

ПРАКТИЧЕСКИЙ АСПЕКТ АЛГОРИТМА РЕШЕНИЙ КОРРЕЛЯЦИИ

При решении задачи необходимо:

1. вычислить коэффициент корреляции по методу квадратов (Пирсону) и методу рангов (Спирмену)
2. определить характер (направление) и силу связи между соответствующими признаками (явлениями)
3. вычислить ошибку коэффициента корреляции
4. вычислить достоверность коэффициента корреляции
5. оценить коэффициент корреляции по силе и направлению связи

Условие для решения типовой задачи

Зависимость между температурой тела и частотой пульса

	Таблица 1
Т тела (x)	Частота пульса (y)
36	60
36	65
36	70
38	80
40	90
40	100

Метод квадратов (Пирсона)

t° тела (x)	Частота пульса (y)	d _x	d _y	d _x × d _y	d _x ²	d _y ²
1.	2.	3.	4.	5.	6.	7.
36	60	-1,67	-17,5	29,23	2,79	306,25
36	65	-1,67	-12,5	20,88	2,79	156,25
36	70	-1,67	-7,5	12,53	2,79	56,25
38	80	+0,33	+2,5	0,83	0,11	6,25
40	90	+2,33	+12,5	29,13	5,43	156,25
40	100	+2,33	+22,5	52,43	5,43	506,25
Σx = 226 Mx=37,67	Σy = 465 My=77,5	Σd _x = 0	Σd _y = 0	Σd _x × d _y = = 145,03	Σd _x ² = = 19,34	Σd _y ² = = 1187,5

). Определяем средние величины для признаков x и y (графы 1 и 2) :

$$M_x = \frac{226}{6} = 37,67 \quad M_y = \frac{465}{6} = 77,5$$

2). Находим d – отклонение каждой варианты от средней величины для ряда x ($d_x = x - M_x$) и для ряда y ($d_y = y - M_y$). Полученные результаты заносим в таблицу (графы 3 и 4).

3). Полученные отклонения умножаем друг на друга ($d_x \times d_y$) и суммируем ($\sum d_x \times d_y$). Полученные результаты заносим в таблицу (графа 5).

4). Каждое отклонение возводим в квадрат и суммируем по ряду x - $\sum d_x^2$ и по ряду y - $\sum d_y^2$. Полученные результаты заносим в таблицу (графы 6 и 7).

5). Рассчитываем коэффициент корреляции (r_{xy}) по формуле

$$r_{xy} = \frac{\sum d_x \times d_y}{\sqrt{\sum d_x^2 \times \sum d_y^2}},$$

подставив в нее полученные результаты (из таблицы 2):

$$r_{xy} = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}} = \frac{145,03}{\sqrt{19,34 \times 1187,5}} = \frac{145,03}{\sqrt{22966,25}} = \frac{145,03}{151,56} = +0,96$$

6). Определяем ошибку коэффициента корреляции ($m_{r_{xy}}$) по формуле:

$$m_{r_{xy}} = \pm \sqrt{\frac{1 - r_{xy}^2}{n - 2}}, \text{ где } r_{xy} \text{ – коэффициент корреляции, } n \text{ – число парных вариантов}$$

$$m_{r_{xy}} = \sqrt{\frac{1 - 0,96^2}{6 - 2}} = \sqrt{\frac{1 - 0,92}{4}} = \sqrt{\frac{0,08}{4}} = \sqrt{0,02} = \pm 0,14$$

7). Определяем достоверность коэффициента корреляции по формуле:

$$t = \frac{r_{xy}}{m_r}, \text{ где } r_{xy} \text{ – коэффициент корреляции, } m_r \text{ – ошибка коэффициента корреляции}$$

$$t = \frac{0,96}{0,14} = 6,86$$

Вывод: между температурой тела и частотой пульса существует прямая, сильная связь, т. к. коэффициент корреляции равен +0,96.

Коэффициент корреляции достоверен с вероятностью безошибочного прогноза $P > 99\%$, т.к. утроенная ошибка, равная 0,42 меньше коэффициента корреляции.

Примечание: с достаточной для медицинских исследований надежностью о наличии той или иной степени связи можно утверждать только тогда, когда величина коэффициента корреляции превышает или равняется величине трех своих ошибок ($r \geq 3m_r$).

Уровень молочной кислоты в крови и длительность охлаждения организма

дни охлаждения (x)	молочная кислота, в мг% (y)
1	7,0
2	7,0
3	7,2
4	7,1
5	8,5
6	8,9
7	8,7
8	9,0
9	9,5
10	9,3

Средняя температура сезона и заболеваемость дизентерией

года	средняя температура сезона, С ⁰ (x)	заболеваемость дизентерией в условн.ед. (y)
2007	14,3	88
2008	15,0	77
2009	14,6	60
2010	13,2	67
2011	15,2	117
2012	15,0	67
2013	14,1	68
2014	13,2	59
2015	17,7	31
2016	14,8	70
2017	17,8	75

Частота раннего прикорма и желудочно-кишечных инфекций у детей в возрасте до 1 года

районы	частота раннего прикорма (на 100 детей до 1 года)	заболеваемость желудочно-кишечными инфекциями (на 100 детей)
А	8,0	15,0
Б	12,0	20,0
В	16,0	30,0
Г	20,0	25,0
Д	25,0	35,0
Е	24,0	34,0
Ж	24,0	35,0
З	28,0	38,0

Варианты заданий на вычисления линейной корреляции

Рассчитать корреляцию и взаимосвязь

Уровни систолического и диастолического давления у 7 лиц призывного возраста проходящих медицинскую комиссию:

Систолическое давление (в мм от. ст)	120	115	125	110	120	130
--------------------------------------	-----	-----	-----	-----	-----	-----

Диастолическое давление (в мм рт. ст)	80	70	60	65	75	85
---------------------------------------	----	----	----	----	----	----

Вычислить коэффициент корреляции методом линейной корреляции и оценить его репрезентативность, используя исходные данные, приведенные в таблице

Рост и масса тела десяти школьников 2 класса:

Рост тела (см)	145	112	120	127	140	135	140
Масса тела (кг)	20	22	30	35	27	30	37