Анализ корреляционного отношения

Точечная оценка показателя $r_{\eta\xi}$. Пусть *экспериментальные данные* представлены в форме (6.10), т.е. сгруппированы по значениям x_i случайной величины ξ.

Тогда за значение *точечной оценки* величины σ_f^2 принимают

$$\widehat{\sigma}_f^2 = \frac{1}{n}\sum_{i=1}^{m} n_i(\overline{y}_i - \overline{y})^2.$$

Значение точечной оценки дисперсии σ_η^2 находим по известной формуле (см. 2.1)

$$\widehat{\sigma}_\eta^2 = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \overline{y})^2.$$

Отсюда на основании (6.5) получаем значение точечной оценки показателя $r_{\eta\xi}$:

$$\widehat{r}_{\eta\xi} = \sqrt{\frac{\widehat{\sigma}_f^2}{\widehat{\sigma}_\eta^2}}. \tag{6.20}$$

Напомним, что точечная оценка $\widehat{r}(\vec{X}_n, \vec{Y}_n)$ определяет степень зависимости случайной величины η от случайной величины ξ. Аналогично можно ввести точечную оценку $\widehat{r}_{\xi\eta}$ для *корреляционного отношения* $r_{\xi\eta}$.

Пусть экспериментальные данные получены в форме (6.9) и не допускают удовлетворительной группировки по оси значений ξ (так как недостаточно велико n или точки (x_i, y_i) слишком „разрежены“ на плоскости).

В этом случае нужно выдвинуть некоторое предположение (*статистическую гипотезу*) о виде функции регрессии $\mathrm{M}(\eta|\xi = x) = f(x)$. Проверка таких гипотез будет рассмотрена ниже (см. 7).

Допустим, что параметрический вид этой функции задан, т.е. принято предположение о том, что

$$f(x) = f(x; \theta_1, \ldots, \theta_k)$$

и найдены значения $\widehat{\theta}_i$ оценок параметров θ_i, $i = \overline{1, k}$ (см. 7). Тогда значение точечной оценки $\widehat{\sigma}_\eta^2$ для дисперсии σ_η^2 находим по формуле

$$\widehat{\sigma}_\eta^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2,$$

а значение $\widehat{\overline{\sigma}}_\eta^2$ оценки $\overline{\sigma}_\eta^2$ можно записать в виде

$$\widehat{\overline{\sigma}}_\eta^2 = \frac{1}{n-k}\sum_{i=1}^{n}\Big(y_i - f\big(x_i;\widehat{\theta}_1,\ldots,\widehat{\theta}_k\big)\Big)^2. \tag{6.21}$$

Следовательно, согласно (6.5), точечную оценку показателя $r_{\eta\xi}$ можно определить равенством

$$\widehat{r}_{\eta\xi} = \sqrt{\frac{1-\widehat{\overline{\sigma}}_\eta^2}{\widehat{\sigma}_\eta^2}}. \tag{6.22}$$

Пример 6.7. Пусть в результате обработки $n = 132$ экспериментальных точек $(x_i,\ y_i)$, $i = \overline{1,n}$, получено *выборочное значение* корреляционного отношения $\widehat{r}_{\eta\xi} = 0{,}60$, причем промежуток, содержащий все *выборочные значения* случайной величины ξ, был разбит на $m = 12$ равных интервалов (см. **1.3**). Найдем значения границ доверительного интервала $(\underline{r},\ \overline{r})$ для показателя $r_{\eta\xi}$ с уровнем доверия $\gamma = 0{,}9$ и проверим значимость этого показателя на уровне значимости $\alpha = 0{,}1$.

Сначала определим по формуле (6.23) число степеней свободы r_1^* (округляя до целого числа):

$$r_1^* = \frac{(12-1+132\cdot 0{,}36)^2}{12-1+2\cdot 132\cdot 0{,}36} \approx 27.$$

По таблице квантилей распределения Фишера с числом степеней свободы $r_1^* = 27$ и $r_2 = n - m = 132 - 12 = 120$ (см. табл. П.4) находим квантили уровней $\alpha/2 = (1-\gamma)/2 = 0{,}05$ и $1-\alpha/2 = 0{,}95$:

$$f_{0,05}(27,120) = 1{,}58;$$

$$f_{0,95}(27,120) = \frac{1}{f_{0,05}(120,27)} = \frac{1}{1{,}73} \approx 0{,}58.$$

По формулам (6.24), (6.25) находим значения границ доверительного интервала:

$$\underline{r} = \sqrt{\frac{120\cdot 0{,}36}{132\cdot 0{,}64\cdot 1{,}58} - \frac{11}{132}} = 0{,}49,$$

$$\overline{r} = \sqrt{\frac{120\cdot 0{,}36}{132\cdot 0{,}64\cdot 0{,}58} - \frac{11}{132}} = 0{,}93.$$

Таким образом, с вероятностью $\gamma = 0{,}9$ истинное значение показателя $r_{\eta\xi}$ (при точечной оценке $\widehat{r}_{\eta\xi} = 0{,}60$) заключено в пределах $0{,}49 < r_{\eta\xi} < 0{,}93$.

Для проверки значимости $r_{\eta\xi}$ (хотя она и так очевидна) найдем квантиль распределения Фишера $f_{1-\alpha}(r_1, r_2)$ при $\alpha = 0{,}1$, $r_1 = 11$, $r_2 = 120$. Поскольку $f_{0,9}(120, 11) = 1{,}58$, то $f_{0,1}(11, 120) = 1/f_{0,9}(120, 11) = 0{,}63$. Значение статистики W_0 равно $6{,}1 > f_{0,1} = 0{,}63$, следовательно, гипотеза H_0: $r_{\eta\xi} = 0$ уверенно отклоняется, т.е. между переменными ξ и η имеет место стохастическая связь.

6.5. Анализ множественных связей

Перейдем к рассмотрению *стохастических связей* между совокупностью $p+1$ случайных величин X_0, X_1, ..., X_p, где *переменные* X_1, ..., X_p являются *входными*, а *переменное* $X_0 = Y$ — *выходным*. Такое выделение переменного X_0 не является обязательным, т.е. все переменные могут быть входными, или выходных переменных может быть несколько, но выделенный случай является, по-видимому, наиболее типичным.

Предположим, что случайный вектор $(X_0, X_1, \ldots, X_p)$ имеет нормальный закон распределения, определяемый вектором математических ожиданий $\vec{\mu} = (\mu_0, \mu_1, \ldots, \mu_p)$ и ковариационной матрицей $\Sigma = (\sigma_{ij})$. Таким образом, известна корреляционная матрица

$$P = \begin{pmatrix} 1 & \rho_{01} & \rho_{02} & \cdots & \rho_{0p} \\ \rho_{10} & 1 & \rho_{12} & \cdots & \rho_{1p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{p0} & \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix}, \qquad (6.27)$$

где ρ_{ij} является коэффициентом корреляции между случайными величинами X_i и X_j, $i, j = \overline{0, p}$.

Частные коэффициенты корреляции. При рассмотрении трех и более случайных величин X_0, X_1, ..., X_p коэффициенты корреляции любой пары из этих случайных величин могут не дать правильного представления о степени связи между всеми случайными величинами. Это объясняется тем, что на закон распределения вероятностей исследуемой пары случайных величин могут оказывать влияние и другие рассматриваемые случайные величины (см. примеры 6.1–6.3).

Это обстоятельство делает необходимым введение показателей стохастической связи между парой случайных величин X_i и X_j ($i=\overline{0,p}$, $j=\overline{0,p}$, $i \neq j$) при условии, что значения других случайных величин зафиксированы. В этом случае говорят о статистическом анализе ***частных связей***.

Частный коэффициент корреляции — мера линейной стохастической зависимости между двумя случайными величинами из некоторой совокупности случайных величин X_0, X_1, ..., X_p, когда исключено влияние остальных, т.е. (для пары X_i и X_j)

$$\rho_{ij(J(i,j))} = \frac{\mathbf{M}\big((Z_i - \mathbf{M}Z_i)(Z_j - \mathbf{M}Z_j)\big)}{\sqrt{\mathbf{D}Z_i\,\mathbf{D}Z_j}}, \tag{6.28}$$

где

$$Z_i = \alpha_0^i + \sum_{k\in J(i,j)} \alpha_k^i X_k, \qquad Z_j = \beta_0^j + \sum_{k\in J(i,j)} \beta_k^j X_k$$

и

$$J(i,j) = \{0, 1, \ldots, p\} \setminus \{i, j\}.$$

При этом i, j называют ***первичными индексами***, а остальные — ***вторичными***. Коэффициенты α_0^i, β_0^j, α_k^i, β_k^j, $k \in J(i,j)$, находят из условия минимизации следующих функций:

$$\mathbf{M}\Big(X_i - \alpha_0^i - \sum_{k\in J(i,j)} \alpha_k^i X_k\Big)^2, \qquad \mathbf{M}\Big(X_j - \beta_0^j - \sum_{k\in J(i,j)} \beta_k^j X_k\Big)^2.$$

Если случайный вектор (X_0, ..., X_p) распределен по нормальному закону, то частный коэффициент корреляции между случайными величинами X_i и X_j вычисляют по формуле*

$$\rho_{ij(J(i,j))} = -\frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}, \tag{6.29}$$

где Σ_{ij} — алгебраическое дополнение для элемента ρ_{ij} корреляционной матрицы (6.27).

Например, при $i=0$, $j=1$ по формуле (6.29) имеем

$$\rho_{01(2)} = \frac{\rho_{01} - \rho_{02}\rho_{12}}{\sqrt{(1-\rho_{02}^2)(1-\rho_{12}^2)}}.$$

Из формулы (6.29) следует, что для вычисления частных коэффициентов корреляции нужны лишь все коэффициенты корреляции случайных величин X_i, X_j, $i \neq j$.

Численные расчеты могут быть упрощены, если использовать рекуррентные соотношения**

$$\rho_{01(2,3,\ldots,k+1)} = \frac{\rho_{01(2,\ldots,k)} - \rho_{0(k+1)(2,\ldots,k)}\rho_{1(k+1)(2,\ldots,k)}}{\sqrt{\left(1-\rho^2_{0(k+1)(2,\ldots,k)}\right)\left(1-\rho^2_{1(k+1)(2,\ldots,k)}\right)}}. \qquad (6.30)$$

Согласно (6.30), любой частный коэффициент корреляции может быть выражен через частные коэффициенты с меньшим на единицу числом вторичных индексов.

Замечание 6.3. Практика многомерного статистического анализа показала, что частные коэффициенты корреляции, определенные соотношениями (6.28)–(6.30), — вполне приемлемые характеристики линейной связи и в том случае, когда распределение анализируемых переменных X_0, X_1, ..., X_p отличается от нормального.

Пример 6.8. По итогам работы 37 однотипных прядильных фабрик в течение года были измерены следующие показатели: $X_0 = Y$ — среднемесячная характеристика качества пряжи (в баллах), X_1 — среднемесячное количество профилактических наладок автоматической линии, X_2 — среднемесячное число обрывов нити.

По матрице исходных данных X_{0i}, X_{1i}, X_{2i}, $i=\overline{1,37}$, были подсчитаны выборочные коэффициенты корреляции $\widehat{\rho}_{ij}$ по формуле (6.12):

$$\widehat{\rho}_{01} = 0{,}105; \qquad \widehat{\rho}_{02} = 0{,}024; \qquad \widehat{\rho}_{12} = 0{,}966.$$

Значения $\widehat{\rho}_{01}$ и $\widehat{\rho}_{02}$ дали основание предполагать, что случайные величины X_0, X_i, $i=1,2$, некоррелированные.

Гипотезы о равенстве нулю ρ_{01} и ρ_{02} были приняты на уровне значимости $\alpha = 0{,}1$. Это свидетельствует об отсутствии стохастической связи между X_0 (качество ткани) и X_1, X_2, но не согласуется с профессиональными представлениями технологов.

Однако расчет значений частных коэффициентов корреляции по формуле (6.29) дает $\widehat{\rho}_{01(2)} = 0{,}907$ и $\widehat{\rho}_{02(1)} = -0{,}906$, что вполне соответствует представлениям специалистов о характере связей между рассмотренными показателями.

Построение доверительных интервалов для истинных значений $\rho_{01(2)}$ и $\rho_{02(1)}$, согласно формулам (6.16), с учетом того, что объем выборки $n = 37$ должен быть уменьшен на 1 (ибо число „мешающих" переменных в данном случае равно $k = 1$), дает следующие результаты (на уровне доверия $\gamma = 0{,}9$):

$$0{,}821 < r_{01(2)} < 0{,}950; \qquad -0{,}950 < r_{02(1)} < -0{,}819.$$

Пример 6.9. С целью исследования влияния погодных условий (X_1 — весеннее количество осадков, см; X_2 — накопленная за весну сумма температур, $^\circ$ C) на урожайность (в ц/га) кормовых трав X_0 в районе с одинаковыми метеорологическими условиями были получены выборочные значения вектора (X_0, X_1, X_2) на $n = 20$ участках. По этим ***экспериментальным данным*** (X_{0i}, X_{1i}, X_{2i}), $i = \overline{1, 20}$, были вычислены значения коэффициентов корреляции $\widehat{\rho}_{01} = 0{,}80$; $\widehat{\rho}_{02} = -0{,}40$; $\widehat{\rho}_{12} = -0{,}56$.

Значение $\widehat{\rho}_{02} = -0{,}40$ вызывает вопрос: действительно ли высокая температура X_2 отрицательно влияет на урожайность, или здесь сказывается влияние „мешающего" фактора — количества осадков?

Вычисление значений частных коэффициентов корреляции по формуле (6.29) дает следующие значения:

$$\widehat{\rho}_{01(2))} = 0{,}759; \qquad \widehat{\rho}_{02(1)} = 0{,}097; \qquad \widehat{\rho}_{12(0)} = -0{,}436.$$

Как видим, если исключить одновременное влияние количества осадков X_1 на X_0 (с ростом X_1 урожайность повышается) и на X_2 (с ростом X_1 температура X_2 понижается), то мы уже не обнаружим отрицательной корреляции между температурой X_2 и урожайностью X_0, ибо $\widehat{\rho}_{02(1)} = 0{,}097$, что не является значимой степенью стохастической связи.

Пример 6.11. Двумерная случайная величина имеет нормальный закон распределения. Определим ***доверительный интервал*** для коэффициента корреляции ρ с ***коэффициентом доверия*** $\gamma = 0{,}99$, если значение $\widehat{\rho}$, найденное по ***выборке объема*** $n = 300$, равно 0,14.

Воспользуемся тем, что при больших объемах выборки ***оценка*** $\widehat{\rho}(\vec{X}_n, \vec{Y}_n)$ распределена почти по нормальному закону с параметрами $\rho - \rho(1 - \rho^2)/2n$ и $(1 - \rho^2)^2/n$ (см. **6.3**). По таблице квантилей нормального распределения (см. табл. П.2) найдем квантиль $u_{(1+\gamma)/2} = u_{0{,}995} = 2{,}575$. Имеем

$$\mathbf{P}\left\{\left|\widehat{\rho}(\vec{X}_n,\vec{Y}_n)-\frac{\rho-\rho(1-\rho^2)}{2n}\right|\frac{\sqrt{n}}{1-\rho^2}<u_{0,995}\right\}=0{,}99.$$

Отсюда

$$\widehat{\rho}(\vec{X}_n,\vec{Y}_n)+\frac{\rho(1-\rho^2)}{2n}-u_{0,995}\frac{1-\rho^2}{\sqrt{n}}<\rho<$$
$$<\widehat{\rho}(\vec{X}_n,\vec{Y}_n)+\frac{\rho(1-\rho^2)}{2n}+u_{0,995}\frac{1-\rho^2}{\sqrt{n}}.$$

Заменяя в левой и правой частях неравенств ρ на $\widehat{\rho}$ и подставляя значение $u_{0,995}=2{,}575$, для данной выборки получаем границы доверительного интервала в виде

$$\underline{\rho}=0{,}14+\frac{0{,}14(1-0{,}14^2)}{600}-2{,}575\frac{1-0{,}14^2}{\sqrt{300}},$$
$$\overline{\rho}=0{,}14+\frac{0{,}14(1-0{,}14^2)}{600}+2{,}575\frac{1-0{,}14^2}{\sqrt{300}}.$$

После вычислений окончательно получаем доверительный интервал $(0{,}13,\ 0{,}16)$.

Пример 6.12. Двумерная случайная величина имеет нормальный закон распределения. Построим доверительный интервал для коэффициента корреляции ρ с коэффициентом доверия $\gamma=0{,}95$, если значение $\widehat{\rho}(\vec{X}_n,\vec{Y}_n)$, найденное по выборке объема $n=12$, равно $-0{,}65$.

Поскольку объем выборки мал, используем случайную величину

$$Z=\frac{1}{2}\ln\frac{1+\widehat{\rho}(\vec{X}_n,\vec{Y}_n)}{1-\widehat{\rho}(\vec{X}_n,\vec{Y}_n)},$$

которая имеет приближенно нормальный закон распределения с параметрами

$$\mu=\frac{1}{2}\ln\frac{1+\rho}{1-\rho}+\frac{\rho}{2(n-1)},\qquad \sigma=\frac{1}{n-3}.$$

Используя таблицу квантилей нормального распределения (см. табл. П.2), находим квантиль $u_{(1+\gamma)/2}=u_{0,975}=1{,}96$. Имеем

$$\mathbf{P}\left\{\left|\frac{1}{2}\ln\frac{1+\widehat{\rho}(\vec{X}_n,\vec{Y}_n)}{1-\widehat{\rho}(\vec{X}_n,\vec{Y}_n)}-\frac{1}{2}\ln\frac{1+\rho}{1-\rho}-\frac{\rho}{2(n-1)}\right|<\frac{1{,}96}{\sqrt{n-3}}\right\}=0{,}95,$$

откуда

$$\frac{1}{2}\ln\frac{1+\widehat{\rho}(\vec{X}_n,\vec{Y}_n)}{1-\widehat{\rho}(\vec{X}_n,\vec{Y}_n)}-\frac{1{,}96}{\sqrt{n-3}}<\frac{1}{2}\ln\frac{1+\rho}{1-\rho}+\frac{\rho}{2(n-1)}<$$

$$<\frac{1}{2}\ln\frac{1+\widehat{\rho}(\vec{X}_n,\vec{Y}_n)}{1-\widehat{\rho}(\vec{X}_n,\vec{Y}_n)}+\frac{1{,}96}{\sqrt{n-3}}.$$

Учитывая условия задачи, получаем

$$\frac{1}{2}\ln\frac{0{,}35}{1{,}65}-\frac{1{,}96}{3}<\frac{1}{2}\ln\frac{1+\rho}{1-\rho}+\frac{\rho}{22}<\frac{1}{2}\ln\frac{0{,}35}{1{,}65}+\frac{1{,}96}{3},$$

или

$$-\ln\frac{33}{7}-1{,}31<\ln\frac{1+\rho}{1-\rho}+\frac{\rho}{11}<-\ln\frac{33}{7}+1{,}31.$$

Решая уравнения

$$\ln\frac{1+\rho}{1-\rho}+\frac{\rho}{11}=1{,}31-\ln\frac{33}{7},$$
$$\ln\frac{1+\rho}{1-\rho}+\frac{\rho}{11}=-1{,}31-\ln\frac{33}{7},$$

находим нижнюю $\underline{\rho}$ и верхнюю $\overline{\rho}$ границы доверительного интервала: $\overline{\rho}\approx -0{,}12$, $\underline{\rho}\approx -0{,}88$. Таким образом, доверительный интервал для ρ имеет вид $(-0{,}988,\ -0{,}12)$.

Заметим, что границы доверительного интервала можно определить с помощью (6.16).

Пример 6.14. По выборке объема $n=28$ из двумерной *генеральной совокупности*, распределенной по нормальному закону, найдено значение оценки $\widehat{\rho}=0{,}88$ коэффициента корреляции. Проверим гипотезу $H_0:\rho\geqslant 0{,}90$ при *альтернативной гипотезе* H_1: $\rho<0{,}90$ на уровне значимости $\alpha=0{,}01$.

Для проверки гипотезы H_0 воспользуемся статистикой

$$Z(\vec{X}_n,\vec{Y}_n)=\frac{1}{2}\ln\frac{1+\widehat{\rho}(\vec{X}_n,\vec{Y}_n)}{1-\widehat{\rho}(\vec{X}_n,\vec{Y}_n)}$$

(см. (6.15)), для которой имеем

$$\mathbf{P}\left\{\left(Z(\vec{X}_n,\vec{Y}_n)-\frac{1}{2}\ln\frac{1+\rho}{1-\rho}-\frac{\rho}{2(n-1)}\right)\sqrt{n-3}<u_{0,01}\right\}=0{,}01.$$

С помощью таблицы квантилей нормального распределения (см. табл. П.3) находим квантиль $u_{0,01}=0{,}504$, а затем — границу *критической области* для $Z(\vec{X}_n,\vec{Y}_n)$:

$$\frac{1}{2}\ln\frac{1+\rho}{1-\rho}+\frac{\rho}{2(n-1)}+\frac{u_{0,01}}{\sqrt{n-3}}=\frac{1}{2}\ln\frac{1+0,9}{0,1}+\frac{0,9}{2\cdot 27}+\frac{0,504}{5}=$$
$$=\frac{1}{2}\ln 19+\frac{0,1}{6}+0,1008\approx\frac{1}{2}\ln 19+0,118.$$

Вычислим выборочное значение

$$Z_в=\frac{1}{2}\ln\frac{1+0,88}{1-0,88}\approx\frac{1}{2}\ln 15,7.$$

Поскольку выборочное значение попало в критическую область ($Z_в<\frac{1}{2}\ln 19+0,118$), гипотезу H_0 отклоняем.

Пример 6.15. По выборке объема $n=19$, заданной в виде таблицы (табл. 6.3), найдем значение оценки *корреляционного*

Таблица 6.3

x	0	0	0	1	2	2	3	3	4	4
y	22,8	21,9	22,1	24,5	26,0	26,1	26,8	27,3	28,2	28,5
x	5	6	6	6	7	8	8	9	1	
y	28,9	30,0	30,3	29,8	30,4	31,4	31,5	31,8	33,7	

отношения и границы соответствующего доверительного интервала с коэффициентом доверия $\gamma=0,8$.

Вычисляем выборочное среднее

$$\overline{y}=\frac{1}{19}\Big(22,8+21,9+22,1+24,5+$$
$$+26,0+26,1+26,8+27,3+28,2+28,5+28,9+30,0+$$
$$+30,3+29,8+30,4+31,4+31,5+31,8+33,1\Big)\approx 28,0$$

и выборочную дисперсию

$$\widehat{\sigma}^2=\frac{1}{19}\Big(22,8^2+21,9^2+22,1^2+24,5^2+26,0^2+26,1^2+26,8^2+$$
$$+27,3^2+28,2^2+28,5^2+28,9^2+30,0^2+30,3^2+29,8^2+30,4^2+$$
$$+31,4^2+31,5^2+31,8^2+33,1^2\Big)-28^2\approx 292,99-782,32=10,67.$$

Чтобы вычислить значение $\widehat{\sigma}_f^2$, составим *статистический ряд* (табл. 6.4). С помощью этого ряда находим

$$\widehat{\sigma}_f^2=\frac{1}{19}\Big(3(22,3-28,0)^2+(24,5-28,0)^2+2(26,0-28,0)^2+$$
$$+2(27,0-28,0)^2+2(28,4-28,0)^2+(28,9-28,0)^2+$$
$$+3(30,0-28,0)^2+(30,4-28,0)^2+2(31,5-28,0)^2+$$
$$+(31,8-28,0)^2+(33,1-28,0)^2\Big)\approx 8,18.$$

Согласно формуле (6.20),

$$\widehat{r}_{\xi\eta} = \sqrt{\frac{8{,}18}{10{,}67}} \approx \sqrt{0{,}77} \approx 0{,}9.$$

Таблица 6.4

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	3	1	2	2	2	1	3	1	2	1	1
y_i	22,3	24,5	26,0	27,0	28,4	28,9	30,0	30,4	31,5	31,8	33,1

Чтобы определить границы доверительного интервала, предварительно найдем степени свободы

$$r_1^* = \frac{(11-1+19\cdot 0{,}77)^2}{11-1+2\cdot 19\cdot 0{,}77} + \frac{(10+19\cdot 0{,}77)^2}{10+38\cdot 0{,}77} \approx 15$$

и $r_2 = n - m = 19 - 11 = 8$, а также квантили распределения Фишера

$$F_{1-0{,}1} = F_{0{,}9} = 2{,}46, \qquad F_{0{,}1} = \frac{1}{F_{0{,}9}} \approx 0{,}41.$$

Далее с помощью формул (см. (6.24), (6.25)) получаем

$$\underline{r}_{\eta\xi} = \sqrt{\frac{(19-11)\cdot 0{,}77}{19(1-0{,}77)\cdot 2{,}46} - \frac{10}{19}} = \sqrt{\frac{10{,}90-10}{19}} \approx \sqrt{0{,}05} \approx 0{,}2,$$

$$\overline{r}_{\eta\xi} = \sqrt{\frac{(19-11)\cdot 0{,}77}{19(1-0{,}77)\cdot 0{,}41} - \frac{10}{19}} = \sqrt{\frac{10{,}90-10}{19}} \approx \sqrt{2{,}91} \approx 1{,}7.$$

Итак, $0{,}2 < r_{\eta\xi} < 1$.

Условия задач.

6.16. По выборке объема $n = 20$ (табл. 6.7) найдите значение оценки корреляционного отношения.

Таблица 6.7

x	1,0	1,0	1,5	2,0	2,0	2,5	2,5	3,0	3,0	3,5
y	0,2	0,3	0,3	0,3	0,4	0,4	0,5	0,5	0,5	0,8
x	3,5	4,0	4,0	4,5	5,0	5,5	5,5	5,5	6,0	6,5
y	0,7	1,1	1,0	1,2	1,7	2,3	2,2	2,4	2,7	3,3

Ответ: $\widehat{r}_{\eta\xi} = 0{,}95$.

6.17. По результатам 10 наблюдений, заданным таблицей (табл. 6.8), найдите:

а) значения оценок коэффициентов корреляции $\widehat{\rho}_{01}$, $\widehat{\rho}_{02}$, $\widehat{\rho}_{12}$;

б) значения оценок частных коэффициентов корреляции $\widehat{\rho}_{01(2)}$ и $\widehat{\rho}_{02(1)}$;

в) значения границ доверительного интервала для $\rho_{01(2)}$ и $\rho_{02(1)}$ с коэффициентом доверия 0,95;

г) значения оценки коэффициента детерминации.

Таблица 6.8

x_1	1	4	0	5	−3	3	−5	−1	2	−2
x_2	4	−6	2	−4	12	−2	14	6	0	8
y	−4	−5	4	−1	4	0	5	1	2	7

Ответ: а) $\widehat{\rho}_{12} = -0{,}98$; $\widehat{\rho}_{01} = -0{,}73$; $\widehat{\rho}_{02} = 0{,}69$; б) $\widehat{\rho}_{01(2)} = -0{,}36$; $\widehat{\rho}_{02(1)} = -0{,}15$; в) $-0{,}3 \pm 0{,}57$; $-0{,}15 \pm 0{,}64$; г) 0,54.