpha/2}$;

5) если $F^*>F_{\nu_1,\nu_2,1-\alpha/2}$, то отвергнуть гипотезу $H_0:\beta_1=0$, в противном случае принять коэффициент регрессии равным нулю и прекратить вычисления.

5.2.3. Интервальная оценка параметров линейной регрессии. Для ее нахождения необходимо:

1) по точечным оценкам параметров выборок m_X, $\hat{s}_X$ и остаточной дисперсии $\hat{s}_d^2$ по (5.13) вычислить для b_0 и b_1 значения стандартных ошибок $S_m(b_0)$, $S_m(b_1)$;

2) для уровня значимости α и числа степеней свободы ν=n-2 по табл.7.4 определить квантиль распределения Стьюдента $t_{\nu,1-\alpha/2}$;

3) вычислить по (5.14) интервальные оценки коэффициентов линейной регрессии.

5.3. ПРИМЕР РАСЧЕТА

Зависимость веса ребенка Y, унции (1 унция=28,35 г), от возраста X, недели, представлены следующими данными:

X	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Y	110	114	128	147	163	172	186	198	208	213	232	240	254	261

5.3.1. Выборочное уравнение регрессии. Строим корреляционное поле, представленное на рис.5.1. Для исходных данных по (1.3) вычисляем: m_X=6,5, m_Y=187,5, $\hat{s}_X$=4,183, $\hat{s}_Y$=50,515.

Данные для вычисления выборочного коэффициента ковариации $\hat{s}_{X,Y}$, вычисленные по (4.1), представлены в табл. 5.1.

Таблица 5.1

i	x_i	$(x_i - m_X)$	y_i	$(y_i - m_Y)$	$(x_i - m_X)\times(y_i - m_Y)$
1	0	−6,5	110	−77,57	504,25
2	1	−5,5	114	−73,57	404,635
3	2	−4,5	128	−59,57	268,065
4	3	−3,5	147	−40,57	141,995
5	4	−2,5	163	−24,57	61,425
6	5	−1,5	172	−15,57	23,355
7	6	−0,5	186	− 1,57	0,785
8	7	0,5	198	10,43	5,215
9	8	1,5	208	20,43	30,645
10	9	2,5	213	25,43	63,575
11	10	3,5	232	44,43	155,505
12	11	4,5	240	52,43	235,935
13	12	5,5	254	66,43	365,365
14	13	6,5	261	73,43	477,295
					2738,00

По этим данным получаем $\hat{s}_{X,Y}$=2738,0/13 ≈210,615, а по (4.2) вычисляем значение выборочного коэффициента корреляции

$$r_{X,Y} = \frac{\hat{s}_{X,Y}}{\hat{s}_X \hat{s}_Y} = \frac{210{,}615}{4{,}183 \cdot 50{,}515} \approx 0{,}9967 .$$

Тогда точечные оценки коэффициентов уравнения линейной регрессии согласно (5.6) равны

$$b_0 = m_Y - r_{X,Y}\frac{\hat{s}_Y}{\hat{s}_X} m_X = 187{,}57 - 0{,}9967\frac{50{,}515}{4{,}183}6{,}5 \approx 109{,}336 ,$$

$$b_1 = r_{X,Y}\frac{\hat{s}_Y}{\hat{s}_X} = 0{,}9967\frac{50{,}515}{4{,}183} \approx 12{,}036 .$$

Таким образом, уравнение линейной регрессии имеет вид

$$\bar{y} = 109{,}336 + 12{,}036\,x\,,$$

что говорит о существовании линейной положительной связи между возрастом и весом детей в первые недели развития.

5.3.2. Проверка гипотезы о равенстве коэффициента регрессии нулю $H_0\colon \beta_1 = 0$. Для подтверждения достоверности такой связи покажем, что коэффициент регрессии b_1 статистически значимо отличается от нуля. Сформулируем гипотезу $H_0\colon \beta_1 = 0$. Необходимые данные для вычисления остаточной дисперсии $\hat{s}_d^2$ приведены в табл. 5.2, откуда по (5.9)

$$\hat{s}_d^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left[y_i - \bar{y}(x_i)\right]^2 \approx \frac{221{,}409}{12} \approx 18{,}451.$$

Таблица 5.2

i	x_i	y_i	$\bar{y}_i$	$d_i^2 = [y_i - \bar{y}_i]^2$	$[\bar{y}_i - m_Y]^2$
1	0	110	109,34	0,441	6119,93
2	1	114	121,372	54,346	4382,18
3	2	128	133,408	29,246	2933,52
4	3	147	145,444	2,42	1774,94
5	4	163	157,48	30,47	905,408
6	5	172	169,516	6,17	325,95
7	6	186	181,552	19,78	36,22
8	7	198	193,588	19,73	36,22
9	8	208	205,624	5,645	325,95
10	9	213	217,66	21,72	905,408
11	10	232	229,696	5,308	1774,94
12	11	240	241,732	2,999	2933,52
13	12	254	253,768	0,054	4382,189
14	13	261	265,804	23,08	6120,56
				221,409	32956,926

В табл.5.2 представлены также и значения $[\bar{y}_i - m_y]^2$, необходимые для вычисления выборочного значения статистики критерия Фишера F^* F^*≈32956,926/18,451≈1786,19. По табл. 7.3 для α=0,05 и ν_1=1, ν_2=12 находим квантиль $F_{\nu_1,\nu_2,1-\alpha/2}$≈6,55. Поскольку F^*>$F_{\nu_1,\nu_2,1-\alpha/2}$, то отклоняем

гипотезу H_0: $\beta_1 = 0$. Таким образом, коэффициент линейной регрессии b_1 значимо отличается от нуля.

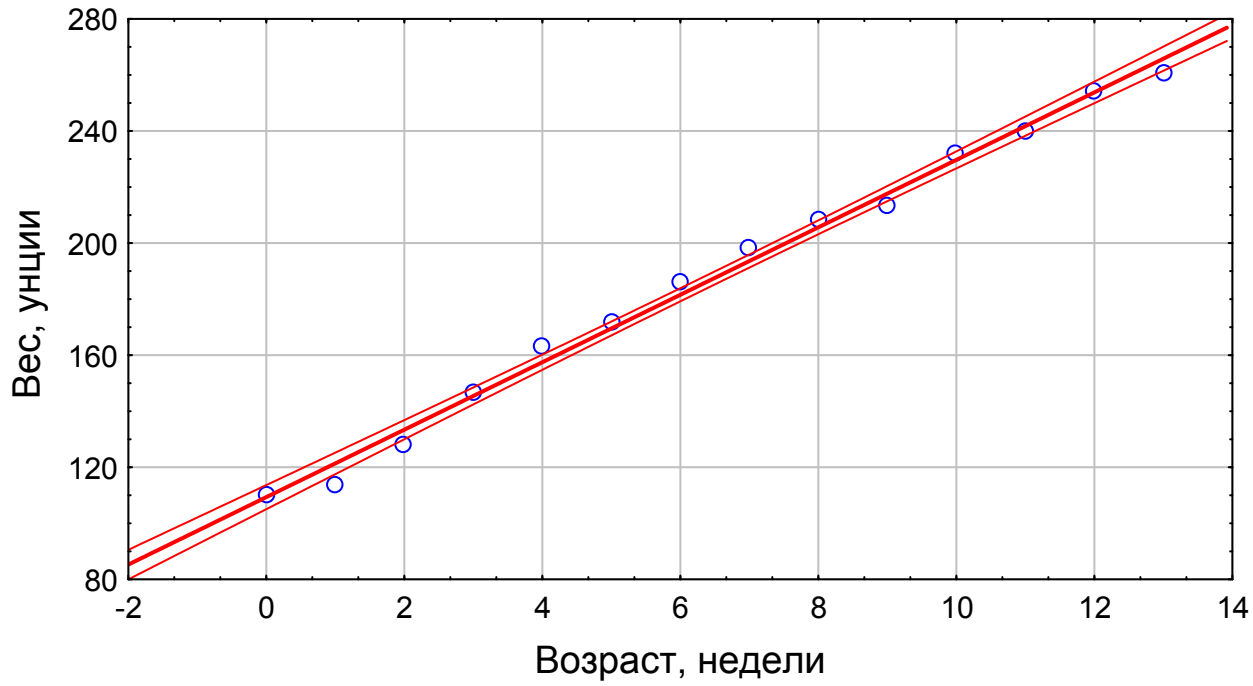

Рис 5.1. Корреляционное поле и линия регрессии

5.3.3. Интервальная оценка параметров линейной регрессии. Используя значения точечных оценок m_X, $\hat{s}_Y$ и остаточной дисперсии $\hat{s}_d^2$, вычисляем по (5.13) стандартные ошибки $S_m(b_0)$, $S_m(b_1)$ коэффициентов уравнения линейной регрессии b_0 и b_1

$$S_m(b_0) = \frac{\hat{s}_d^2\left[(n-1)\hat{s}_X^2 + nm_X^2\right]}{n(n-1)\hat{s}_X^2} \approx \frac{18{,}451\left(13\cdot 4{,}183^2 + 14\cdot 6{,}5^2\right)}{14\cdot 13\cdot 4{,}183^2} \approx 4{,}745 \quad ,$$

$$S_m(b_1) = \frac{\hat{s}_d^2}{(n-1)\hat{s}_X^2} \approx \frac{18{,}451}{13\cdot 4{,}183^2} \approx 0{,}081.$$

Для α=0,05 и ν=n-2=12 определяем по табл. 7.4 значение квантиля распределения Стьюдента $t_{\nu,1-\alpha/2}$≈2,18 .Тогда

$$S_m(b_0)\ t_{n-2,1-\alpha/2} \approx 4{,}745\cdot 2{,}18 \approx 10{,}344,$$

$$S_m(b_1)\ t_{n-2,1-\alpha/2} \approx 0{,}081\cdot 2{,}18 \approx 0{,}177.$$

Поэтому интервальные оценки коэффициентов уравнения линейной регрессии согласно (5.14) имеют вид

$$98{,}99 < \beta_0 < 119{,}68, \qquad 11{,}86 < \beta_1 < 12{,}21.$$

6. ЗАДАЧИ К КУРСОВОМУ ЗАДАНИЮ

6.1. ТОЧЕЧНАЯ И ИНТЕРВАЛЬНАЯ ОЦЕНКА ЧИСЛОВЫХ ХАРАКТЕРИСТИК РАСПРЕДЕЛЕНИЯ, ЭМПИРИЧЕСКИЕ ФУНКЦИЯ И ЗАКОН РАСПРЕДЕЛЕНИЯ

1. Постройте вариационный и интервальный ряды, гистограмму и эмпирическую функцию распределения для указанной переменной X из выборки.

2. Проверьте гипотезу о согласовании эмпирического распределения с нормальным.

3. Выполните точечную и интервальную оценку $m, \hat{s}$ параметров μ, σ распределения.

6.1.1. Распределение по весу ягод клубники, г (объем выборки n=65):

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

6.1.2. Распределение по весу ягод земляники, г (объем выборки n=121):

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

6.1.3. Концентрация витамина C в образцах свежего томатного сока, мг/100г (объем выборки n=50):

37 41 43 39 39 36 38 40 44 40 38 41 42 38 36 39 41 42 40 45 41
38 34 37 39 37 42 40 47 41 39 43 39 35 40 34 39 46 33 37 41 39
39 42 40 43 44 42 38 46 37

6.1.4. Концентрация витамина C в образцах консервированного томатного сока, мг/100 г (объем выборки n=56):

16 22 21 20 23 21 19 15 13 23 17 20 29 18 22 16 25 23 26 20 17
19 27 22 23 22 21 18 21 24 20 23 20 19 22 21 24 21 19 18 25 21
20 23 21 19 24 22 29 21 19 24 22 19 25 15

6.1.5. Длина межузлий у растений диплоидной ржи, см (объем выборки n=51):

11,5 10,9 11,6 11,9 11,5 11,5 11,1 12,4 10,8 11,3 11,7 12,1 11,5 11,3
11,8 11,6 11,4 11,6 11,0 12,3 11,6 11,4 11,6 11,3 12,2 11,5 11,4 11,6
11,1 11,6 11,4 12,0 11,7 11,4 11,2 11,5 11,8 12,1 11,6 11,4 11,7 12,0
11,5 11,8 11,5 11,3 11,4 11,6 11,2 11,7 11,3

6.1.6. Длина межузлий у растений тетраплоидной ржи, см (объем выборки n=60):

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

6.1.7. Длина крыла полевого жаворонка, мм (объем выборки n=81):

99 102 105 97 103 97 101 94 105 110 86 112 111 103 115
99 102 104 95 100 105 102 121 92 96 87 99 102 107 104
98 112 89 101 99 86 118 105 101 88 99 105 101 98 84
102 104 97 101 101 102 116 95 89 88 109 111 102 95 96
91 95 115 102 107 111 87 102 85 98 107 120 100 112 115
116 109 97 100 102 96 101 104 102 106 106 106 100 89 106
102

6.1.8. Длина крыла скворца, мм (объем выборки n=68):

121 119 123 117 123 120 122 122 118 123 120 119 124 120 122
122 121 119 121 121 124 118 121 120 125 120 121 122 120 124
121 120 121 119 123 119 123 121 120 127 119 121 123 122 121
120 120 122 125 123 124 121 121 120 123 123 120 122 121 123
121 124 119 121 123 122 118 120

6.1.9. Содержание кофеина в партии кофе A, $\%\cdot 10^3$ (объем выборки n=97):

429 443 451 455 440 436 433 396 383 401 403 399 423 419 421
405 393 361 409 401 398 391 441 424 393 406 398 396 417 417
422 408 387 375 413 404 399 375 446 420 389 407 397 392 432
418 412 386 399 410 401 388 375 386 443 416 409 395 410 396
391 430 420 405 398 383 406 398 383 406 398 383 406 398 393
389 393 405 409 400 410 392 388 418 417 405 396 389 417 405
396 389 417 382 401 379 394

6.1.10. Содержание кофеина в партии кофе B, $\%\cdot 10^3$ (объем выборки n=85):

433 423 412 411 426 426 427 431 436 432 408 414 417 400 404
413 403 402 384 392 412 411 414 412 404 409 407 407 402 409
391 395 391 387 398 395 405 414 425 433 420 418 404 403 418
421 422 428 429 428 411 407 411 413 409 392 391 399 407 406
393 396 387 380 359 421 407 384 393 410 399 376 389 395 397

384 400 402 398 397 389 391 385 395 398

6.1.11. Продолжительность развития эмбрионов кроликов-альбиносов, дни (объем выборки n=55):

30 32 30 30 31 33 31 29 32 30 30 31 33 31 29 31 35 30
30 31 31 32 35 30 30 31 30 30 32 30 32 31 29 30 31 34
31 30 36 30 32 29 33 30 31 32 30 34 30 32 29 36 30 31
31

6.1.12. Продолжительность развития эмбрионов кроликов породы шиншила, дни (объем выборки n=48):

32 34 30 30 31 32 30 30 29 33 30 31 33 32 31 31 32 32
31 31 30 32 30 30 30 32 31 29 31 31 30 32 30 33 32 30
31 31 29 30 30 32 30 31 32 30 31 32

6.1.13. Интервалы между разрядами нейрона зрительной коры кролика, мкс (объем выборки n=64):

200 8 14 53 256 56 74 35 23 19 39 15 11 18 63 19 30
30 15 5 13 50 25 4 42 26 16 46 36 27 50 10 130 70
43 5 45 2 8 19 44 54 49 32 55 53 47 73 11 58 121
144 53 265 4 57 75 108 65 14 11 7 98 140

6.2. СРАВНЕНИЕ ПАРАМЕТРОВ ДВУХ НЕЗАВИСИМЫХ РАСПЕРЕДЕЛЕНИЙ

Сформулируйте и проверьте соответствующую статистическую гипотезу для указанных независимых переменных X и Y. Обоснуйте выбор критерия проверки и уровня значимости α. Предварительно проведите проверку нормальности распределения случайных величин X и Y, если объемы выборок больше 20. Для малых выборок сделайте предположение о нормальности.

6.2.1. Систолическое давление в аорте (X) и больших артериях конечностей (Y), мм рт. ст. взрослых людей среднего возраста:

X (объем выборки n=50)

120 122 115 118 125 110 120 124 113 120 108 123 132 115 119 104 117 111
124 116 116 128 115 124 130 120 111 116 118 129 123 115 121 131 117 110
125 119 128 124 120 105 125 109 112 114 126 130 115

Y (объем выборки n=45)

106 114 103 109 115 105 94 107 105 121 112 108 95 109 94 106 110 118
104 110 86 103 100 114 86 110 105 114 114 95 107 120 101 108 95 106
102 119 108 104 111 88 115 110 124

6.2.2. Данные о количестве воды, которую выпивает человек в течение суток в покое (X) и при физической работе в условиях жаркого климата (Y), л:

X (объем выборки n=53)

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

Y (объем выборки n=68)

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

6.2.3. Концентрация гемоглобина в крови мальчиков 7 лет (X), и в крови девочек 7 лет (Y), г/100 мл:

X (объем выборки n=55)

13,1	13,3	12,7	13,1	13,2	13,0	12,2	13,6	13,3	13,1	13,2	13,2	12,9	13,4
12,9	13,4	12,9	13,5	13,2	14,0	13,0	13,3	13,1	13,3	12,9	13,2	13,4	13,8
13,2	13,0	13,5	12,8	13,2	13,1	13,2	12,9	13,3	13,6	12,8	13,2	13,4	13,0
13,7	13,4	13,9	13,3	13,1	13,1	12,6	13,5	13,3	13,1	13,8	13,1	13,4	

Y (объем выборки n=53)

13,6	13,1	13,1	13,1	12,8	13,0	13,3	13,2	13,3	13,1	12,9	13,5	13,3	13,0
12,6	13,1	13,4	13,2	13,7	13,1	13,0	12,7	13,2	13,8	13,2	13,0	13,3	12,8
13,2	13,5	13,6	13,1	13,4	13,0	13,0	13,6	13,1	13,5	13,3	13,1	13,5	13,4
13,0	12,9	13,3	13,2	12,9	13,4	13,2	13,4	12,9	13,1	12,5			

6.2.4. Концентрация гемоглобина в крови мальчиков 14-16 лет (X) и в крови девочек 14-16 лет (Y), г/100 мл:

X (объем выборки n=50)

14,9	14,0	14,8	14,7	15,1	14,9	14,8	15,2	14,8	14,9	15,1	15,0	14,6	15,0
14,9	14,7	14,2	14,8	14,9	14,7	15,0	14,8	14,9	14,7	15,5	14,3	14,8	14,7
14,9	14,3	15,4	14,8	14,9	15,0	14,6	15,1	14,8	14,7	15,1	14,5	14,9	15,0
15,2	15,7	14,9	14,7	15,0	14,8	15,3	14,7						

Y (объем выборки n=65)

14,3	14,6	14,4	13,9	14,6	14,4	15,0	14,2	14,1	13,5	14,3	14,1	14,5	14,4
14,2	14,6	13,7	14,2	14,4	14,5	14,5	14,9	14,2	14,0	14,7	14,5	14,3	14,7
14,5	14,3	14,0	13,8	14,4	14,3	14,7	14,4	14,2	13,9	14,4	14,6	14,5	14,3
14,3	13,6	14,4	14,1	14,6	14,0	14,6	14,4	14,2	14,7	14,4	14,2	14,1	14,1
14,5	14,3	14,3	14,8	14,8	14,5	14,0	14,7	14,7					

6.2.5. Концентрация гемоглобина в крови мальчиков 7 лет (X) и 14-16 лет (Y), г/100 мл:

X (объем выборки n=55)

13,1	13,3	12,7	13,1	13,2	13,0	12,2	13,6	13,3	13,1	13,2	13,2	12,9
13,4	12,9	13,4	12,9	13,5	13,2	14,0	13,0	13,3	13,1	13,3	12,9	13,2
13,4	13,8	13,2	13,0	13,5	12,8	13,2	13,1	13,2	12,9	13,3	13,6	12,8

13,2 13,4 13,0 13,7 13,4 13,9 13,3 13,1 13,1 12,6 13,5 13,3 13,1
13,8 13,1 13,4

Y (объем выборки n=50)

14,9 14,0 14,8 14,7 15,1 14,9 14,8 15,2 14,8 14,9 15,1 15,0 14,6
15,0 14,9 14,7 14,2 14,8 14,9 14,7 15,0 14,8 14,9 14,7 15,5 14,3
14,8 14,7 14,9 14,3 15,4 14,8 14,9 15,0 14,6 15,1 14,8 14,7 15,1
14,5 14,9 15,0 15,2 15,7 14,9 14,7 15,0 14,8 15,3 14,7

6.2.6. Концентрация гемоглобина в крови девочек 7 лет (X) и 14-16 лет (Y), г/100 мл:

X (объем выборки n=53)

13,6 13,1 13,1 13,1 12,8 13,0 13,3 13,2 13,3 13,1 12,9 13,5 13,3
13,0 12,6 13,1 13,4 13,2 13,7 13,1 13,0 12,7 13,2 13,8 13,2 13,0
13,3 12,8 13,2 13,5 13,6 13,1 13,4 13,0 13,0 13,6 13,1 13,5 13,3
13,1 13,5 13,4 13,0 12,9 13,3 13,2 12,9 13,4 13,2 13,4 12,9 13,1 12,5

Y (объем выборки n=65)

14,3 14,6 14,4 13,9 14,6 14,4 15,0 14,2 14,1 13,5 14,3 14,1 14,5 14,4
14,2 14,6 13,7 14,2 14,4 14,5 14,5 14,9 14,2 14,0 14,7 14,5 14,3 14,7
14,5 14,3 14,0 13,8 14,4 14,3 14,7 14,4 14,2 13,9 14,4 14,6 14,5 14,3
14,3 13,6 14,4 14,1 14,6 14,0 14,6 14,4 14,2 14,7 14,4 14,2 14,1 14,1
14,5 14,3 14,3 14,8 14,8 14,5 14,0 14,7 14,7

6.2.7. Концентрация холестерина в крови белых американцев (X) и представителей некоторых племен Нигерии (Y), мг/100 мл:

X (объем выборки n=50)

162 203 214 187 214 210 195 242 216 198 210 214 208 226 218
155 184 215 234 205 133 232 147 206 225 242 216 191 211 255
177 198 210 182 224 227 173 251 166 208 195 183 226 243 194
217 238 209 240 212

Y (объем выборки n=65)

100 118 85 121 107 175 155 137 110 98 125 184 130 95 150
105 173 161 132 119 152 111 172 112 158 167 180 122 87 118
135 102 88 171 155 135 124 110 121 134 126 160 125 165 138
118 169 123 120 140 115 142 131 141 138 184 130 144 131 145
116 140 107 142 122

6.2.8. Коэффициент отражения волос у европейцев-блондинов (X) и у европейцев-брюнетов (Y) при длине световой волны 650 мкм, %:

X (объем выборки n=64)

25,1 23,7 25,7 24,1 28,0 24,5 25,3 25,8 24,6 25,4 24,6 27,0 26,4
26,8 24,7 25,3 25,1 26,4 25,5 25,0 23,5 26,1 25,6 26,0 24,9 25,5
24,8 27,0 25,6 25,3 24,8 25,0 25,1 25,6 27,5 24,9 25,1 24,9 25,5
25,9 26,0 24,5 26,5 25,2 25,3 24,8 27,4 26,8 24,9 23,6 26,8 24,6
25,7 25,1 26,2 24,5 25,7 25,3 24,5 24,0 24,9 26,1 23,0 23,5

Y (объем выборки n=45):

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

6.2.9. Коэффициент отражения волос у европейцев-блондинов (X) и у европейцев-альбиносов (Y) при длине световой волны 650 мкм, %:

X (объем выборки n=64):

25,1 23,7 25,7 24,1 28,0 24,5 25,3 25,8 24,6 25,4 24,6 27,0 26,4
26,8 24,7 25,3 25,1 26,4 25,5 25,0 23,5 26,1 25,6 26,0 24,9 25,5
24,8 27,0 25,6 25,3 24,8 25,0 25,1 25,6 27,5 24,9 25,1 24,9 25,5
25,9 26,0 24,5 26,5 25,2 25,3 24,8 27,4 26,8 24,9 23,6 26,8 24,6
25,7 25,1 26,2 24,5 25,7 25,3 24,5 24,0 24,9 26,1 23,0 23,5

Y (объем выборки n=50):

54,0 54,9 54,8 55,0 53,8 57,0 55,2 54,6 55,0 54,1 52,8 53,5 54,6
56,0 55,5 54,4 57,4 54,4 54,9 53,9 56,8 54,7 53,8 54,7 56,9 54,2
54,3 55,1 54,8 54,0 52,5 57,6 56,0 54,0 55,0 55,6 54,4 55,1 53,5
55,0 55,9 55,2 53,5 54,4 55,2 54,4 54,0 56,8 55,9 55,0

6.2.10. Коэффициент отражения волос у европейцев-блондинов (X) и у африканских негров (Y) при длине световой волны 650 мкм , %:

X (объем выборки n=64)

25,1 23,7 25,7 24,1 28,0 24,5 25,3 25,8 24,6 25,4 24,6 27,0 26,4
26,8 24,7 25,3 25,1 26,4 25,5 25,0 23,5 26,1 25,6 26,0 24,9 25,5
24,8 27,0 25,6 25,3 24,8 25,0 25,1 25,6 27,5 24,9 25,1 24,9 25,5
25,9 26,0 24,5 26,5 25,2 25,3 24,8 27,4 26,8 24,9 23,6 26,8 24,6
25,7 25,1 26,2 24,5 25,7 25,3 24,5 24,0 24,9 26,1 23,0 23,5

Y (объем выборки n=55)

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

6.2.11. Коэффициент отражения волос у европейцев-брюнетов (X) и у европейцев-альбиносов (Y) при длине световой волны 650 мкм, %:

X (объем выборки n=45)

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

Y (объем выборки n=50)

54,0 54,9 54,8 55,0 53,8 57,0 55,2 54,6 55,0 54,1 52,8 53,5 54,6
56,0 55,5 54,4 57,4 54,4 54,9 53,9 56,8 54,7 53,8 54,7 56,9 54,2
54,3 55,1 54,8 54,0 52,5 57,6 56,0 54,0 55,0 55,6 54,4 55,1 53,5
55,0 55,9 55,2 53,5 54,4 55,2 54,4 54,0 56,8 55,9 55,0

6.2.12. Коэффициент отражения волос у европейцев-брюнетов (X) и у африканских негров (Y) при длине световой волны 650 мкм, %:

X (объем выборки n=45)

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

Y (объем выборки n=55)

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

6.2.13. Коэффициент отражения волос у европейцев-альбиносов (X) и у африканских негров (Y) при длине световой волны 650 мкм , %:

X (объем выборки n=50)

54,0 54,9 54,8 55,0 53,8 57,0 55,2 54,6 55,0 54,1 52,8 53,5 54,6
56,0 55,5 54,4 57,4 54,4 54,9 53,9 56,8 54,7 53,8 54,7 56,9 54,2
54,3 55,1 54,8 54,0 52,5 57,6 56,0 54,0 55,0 55,6 54,4 55,1 53,5
55,0 55,9 55,2 53,5 54,4 55,2 54,4 54,0 56,8 55,9 55,0

Y (объем выборки n=55)

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

6.3. СРАВНЕНИЕ ПАРАМЕТРОВ ДВУХ ЗАВИСИМЫХ РАСПРЕДЕЛЕНИЙ

Сформулируйте и проверьте соответствующую статистическую гипотезу для указанных независимых переменных X и Y. Обоснуйте выбор критерия проверки и уровня значимости α. Предварительно проведите проверку нормальности распределения случайных величин X и Y, если объемы выборок больше 20. Для малых выборок сделайте предположение о нормальности.

6.3.1. Скорость кровотока, мл/с, у спортсменов-лыжников до (X) и после (Y) бега на 15 км:

X 8,6 7,4 8,3 6,2 5,5 5,1 5,7 8,2 8,7 9,7 7,0 6,5
Y 5,8 5,2 6,6 4,8 4,4 4,1 5,7 7,3 7,6 7,3 4,7 7,4

6.3.2. При прочтении заикающимися одного и того же текста несколько раз подряд число заиканий уменьшается. Электрошок как метод борьбы с заиканиями претендует на ускорение этого процесса (аккомодацию). Цель описанного эксперимента − ответить на вопрос: так ли это? Восемнадцать заикающихся студентов читали вслух последовательно до пяти

раз в двух различных ситуациях два различных фрагмента текста. Одна ситуация была контрольной и никаких действий не предусматривала (X). В другой ситуации они испытывали действие электрошока в момент заикания (Y). В результате эксперимента по специальной методике получены следующие оценки аккомодации (в баллах, высокий балл отвечает лучшей аккомодации):

X 57 59 41 51 43 49 48 56 44 50 44 50 70 42 68 54 38 48
Y 51 56 44 44 50 54 50 40 50 50 56 46 74 57 74 48 48 44

Замечание. На Украине электрошок с лечебной целью не используется.

6.3.3. Содержание остаточного азота в крови больных, мг/% до (X) и после (Y) операции:

X 38 42 40 38 38 38 40 42 38 40 36 34 38 40 40 42 40 40 42 40
Y 36 40 40 36 40 38 40 36 40 36 44 30 40 42 40 36 34 38 36 38

6.3.4. Содержание адреналина, мг/л у онкологических больных до (X) и после (Y) приема препарата, который стимулирует симпатоадреналовую систему:

X 0,5 0,4 0,9 1,1 0,8 0,5 0,9 1,0 0,6 0,7
Y 0,6 0,2 0,5 0,2 0,6 0,7 0,4 0,3 0,6 0,3

6.3.5. Количество метилникотинамида, мг выведенного из организма больных на протяжении первого (X) и второго (Y) дня после операции:

X 1,96 3,12 5,30 44,80 1,55 5,06 11,72 10,20
Y 4,20 22,13 26,50 25,50 56,40 79,12 12,13 123,20

6.3.6. Систолическое давление, мм рт. ст. у пациентов до (X) и после (Y) приема каптоприла:

X 210 169 187 160 167 176 185 206 173 146 174 201 198 148 154
Y 201 165 166 157 147 145 168 180 147 136 151 168 179 129 131

6.3.7. Диастолическое давление, мм рт. ст. у пациентов до (X) и после (Y) приема каптоприла:

X 130 122 124 104 112 101 121 124 115 102 98 119 106 107 100
Y 125 121 121 106 101 85 98 105 103 98 90 98 110 103 82

6.3.8. При изучении регуляции тонуса кровеносных сосудов подопытное животное (кошку) подключали к системе искусственного кровообращения и измеряли артериальное давление (X), мм рт. ст. затем прерывали и снова возобновляли кровоток, повторно измеряя кровяное давление (Y):

X 80 90 90 90 90 100 70 70 70 90 90 90 80 80 90 90 90 90
Y 85 77 95 70 80 85 75 70 70 85 85 88 70 65 90 92 110 110

6.3.9. При изучении влияния двух рационов питания на привес свиней подбирались пары животных с одинаковым весом, для которых выбор рациона осуществлялся наугад. Получены следующие данные о привесе от первого (X) и второго (Y) рационов, фунты (1 фунт≈0,4536 кг):

X	17	22	22	15	24	22	21	21	17	21
Y	20	21	21	24	24	22	23	22	22	23

6.3.10. При изучении влияния витамина B_{12} на стимулирование обмена веществ у свиней подбирались пары животных из одного помета. Получены следующие привесы для свиней, не получавших (X) и получавших (Y) добавку B_{12} к корму, фунт/сутки (1 фунт ≈0,4536 кг):

X	1,56	1,52	1,52	1,49	1,59	1,60	1,56
Y	1,60	1,68	1,75	1,64	1,75	1,79	1,77

6.3.11. Данные о дополнительных часах сна, полученные в результате применения двух снотворных препаратов (X и Y) для одних и тех же пациентов:

X	+0,7	-1,6	-0,2	-1,2	-0,1	+3,4	+3,7	+0,8	0,0	+2,0
Y	+1,9	+0,8	+1,1	+0,1	-0,1	+4,4	+5,5	+1,6	+4,6	+3,4

6.3.12. При исследовании влияния гипофиза на чувствительность органов лягушки к ацетилхолину испытывались 20 одинаковых по весу и полу пар. После удаления гипофиза у одной из каждой пары лягушек получили следующие значения чувствительности для животных с гипофизом (X) и без него (Y), ед. концентрации вещества:

X	47	32	40	27	38	20	43	28	43	39	43	28	35	33	43
	32	36	21	38	26										
Y	28	20	45	35	36	22	34	32	36	37	34	29	35	26	33
	24	37	34	28	26										

6.3.13. Данные о времени реакции, мс левым (X) и правым (Y) зрительными полями испытуемых на числа, показываемые в случайном порядке:

X	564	521	495	564	560	781	545	478	580	484	539	467
Y	557	505	465	562	544	448	531	458	560	485	520	445

6.4. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ЗАВИСИМЫХ РАСПРЕДЕЛЕНИЙ

Для указанных переменных X и Y постройте диаграмму рассеяния и оцените связь между ними. Выполните точечное и интервальное оценивание коэффициента корреляции.

6.4.1. X − число эритроцитов, Y − концентрация гемоглобина (табл.6.1).

6.4.2. X − число эритроцитов, Y − оседание крови (табл. 6.1).

6.4.3. X − концентрация гемоглобина, Y − оседание крови (табл. 6.1).

Таблица 6.1

1	2	3	1	2	3	1	2	3	1	2	3
0,80	22	8	3,30	82	30	4,33	82	34	4,36	94	44
0,71	45	18	4,10	81	32	3,80	79	35	1,30	27	12
2,63	61	24	3,29	82	32	3,82	87	36	2,50	50	20

Продолжение табл 6.1

1	2	3	1	2	3	1	2	3	1	2	3
3,19	66	26	3,46	77	32	3,81	87	37	2,80	63	26
2,80	72	28	3,32	80	33	4,20	87	37	3,10	71	28
3,14	83	29	3,11	82	33	4,47	90	38	2,87	70	29
3,21	73	30	3,28	79	34	3,71	97	40	3,68	72	30
3,28	82	30	3,66	84	34	4,22	96	40	3,59	76	30
3,63	78	30	3,90	75	34	3,90	92	40	3,40	71	30

Примечание: 1 − число эритроцитов, млн.; 2 − концентрация гемоглобина,%; 3 − оседание крови за 24 часа, мм.

6.4.4. X − возраст, Y − систолическое давление у женщин (табл. 6.2).

6.4.5. X − возраст, Y − диастолическое давление у женщин (табл. 6.2).

Таблица 6.2

Возраст, годы	Кровяное давление	
	систолическое	диастолическое
1	95	65
2	92	60
3	100	64
5	92	62
10	103	70
12	106	72
15	112	76
20-24	116	72
25-29	117	74
30-34	120	75
35-39	124	78
40-44	127	80
45-49	131	82
50-54	137	84
55-59	139	84
60-64	150	85
65-69	154	85
70-74	159	85
75-79	158	84
80-84	157	83
85-89	154	82
90-94	150	79

6.4.6. Данные о росте X, см и массе Y, кг взрослых мужчин:

X 165 176 175 168 167 172 175 180 179 173 166 178 169 169 170 176 180 169 177 176

Y 56 75 70 61 61 63 72 80 76 68 58 76 60 64 63 71 78 63 75 71

6.4.7. Данные о составе ДНК (X) и РНК (Y) различных бактерий, выраженные соответственно как отношение {Г+Ц}/{А+Т} и {Г+Ц}/{А+У}:

Вид бактерии	X	Y
Clostridium perfringens	0,45	1,06
Staphyloccocus pyogenes	0,53	1,05
Pasteurella tularensis	0,53	1,03
Proteus vulgaris	0,68	1,22
Escherichia coli	1,09	1,21
Proteus morganii	1,13	1,21
Shigella dysenteria	1,15	1,21
Salmonella typhosa	1,14	1,21
Salmonella typhimurius	1,18	1,21
Erninia carotorova	1,17	1,14
Corgnebacterium diphtheria	1,20	1,24
Azotobacter agile	1,21	1,33
Azotobacter vinelandii	1,28	1,27
Azotobacter chroococcum	1,34	1,23
Aerobacter aerogenes	1,31	1,19
Micobacterium vadosum	1,37	1,23
Brucella abortus	1,37	1,23
Alcaligenes faecalis	2,00	1,22
Psendomonas aeruginosa	2,03	1,24
Micobacter tuberculosis BBG	2,08	1,45
Sarcina luteu	2,57	1,32
Actynomyces globisporus streptomycini	2,73	1,29

6.4.8. Структурными элементами белков служат 20 канонических аминокислот, важными характеристиками которых являются полярность (X) и гидрофобность (Y):

Аминокислота	X	Y	Аминокислота	X	Y
Аланин	0,00	0,87	Лейцин	0,13	2,17
Аргинин	52,00	0,85	Лизин	49,50	1,64
Аспарагин	3,38	0,09	Метионин	1,43	1,67
Аспарагиновая кислота	49,70	0,66	Пролин	1,58	2,77
Валин	0,13	1,87	Серин	1,67	0,07
Гистидин	51,60	0,87	Тирозин	1,61	2,67
Глицин	0,00	0,10	Треонин	1,66	0,07
Глутамин	3,53	0,00	Триптофан	2,10	3,77
Глутаминовая кислота	49,90	0,67	Фенилаланин	0,35	2,82
Изолейцин	0,13	3,15	Цистеин	1,48	1,52

6.4.9. Данные о величине падения потенциала головного мозга (X), mv и количестве выделяемого желудочного сока, (Y), мл собаки (n=10 опытов):

X	40	52	57	42	39	51	56	37	43	63
Y	3	4	5	3	4	5	3	1	4	6

6.4.10. Данные об уровне содержания хлоридов (X) и калия (Y) в плазме больных уролитиазом, мг/% :

X 460 440 520 510 480 470 500 490 460 510 500 439

Y 158 150 170 170 158 166 160 162 160 156 170

6.4.11. X - рост мальчиков, Y - вес мальчиков (табл. 6.3).

6.4.12. X - рост девочек; Y - вес девочек (табл. 6.3).

Таблица 6.3

Возраст	Мальчики		Девочки	
	Рост, см	Вес, кг	Рост, см	Вес, кг
при рождении	49,0- 52,5	2,80- 3,90	47,0- 52,0	2,70- 3,96
1 месяц	52,5- 56,8	3,57- 4,58	51,4- 55,5	3,44- 4,42
3	59,0- 63,2	5,40- 6,66	58,0- 61,6	4,89- 6,23
6	65,1- 71,3	7,00- 8,77	64,0- 69,3	6,53- 8,27
9	71,0- 75,0	8,30-10,36	68,6- 73,5	7,62- 9,71
12	73,4- 79,0	9,17-11,13	72,0- 77,0	8,40-11,05
18	79,9- 86,0	10,36-12,91	78,5- 85,0	10,15-12,42
2 года	81,7- 92,3	9,7 -15,1	81,4- 91,6	10,1 -13,7
3	91,0-101,6	11,9 -17,3	89,6-100,8	11,7 -16,9
4	97,9-109,1	14,0 -19,4	96,5-108,1	13,3 -19,3
5	104,2-116,2	16,1- 21,7	103,5-114,9	15,2 -21,4
6	110,2-122,8	18,0 -24,0	109,4-122,4	16,8 -24,4
7	115,8-129,0	19,8 -26,6	115,3-129,5	18,4 -28,6
8	121,3-134,7	20,8 -29,8	121,7-134,7	20,7 -32,7
9	126,7-139,9	21,7 -33,3	126,5-139,9	22,6 -35,0
10	131,2-145,4	23,5 -37,5	130,2-144,8	24,8 -38,2
11	134,9-150,5	27,1 -42,9	133,4-150,2	26,5 -43,2
12	138,2-155,8	30,6 -48,4	138,7-158,3	29,4 -48,6
13	143,3-163,5	34,2 -54,0	144,8-163,4	33,6 -55,0
14	150,2-171,0	38,3 -58,9	150,8-165,6	39,4 -60,8
15	157,1-176,5	42,5 -63,7	153,8-168,4	44,3 -62,1
16	163,3-180,1	74,4 -68,0	155,3-170,5	46,4 -63,8
17	166,7-181,9	52,1 -69,5	155,8-171,8	47,7 -65,3

6.4.13. Скорость кровотока, мл/час определяли двумя методами: обычным (X), или путем непосредственного измерения, и новым (Y), технически более простым:

X 1190 1455 1550 1730 1745 1770 1900 1920 1960 2295 2335
2490 2720 2710 2530 2900 2760 3010

Y 1115 1425 1515 1795 1715 1710 1830 1920 1970 2300 2280
2520 2630 2740 2390 2800 2630 2970

6.5. РЕГРЕССИОННЫЙ АНАЛИЗ ЗАВИСИМЫХ РАСПРЕДЕЛЕНИЙ

Для указанных переменных X и Y постройте диаграмму рассеяния и проверьте гипотезу о линейной связи между переменными.

6.5.1. При калибровке прибора для измерения концентрации молочной кислоты в крови исследователь использует образцы с известной концентрацией (X), мМ, а потом расчитывает концентрацию (Y), мМ, калибруемым прибором (объем выборки n=20):

X	1	1	1	1	3	3	3	3	3	5
	5	5	10	10	10	10	15	15	15	15
Y	1,1	0,7	1,8	0,4	3,0	1,4	4,9	4,4	4,5	7,3
	8,2	6,2	12,0	13,1	12,6	13,2	18,7	19,7	17,7	17,1

6.5.2. Скорость кровотока, мл/час., определяли двумя методами: обычным (X), т.е. путем непосредственного измерения, и новым (Y), технически более простым:

X	1190	1455	1550	1730	1745	1770	1900	1920	1960
	2295	2335	2490	2720	2710	2530	2900	2760	3010
Y	1115	1425	1515	1795	1715	1710	1830	1920	1970
	2300	2280	2520	2630	2740	2390	2800	2630	2970

6.5.3. X – возраст, Y – основной обмен у мужчин, ккал/м^2/г (табл. 6.4).

6.5.4. X – возраст, Y – основной обмен у женщин, ккал/м^2/г (табл. 6.4).

Таблица 6.4

Возраст, годы	Основной обмен, ккал/м^2/г	
	Мужчины	Женщины
1	53	53
2	52,4	52,4
5	49,3	48,4
8	46,3	43,8
10	44,0	42,5
12	42,5	41,3
15	41,8	37,9
20	38,6	35,3
25	37,5	35,2
30	36,8	35,1
40	36,2	34,9
50	35,8	33,9
60	34,9	32,7
70	33,8	31,7
75	33,2	31,3

6.5.5. X − возраст, Y − суточная потребность человека во сне (табл. 6.5):

Таблица 6.5

Возраст	Суточная потребность во сне, часы
1-15 дней	16
3-5 месяцев	14
2− 3 года	13
5− 9 лет	11
10−14 лет	9
19−30 лет	8
33−45 лет	7
50−70 лет	6

6.5.6. Данные о потреблении кислорода пиявками (Y), г/кг/час, в зависимости от температуры (X), ^{0}C:

X 5,5 5,6 6,2 8,4 9,0 10,5 16,1 16,6 17,1 18,8 19,8 20,0 20,7 26,5

Y 16,1 14,9 18,8 32,5 32,1 37,1 88,5 91,0 94,0 122,0 162,0 167,0 187,0 436,0

6.5.7. X − возраст, Y − рост мальчиков (табл. 6.3).

6.5.8. X − возраст, Y − вес мальчиков (табл. 6.3).

6.5.9. X − возраст, Y − рост девочек (табл. 6.3).

6.5.10. X − возраст, Y − вес девочек (табл. 6.3).

6.5.11. При экспресс-анализе концентрации кофеина (X), мг/100мл, отвечает оптическая плотность (Y):

X 0,257 0,498 0,506 0,514 0,747 0,760 0,770 0,996 1,013 1,027 1,246 1,266 1,284 1,494 1,519 1,541 1,798 2,054 2,311 2,568

Y 0,131 0,262 0,265 0,263 0,384 0,393 0,396 0,512 0,518 0,523 0,633 0,643 0,650 0,760 0,768 0,775 0,903 1,040 1,160 1,290

6.5.12. X − возраст, Y − поле зрения мужчин (табл. 6.6).

6.5.13. X − возраст, Y − поле зрения женщин (табл. 6.6).

Таблица 6.6

Возраст, годы	Поле зрения, град	
	Мужчины	Женщины
16-19	174,5	176,4
20-29	174,8	176,2
30-39	173,8	175,3
40-49	171,7	172,7
50-59	167,1	170,1
60-69	160,1	162,3
70-79	151,2	155,5
80−	139,8	138,5

7. СПРАВОЧНЫЕ СВЕДЕНИЯ

7.1. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ $N(\mu,\sigma)$

Непрерывная случайная величина $Y \equiv \tilde{y}$ распределена по нормальному закону с двумя параметрами μ (математическим ожиданием $M[\tilde{y}]$) и σ (среднеквадратическим отклонением, $\sigma = \sqrt{D[\tilde{y}]}$), если ее ***функция плотности*** имеет вид:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}, \qquad (7.1)$$

$$-\infty < y < +\infty, \ -\infty < \mu < +\infty, \ \sigma > 0.$$

При изменении значения математического ожидания μ кривая $f(y)$, не изменяя своей формы, смещается вдоль оси абсцисс (рис. 7.1).

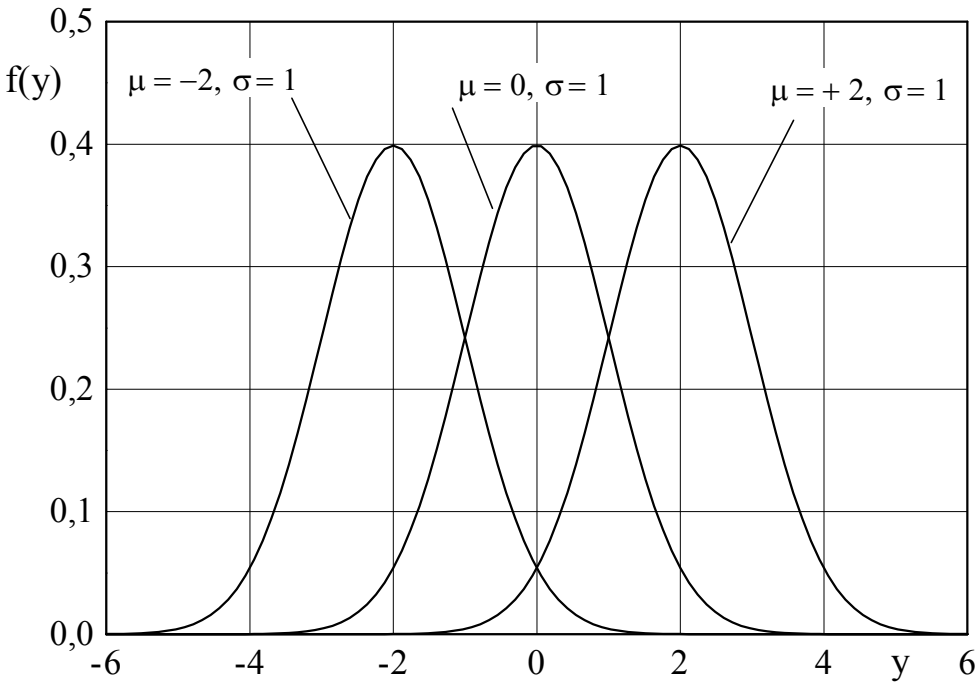

Рис.7.1. Влияние параметра μ на график функции $f(y)$

Изменение значения среднеквадратического отклонения σ равносильно изменению масштаба кривой по обеим осям. Например, при удвоении значения σ масштаб по оси абсцисс увеличится, а по оси орди-

нат − уменьшится в два раза (рис.7.2). Кривая плотности при этом становится более плоской, растягиваясь вдоль оси абсцисс. При уменьшении значения σ − вытягивается вверх, одновременно сжимаясь с боков. В обоих случаях кривая $f(y)$ ограничивает над осью y единичную площадь

$$\int_{-\infty}^{+\infty} f(y)\,dy = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\} dy = 1. \qquad (7.2)$$

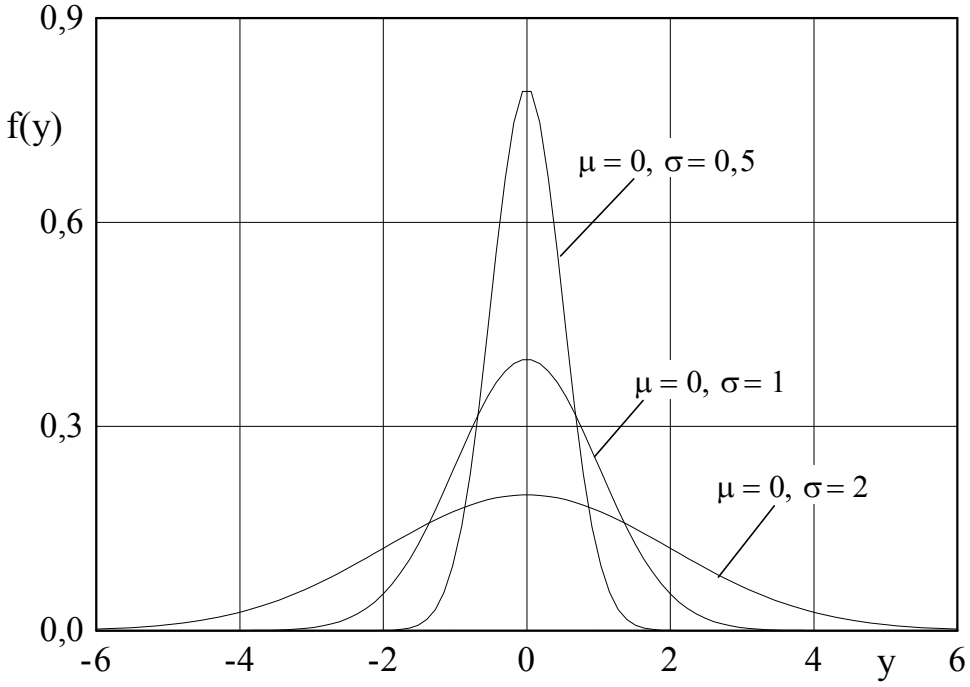

Рис.7.2. Влияние параметра σ на график функции $f(y)$

По определению функции плотности $f(y)$ вероятность принятия нормально распределенной случайной величиной $\tilde{y}$ значений в промежутке (y_1, y_2) равна

$$P\left(\left\{y_1 < \tilde{y} < y_2\right\}\right) = \int_{y_1}^{y_2} f(y)\,dy = \frac{1}{\sigma\sqrt{2\pi}} \int_{y_1}^{y_2} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\} dy. \qquad (7.3)$$

Произведя в интеграле (7.3) замену переменной $z = (y-\mu)/\sigma$ (“стандартизация” случайной величины $\tilde{y}$) и изменив пределы интегрирования, можно получить

$$P\left(\left\{y_1 < \tilde{y} < y_2\right\}\right) = \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{y_1-\mu}{\sigma}}^{\frac{y_2-\mu}{\sigma}} e^{-z^2/2}\, dz = \int\limits_{z_1}^{z_2} f(z)\, dz = P\left(\left\{z_1 < \tilde{z} < z_2\right\}\right), \tag{7.4}$$

$$z_1 = \frac{y_1 - \mu}{\sigma}, \quad z_2 = \frac{y_2 - \mu}{\sigma},$$

где $Z \equiv \tilde{z}$ − нормально распределенная случайная величина с параметрами μ=0, σ=1, имеющая функцию плотности

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \tag{7.5}$$

Для величин $\tilde{y}$ и $\tilde{z}$ используются следующие сокращенные обозначения: $\tilde{y}$~$N(\mu, \sigma)$, $\tilde{z}$~$N(0, 1)$. Неопределенный интеграл $\int e^{-z^2/2}\, dz$ в (7.4) не выражается через элементарные функции ("не берется"), но его можно выразить через специальную функцию

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int\limits_0^z e^{z^2/2}\, dz, \tag{7.6}$$

называемую ***функцией Лапласа*** (или интегралом вероятностей), график которой показан на рис.7.3. С помощью этой функции искомая вероятность выражается простой формулой:

$$P\left(\left\{y_1 < \tilde{y} < y_2\right\}\right) = \Phi\left(\frac{y_2 - \mu}{\sigma}\right) - \Phi\left(\frac{y_1 - \mu}{\sigma}\right) = \Phi\left(z_2\right) - \Phi\left(z_1\right). \tag{7.7}$$

Функция Лапласа обладает следующими свойствами:

1) $\Phi(0)=0$;

2) $\Phi(-z)= -\Phi(z)$, то есть $\Phi(z)$ − нечетная функция;

3) $\Phi(+\infty)=0{,}5$, откуда по свойству 2 $\Phi(-\infty)= -0{,}5$.

Наиболее просто выражается через функцию Лапласа вероятность того, что нормально распределенная случайная величина $\tilde{y}$~$N(\mu, \sigma)$ принимает значения из промежутка длиной $2h$, симметричного относительно математического ожидания μ:

$$P\left(\left\{\mu - h < \tilde{y} < \mu + h\right\}\right) = P\left(\left\{\,|\,\tilde{y} - \mu\,| < h\,\right\}\right) =$$
$$= \Phi\left(\frac{\mu - \mu + h}{\sigma}\right) - \Phi\left(\frac{\mu - \mu - h}{\sigma}\right) = \Phi\left(\frac{h}{\sigma}\right) - \Phi\left(-\frac{h}{\sigma}\right) \qquad (7.8)$$

или, принимая во внимание свойство 2 функции Лапласа:

$$P\left(\left\{\,|\,\tilde{y} - \mu\,| < h\,\right\}\right) = 2\,\Phi\left(\frac{h}{\sigma}\right). \qquad (7.9)$$

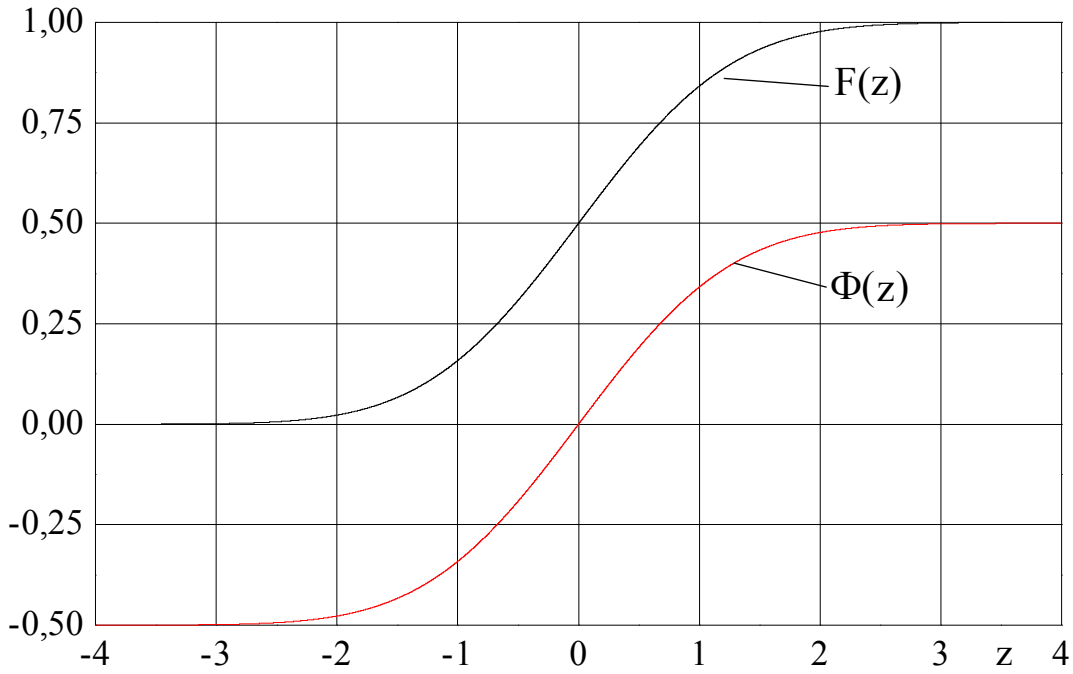

Рис. 7.3. Графики функции Лапласа $\Phi(z)$ и функции распределения $F(z)$

Через функцию Лапласа выражается и ***функция распределения*** $F(y)$ случайной величины $\tilde{y} \sim N(\mu,\sigma)$, если принять, что $y_1 \to -\infty$, $y_2 = y$, и учесть свойство 3 функции Лапласа:

$$F(y) = \begin{cases} \dfrac{1}{2} + \Phi\left(\dfrac{y-\mu}{\sigma}\right) = \dfrac{1}{2} + \Phi(z), & y \ge \mu\ (z \ge 0), \\ \dfrac{1}{2} + \Phi\left(\dfrac{y-\mu}{\sigma}\right) = \dfrac{1}{2} - \Phi\left(-\dfrac{y-\mu}{\sigma}\right) = \dfrac{1}{2} - \Phi(-z), & y < \mu\ \ (z < 0). \end{cases} \qquad (7.10)$$

График функции распределения $F(z)$ показан на рис. 7.3.

Для промежутков $\mu\pm1\sigma$, $\mu\pm2\sigma$, $\mu\pm3\sigma$ вероятности, вычисляемые по (7.7), равны соответственно 0,6827, 0,9545, 0,9973, что составляет содержание ***правила 3σ*** (рис.7.4):

$$P(\{\mu-1\sigma < \tilde{y} < \mu+1\sigma\}) = P(\{-1 < \tilde{z} < +1\}) = \Phi(+1) - \Phi(-1) = 2\,\Phi(1) \approx 0{,}6827;$$
$$P(\{\mu-2\sigma < \tilde{y} < \mu+2\sigma\}) = P(\{-2 < \tilde{z} < +2\}) = \Phi(+2) - \Phi(-2) = 2\,\Phi(2) \approx 0{,}9545;$$
$$P(\{\mu-3\sigma < \tilde{y} < \mu+3\sigma\}) = P(\{-3 < \tilde{z} < +3\}) = \Phi(+3) - \Phi(-3) = 2\,\Phi(3) \approx 0{,}9973.$$

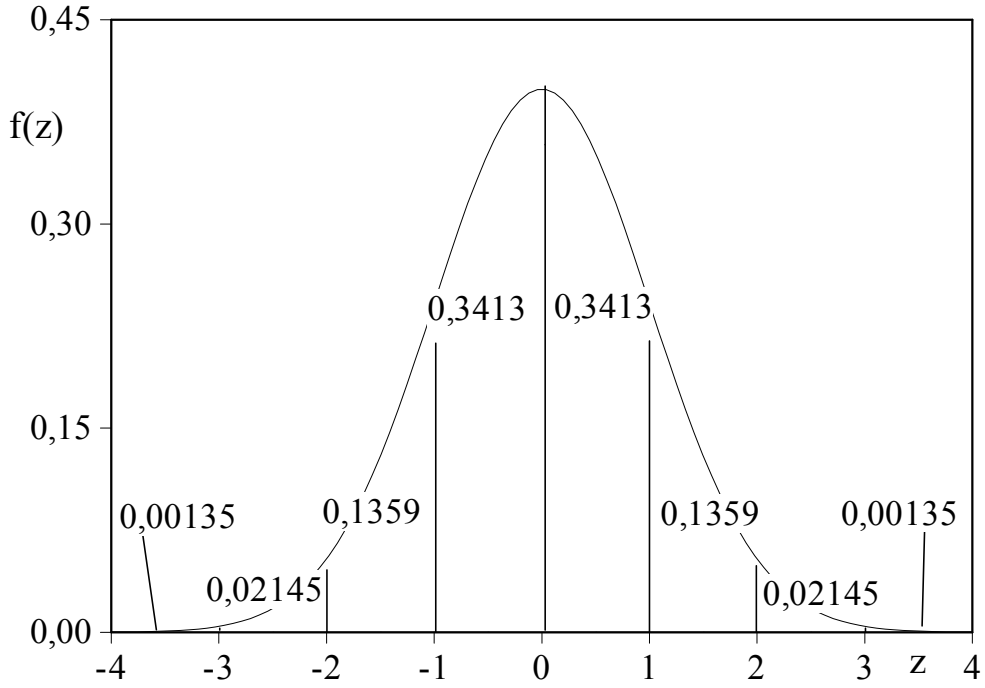

Рис. 7.4. Графическая иллюстрация правила 3σ

Значения функции Лапласа $\Phi(z)$ для положительных значений аргумента z приведены в табл. 7.1. Для отрицательных значений z следует воспользоваться свойством 2.

7.2. РАСПРЕДЕЛЕНИЕ ПИРСОНА χ^2

Пусть $\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_\nu$ − попарно независимые случайные величины, распределенные по закону $N(0, 1)$ (см. 7.1), тогда по определению случайная величина

$$\tilde{\chi}_{\nu}^{2} = \sum_{i=1}^{\nu} \tilde{z}_{i}^{2} \qquad (7.11)$$

имеет распределение Пирсона $\tilde{\chi}_{\nu}^{2}$ с параметром ν (“хи”-квадрат с числом степеней свободы “ню”), плотность распределения которого

$$f(\chi_{\nu}^{2}) = \frac{\left(\chi_{\nu}^{2}\right)^{\nu/2-1} e^{-\chi_{\nu}^{2}/2}}{2^{\nu/2}\,\Gamma(\nu/2)}. \qquad (7.12)$$

Числовые характеристики распределения $\tilde{\chi}_{\nu}^{2}$: математическое ожидание $M[\tilde{\chi}_{\nu}^{2}] = \nu$ и дисперсия $D[\tilde{\chi}_{\nu}^{2}] = 2\nu$. При малых значениях параметра ν функция плотности имеет длинный правый хвост, а при увеличении значения ν (ν~100 и более) асимптотически становится почти симметричной и хорошо аппроксимируемой нормальным распределением с математическим ожиданием $M[\tilde{\chi}_{\nu}^{2}]$ и дисперсией $D[\tilde{\chi}_{\nu}^{2}]$.

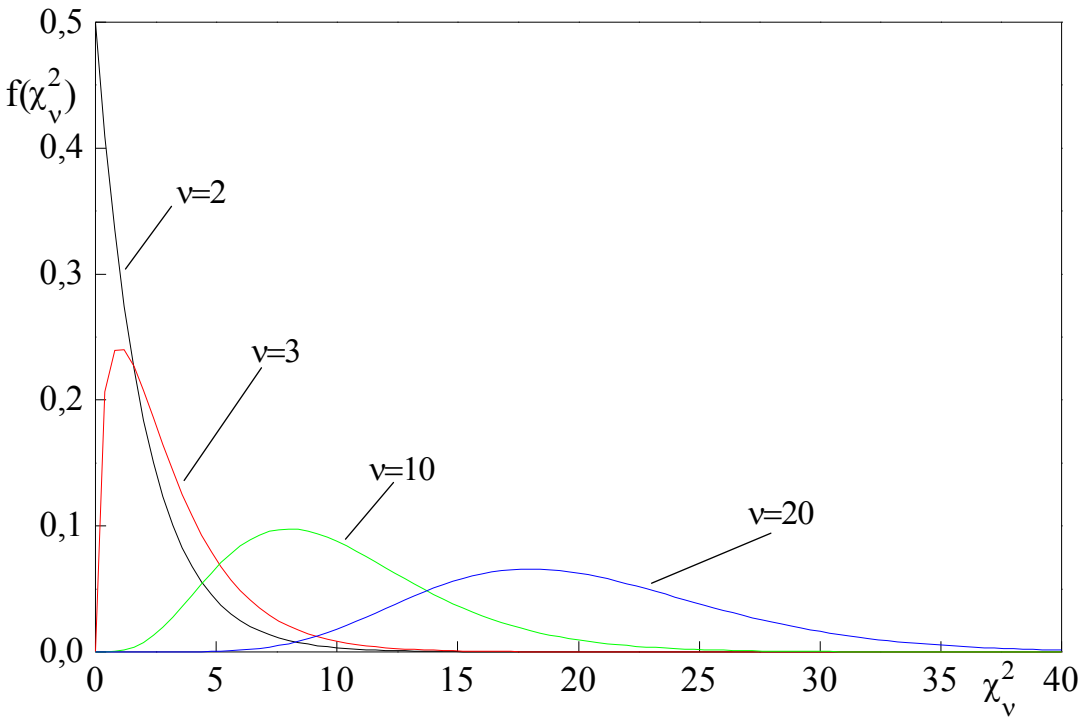

Рис. 7.5. Графики функций плотности $f(\tilde{\chi}_{\nu}^{2})$

Значения функции Лапласа $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{t^2}{2}} dt$ Таблица 7.1

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,34134	0,34375	0,34614	0,34850	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41309	0,41466	0,41621	0,41774
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42786	0,42922	0,43056	0,43189
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169

Окончание табл. 7.1

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997
4,0	0,49998	0,49998	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999

Значения квантилей $\chi^2_{\nu,1-\alpha}$ распределения Пирсона Таблица 7.2

Число степеней свободы ν	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0
35	57,3	53,2	49,8	22,5	20,6	18,5
40	63,7	59,3	55,8	26,5	24,4	22,2
45	70,0	65,4	61,7	30,6	28,4	25,9
50	76,2	71,4	67,5	34,8	32,4	29,7
75	106,4	100,8	96,2	56,1	52,9	49,5
100	135,6	129,6	124,3	77,9	74,2	70,1

Функция распределения $F(\chi_\nu^2)$ (рис. 7.6), как и функция распределения $F(z)$ и функция Лапласа $\Phi(z)$ нормального распределения, не представима в замкнутом виде, поэтому следует пользоваться значениями, приведенными в табл. 7.2.

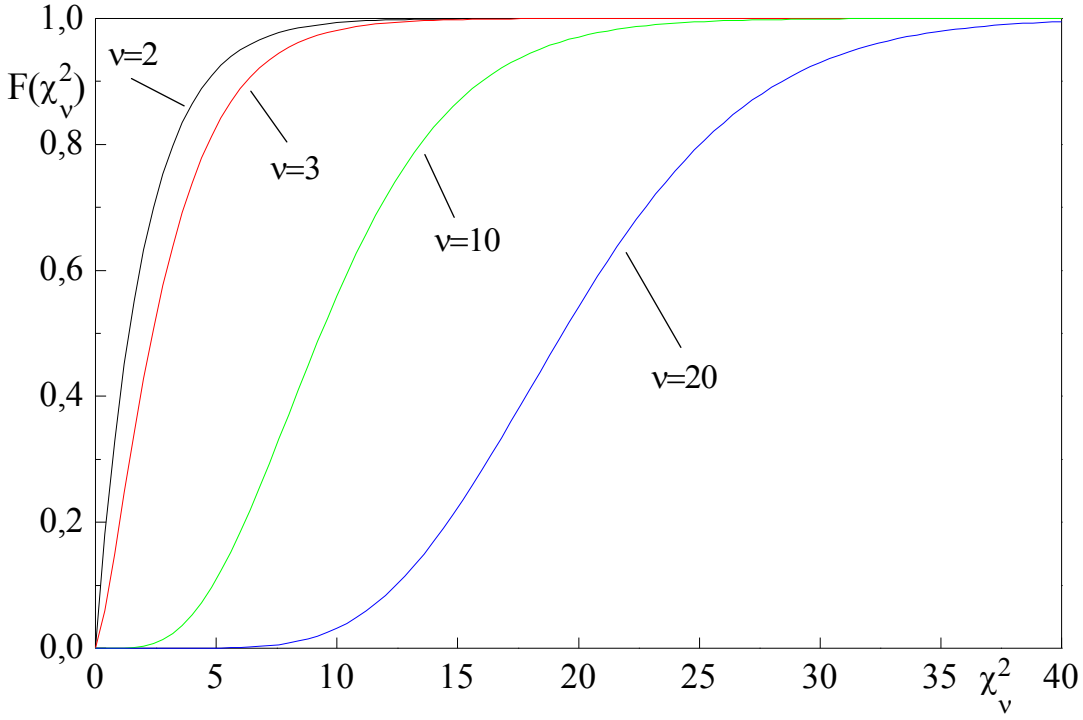

Рис.7.6. Графики функций распределения $F(\chi_\nu^2)$

7.3. РАСПРЕДЕЛЕНИЕ ФИШЕРА F

Пусть $\tilde{\chi}_{\nu_1}^2$ и $\tilde{\chi}_{\nu_2}^2$ - случайные величины, распределенные по закону $\tilde{\chi}_\nu^2$ с числом степеней свободы соответственно ν_1 и ν_2, тогда по определению случайная величина

$$\tilde{F}_{\nu_1,\nu_2} = \frac{\tilde{\chi}_{\nu_1}^2 / \nu_1}{\tilde{\chi}_{\nu_2}^2 / \nu_2} \tag{7.13}$$

имеет распределение Фишера $\tilde{F}_{\nu_1,\nu_2}$ с параметрами ν_1 и ν_2 (числа степеней свободы числителя и знаменателя), плотность распределения которого

$$f(F_{\nu_1,\nu_2}) = \frac{\Gamma((\nu_1+\nu_2)/2)}{\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)}\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}\frac{(F_{\nu_1,\nu_2})^{\nu_1/2-1}}{\left(1+\frac{\nu_1}{\nu_2}F_{\nu_1,\nu_2}\right)^{(\nu_1+\nu_2)/2}}. \qquad (7.14)$$

Числовые характеристики распределения $\widetilde{F}_{\nu_1,\nu_2}$: математическое ожидание $M[\widetilde{F}_{\nu_1,\nu_2}] = \nu_1/(\nu_2-2)$ и дисперсия $D[\widetilde{F}_{\nu_1,\nu_2}] = 2\nu_2^2(\nu_1+\nu_2-2)/[\nu_1(\nu_2-2)^2(\nu_2-4)]$. При малых значениях параметров ν_1 и ν_2 функция плотности имеет длинный правый хвост, а при увеличении значений ν_1 и ν_2 асимптотически становится почти симметричной и хорошо аппроксимируемой нормальным распределением с математическим ожиданием $M[\widetilde{F}_{\nu_1,\nu_2}]$ и дисперсией $D[\widetilde{F}_{\nu_1,\nu_2}]$.

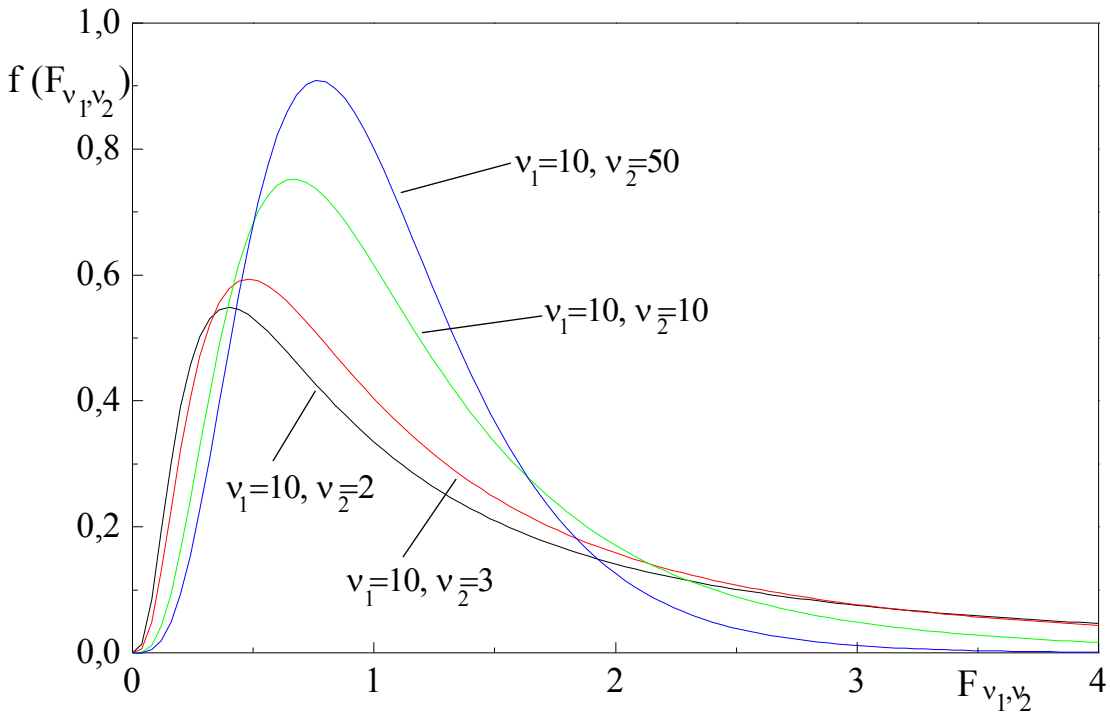

Рис. 7.7. График функции плотности $f(F_{\nu_1,\nu_2})$

Функция распределения $F(F_{\nu_1,\nu_2})$ не представима в замкнутом виде, поэтому следует пользоваться значениями, приведенными в табл. 7.3. Необходимо различать обозначения функции распределения F и значения случайной величины F_{ν_1,ν_2}.

Значения квантилей $F_{\nu_1,\nu_2,1-\alpha/2}$ распределения Фишера для уровня значимости α=0,05 Таблица 7.3

$\nu_2 \setminus \nu_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7	963,3	968,6	976,7	984,9	993,1	1001	1018
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,43	39,45	39,46	39,50
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,34	14,25	14,17	14,08	13,90
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,75	8,66	8,56	8,46	8,26
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,52	6,43	6,33	6,23	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,37	5,27	5,17	5,07	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,67	4,57	4,47	4,36	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,20	4,10	4,00	3,89	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,87	3,77	3,67	3,56	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,62	3,52	3,42	3,31	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,43	3,33	3,23	3,12	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,28	3,18	3,07	2,96	2,72
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,15	3,05	2,95	2,84	2,60
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	3,05	2,95	2,84	2,73	2,49
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,96	2,86	2,76	2,64	2,40
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,89	2,79	2,68	2,57	2,32
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,82	2,72	2,62	2,50	2,25
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,77	2,67	2,56	2,44	2,19
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,72	2,62	2,51	2,39	2,13
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,68	2,57	2,46	2,35	2,09

Окончание табл.7.3

$\nu_2 \setminus \nu_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73	2,64	2,53	2,42	2,31	2,04
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70	2,60	2,50	2,39	2,27	2,00
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,57	2,47	2,36	2,24	1,97
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,54	2,44	2,33	2,21	1,94
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,51	2,41	2,30	2,18	1,91
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,49	2,39	2,28	2,16	1,88
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57	2,47	2,36	2,25	2,13	1,85
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,45	2,34	2,23	2,11	1,83
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53	2,43	2,32	2,21	2,09	1,81
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,41	2,31	2,20	2,07	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,29	2,18	2,07	1,94	1,64
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,17	2,06	1,94	1,82	1,48
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	2,05	1,94	1,82	1,69	1,31
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,94	1,83	1,71	1,57	1,00

7.4. РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА t

Пусть $\tilde{z}$ - случайная величина, распределенная по закону $N(0,\ 1)$, $\tilde{\chi}^2_\nu$ – случайная величина, распределенная по закону $\tilde{\chi}^2_\nu$ с числом степеней свободы ν, тогда по определению случайная величина

$$\tilde{t}_\nu = \frac{\tilde{z}}{\sqrt{\tilde{\chi}^2_\nu / \nu}} \tag{7.15}$$

имеет распределение Стьюдента с параметром ν (число степеней свободы), плотность распределения которого

$$f(t_\nu) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)\left(1 + t^2/\nu\right)^{(\nu+1)/2}}. \tag{7.16}$$

Числовые характеристики распределения $\tilde{t}_\nu$: математическое ожидание $M[\tilde{t}_\nu] = 0$ и дисперсия $D[\tilde{t}_\nu] = \nu/(\nu-2)$. При малых значениях параметра ν функция $f(t_\nu)$ имеет длинный правый хвост, а при увеличении значения ν (ν~30 и более) аппроксимируется нормальным распределением $\tilde{z}$ ~ N(0,1).

Функция распределения $F(t_\nu)$ не представима в замкнутом виде, поэтому следует пользоваться значениями, приведенными в табл. 7.4.

7.5. ГРУППИРОВКА ВЫБОРОЧНЫХ ДАННЫХ

При первичной статистической обработке выборочных данных число интервалов группировки подсчитывается по одной из формул (1.8) – (1.10), причем результат округляется до целого значения. Для упрощения вычислений следует воспользоваться графическими изображениями зависимостей числа интервалов от объема выборки $K_{1-3}(n)$ (рис. 7.9).

7.6. ПРЕОБРАЗОВАНИЕ ФИШЕРА

При проверке статистических гипотез о коэффициенте корреляции для упрощения вычислений при прямом преобразовании Фишера (5.9) $\tilde{r} \mapsto \tilde{y} \mapsto \tilde{z}$ следует воспользоваться графиком функции $y(r)$ (рис. 7.10) или табл. 7.5, а при обратном преобразовании $\tilde{z} \mapsto \tilde{y} \mapsto \tilde{r}$ - графиком функции $th(y)$ (рис. 7.11) или табл. 7.6.

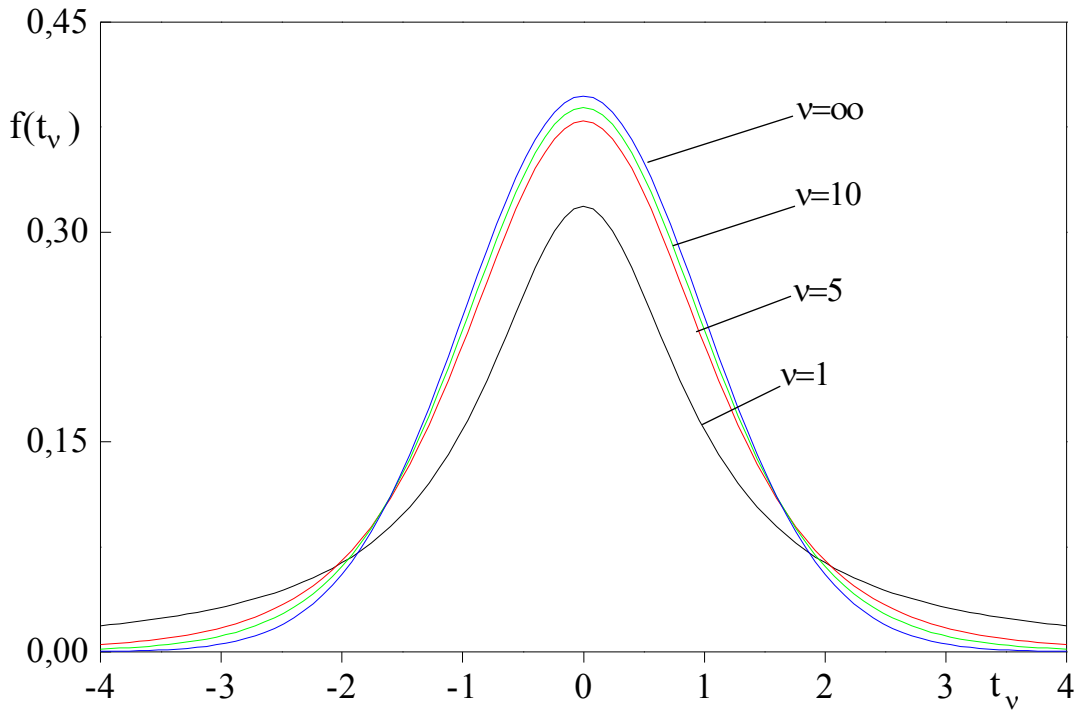

Рис. 7.8. Графики функций плотности $f(t_\nu)$

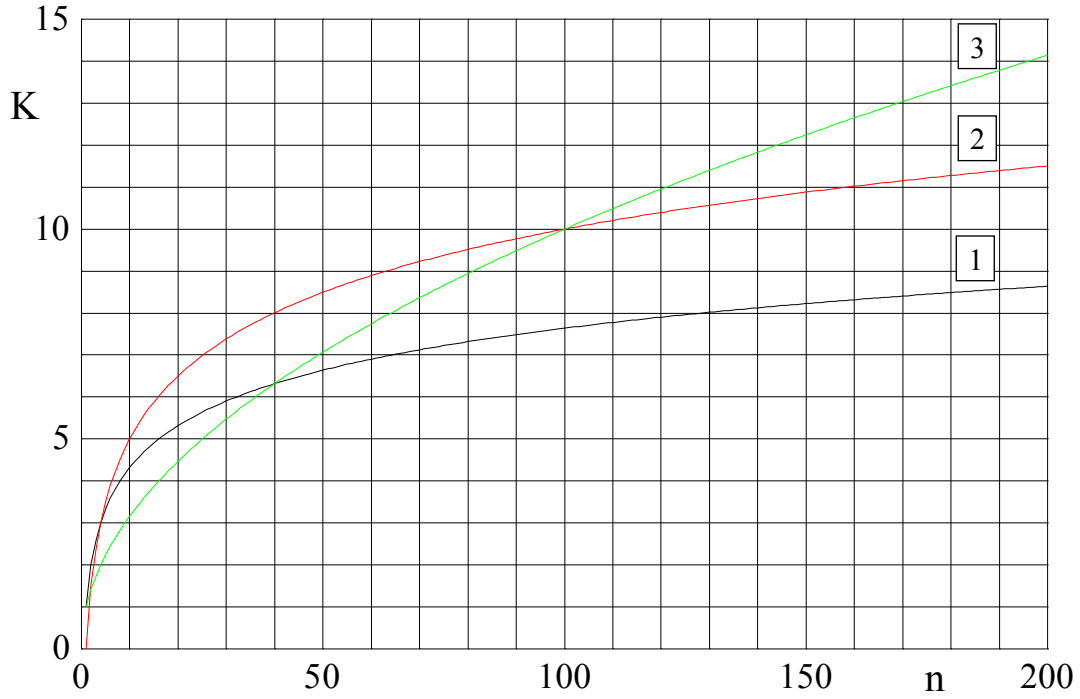

Рис. 7.9. Зависимости числа интервалов от объема выборки

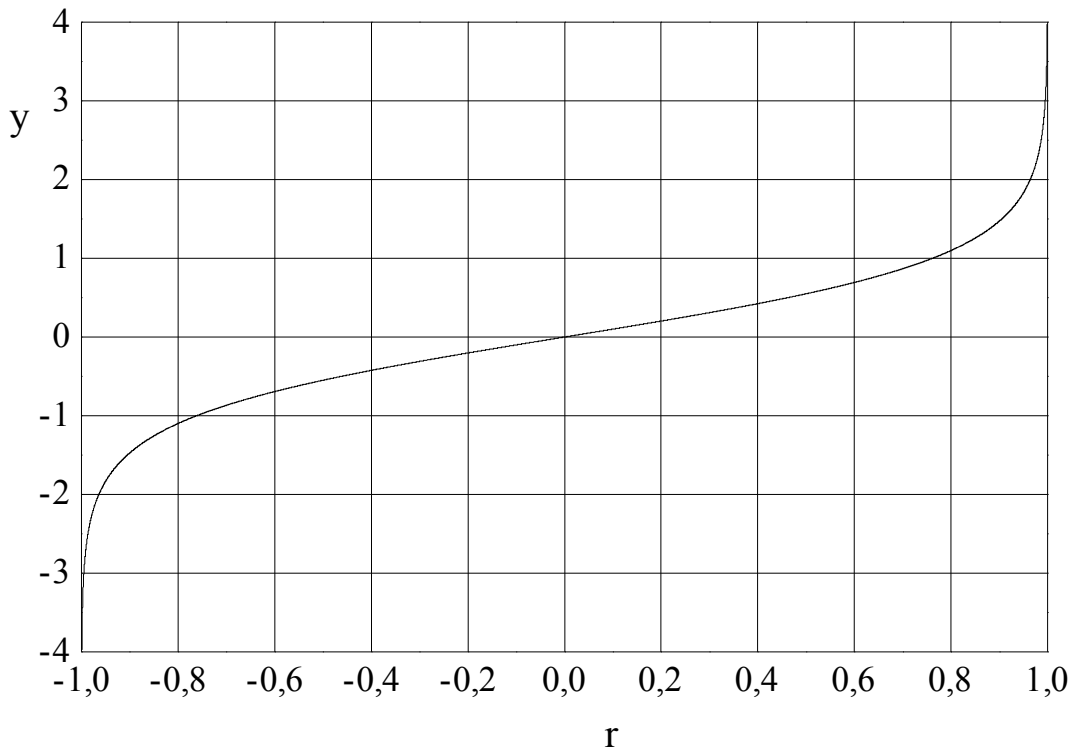

Рис.7.10. График функции y (r)

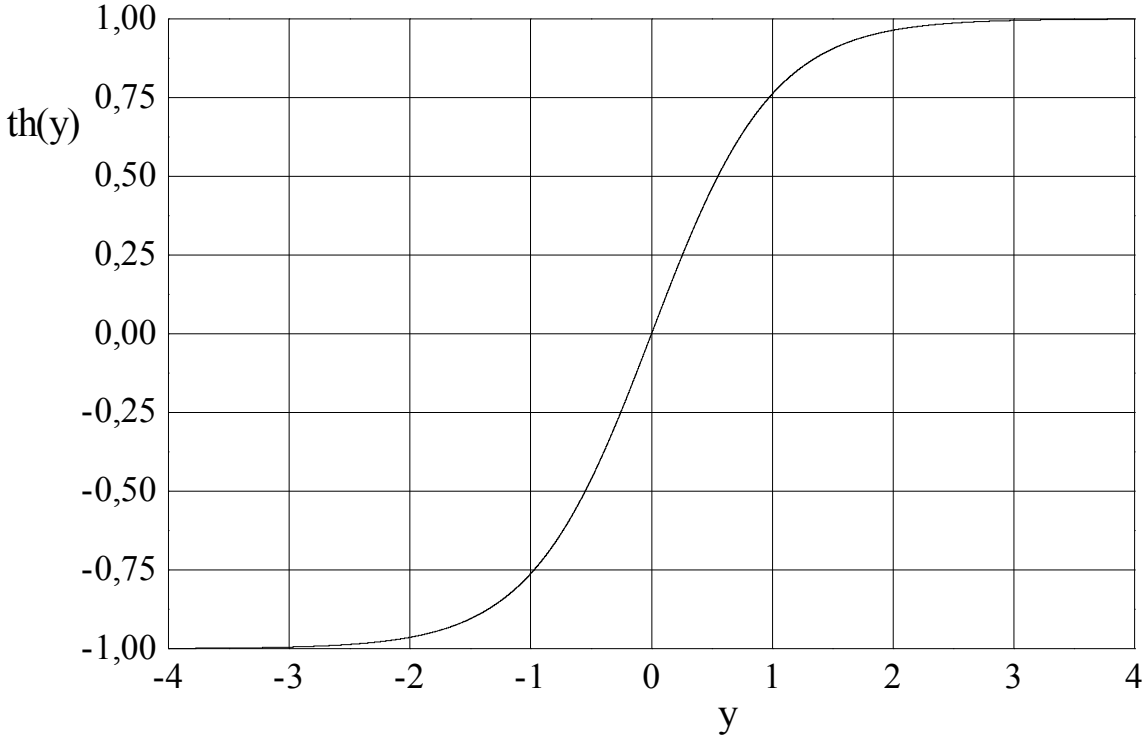

Рис.7.11. График функции th(y)

Значения квантилей $t_{\nu,1-\alpha/2}$ распределения Стьюдента Таблица 7.4

Число степеней свободы	Уровень значимости α					
ν	0,10	0,05	0,02	0,01	0,002	0,001
1	6,314	12,706	31,821	63,657	318,309	636,619
2	2,920	4,303	6,965	9,925	22,327	31,598
3	2,353	3,182	4,541	5,841	10,215	12,941
4	2,132	2,776	3,747	4,604	7,173	8,610
5	2,015	2,571	3,365	4,032	5,893	6,869
6	1,943	2,447	3,143	3,707	5,208	5,959
7	1,895	2,365	2,998	3,499	4,785	5,408
8	1,860	2,306	2,896	3,355	4,501	5,041
9	1,833	2,262	2,821	3,250	4,297	4,781
10	1,812	2,228	2,764	3,169	4,144	4,587
11	1,796	2,201	2,718	3,106	4,025	4,437
12	1,782	2,179	2,681	3,055	3,930	4,318
13	1,771	2,160	2,650	3,012	3,852	4,221
14	1,761	2,145	2,624	2,977	3,787	4,140
15	1,753	2,131	2,602	2,947	3,733	4,073
16	1,746	2,120	2,583	2,921	3,686	4,015
17	1,740	2,110	2,567	2,898	3,646	3,965
18	1,734	2,101	2,552	2,878	3,610	3,922
19	1,729	2,093	2,539	2,861	3,579	3,883
20	1,725	2,086	2,528	2,845	3,552	3,850
21	1,721	2,080	2,518	2,831	3,527	3,819
22	1,717	2,074	2,508	2,819	3,505	3,792
23	1,714	2,069	2,500	2,807	3,485	3,767
24	1,711	2,064	2,492	2,797	3,467	3,745
25	1,708	2,060	2,485	2,787	3,450	3,725
26	1,706	2,056	2,479	2,779	3,435	3,707
27	1,703	2,052	2,473	2,771	3,421	3,690
28	1,701	2,048	2,467	2,763	3,408	3,674
29	1,699	2,045	2,462	2,756	3,396	3,659
30	1,697	2,042	2,457	2,750	3,385	3,646
∞	1,645	1,960	2,326	2,576	3,090	3,291

Значения преобразования Фишера y(r) Таблица 7.5

r	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0601	0,0701	0,0802	0,0902
0,1	0,1003	0,1104	0,1206	0,1307	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
0,2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2554	0,2661	0,2769	0,2877	0,2986
0,3	0,3095	0,3205	0,3316	0,3428	0,3541	0,3654	0,3769	0,3884	0,4001	0,4118
0,4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361
0,5	0,5493	0,5627	0,5763	0,5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777
0,6	0,6931	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
0,7	0,8673	0,8872	0,9076	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
0,8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
0,9	1,4722	1,5275	1,5890	1,6584	1,7380	1,8318	1,9459	2,0923	2,2976	2,6467

Значения функции th(y) Таблица 7.6

y	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0599	0,0699	0,0798	0,0898
0,1	0,0997	0,1096	0,1194	0,1293	0,1391	0,1489	0,1586	0,1684	0,1781	0,1877
0,2	0,1974	0,2070	0,2165	0,2260	0,2355	0,2449	0,2543	0,2636	0,2729	0,2821
0,3	0,2913	0,3004	0,3095	0,3185	0,3275	0,3364	0,3452	0,3540	0,3627	0,3714
0,4	0,3799	0,3885	0,3969	0,4053	0,4136	0,4219	0,4301	0,4382	0,4462	0,4542
0,5	0,4621	0,4699	0,4777	0,4854	0,4930	0,5005	0,5080	0,5154	0,5227	0,5299
0,6	0,5370	0,5441	0,5511	0,5581	0,5649	0,5717	0,5784	0,5850	0,5915	0,5980
0,7	0,6044	0,6107	0,6169	0,6231	0,6291	0,6351	0,6411	0,6469	0,6527	0,6584
0,8	0,6640	0,6696	0,6751	0,6805	0,6858	0,6911	0,6963	0,7014	0,7064	0,7114
0,9	0,7163	0,7211	0,7259	0,7306	0,7352	0,7398	0,7443	0,7487	0,7531	0,7574
1,0	0,7616	0,7658	0,7699	0,7739	0,7779	0,7818	0,7857	0,7895	0,7932	0,7969
1,1	0,8005	0,8041	0,8076	0,8110	0,8144	0,8178	0,8210	0,8243	0,8275	0,8306
1,2	0,8337	0,8367	0,8397	0,8426	0,8455	0,8483	0,8511	0,8538	0,8565	0,8591
1,3	0,8617	0,8643	0,8668	0,8692	0,8717	0,8741	0,8764	0,8787	0,8810	0,8832
1,4	0,8854	0,8875	0,8896	0,8917	0,8937	0,8957	0,8977	0,8996	0,9015	0,9033
1,5	0,9051	0,9069	0,9087	0,9104	0,9121	0,9138	0,9154	0,9170	0,9186	0,9201
1,6	0,9217	0,9232	0,9246	0,9261	0,9275	0,9289	0,9302	0,9316	0,9329	0,9341
1,7	0,9354	0,9366	0,9379	0,9391	0,9402	0,9414	0,9425	0,9436	0,9447	0,9458
1,8	0,9468	0,9478	0,9488	0,9498	0,9508	0,9517	0,9527	0,9536	0,9545	0,9554
1,9	0,9562	0,9571	0,9579	0,9587	0,9595	0,9603	0,9611	0,9618	0,9626	0,9633
2,0	0,9640	0,9647	0,9654	0,9661	0,9667	0,9674	0,9680	0,9687	0,9693	0,9699
2,1	0,9705	0,9710	0,9716	0,9721	0,9727	0,9732	0,9737	0,9743	0,9748	0,9753
2,2	0,9757	0,9762	0,9767	0,9771	0,9776	0,9780	0,9785	0,9789	0,9793	0,9797
2,3	0,9801	0,9805	0,9809	0,9812	0,9816	0,9820	0,9823	0,9827	0,9830	0,9833
2,4	0,9837	0,9840	0,9843	0,9846	0,9849	0,9852	0,9855	0,9858	0,9861	0,9863
2,5	0,9866	0,9869	0,9871	0,9874	0,9876	0,9879	0,9881	0,9884	0,9886	0,9888
2,6	0,9890	0,9892	0,9895	0,9897	0,9899	0,9901	0,9903	0,9905	0,9906	0,9908
2,7	0,9910	0,9912	0,9914	0,9915	0,9917	0,9919	0,9920	0,9922	0,9923	0,9925
2,8	0,9926	0,9928	0,9929	0,9931	0,9932	0,9933	0,9935	0,9936	0,9937	0,9938
2,9	0,9940	0,9941	0,9942	0,9943	0,9944	0,9945	0,9946	0,9947	0,9949	0,9950

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Выборочная плотность 7
Выборочный (эмпирический) закон распределения 8
Выборочный коэффициент корреляции 26

Гипотеза 11,18
Гистограмма 14

Дисперсия 5
Доверительная вероятность 5
Доверительный интервал для дисперсии 6
Доверительный интервал для коэффициента корреляции 31
Доверительный интервал для коэффициента регрессии 38
Доверительный интервал для математического ожидания 6

Интервальная оценка параметров распределения 6

Квантиль нормального распределения 59
Квантиль распределения Пирсона χ^2 61
Квантиль распределения Стьюдента t 71
Квантиль распределения Фишера F 65
Класс (интервал) группировки 7
Ковариация 26
Корреляционное поле (диаграмма рассеяния) 31
Коэффициент корреляции 27
Коэффициент линейной регрессии 36
Критерий проверки гипотез 10, 17
Критерий Уэлча 18
Критическая область 10

Математическое ожидание 5

Относительная частота 7

Правило 3σ 61
Преобразование Фишера 30

Регрессия 34

Среднее квадратическое отклонение 5
Статистика 6

Точечная оценка параметров распределения 5

Уровень значимости 5

Функция Лапласа 59

Частота 7
Число классов при группировке 70

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ: Пер. с англ. − М.: Мир, 1982. – 488 с.

2. Борщ В.Л., Совіт Ю.П. Збірник задач з математичної статистики (для студентів біолого-медичного інституту). − Дніпропетровськ: ДДУ, 1997. – 80 с.

3. Биометрия / Н.В. Глотов, Л.А. Животовский, Н.В. Хованов, Н.Н. Хромов-Борисов. − Л.: Изд-во ЛГУ, 1982. – 263 с.

4. Владимирский Б.М. Математические методы в биологии. − Ростов: Изд-во Ростов. ун-та, 1983. – 304 с.

5. Герасимович А.И. Математическая статистика. − Мн.: Высш. шк., 1983. – 279 с.

6. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии: Пер. с англ. − М.: Прогресс, 1976. – 495 с.

7. Глотов Н.В., Филатов А.А., Хромов-Борисов Н.Н. Сборник задач по биометрии. − Л.: Изд-во ЛГУ, 1985. – 81 с.

8. Данко П.Е., Попов А.Г., Кожевникова Т.Я. Высшая математика в упражнениях и задачах: В 2-х ч. − Ч 2: Учеб. пособие для втузов: В 2 ч. − М.: Высш. шк., 1997. − Ч 2. – 416 с.

9. Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке. Методы обработки данных: Пер. с англ. − М.: Мир, 1980. – 520 с.

10. Жалдак М.І., Кузьміна Н.М., Берлінська С.Ю. Теорія ймовірностей і математична статистика з елементами інформаційної технології. − К.: Вища шк., 1995. – 351 с.

11. Лакин Г.Ф. Биометрия. − М.: Высш. шк., 1980. – 293 с.

12. Математическая статистика / В.Н. Иванова, В.Н. Калинина, Л.А. Нешумова, И.С.Решетникова. − М.: Высш. шк., 1981. – 327 с.

13. Методические указания к изучению курса “Математические методы в биологии” / А.Ф. Приставка, Ю.А. Агеев, А.В. Пашков. − Днепропетровск: ДГУ, 1986. − 56 с.

14. Приставка А.Ф., Передерий А.И., Измайлова С.В. Вычислительные схемы и программное обеспечение критериев однородности и независимости. − Днепропетровск, ДГУ, 1995. − 60с.

15. Рокицкий П.Ф. Биологическая статистика. − Минск: Вышэйшая шк., 1973. − 320 с.

16. Тейлор Дж. Введение в теорию ошибок. − М.: Мир, 1985. − 272 с.

17. Турчин В.М. Математична статистика в прикладах та задачах. − Дніпропетровськ: Вид-во ДДУ, 1992. − 164 с.

18. Турчин В.М. Математическая статистика в примерах и задачах: Учеб. пособие. − Днепропетровск: Изд-во ДГУ, 1992. − 164с.

19. Турчин В.М. Математическая статистика. Лабораторный практикум: Учеб. пособие. − Днепропетровск: Изд-во ДГУ, 1992. − 184с.

20. Турчин В.М. Теория вероятностей и математическая статистика: Учеб. пособие: В 2 т. − Днепропетровск: Изд-во ДГУ, 1996. Т.2: Математическая статистика. − 304с.

21. Урбах В.Ю. Статистический анализ в биологических и медицинских исследованиях. − М.: Медицина, 1975. − 294 с.

22. Фишер Р.А. Статистические методы для исследователей. − М.: Госстатиздат. 1958. − 268 с.

СОДЕРЖАНИЕ

Предисловие 3
1. Точечная и интервальная оценка числовых характеристик распределения, эмпирические функция и закон распределения . 4
1.1. Теоретические сведения 4
1.2. Алгоритм решения 11
1.3. Пример расчета 12
2. Сравнение параметров двух независимых распределений . 15
2.1. Теоретические сведения 15
2.2. Алгоритм решения 19
2.3. Пример расчета 20
3. Сравнение параметров двух зависимых распределений . . 21
3.1. Теоретические сведения 21
3.2. Алгоритм решения 24
3.3. Пример расчета 24
4. Корреляционный анализ зависимых распределений . . . 25
4.1. Теоретические сведения 25
4.2. Алгоритм решения 31
4.3. Пример расчета 32
5. Регрессионный анализ зависимых распределений 34
5.1. Теоретические сведения 34
5.2. Алгоритм решения 38
5.3. Пример расчета 39
6. Задачи к курсовому заданию 43
6.1. Точечная и интервальная оценка числовых характеристик распределения, эмпирические функция и закон распределения 43
6.2. Сравнение параметров двух независимых распределений 45
6.3. Сравнение параметров двух зависимых распределений 49
6.4. Корреляционный анализ зависимых распределений . 51
6.5. Регрессионный анализ зависимых распределений . . 55
7. Справочные сведения 57
7.1. Нормальное распределение $N(\mu,\sigma)$ 57
7.2. Распределение Пирсона χ^2 61
7.3. Распределение Фишера F 65
7.4. Распределение Стьюдента t 70
7.5. Группировка выборочных данных 70
7.6. Преобразование Фишера 70
Предметный указатель 76
Библиографические ссылки 77

Дод. план 1998

Навчальне видання
Володимир Леонідович Борщ
Юрій Петрович Совіт

**МАТЕМАТИЧНІ МЕТОДИ В БІОЛОГІЇ
В ПРИКЛАДАХ І ЗАДАЧАХ**

(для студентів-іноземців біолого-медичного інституту)

Навчальний посібник
Російською мовою
Редактор А.И.Колісник
Техредактор В.М.Пластун
Корректор Л.В.Куценко

Підписано до друку 28.12.98. Формат 60×84/16. Папір друкарський. Друк плоский. Умов. друк. арк. . Облік.-друк. арк. . Тираж 200 прим. Замовлення 133.

Редакційно-видавничий відділ ДДУ, 320625, МСП, м.Дніпропетровськ-10, пр. Гагаріна, 72
Ротапринт ДДУ, 320625, м.Дніпропетровськ, вул.Козакова, 4 б.