

ЛИНЕЙНАЯ РЕГРЕССИЯ: РЕШЕНИЕ ЗАДАЧ

ЗАДАНИЕ.

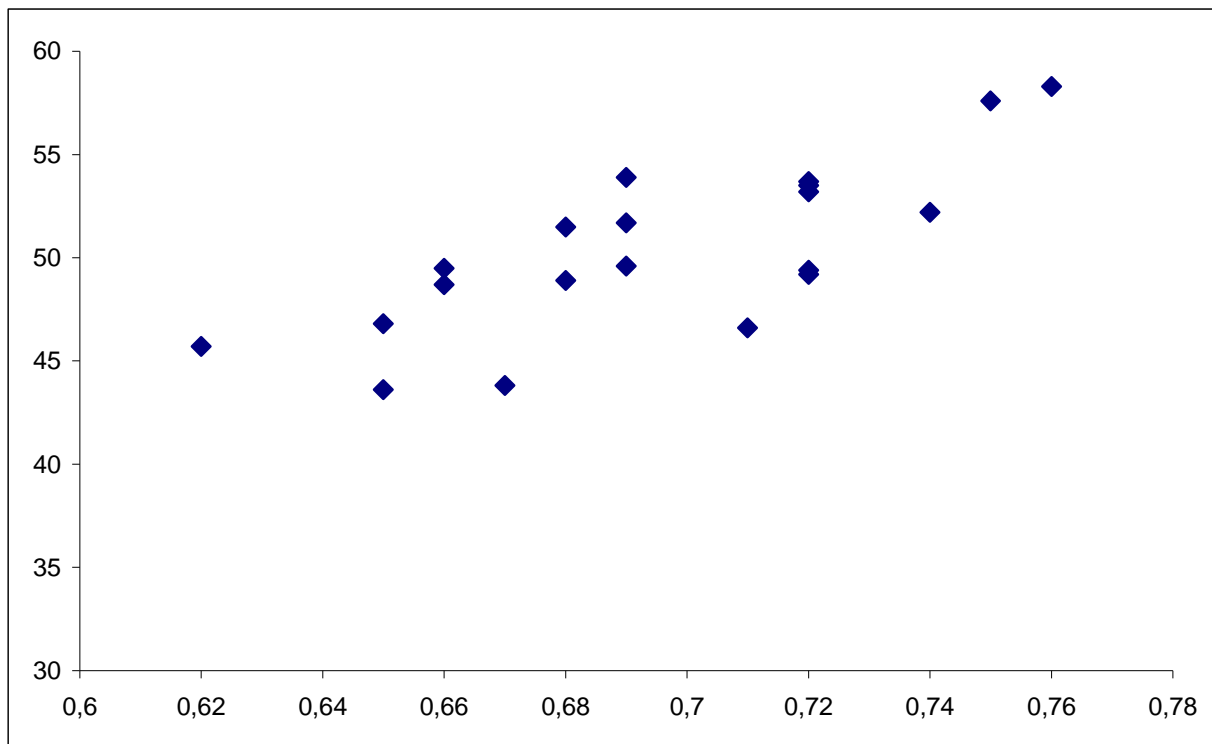
Вычислить коэффициент уравнения регрессии. Определить выборочный коэффициент корреляции между плотностью древесины маньчжурского ясеня и его прочностью.

Решая задачу необходимо построить поле корреляции, по виду поля определить вид зависимости, написать общий вид уравнения регрессии Y на X , определить коэффициенты уравнения регрессии и вычислить коэффициенты корреляции между двумя заданными величинами.

Строка	Плотн., кг/м ³	Проч., МПа
1	0,69	53,9
2	0,68	48,9
3	0,65	46,8
4	0,74	52,2
5	0,72	53,5
6	0,66	48,7
7	0,72	49,2
8	0,72	53,7
9	0,72	53,2
10	0,75	57,6
11	0,76	58,3
12	0,62	45,7
13	0,69	49,6
14	0,72	49,4
15	0,66	49,5
16	0,69	51,7
17	0,65	43,6
18	0,68	51,5
19	0,71	46,6
20	0,67	43,8

РЕШЕНИЕ.

Построим поле корреляции по данным таблицы, обозначив плотность за X , прочность за Y .



Можно сделать вывод, что наблюдается прямая линейная зависимость. Найдем уравнение линейной регрессии Y на X .

Вычислим характеристики X и Y (выборочная средняя и выборочное среднее квадратическое отклонение). Вычисления проведем в таблице:

	x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$x_i y_i$
	0,69	53,9	2,5E-05	12,4609	37,191
	0,68	48,9	0,000225	2,1609	33,252
	0,65	46,8	0,002025	12,7449	30,42
	0,74	52,2	0,002025	3,3489	38,628
	0,72	53,5	0,000625	9,7969	38,52
	0,66	48,7	0,001225	2,7889	32,142
	0,72	49,2	0,000625	1,3689	35,424
	0,72	53,7	0,000625	11,0889	38,664
	0,72	53,2	0,000625	8,0089	38,304
	0,75	57,6	0,003025	52,2729	43,2
	0,76	58,3	0,004225	62,8849	44,308
	0,62	45,7	0,005625	21,8089	28,334
	0,69	49,6	2,5E-05	0,5929	34,224
	0,72	49,4	0,000625	0,9409	35,568
	0,66	49,5	0,001225	0,7569	32,67
	0,69	51,7	2,5E-05	1,7689	35,673
	0,65	43,6	0,002025	45,8329	28,34

	0,68	51,5	0,000225	1,2769	35,02
	0,71	46,6	0,000225	14,2129	33,086
	0,67	43,8	0,000625	43,1649	29,346
Сумма	13,9	1007,4	0,0259	309,282	702,314
Среднее	0,695	50,37	0,001295	15,4641	35,1157

Выборочная средняя $\bar{x} = \frac{1}{n} \sum x_i = \frac{1}{20} 13,9 = 0,695$

Выборочное среднее квадратическое отклонение $\sigma_x = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2} = \sqrt{\frac{1}{20} 0,0259} = 0,036$

Выборочная средняя $\bar{y} = \frac{1}{n} \sum y_i = \frac{1}{20} 1007,4 = 50,37$

Выборочное среднее квадратическое отклонение

$$\sigma_y = \sqrt{\frac{1}{n} \sum (y_i - \bar{y})^2} = \sqrt{\frac{1}{20} 309,282} = 3,932$$

Найдем сумму $\sum x_i y_i = 702,314$.

Вычислим коэффициент линейной корреляции по формуле $r_g = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n \sigma_x \sigma_y}$.

Получим: $r_g = \frac{702,314 - 20 \cdot 0,695 \cdot 50,37}{20 \cdot 0,036 \cdot 3,932} \approx 0,767$.

Направление связи прямое (так как коэффициент положительный), связь тесная.

Уравнение регрессии Y на X имеет вид $\bar{y}_x - \bar{y} = r_g \frac{\sigma_y}{\sigma_x} (x - \bar{x})$. Подставляем все величины:

$$\bar{y}_x - 50,37 = 0,767 \frac{3,932}{0,036} (x - 0,695),$$

$$\bar{y}_x = 83,822x - 7,887.$$

Построим на графике линию регрессии:

