Задача 6. В таблице приведено распределение 120 коров по дневному надою Y (в кг) и по жирности X (в %):

$x \setminus y$	7	9	11	13	15	Итого
3,3				8		8
3,5			2	16	8	26
3,7		4	16	10	2	32
3,9	2	6	10	2		20
4,1	8	6	20			34
Итого	10	16	48	36	10	120

Необходимо:

1) Вычислить групповые средние $\overline{x}_i$ и $\overline{y}_j$ и построить эмпирические линии регрессии;

2) Предполагая, что между переменными X и Y существует линейная корреляционная зависимость:

а) найти уравнения прямых регрессии, построить их графики на одном чертеже с эмпирическими линиями регрессии и дать экономическую интерпретацию полученных уравнений;

б) вычислить коэффициент корреляции, на уровне значимости α = 0,05 оценить его значимость и сделать вывод о тесноте и направлении связи между переменными X и Y;

в) используя соответствующее уравнение регрессии, оценить средний процент жирности молока для коров, дневной удой которых составляет 12 кг.

Решение.

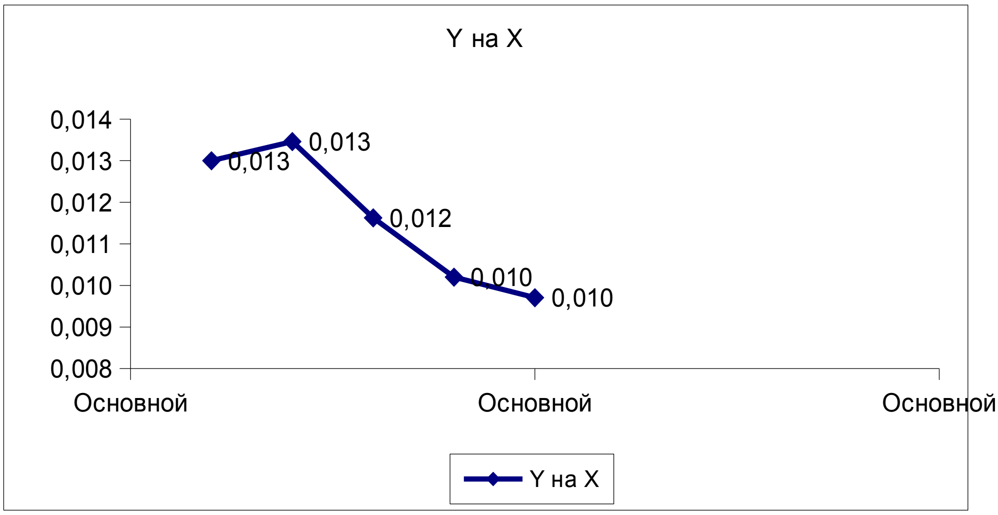
1) Найдем групповые средние по формулам: $\overline{x_j} = \dfrac{\sum\limits_{i=1}^{5} x_i n_{ij}}{n_j}$; $\overline{y_i} = \dfrac{\sum\limits_{j=1}^{5} y_j n_{ij}}{n_i}$.

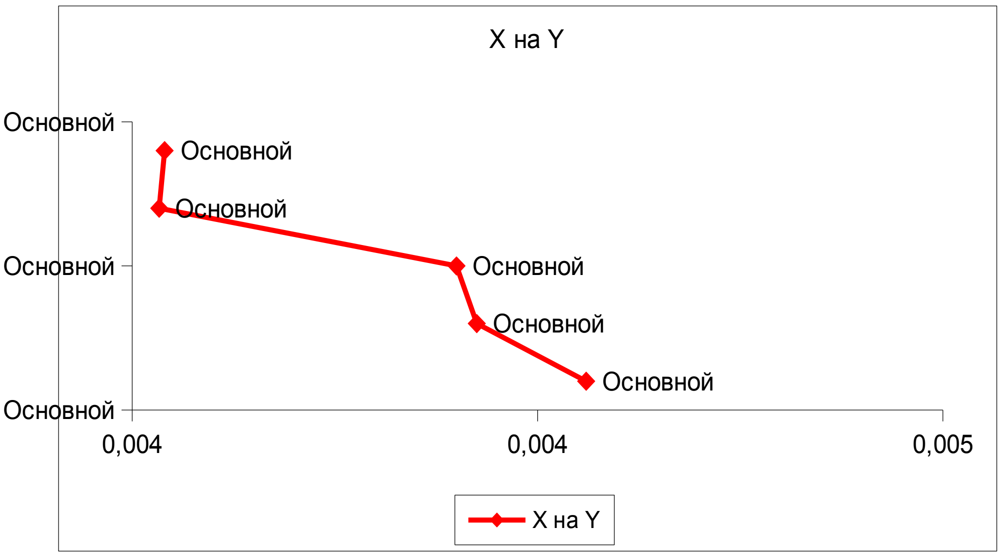
Вычисления проведем в Excel, получаем:

$\overline{x_j}$	4,060	3,925	3,900	3,533	3,540
y_j	7	9	11	13	15

x_i	$\overline{y_i}$
3,3	13,000
3,5	13,462
3,7	11,625
3,9	10,200
4,1	9,706

Построим эмпирические линии регрессии (Y на X, X на Y).





Из вида эмпирических линий регрессии можно заключить, что между переменными наблюдается линейная зависимость.

Найдем уравнения прямых линий регрессии. Вычислим необходимые величины (расчеты в таблицах ниже):

x_i	3,3	3,5	3,7	3,9	4,1	Сумма
n_i	8	26	32	20	34	120
$x_i \cdot n_i$	26,4	91	118,4	78	139,4	453,2
$x_i^2 \cdot n_i$	87,12	318,5	438,08	304,2	571,54	1719,44

y_j	7	9	11	13	15	Сумма
n_j	10	16	48	36	10	120
$y_j \cdot n_j$	70	144	528	468	150	1360
$y_j^2 \cdot n_j$	490	1296	5808	6084	2250	15928

$$\sum_{i=1}^{5} x_i n_i = 453,2, \quad \sum_{i=1}^{5} x_i^2 n_i = 1719,44,$$

$$\overline{x} = \frac{\sum_{i=1}^{5} x_i n_i}{n} = \frac{453,2}{120} = 3,777,$$

$$s_x^2 = \frac{\sum_{i=1}^{5} x_i^2 n_i}{n} - \overline{x}^2 = \frac{1719,44}{120} - 3,777^2 = 0,065$$

$$\sum_{i=1}^{5} y_j n_j = 1360, \quad \sum_{j=1}^{5} y_j^2 n_j = 15928,$$

$$\overline{y} = \frac{\sum_{i=1}^{5} y_j n_j}{n} = \frac{1360}{120} = 11,333,$$

$$s_y^2 = \frac{\sum_{j=1}^{5} y_j^2 n_j}{n} - \overline{y}^2 = \frac{15928}{120} - 11,333^2 = 4,289$$.

$$\sum_{i=1}^{5}\sum_{j=1}^{5} x_i y_j n_{ij} = 5093,2$$

$$\mu = \overline{xy} - \overline{x} \cdot \overline{y} = \frac{\sum_{i=1}^{5}\sum_{j=1}^{5} x_i y_j n_{ij}}{n} - \overline{x} \cdot \overline{y} = \frac{5093,2}{120} - 3,777 \cdot 11,333 = -0,359$$

$$b_{yx} = \frac{\mu}{s_x^2} = \frac{-0,359}{0,065} = -5,483$$

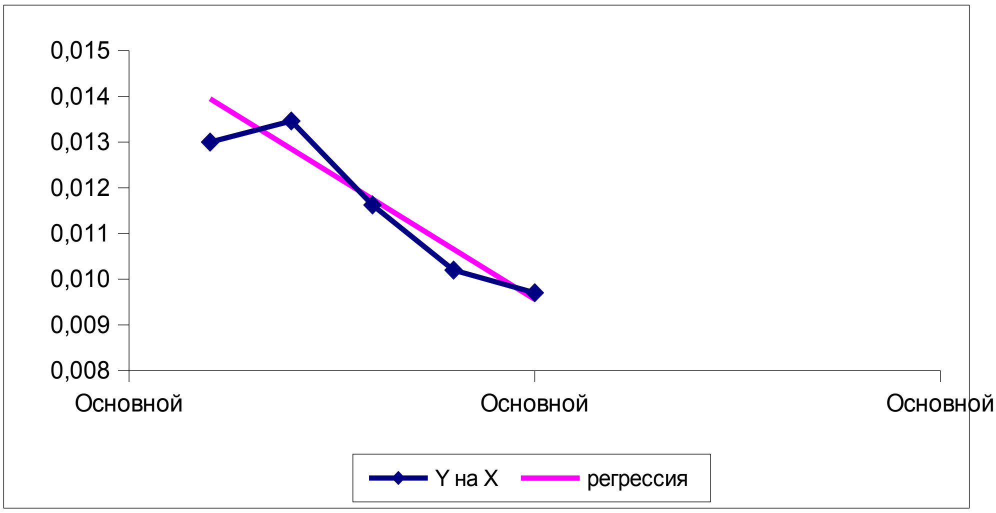
$$b_{xy} = \frac{\mu}{s_y^2} = \frac{-0,359}{4,289} = -0,084$$

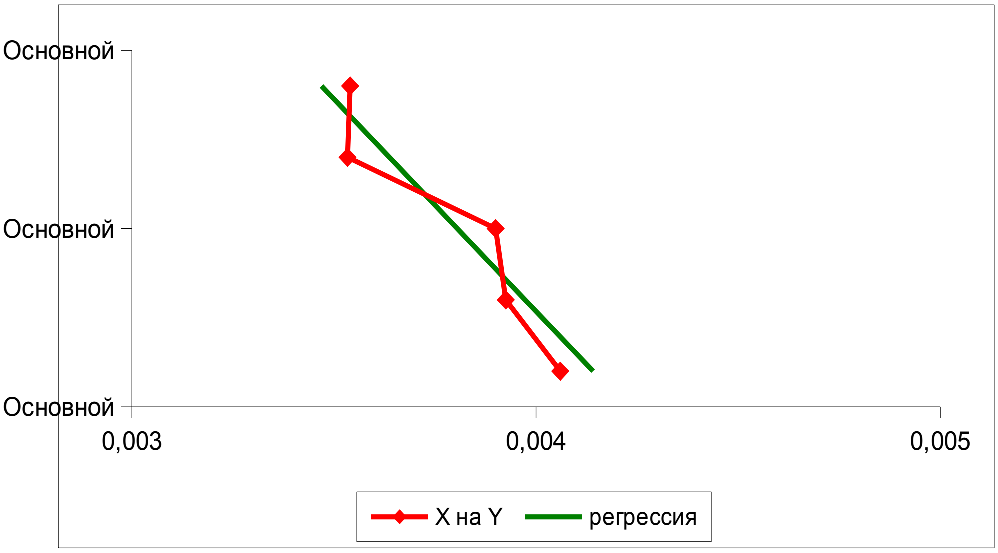
Уравнения прямых регрессии:

$$y_x - \overline{y} = b_{yx}(x - \overline{x}),$$
$$y_x - 11,333 = -5,483(x - 3,777),$$
$$y_x = -5,483x + 32,042.$$

$$x_y - \overline{x} = b_{xy}(y - \overline{y}),$$
$$x_y - 3,777 = -0,084(y - 11,333),$$
$$x_y = -0,084y + 4,729.$$

Построим графики линий регрессии на одном чертеже с эмпирическими линиями регрессии.





Экономическая интерпретация полученных уравнений:

$y_x = -5,483x + 32,042$ - при увеличении жирности молока на 1%, дневной надой уменьшается в среднем на 5,483 кг.

$x_y = -0,084y + 4,729$ - при увеличении дневного надоя на 1 кг, жирность молока уменьшается в среднем на 0,084 %.

Вычислим коэффициент корреляции $r = \sqrt{b_{yx} b_{xy}} = \sqrt{5,483 \cdot 0,084} \approx 0,679$

На уровне значимости $\alpha = 0,05$ оценим значимость коэффициента корреляции. Вычислим значение критерия $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,679\sqrt{120-2}}{\sqrt{1-0,679^2}} \approx 10,047$

По таблице критерия Стьюдента для уровня значимости 0,05 находим $t_{0,95;118} = 1,98$. Так как наблюдаемое значение 10,047 больше критического, коэффициент корреляции значим.

Связь между переменными X и У тесная, обратная.

Используя соответствующее уравнение регрессии, оценим средний процент жирности молока для коров, дневной удой которых составляет 12 кг:

$x_y(12) = -0,084 \cdot 12 + 4,729 \approx 3,721\%$.