

ПРЯМОЛИНЕЙНАЯ КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

Пример. Зависимость урожайности кукурузы (Y, т/га) от засоренности посевов (X, шт/м²).

В результате наблюдений были получены следующие данные:

Количество сорных растений на 1 м ² (X)	12,0	18,1	25,2	30,0	32,5	40,0	45,4	50,6	51,0	62,6
Урожайность, т/га (Y)	6,6	6,4	6,3	6,2	6,2	6,0	5,8	5,2	4,7	4,0

Следует определить, существует ли зависимость между засоренностью посевов и урожайностью кукурузы.

1. Рассчитываем средние арифметические по ряду X (количество сорных растений на 1 м²) и ряду Y (урожайность кукурузы).

$$\bar{x} = \frac{\sum X}{n} = \frac{12,0 + 18,1 + 25,2 + 30,0 + 32,5 + 40,0 + 45,4 + 50,6 + 51,0 + 62,6}{10} = \frac{367,4}{10} = 36,7$$

$$\bar{y} = \frac{\sum Y}{n} = \frac{6,6 + 6,4 + 6,3 + 6,2 + 6,2 + 6,0 + 5,8 + 5,2 + 4,7 + 4,0}{10} = \frac{57,4}{10} = 5,7$$

2. Заполняем вспомогательную таблицу

Признаки		Отклонения		Квадраты отклонений		Произведения (Y - \bar{y}) · (X - \bar{x})
Y	X	(Y - \bar{y})	(X - \bar{x})	(Y - \bar{y}) ²	(X - \bar{x}) ²	
6,6	12,0	0,9	-24,7	0,81	610,1	-22,2
6,4	18,1	0,7	-18,6	0,49	346,0	-13,0
6,3	25,2	0,6	-11,5	0,36	132,2	-6,9
6,2	30,0	0,5	-6,7	0,25	44,9	-3,4
6,2	32,5	0,5	-4,2	0,25	17,6	- 2,1
6,0	40,0	0,3	3,3	0,09	10,9	1,0
5,8	45,4	0,1	8,7	0,01	75,7	0,9
5,2	50,6	-0,5	13,9	0,25	193,2	-7,0
4,7	51,0	-1,0	14,3	1,00	204,5	-14,3
4,0	62,6	-1,7	25,9	2,89	670,8	- 44,0
-	-	-	-	Σ=6,40	Σ=2305,9	Σ=-111

3. Рассчитываем коэффициенты:

корреляции

$$r = \frac{\sum(Y - \bar{y}) \cdot (X - \bar{x})}{\sqrt{\sum(Y - \bar{y})^2 \cdot \sum(X - \bar{x})^2}} = \frac{-111}{\sqrt{6,40 \cdot 2305,9}} = \frac{-111}{\sqrt{14757,8}} = \frac{-111}{121,5} = -0,91$$

Коэффициент корреляции имеет знак «-», следовательно, корреляционная связь *обратная*, или *отрицательная*; числовое значение коэффициента корреляции больше 0,7, следовательно, корреляционная связь *сильная*.

детерминации $d_{yx} = r^2 = -0,91^2 = 0,83$ или 83%.

Коэффициент детерминации равен 83%, следовательно, варьирование результативного признака Y на 83% определяется влиянием факториального признака X.

регрессии: для расчета уравнения регрессии необходимо рассчитать один или два коэффициента регрессии: b_{yx} и b_{xy} .

Коэффициент регрессии b_{yx} показывает в каком направлении и на какую величину изменяется признак Y при увеличении признака X на единицу измерения. В нашем примере коэффициент регрессии b_{yx} показывает на какую величину изменяется урожайность кукурузы при увеличении количества сорных растений на 1 штуку на 1 м².

Коэффициент регрессии b_{xy} показывает в каком направлении и на какую величину изменяется признак X при увеличении признака Y на единицу измерения. В нашем примере коэффициент регрессии b_{xy} показывает на какую величину изменяется количество сорных растений на 1 м² при увеличении урожайности кукурузы на 1 т/га.

В нашем примере нет смысла рассчитывать коэффициент регрессии b_{xy} .

$$b_{yx} = \frac{\sum(Y - \bar{y}) \cdot (X - \bar{x})}{\sum(X - \bar{x})^2} = \frac{-111}{2305,9} = -0,05$$

Таким образом, при увеличении количество сорных растений на 1 штуку на 1 м² урожайность кукурузы снижается на 0,05 ц/га.

4. Проводим оценку существенности коэффициента корреляции.

ошибка коэффициента корреляции

$$s_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-0,91^2}{10-2}} = \sqrt{\frac{1-0,83}{8}} = \sqrt{\frac{0,17}{8}} = \sqrt{0,021} = 0,14$$

критерий существенности $t_{\text{факт}} = \frac{r}{s_r} = \frac{-0,91}{0,14} = -6,5$

Знак критерия существенности не несет смысловой нагрузки, поэтому мы его не учитываем.

При $v = n - 2 = 10 - 2 = 8$ $t_{05} = 2,31$

Критерий существенности больше чем критерий Стьюдента, следовательно между изучаемыми признаками существует *значимая* или *существенная* корреляционная зависимость.

5. Выводим уравнение линейной регрессии

$$Y = \bar{y} + b_{yx} \cdot (X - \bar{x}) = 5,7 - 0,05(X - 36,7) = 5,7 - 0,05X + 1,8 = 7,5 - 0,05X$$

Таким образом, уравнение регрессии имеет следующий вид:

$$Y = 7,5 - 0,05X$$

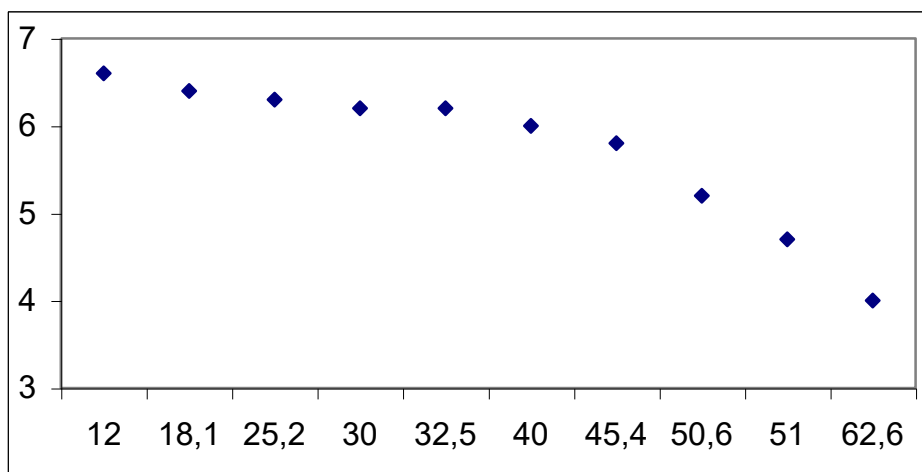
Подставляя в полученное уравнение экстремальные значения X (т.е. X_{min} и X_{max}), определяем соответствующие им значения Y.

В нашем примере X_{min}=12,0, подставляем его в уравнение регрессии Y=7,5-0,05X=7,5-0,05×12,0=7,5-0,6=6,9, т.е. Y=6,9;

X_{max}=62,6, подставляем его в уравнение регрессии Y=7,5-0,05X=7,5-0,05×62,6=7,5-3,1=4,4, т.е. Y=4,4

6. Строим теоретическую линию регрессии Y по X.

На корреляционном поле размещаем точечные фактические значения данных эксперимента. На основании полученных точек (X_{min}, Y и X_{max}, Y) строим теоретическую линию регрессии. Теоретическая линия регрессии не должна выходить за пределы расчетных точек, т.к. полученное уравнение верно лишь для изучаемого диапазона значений факториального признака X.



Вывод. Установлена обратная значимая корреляционная зависимость между изучаемыми признаками. С увеличением засоренности урожайность кукурузы снижается.

Задание. Согласно N примера и задания провести корреляционно-регрессионный анализ.

Пример 1. Число развитых колосков в колосе (X, шт) и число зерен в колосе (Y, шт)

1 задание		2 задание		3 задание		4 задание		5 задание	
Y	X	Y	X	Y	X	Y	X	Y	X
38	18	28	19	30	13	32	17	18	10
29	12	39	15	22	11	21	9	28	13
32	13	25	13	28	9	34	15	32	15
42	17	29	13	38	15	27	12	41	17
30	16	27	14	32	13	41	17	32	14
33	15	26	12	35	15	31	14	28	13
36	16	34	15	26	13	28	13	31	14
41	16	40	16	31	14	31	14	29	13
30	13	34	14	22	10	36	16	31	15
28	13	29	16	31	14	31	14	18	9

Пример 2. Зависимость урожайности озимой пшеницы (Y, ц/га) от пораженности бурой ржавчиной (X, %)

1 задание		2 задание		3 задание		4 задание		5 задание	
Y	X	Y	X	Y	X	Y	X	Y	X
61,1	20	49,3	36	50,3	22	49,7	21	55,1	19
50,3	27	52,5	22	50,4	18	30,2	64	53,0	25
51,1	21	55,3	8	43,3	54	32,5	55	50,2	41
48,4	46	43,8	54	46,5	45	48,1	26	51,1	35
49,0	31	48,1	43	55,3	10	41,0	49	50,7	42
49,0	33	39,6	64	50,3	20	53,0	15	51,5	34
48,5	37	41,3	57	40,2	60	51,5	24	45,8	54
49,0	32	55,3	12	55,2	10	45,2	34	50,5	43
41,6	60	48,0	43	43,8	54	42,0	34	58,8	12
45,0	54	40,6	56	48,5	36	41,6	49	40,2	60

Пример 3. Зависимость прибавки урожая картофеля при внесении полного минерального удобрения (Y, ц/га) от количества осадков за май-июль (X, мм)

1 задание		2 задание		3 задание		4 задание		5 задание	
Y	X	Y	X	Y	X	Y	X	Y	X
150	235	46	113	45	121	203	245	30	60
136	204	125	212	84	142	65	153	33	74
42	120	77	145	83	138	74	138	47	94
216	238	201	247	143	221	34	84	42	99
37	96	112	176	136	198	90	150	88	150
95	145	112	188	100	168	80	137	60	143
82	140	37	88	95	140	33	74	144	220
48	119	42	95	103	178	46	112	45	108
42	115	39	83	31	78	103	165	27	64
96	156	54	113	36	85	136	198	42	100

Выполнение работы.

1. Записать название примера, номер задания.

2. Заполнить вспомогательную таблицу

Признаки		Отклонения		Квадраты отклонений		Произведения $(Y - \bar{y}) \cdot (X - \bar{x})$
Y	X	$(Y - \bar{y})$	$(X - \bar{x})$	$(Y - \bar{y})^2$	$(X - \bar{x})^2$	
Σ	Σ	-	-	Σ	Σ	Σ

3. Рассчитать средние арифметические по ряду X и Y.

n = _____ n – количество парных значений

$$\bar{x} = \sum X \div n =$$

$$\bar{y} = \sum Y \div n =$$

4. Рассчитать коэффициенты:

корреляции $r = \frac{\sum (Y - \bar{y}) \cdot (X - \bar{x})}{\sqrt{\sum (Y - \bar{y})^2 \cdot \sum (X - \bar{x})^2}} =$

регрессии $b_{yx} = \frac{\sum (Y - \bar{y}) \cdot (X - \bar{x})}{\sum (X - \bar{x})^2} =$

детерминации $d_{yx} = r^2 =$

5. Провести оценку существенности коэффициента корреляции.

ошибка коэффициента корреляции $S_r = \sqrt{\frac{1-r^2}{n-2}} =$

критерий существенности $t_{\text{факт}} = \frac{r}{S_r} =$

При $v = n - 2 =$ $t_{05} =$

6. Вывести уравнение линейной регрессии

$$Y = \bar{y} + b_{yx} \cdot (X - \bar{x}) =$$

Подставляя в полученное уравнение экстремальные значения X, определяем значения Y и строим теоретическую линию регрессии Y по X.



Вывод:

Контрольные вопросы к работе:

1. Какие связи называются корреляционными?
2. Виды корреляции.
3. Какая корреляция называется линейной?
4. Коэффициенты, описывающие линейную корреляцию.
5. Оценка значимости корреляционной зависимости.