

КРИВОЛИНЕЙНАЯ КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

Пример. Определить зависимость урожайности (Y, ц/га) малины от пораженности ее антракнозом (X, %)

Урожайность, ц/га	98	92	85	93	89	75
Развитие болезни, %	5,1	10,2	20,2	8,0	16,1	28,4

Урожайность, ц/га	90	95	90	63	81	68
Развитие болезни, %	14,1	6,4	12,8	36,6	23,8	34,1

Следует определить, существует ли зависимость между урожайностью малины и пораженностью растений антракнозом.

1. Заполняем расчетную таблицу.

Значения признака X ранжируем в порядке возрастания (т.е. размещают в порядке возрастания от самого маленького до самого большого значения). Напротив каждого значения X записываем соответствующее ему значение Y. Т.е. ранжируются только значения признака X.

X	\bar{x}_y	Y	\bar{y}_x	$Y - \bar{y}_x$	$Y - \bar{y}$	$(Y - \bar{y}_x)^2$	$(Y - \bar{y})^2$
5,1	6,5	98	95	3	13	9	169
6,4		95		0	10	0	100
8,0		93		-2	8	4	64
10,2	12,4	92	91	1	7	1	49
12,8		90		-1	5	1	25
14,1		90		-1	5	1	25
16,1	20,0	89	85	4	4	16	16
20,2		85		0	0	0	0
23,8		81		-4	-4	16	16
28,4	33,0	75	69	6	-10	36	100
34,1		68		-1	-17	1	289
36,6		63		-6	-22	36	484
-	-	$\Sigma=1019$	-	-	-	$\Sigma=121$	$\Sigma=1337$

Значения разбиваем на группы. В каждой группе должно быть не менее 2 значений, причем размер может быть неодинаковым. В нашем примере разобьем значения на 4 группы.

Находим среднюю арифметическую по ряду Y и групповые средние

$$\bar{x}_y \text{ и } \bar{y}_x.$$

Средняя арифметическая по ряду Y

$$\bar{y} = \frac{\sum Y}{n} = \frac{98+95+93+92+90+90+89+85+81+75+68+63}{12} = \frac{1019}{12} = 85$$

Средние арифметические по ряду X:

$$\bar{x}_{y1} = \frac{5,1+6,4+8,0}{3} = \frac{19,5}{3} = 6,5$$

$$\bar{x}_{y2} = \frac{10,2+12,8+14,1}{3} = \frac{37,1}{3} = 12,4$$

$$\bar{x}_{y3} = \frac{16,1+20,2+23,8}{3} = \frac{60,1}{3} = 20,0$$

$$\bar{x}_{y4} = \frac{28,4+34,1+36,6}{3} = \frac{99,1}{3} = 33,0$$

Средние арифметические по ряду Y:

$$\bar{y}_{x1} = \frac{98+95+93}{3} = \frac{286}{3} = 95$$

$$\bar{y}_{x2} = \frac{92+90+90}{3} = \frac{272}{3} = 91$$

$$\bar{y}_{x3} = \frac{89+85+81}{3} = \frac{255}{3} = 85$$

$$\bar{y}_{x4} = \frac{75+68+63}{3} = \frac{206}{3} = 69$$

2. Определяем:
корреляционное отношение

$$\eta = \sqrt{\frac{\sum(Y-\bar{y})^2 - \sum(Y-\bar{y}_x)^2}{\sum(Y-\bar{y})^2}} = \sqrt{\frac{1337-121}{1337}} = \sqrt{\frac{1216}{1337}} = \sqrt{0,91} = 0,95$$

ошибку корреляционного отношения

$$S_{\eta} = \sqrt{\frac{1-\eta^2}{n-2}} = \sqrt{\frac{1-0,95^2}{12-2}} = \sqrt{\frac{1-0,91}{10}} = \sqrt{\frac{0,09}{10}} = \sqrt{0,009} = 0,09$$

n – количество парных значений

критерий существенности

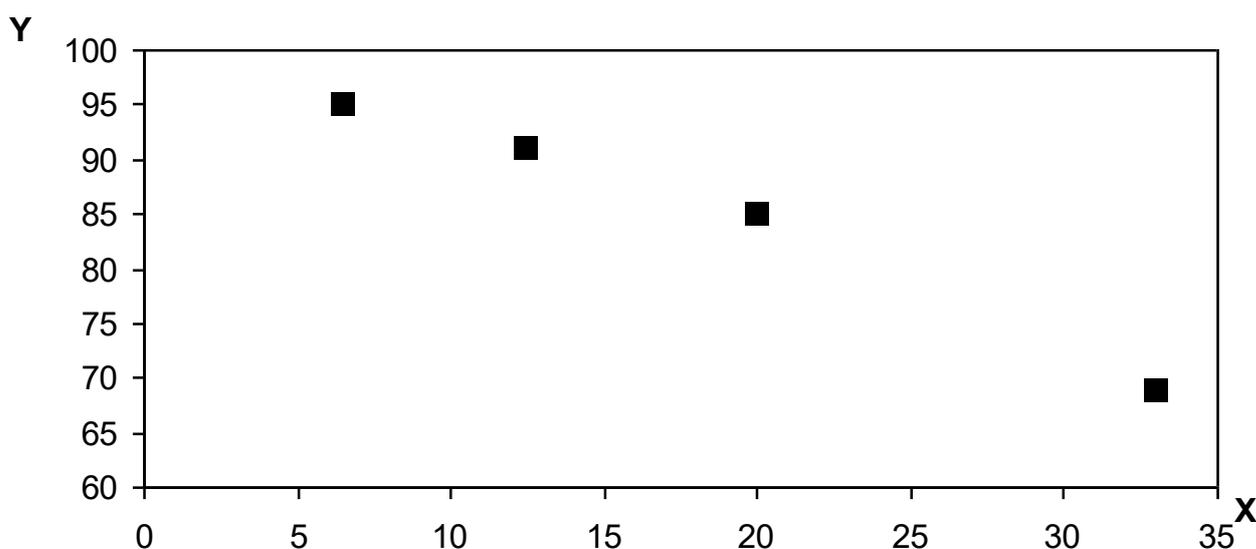
$$t_{\eta} = \frac{\eta}{S_{\eta}} = \frac{0,95}{0,09} = 10,56$$

Сравниваем критерий существенности с критерием Стьюдента, который берем из статистической таблицы на 5% уровне значимости и числе степеней свободы $\nu = n-2$.

$$\text{При } \nu = n-2=12-2=10 \quad t_{05} = 2,23$$

Критерий существенности больше критерия Стьюдента, следовательно между изучаемыми признаками существует **значимая корреляционная зависимость**.

3. На основании групповых средних строим точечный график и определяем теоретическую линию регрессии, проведя кривую между точками.



Вывод. Установлена значимая обратная корреляционная зависимость между урожайностью малины и пораженностью растений антракнозом. С увеличением пораженности антракнозом урожайность малины снижается.

Задание. Согласно N примера рассчитать корреляционное отношение, определить его значимость и представить данные в виде графика.

Пример 1. Определить зависимость урожайности озимой пшеницы (Y, ц/га) от засоренности посевов (X, шт/м²)

Урожайность, ц/га	50,5	49,8	40,7	44,5	42,5	39,1	47,6	38,5	46,8
Засоренность, шт/м ²	5	10	163	50	100	258	20	286	38

Урожайность, ц/га	43,8	43,1	43,7	44,1	39,9	48,3	43,3	39,5	43,5
Засоренность, шт/м ²	75	120	68	75	200	30	145	217	87

Пример 2. Определить зависимость между заселенностью красным плодовым клещом (Y, балл) и толщиной штамба (X, мм)

Заселенность, балл	0,3	0,7	3,0	2,5	0,5	1,0	2,0	1,5	3,5
Толщина штамба, мм	73	68	63	65	70	68	66	66	62

Заселенность, балл	4,8	4,2	4,0	0,8	4,6	4,9	0,7	1,3	3,7
Толщина штамба, мм	59	60	62	65	58	60	67	67	61

Пример 3. Определить степень поврежденности побегов черной смородины листовой галлицей (Y, %) в зависимости от количества эфирных желез на листе (X, шт.).

Степень повреждения, %	57,5	72,5	57,4	40,0	67,5	27,5	45,0	27,5	52,5
Количество эфирных желез, шт	48	50	53	54	57	61	61	61	62

Степень повреждения, %	57,5	25,5	22,5	5,0	16,0	2,5	5,0	5,0	2,5
Количество эфирных желез, шт	69	72	98	112	118	119	126	188	165

критерий существенности

$$t_{\eta} = \frac{\eta}{S_{\eta}}$$

при $v = n - 2$

$t_{05} =$

4. Построить теоретическую линию регрессии по групповым средним X и Y.



Вывод:

Контрольные вопросы к работе:

1. Какая корреляция называется нелинейной?
2. Какой коэффициент описывает нелинейную зависимость? Почему он является универсальным?
3. Как определить форму связи?