

Практическое занятие Корреляционный анализ

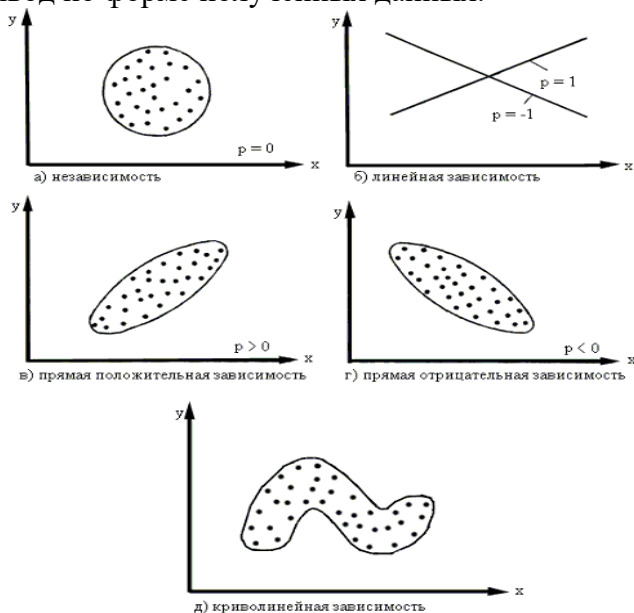
Задание 1.

Вариант 2. Определить форму и направление взаимосвязи между затратами на качество и уровнем брака:

Р, тыс тг ~	107,5	110	110	115	115	107,5	107,5	120	122,5	112,5	120	110
С, тыс тг	57	60	58	61	63	58	55	64	65	64	66	61

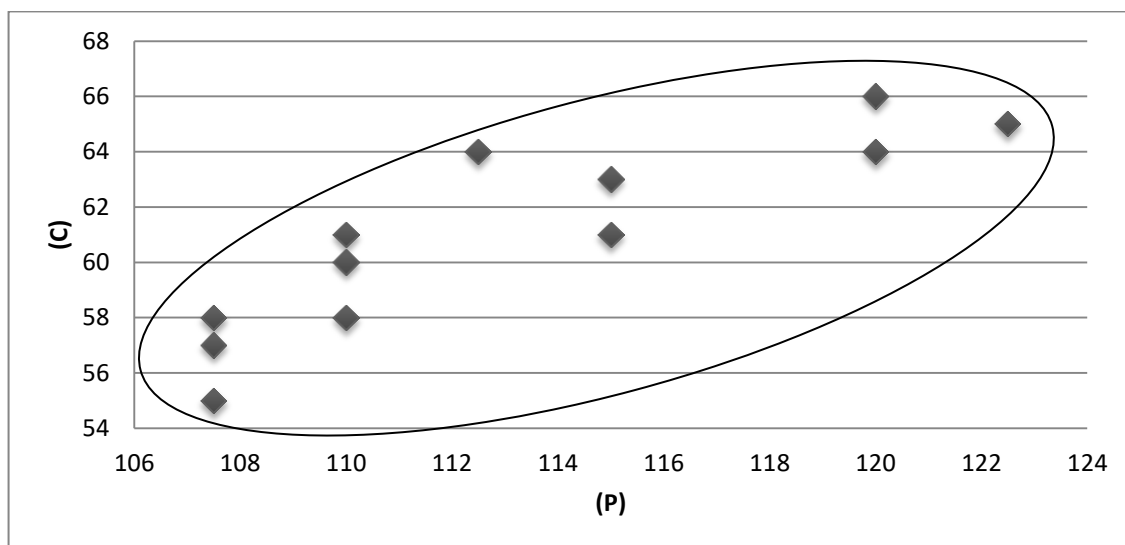
Методические рекомендации к расчету.

Построить на графике взаимосвязь уровня брака (у) от затрат на качество продукции (х). Сделать вывод по форме полученных данных.



Решение:

Диаграмма рассеяния для данных пар значений выглядит следующим образом:



По диаграмме можно сделать вывод, что между затратами на качество и уровнем брака есть некоторая взаимосвязь. Точки на графике расположены в форме восходящего овала ($p > 0$), что говорит о прямой положительной зависимости между затратами на качество и уровнем брака.

Задание 2. Нормированный коэффициент корреляции Браве-Пирсона

Вариант 2. Определить достоверность взаимосвязи между стоимостью и производительностью у 8 установок с помощью расчета нормированного коэффициента корреляции, если данные выборки таковы:

Стоимость ~ 66; 80; 73; 74; 85; 79; 68; 71.

Производительность: ~ 70; 85; 78; 78; 90; 84; 66; 72. $r_T=0,71$

Методические рекомендации к расчету.

Расчет нормированного коэффициента корреляции Пирсона произвести по формуле

$$r_{xy}^P = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$$

Данные тестирования занести в рабочую таблицу и сделать необходимые расчеты

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	y_i	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
\bar{x}	$\sum (x_i - \bar{x})$	$\sum (x_i - \bar{x})^2$	\bar{y}	$\sum (y_i - \bar{y})$	$\sum (y_i - \bar{y})^2$	$\sum (x_i - \bar{x})(y_i - \bar{y})$

Рассчитать число степеней свободы по формуле $k = n - 2$

Сравнить расчетное значение нормированного коэффициента корреляции с табличным значением при $\alpha = 5\%$ и сделать вывод.

Решение:

Вычислим среднее арифметическое значение стоимости и производительности:

Среднее значение стоимости:

$$(\bar{x}) = (66 + 80 + 73 + 74 + 85 + 79 + 68 + 71) / 8 = 74.5$$

Среднее значение производительности:

$$(\bar{y}) = (70 + 85 + 78 + 78 + 90 + 84 + 66 + 72) / 8 = 78.875$$

Используя полученные значения вычисляем параметры, необходимые для расчёта нормированного коэффициента корреляции Пирсона:

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	y_i	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
66	-8,5	72,25	70	-7,875	62,0156	66,9375
80	5,5	30,25	85	7,125	50,7656	39,1875
73	-1,5	2,25	78	0,125	0,01563	-0,1875
74	-0,5	0,25	78	0,125	0,01563	-0,0625
85	10,5	110,25	90	12,125	147,016	127,313
79	4,5	20,25	84	6,125	37,5156	27,5625
68	-6,5	42,25	66	-11,875	141,016	77,1875
71	-3,5	12,25	72	-5,875	34,5156	20,5625
\bar{x}	$\sum (x_i - \bar{x})$	$\sum (x_i - \bar{x})^2$	\bar{y}	$\sum (y_i - \bar{y})$	$\sum (y_i - \bar{y})^2$	$\sum (x_i - \bar{x})(y_i - \bar{y})$
74.5	0	290	77.875	0	472,875	358,5

Нормированный коэффициент корреляции Пирсона, рассчитанный по вышеприведённым данным:

$$r_{xy}^P = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} = 0,97$$

Количество степеней свободы равно $n-2$, где n – число элементов в каждой выборке, то есть $k = 8-2 = 6$.

Теперь необходимо сравнить полученное значение коэффициента корреляции с табличным значением при уровне значимости $\alpha = 0.05$ и числе степеней свободы $k = 6$. Используя таблицу критических значений коэффициента корреляции Пирсона, найдем критическое значение для $\alpha = 0.05$ и $k = 6$, которое равно 0.71. Таким образом, полученное значение коэффициента корреляции $r = 0,97$ значительно больше критического значения 0.71, что свидетельствует о высокой степени взаимосвязи между стоимостью и производительностью установок, близкой к прямолинейной.

Итак, можно сделать вывод, что нормированный коэффициент корреляции, равный 0,97, говорит о наличии положительной линейной связи между стоимостью и производительностью установок. Это означает, что более дорогие установки имеют более высокую производительность, а менее дорогие установки – более низкую производительность.

Однако, следует помнить, что коэффициент корреляции не дает никакой информации о причинно-следственных связях между переменными. Также необходимо учитывать, что выборка состоит только из 8 наблюдений, что может быть недостаточно для общего вывода о связи между стоимостью и производительностью установок во всей генеральной совокупности.

Задание 3. Коэффициент ранговой корреляции Спирмена

Вариант 2. Определить достоверность взаимосвязи между стоимостью и производительностью 9 установок с помощью расчета рангового коэффициента корреляции, если данные выборки таковы:

Стоимость ~ 156; 130; 143; 124; 135; 125; 138; 141; 139.

Производительность ~ 16; 15; 20; 20; 16; 15; 15; 20; 15. . $r_T=0,68$

Методические рекомендации к расчету.

Расчет рангового коэффициента корреляции Спирмена произвести по формуле

$$r_{xy}^S = 1 - \frac{6 \cdot \sum (d_x - d_y)^2}{n \cdot (n^2 - 1)},$$

где: d_x и d_y — ранги показателей x и y ;

n — число коррелируемых пар или исследуемых.

Данные тестирования занести в рабочую таблицу и сделать необходимые расчеты.

x_i	d_x	y_i	d_y	$(d_x - d_y)$	$(d_x - d_y)^2$
				$\sum (d_x - d_y)$	$\sum (d_x - d_y)^2$

Сравнить расчетное значение рангового коэффициента корреляции с табличным значением при $\alpha = 5\%$ и сделать вывод.

Решение:

Для расчета рангового коэффициента корреляции Спирмена необходимо отсортировать данные для каждой выборки по возрастанию и присвоить ранги. Если значения повторяются, то им присваивается среднее значение ранга.

Результаты ранжирования и параметры, необходимые для расчёта рангового коэффициента корреляции сведены в представленную ниже таблицу:

x_i	d_x	y_i	d_y	$(d_x - d_y)$	$(d_x - d_y)^2$
156 ₉	9	16 ₅	5.5	3,5	12,25
130 ₃	3	15 ₁	2.5	0,5	0,25
143 ₈	8	20 ₇	8	0	0
124 ₁	1	20 ₈	8	-7	49
135 ₄	4	16 ₆	5.5	-1,5	2,25
125 ₂	2	15 ₂	2.5	-0,5	0,25
138 ₅	5	15 ₃	2.5	2,5	6,25
141 ₇	7	20 ₉	8	-1	1
139 ₆	6	15 ₄	2.5	3,5	12,25
				$\sum (d_x - d_y)$	$\sum (d_x - d_y)^2$
				0	83,5

Ранговый коэффициент корреляции Спирмена, рассчитанный по вышеприведённым данным:

$$r_{xy}^S = 1 - \frac{6 \cdot \sum (d_x - d_y)^2}{n \cdot (n^2 - 1)} = 0,66$$

Таким образом, если расчетное значение рангового коэффициента корреляции Спирмена равно 0,66, а табличное критическое значение при $\alpha = 5\%$ равно 0,6, то можно сделать вывод, что расчетное значение почти не превышает критическое значение.

Это означает, что мы не можем отвергнуть нулевую гипотезу о том, что корреляция между двумя переменными равна нулю на уровне значимости 5%. Следовательно, мы не можем заключить, что между двумя переменными существует значимая корреляция.