СТРОГИЕ ОПРЕДЕЛЕНИЯ

3.4. Зависимость и независимость двух случайных величин

Зная законы распределения случайных величин X и Y, входящих в систему (X, Y), можно найти закон распределения системы *только в случае, когда случайные величины X и Y — независимы.* С понятием независимости случайных величин мы уже встречались в п. 2.5: две случайные величины называются *независимыми*, если закон распределения каждой из них не зависит от того, какие значения принимает другая. В противном случае случайные величины называются *зависимыми.*

Коэффициент корреляции r.

О п р е д е л е н и е. ***Ковариацией*** (или ***корреляционным моментом***) K_{xy} *случайных величин X и Y называется математическое ожидание произведения отклонений этих величин от своих математических ожиданий*[1], т.е.

$$K_{xy} = M[(X - M(X))(Y - M(Y))], \qquad \textbf{(5.32)}$$

или

$$K_{xy} = M[(X - a_x)(Y - a_y)],$$

где $a_x = M(X)$, $a_y = M(Y)$.

$$K_{xy} = M(XY) - M(X) \cdot M(Y)$$

Определение. ***Коэффициентом корреляции***[1] *двух случайных величин называется отношение их ковариации к произведению средних квадратических отклонений этих величин*:

$$\rho_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}. \qquad \textbf{(5.38)}$$

Коэффициент корреляции принимает значения на отрезке [−1; 1], т.е.

$$-1 \le \rho \le 1. \qquad \textbf{(5.39)}$$

Теорема.

Если случайные величины X и У независимы, то их коэффициент корреляции равен нулю.

Случайные величины называются *некоррелированными*, если их коэффициент корреляции равен нулю. Таким образом, из *независимости случайных величин следует их некоррелированность.*

Обратное утверждение, вообще говоря, неверно: из *некоррелированности двух случайных величин еще не следует их независимость* (см. пример **5.6**).

3. *Если коэффициент корреляции двух случайных величин равен (по абсолютной величине) единице, то между этими случайными величинами существует линейная функциональная зависимость.*

ОЦЕНКА (НАХОЖДЕНИЕ ПРИБЛИЖЕННЫХ ЗНАЧЕНИЙ) КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ ПО ВЫБОРКЕ.

ВЫЧИСЛЕНИЕ ПРИБЛИЖЕННОГО ЗНАЧЕНИЯ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Мерой силы (тесноты) и направления линейной связи между двумя переменными x и y, вычисленной по ряду из n пар $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, является эмпирический коэффициент корреляции:

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\overline{y}^2\right)}}$$

Где соответствующие средние значения:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Коэффициент ранговой корреляции Спирмена

Для определения взаимозависимости между переменными (также и нечисловой природы) можно пользоваться коэффициентом ранговой корреляции Спирмена. Ряды измерений x_i и y_i, $i = 1, 2, ...$,преобразуются с помощью рангов следующим образом. Каждому значению x_i ставится в соответствие ранг R_i, т.е. номер элемента x_i в вариационном ряду (ряду x_i, упорядоченных по возрастанию); аналогичным образом определяются ранги R'_i элементов y_i. Таким образом,

каждой паре (x_i, y_i) соответствует пара рангов (R_i, R'_i). Ранговый коэффициент корреляции Спирмена вычисляется по формуле:

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} (R_i - R'_i)^2}{n \cdot (n^2 - 1)}. \qquad (4.2)$$

R_i	1	2	3	4	5	6	7	8	9	10
R_i'	6	5	1	4	2	7	8	10	3	9
$R_i - R_i'$	–5	–3	2	0	3	–1	–1	–2	6	1

По формуле $r_s = 1 - \dfrac{6 \cdot \sum_{i=1}^{10} (R_i - R_i')^2}{10 \cdot (10^2 - 1)}$ вычислим ранговый коэффициент корреляции $r_s = 1 - \dfrac{6 \cdot (25+9+4+9+1+1+4+36+1)}{10 \cdot (100-1)} \approx 0,455$.

4.3. Доверительный интервал для коэффициента корреляции r_0

При достаточно больших n (n > 25) статистика **w :**

$$w = \frac{1}{2} \ln \frac{1+r}{1-r}$$

имеет приближенно нормальное распределение с параметрами

$$M[w] = \frac{1}{2} \ln \frac{1+r_0}{1-r_0}, \quad D[w] = \frac{1}{n-3}.$$

Таким образом, случайная величина

$$z = \frac{w - M[w]}{\sqrt{D[w]}} \qquad (4.4)$$

приближенно удовлетворяет нормированному нормальному распределению N(0;1). Для заданной доверительной вероятности γ определяем соответствующее значение квантиля z.

$$\text{th}\left(\frac{1}{2} \ln \frac{1+r}{1-r} - z_\alpha \frac{1}{\sqrt{n-3}}\right) < r_0 < \text{th}\left(\frac{1}{2} \ln \frac{1+r}{1-r} + z_\alpha \frac{1}{\sqrt{n-3}}\right).$$

где **th(y)** $= \dfrac{e^{2y}-1}{e^{2y}+1}$;

Пример 2. Выборочный коэффициент корреляции, вычисленный по выборке объема n=50, равен r_0=0,687. Найти доверительный интервал для r_0 . Принять γ=0,95.

Решение. По таблице функции Лапласа $\Phi_0(x)$ находим соответствующее значение квантиля t: $\Phi_0(Z)=\gamma/2=0{,}475$. Z=1,96.

$$\frac{1}{\sqrt{n-3}}=\frac{1}{\sqrt{47}}=0{,}146$$

$$\mathrm{th}\left(\frac{1}{2}\ln\frac{1+0{,}687}{1-0{,}687}-1{,}96\cdot 0{,}146\right)<r_0<\mathrm{th}\left(\frac{1}{2}\ln\frac{1+0{,}687}{1-0{,}687}+1{,}96\cdot 0{,}146\right)$$

Отсюда получаем

$$0{,}505<r_0<0{,}810$$

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ КОЭФФИЦИЕНТА r НУЛЮ.

Проверка гипотезы H_0 о равенстве нулю коэффициента корреляции r при уровне значимости α = 0,05 .

H_0: коэффициент корреляции ***r*** равен нулю.

H_1: коэффициент корреляции ***r*** не равен нулю.

Порядок выполнения:

Вычисляется **r**

Вычисляется критериальная статистика: T = √(n-3)*ln((1+r)/(1-r))/2

При истинности H_0 статистика T асимптотически (для достаточно больших значений *n)* имеет стандартное нормальное распределение.

Если используется таблица нормального распределения $\Phi_0(x)$, то критическое значение t находится из $\Phi_0(t)=(1-\alpha)/2$.

При использовании таблицы квантилей стандартного нормального распределения находится как квантиль порядка 1-α/2.

Гипотеза H_0 принимается, если $|T| < t$.