

**Задача 1.** По территориям региона приводятся данные за 20XX г. (см. таблицу своего варианта).

**Требуется:**

1. Построить линейное уравнение парной регрессии  $y$  от  $x$ .
2. Рассчитать линейный коэффициент парной корреляции и среднюю ошибку аппроксимации.
3. Оценить статистическую значимость параметров регрессии и корреляции с помощью  $F$ -критерия Фишера и  $t$ -критерия Стьюдента.
4. Выполнить прогноз заработной платы  $y$  при прогнозном значении среднедушевого прожиточного минимума  $x$ , составляющем 107% от среднего уровня.
5. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.
6. На одном графике построить исходные данные и теоретическую прямую.

### Вариант 9

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., $x$	Среднедневная заработная плата, руб., $y$
1	78	133
2	94	139
3	85	141
4	73	127
5	91	154
6	88	142
7	73	122
8	82	135
9	99	142
10	113	168
11	69	124
12	83	130

## Решение

1. Для расчета параметров уравнения линейной регрессии строим расчетную таблицу.

Таблица 1

	$x$	$y$	$yx$	$x^2$	$y^2$	$\hat{y}_x$	$(y - \hat{y}_x)^2$	$A_i$
1	78	133	10374	6084	17689	130,8045	4,820255	1,650758
2	94	139	13066	8836	19321	145,9951	48,93166	5,032458
3	85	141	11985	7225	19881	137,4504	12,59973	2,517453
4	73	127	9271	5329	16129	126,0574	0,888454	0,742187
5	91	154	14014	8281	23716	143,1469	117,7903	7,047484
6	88	142	12496	7744	20164	140,2986	2,89465	1,198146
7	73	122	8906	5329	14884	126,0574	16,46267	3,325756
8	82	135	11070	6724	18225	134,6021	0,158286	0,294705
9	99	142	14058	9801	20164	150,7422	76,42584	6,15647
10	113	168	18984	12769	28224	164,034	15,72928	2,360724
11	69	124	8556	4761	15376	122,2598	3,028416	1,403415
12	83	130	10790	6889	16900	135,5516	30,81985	4,270433
Итого	1028	1657	143570	89772	230673	1657	330,5494	35,99999
Среднее значение	85,66667	138,0833	11964,17	7481	19222,75	138,0833	—	—
$\sigma$	11,9257	12,47971	—	—	—	—	—	—
$\sigma^2$	142,2222	155,7431	—	—	—	—	—	—

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{11964,17 - 85,67 \cdot 138,08}{7481 - 85,67^2} = 0,949$$

$$a = \bar{y} - b \cdot \bar{x} = 138,08 - 0,949 \cdot 85,67 = 56,75$$

Получено уравнение регрессии:  $y = 56,75 + 0,95 \cdot x$

С увеличением среднедушевого прожиточного минимума на 1 руб. среднедневная заработная плата возрастает в среднем на 0,93 руб.

2. Тесноту линейной связи оценит коэффициент корреляции:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = 0,95 \cdot \frac{11,93}{12,48} = 0,907$$

$$r_{xy}^2 = 0,823$$

Это означает, что 82% вариации заработной платы ( $Y$ ) объясняется вариацией фактора  $X$  – среднедушевого прожиточного минимума.

Качество модели определяет средняя ошибка аппроксимации:

$$\bar{A} = \frac{1}{n} \sum A_i = \frac{35,9999}{12} = 2,99\%$$

Качество построенной модели оценивается как хорошее, так как  $\bar{A}$  не превышает 8-10%.

3. Оценку значимости уравнения регрессии в целом проведем с помощью  $F$ -критерия Фишера. Фактическое значение  $F$ -критерия:

$$F_{\text{факт}} = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2) = \frac{0,823}{1 - 0,823} \cdot 10 = 46,54$$

Табличное значение критерия при пятипроцентном уровне значимости и степенях свободы  $k_1 = 1$  и  $k_2 = 12 - 2 = 10$  составляет  $F_{\text{табл}} = 4,96$ . Так как  $F_{\text{факт}} = 46,54 > F_{\text{табл}} = 4,96$ , то уравнение регрессии признается статистически значимым.

4. Оценку статистической значимости параметров регрессии проведем с помощью  $t$ -статистики Стьюдента и путем расчета доверительного интервала каждого из показателей.

Табличное значение  $t$ -критерия для числа степеней свободы  $df = n - 2 = 12 - 2 = 10$  и  $\alpha = 0,05$  составит  $t_{\text{табл}} = 2,23$ .

Определим случайные ошибки  $m_a$ ,  $m_b$ ,  $m_{r_{xy}}$ :

$$m_a = S_{\text{ост}} \cdot \frac{\sqrt{\sum x^2}}{n \cdot \sigma_x} = 5,75 \cdot \frac{\sqrt{89772}}{12 \cdot 11,93} = 12,037$$

$$m_b = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}} = \frac{5,75}{11,93 \cdot \sqrt{12}} = 0,139;$$

$$m_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n - 2}} = \sqrt{\frac{1 - 0,823}{12 - 2}} = 0,133$$

Тогда

$$t_a = \frac{a}{m_a} = \frac{56,75}{12,037} = 4,71;$$

$$t_b = \frac{b}{m_b} = \frac{0,95}{0,139} = 6,82;$$

$$t_{r_{xy}} = \frac{r_{xy}}{m_{r_{xy}}} = \frac{0,907}{0,133} = 6,82$$

Фактические значения  $t$ -статистики превосходят табличное значение:

$$t_a = 4,71 > t_{табл} = 2,23; t_b = 6,82 > t_{табл} = 2,23; t_{r_{xy}} = 6,82 > t_{табл} = 2,23,$$

поэтому параметры  $a$ ,  $b$  и  $r_{xy}$  не случайно отличаются от нуля, а статистически значимы.

5. Рассчитаем доверительные интервалы для параметров регрессии  $a$  и  $b$ . Для этого определим предельную ошибку для каждого показателя:

$$\Delta_a = t_{табл} \cdot m_a = 2,23 \cdot 12,04 = 26,82;$$

$$\Delta_b = t_{табл} \cdot m_b = 2,23 \cdot 0,139 = 0,31$$

Доверительные интервалы

$$y_a = a \pm \Delta_a = 56,75 \pm 26,82;$$

$$y_{a_{мин}} = 56,75 - 26,82 = 29,93;$$

$$y_{a_{макс}} = 56,75 + 26,82 = 83,57;$$

$$y_b = b \pm \Delta_b = 0,95 \pm 0,31;$$

$$y_{b_{мин}} = 0,95 - 0,31 = 0,639;$$

$$y_{b_{макс}} = 0,95 + 0,31 = 1,260.$$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью  $p = 1 - \alpha = 0,95$  параметры  $a$  и  $b$ , находясь

в указанных границах, не принимают нулевых значений, т.е. не являются статистически незначимыми и существенно отличны от нуля.

6. Полученные оценки уравнения регрессии позволяют использовать его для прогноза. Если прогнозное значение прожиточного минимума составит:

$x_p = \bar{x} \cdot 1,07 = 85,67 \cdot 1,07 = 91,66$  руб., тогда прогнозное значение заработной платы составит:  $\bar{y}_p = 56,75 + 0,95 \cdot 91,66 = 143,78$  руб.

7. Ошибка прогноза составит:

$$m_{\hat{y}_p} = S_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}} = 5,75 \cdot \sqrt{1 + \frac{1}{12} + \frac{(91,66 - 85,67)^2}{1706,67}} = 6,04$$

Предельная ошибка прогноза, которая в 95% случаев не будет превышена, составит:

$$\Delta_{\hat{y}_p} = t_{\text{табл}} \cdot m_{\hat{y}_p} = 2,23 \cdot 6,04 = 13,46$$

Доверительный интервал прогноза:

$$y_{\hat{y}_p} = \hat{y}_p \pm \Delta_{\hat{y}_p} = 143,78 \pm 13,46;$$

$$y_{\hat{y}_{p\text{min}}} = 143,78 - 13,46 = 130,31 \text{ руб.};$$

$$y_{\hat{y}_{p\text{max}}} = 143,78 + 13,46 = 157,24 \text{ руб.}$$

Выполненный прогноз среднемесячной заработной платы является надежным ( $p = 1 - \alpha = 1 - 0,05 = 0,95$ ) и находится в пределах от 130,31 руб. до 157,24 руб.

8. В заключение решения задачи построим на одном графике исходные данные и теоретическую прямую (рис. 1):

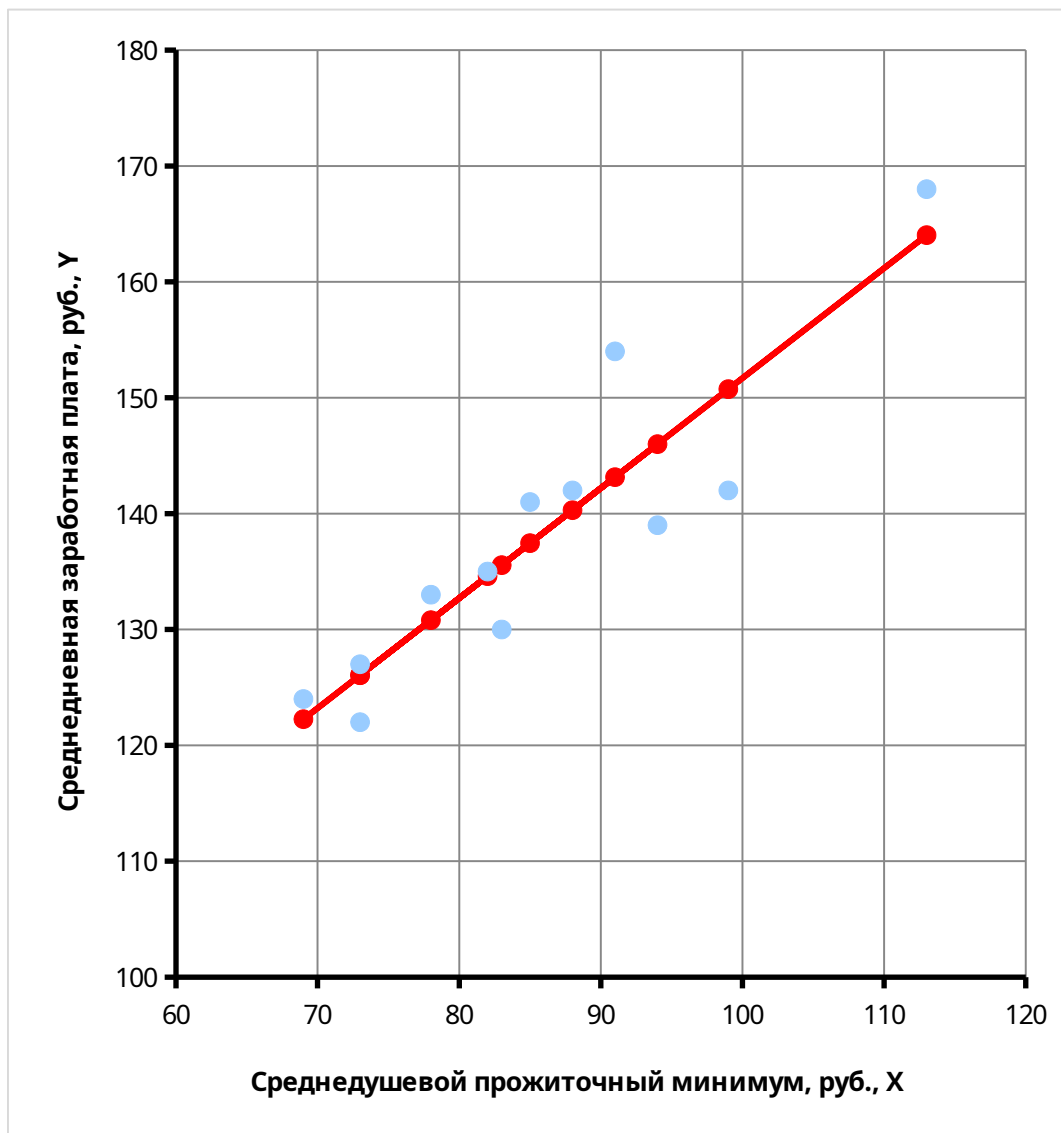


Рис. 1.