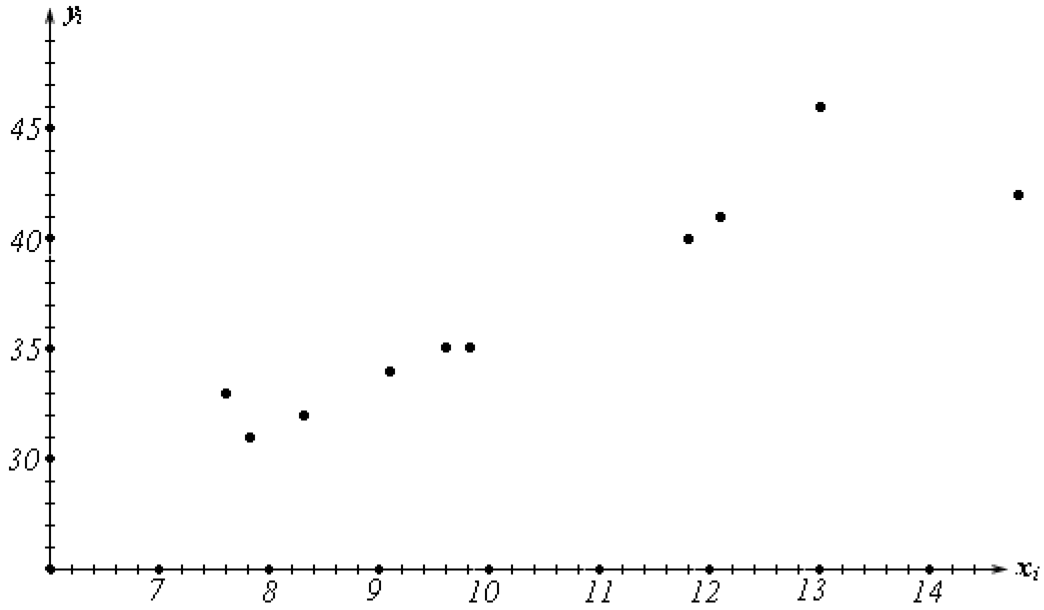
КОРРЕЛЯЦИОННАЯ ЗАВИСИМОСТЬ

№1

Построить корреляционное поле точек и вычислить коэффициент корреляции между ростом *(X)* и массой (*У*) некоторых животных. Исходные данные приведены в выборке объема n = 10.

x_i	31	32	33	34	35	35	40	41	42	46
y_i	7,8	8,1	7,6	9,1	9,6	9,8	11,8	17,1	14,7	13,0

Решение:



Коэффициент корреляции:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{31+32+33+34+35+35+40+41+42+46}{10} = 36{,}9;$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{7{,}8+8{,}3+7{,}6+9{,}1+9{,}6+9{,}8+11{,}8+12{,}1+14{,}7+13{,}0}{10} = 10{,}38;$$

$$\sum_{i=1}^{n}\left(x-\bar{x}\right)\cdot\left(y-\bar{y}\right) = (31-39{,}6)\cdot(7{,}8-10{,}38)+\ldots+(46-39{,}6)\cdot(13-10{,}38) = 99{,}9;$$

$$\sum_{i=1}^{n}\left(x-\bar{x}\right)^2 = (31-39{,}6)^2+(32-39{,}6)^2\ldots+(46-39{,}6)^2 = 224{,}8;$$

$$\sum_{i=1}^{n}\left(y-\bar{y}\right)^2 = (7{,}8-10{,}38)^2 + (8{,}3-10{,}38)^2 \ldots + (13-10{,}38)^2 = 51{,}9$$

$$r = \frac{99{,}9}{\sqrt{224{,}8\cdot 51{,}9}} = 0{,}925$$

Величина r близка к 1, это говорит о тесной связи роста и массы.

№2

Измерена концентрация ($c=Y_i$) алкоголя в крови у n=5 добровольцев с одинаковым весом после нескольких порций алкоголя (X_i). Методом наименьших квадратов определите коэффициенты a и b сглаживающей прямой $\bar{y}_x = ax+b$. Постройте график.

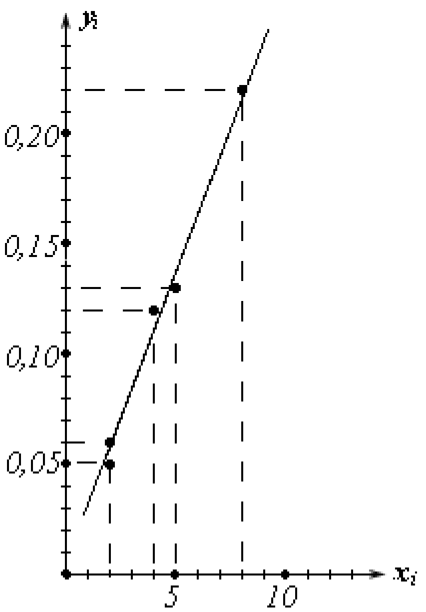
Число порций, X_i	2	2	4	5	8
Концентрация, Y_i	0,05	0,06	0,11	0,13	0,22

Решение:

$$a = \frac{n\sum_{i=1}^{n}x_i y_i - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2 - (\sum_{i=1}^{n}x_i)^2}, \qquad b = \frac{\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}y_i - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_i y_i}{n\sum_{i=1}^{n}x_i^2 - (\sum_{i=1}^{n}x_i)^2}$$

Найдём предварительно $\sum_{i=1}^{n}x_i$, $\sum_{i=1}^{n}y_i$, $\sum_{i=1}^{n}x_i^2$, $\sum_{i=1}^{n}x_i y_i$:

x_i	y_i	x_i^2	$x_i\cdot y_i$
2	0,05	4	0,10
2	0,06	4	0,12
4	0,11	16	0,44
5	0,13	25	0,65
8	0,22	64	1,76
$\sum_{i=1}^{5}x_i$ =21	$\sum_{i=1}^{5}y_i$ =0,57	$\sum_{i=1}^{5}x_i^2$ =113	$\sum_{i=1}^{n}x_i y_i$ =3,07



$$a = \frac{5\cdot 3{,}07 - 21\cdot 0{,}57}{5\cdot 113 - 21^2} = 0{,}027;$$

$$b = \frac{\quad}{5\cdot 113 - 21^2} = -0{,}00048.$$

Искомое уравнение сглаживающей прямой $\bar{y}_x = ax+b$: $\bar{y}_x = 0{,}027x - 0{,}00048$.
График искомой сглаживающей прямой приведён на рисунке:

№3

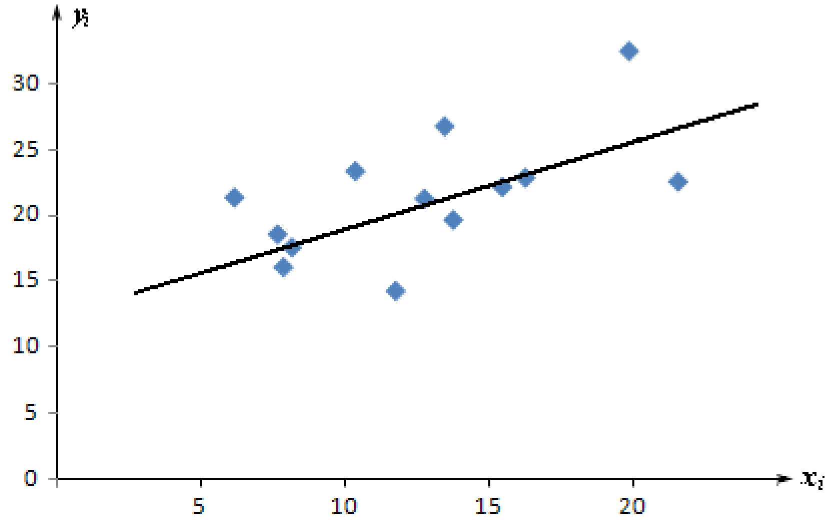
В эксперименте на 13 кошках получены данные об интрасклеральном (x) и внутриглазном давлении (y):

x_i	19,8	7,8	12,7	13,4	10,3	13,7	16,2	15,4	21,5	8,1	11,7	7,6	6,1
y_i	32,5	16,1	21,3	26,8	23,4	19,7	22,9	22,2	22,6	17,6	14,3	18,6	21,4

1. Установите, имеется ли корреляционная связь между x и y; определите коэффициент корреляции r:
2. Определите тесноту корреляционной связи.
3. Составьте уравнение регрессии и найдите ожидаемое значение для y при x=18.

Решение:

Построим график, отложив вдоль оси абсцисс X величину интрасклерального давления x , а вдоль оси ординат Y – величину внутриглазного давления y. Тогда каждой паре значений x и y на графике будет соответствовать определённая точка. По характеру расположения точек можно предположить существование линейной корреляционной связи между x и y.



Вычислим коэффициент линейной корреляции r по формуле

$$r_s = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2}\sqrt{\overline{y^2} - (\bar{y})^2}}$$

$$\bar{x} = \frac{1}{13}(19{,}8 + 7{,}8 + \ldots + 6{,}1) = 12{,}64;$$

$$\bar{y} = \frac{1}{13}(32{,}5 + 16{,}1 + \ldots + 21{,}4) = 21{,}49$$

$$\overline{x^2} = \frac{1}{13}(19{,}8^2 + 7{,}8^2 + \ldots + 6{,}1^2) = 180{,}5;$$

$$\overline{y^2} = \frac{1}{13}(32{,}5^2 + 16{,}1^2 + \ldots + 21{,}4^2) = 482{,}2$$

$$\overline{xy} = \frac{1}{13}(19{,}8 \cdot 32{,}5 + 7{,}8 \cdot 16{,}1 + \ldots + 6{,}1 \cdot 21{,}4) = 283{,}9;$$

$$s_x = \sqrt{180{,}5 - 12{,}64^2} = 4{,}55; \qquad s_y = \sqrt{482{,}2 - 21{,}49^2} = 4{,}51;$$

$$r = \frac{283{,}9 - 12{,}64 \cdot 21{,}49}{4{,}55 \cdot 4{,}51} = 0{,}595$$

Связь *x* и *y* умеренная, положительная.

По формуле $\rho_{yx} = r\frac{s_y}{s_x}$ находим коэффициент регрессии:

$$\rho_{yx} = 0{,}595\,\frac{4{,}51}{4{,}55} = 0{,}589.$$

Далее, подставляя ρ_{yx} в формулу $y - \bar{y} = \rho_{yx}(x - \bar{x})$ и вычисляя b, находим уравнение регрессии: $y = 0{,}589x + 14$

И, наконец, вычисляем ожидаемое значение *y* при *x=18*:

$$y(18) = 0{,}589 \cdot 18 + 14 = 24{,}6$$

№4

В результате регистрации некоторых объектов определенного вида по заданным значениям признаков *x* и *y* получены числа (частоты) совпадений заданных значений этих признаков, помещенные в следующей таблице. По данным этой таблицы:

1) определить условные средние значения величин *x* и *y*, с их помощью получить изображение корреляционного поля и по характеру расположения точек на нем сделать вывод о типе линии регрессионной зависимости между величинами *x* и *y*;

2) найти коэффициенты регрессии *y* на *x* и *x* на *y* по методу наименьших квадратов;

3) составить уравнения прямых регрессии *y* на *x* и *x* на *y*;

4) вычислить коэффициент корреляции этих величин;

5) построить систему координат и в ней прямые регрессий.

y \ x	111	113	115	117	n_y
11	1	1			2
12	1	2	2	1	6
13			1	1	2
n_x	2	3	3	2	

На основании данных, приведенных в таблице, найдем условные средние $\overline{y_{xi}}$ величины y для всех значений x по формуле: $\bar{y}_{xi} = \sum_{j=1}^{t} \frac{n_{ij} y_j}{n_{xi}}$,

$$\bar{y}_{x=111} = \frac{1\cdot 11 + 1\cdot 12}{2} = 11{,}5; \qquad \bar{y}_{x=113} = \frac{1\cdot 11 + 2\cdot 12}{3} = 11{,}7;$$

$$\bar{y}_{x=115} = \frac{2\cdot 12 + 1\cdot 13}{3} = 12{,}3; \qquad \bar{y}_{x=117} = \frac{1\cdot 12 + 1\cdot 13}{2} = 12{,}5.$$

По полученным результатам составим таблицу:

x	111	113	115	117
$\overline{y_x}$	11,5	11,7	12,3	12,5

На основании данных, приведенных в таблице, найдем условные средние $\overline{x_{yj}}$ величины x для всех значений y по формуле: $\bar{x}_{yj} = \sum_{i=1}^{t} \frac{n_{ij} x_i}{n_{yj}}$,

$$\bar{x}_{y=11} = \frac{1\cdot 111 + 1\cdot 113}{2} = 112;$$

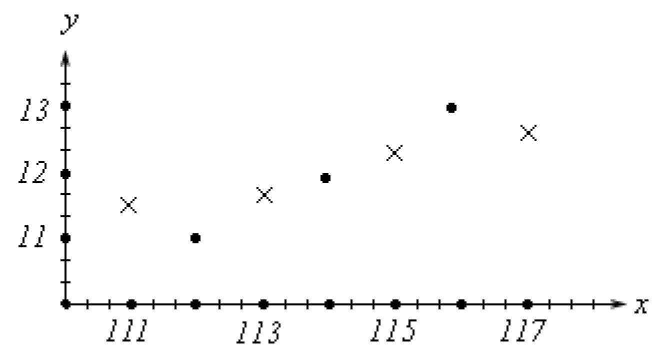
$$\bar{x}_{y=12} = \frac{1\cdot 111 + 2\cdot 113 + 2\cdot 115 + 1\cdot 117}{6} = 114;$$

$$\bar{x}_{y=13} = \frac{1\cdot 115 + 1\cdot 117}{2} = 116.$$

По полученным результатам составим таблицу:

y	11	12	13
$\overline{x_y}$	112	114	116

Данные таблиц отражены на рисунке - корреляционном поле (крестики соответствуют данным первой таблицы, точки – данным второй таблицы).



Как видно из рисунка, характер расположения построенных точек указывает на приблизительную линейную зависимость $\overline{y_x}$ от x и $\overline{x_y}$ от y. Поэтому уравнения регрессии следует искать в виде:

$$\overline{y_x} = \rho_{yx} x + b, \quad (*)$$

$$\overline{x_y} = \rho_{xy} y + d, \quad (**)$$

где ρ_{yx}, ρ_{xy} – выборочные коэффициенты регрессии, составим вспомогательную расчетную таблицу для их нахождения. Так как значения вариант признаков X и Y достаточно велики, то введем условные варианты u и v следующим образом:

$C_x = 113$, $C_y = 12$ – варианты x и y, на которые приходятся условные варианты $u=v=0$;

$h_x = 2$, $h_y = 1$ – длина интервалов вариации значений признаков x и y.

	u	-1	0	1	2			
v	$y \backslash x$	111	113	115	117	n_{yj}	$n_{yj}v_j$	$n_{yj}v_j^2$
-1	11	1	1			2	-2	2
0	12	1	2	2	1	6	0	0
1	13			1	1	2	2	2
	n_{xi}	2	3	3	2	10	0	4
	$n_{xi}u_i$	-2	0	3	4	5		
	$n_{xi}u_i^2$	2	0	3	8	13		
	$\sum n_{ij}v_j$	-1	-1	1	1	0		
	$u_i\sum n_{ij}v_j$	1	0	1	2	4		

По результатам вычислений, сведенным в таблицу, находим:

$$\bar{u} = \frac{\sum n_{xi}u_i}{n} = \frac{5}{10} = 0{,}5,$$

$$\bar{v} = \frac{\sum n_{yj}v_j}{n} = \frac{0}{10} = 0.$$

Тогда выборочные средние признаков x и y:

$$\bar{x} = C_x + \bar{u} \cdot h_x = 113 + 0{,}5 \cdot 2 = 114,$$

$$\bar{y} = C_y + \bar{v} \cdot h_y = 12 + 0 \cdot 1 = 12$$

Для вычисления дисперсий признаков x и y находим дисперсии условных вариант u и v по формулам:

$$S_u^2 = \overline{u^2} - (\bar{u})^2 = \frac{\sum n_{xi} u_i^{\,2}}{n} - \left(\frac{\sum n_{xi} u_i}{n} \right)^2 = \frac{13}{10} - (0{,}5)^2 = 1{,}3 - 0{,}25 = 1{,}05,$$

$$S_v^2 = \overline{v^2} - (\bar{v})^2 = \frac{\sum n_{yj} v_j^{\,2}}{n} - \left(\frac{\sum n_{yj} v_j}{n} \right)^2 = \frac{7}{10} - 0^2 = 0{,}7 - 0 = 0{,}7.$$

Тогда

$$S_u = 1{,}02, \qquad S_v = 0{,}84.$$

$$S_x = S_u \cdot h_x = 1{,}02 \cdot 2 = 2{,}04, \qquad S_y = S_v \cdot h_y = 0{,}84 \cdot 1 = 0{,}84.$$

Коэффициент корреляции признаков x и y совпадает с коэффициентом корреляции условных вариант и вычисляется по формуле:

$$r = \frac{\overline{uv} - \bar{u} \cdot \bar{v}}{S_u \cdot S_v} = \frac{1}{S_u \cdot S_v} \left(\frac{\sum n_{ij} u_i v_j}{n} - \frac{1}{n} \sum n_{xi} u_i \cdot \frac{1}{n} \sum n_{yj} v_j \right) = \frac{1}{1{,}02 \cdot 0{,}84} \left(\frac{4}{10} - 0{,}5 \cdot 0 \right) = \frac{0{,}4}{0{,}86} \approx 0{,}46$$

Следовательно, коэффициенты регрессии:

$$\rho_{yx} = r \frac{S_y}{S_x} = 0{,}46 \cdot \frac{0{,}84}{2{,}04} \approx 0{,}19,$$

$$\rho_{xy} = r \frac{S_x}{S_y} = 0{,}46 \cdot \frac{2{,}04}{0{,}84} \approx 1{,}12;$$

коэффициенты b и d в уравнениях (*) и (**):

$$b = \bar{y} - \rho_{yx} \bar{x} = 12 - 0{,}19 \cdot 114 = -9{,}66,$$

$$d = \bar{x} - \rho_{xy} \bar{y} = 114 - 1{,}12 \cdot 12 = 100{,}56.$$

Таким образом, уравнение прямой регрессии y на x запишется в виде:

$$\overline{y_x} = 0{,}19x - 9{,}66.$$

Прямая регрессии x на y имеет вид:

$$\overline{x_y} = 1{,}12y + 100{,}56.$$

Построим прямые регрессии внутри корреляционного поля: