

Решение задачи по эконометрике (парная регрессия)

Задание

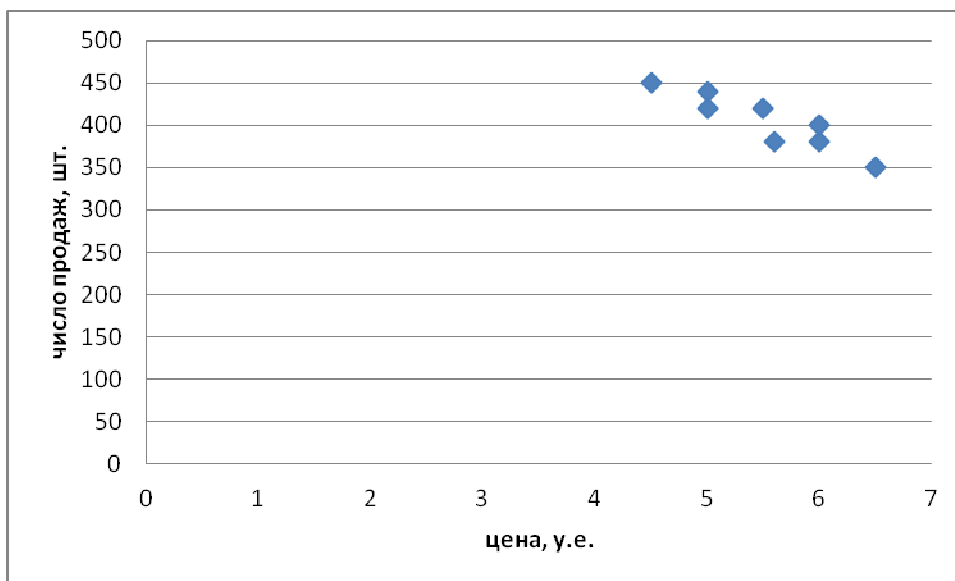
- 1) Постройте поле корреляции результивного и факторного признаков.
- 2) Определите параметры уравнения парной линейной регрессии. Дайте интерпретацию найденных параметров и всего уравнения в целом.
- 3) Постройте теоретическую линию регрессии, совместив ее с полем корреляции. Сделайте выводы.
- 4) Рассчитайте линейный коэффициент корреляции и поясните его смысл. Определите коэффициент детерминации и дайте его интерпретацию.
- 5) С вероятностью 0,95 оцените статистическую значимость коэффициента регрессии и уравнения регрессии в целом. Сделайте выводы.
- 6) С вероятностью 0,95 постройте доверительный интервал для прогноза оценки \hat{y}_i и доверительный интервал генерального значения $\hat{y}_{ген.}$ ($x^* = 5,75$).
- 7) Определите значение коэффициента эластичности и объясните его.

Компания, занимающаяся продажей радиоаппаратуры, установила на видеомagniтофон определенной модели цену, дифференцированную по регионам. Следующие данные показывают цены на видеомagniтофон в 8 различных регионах и соответствующее им число продаж:

Число продаж, шт.	420	380	350	400	440	380	450	420
Цена, у.е.	5,5	6,0	6,5	6,0	5,0	5,6	4,5	5,0

Решение

1) Построим поле корреляции:



Очевидно, что число продаж видеомagniтофонов находится в обратной зависимости от установленных цен. Можно предположить, что зависимость линейная, так как расположение точек на графике по форме близко к прямой линии.

2) Находим коэффициенты постулируемого уравнения регрессии, а также коэффициенты корреляции и детерминации. Для этого можно воспользоваться следующими формулами:

$$\hat{y}_x = a + bx$$

$$b = \frac{\overline{xy} - \bar{x} * \bar{y}}{x^2 - (\bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

Приведём расчётную таблицу:

№	y	x	y ²	x ²	xy	\hat{y}_x	$(y - \hat{y}_x)^2$	$(y - \bar{y})^2$
1	420	5,5	176400	30,25	2310	405,59	207,59	225
2	380	6	144400	36	2280	381,91	3,65	625
3	350	6,5	122500	42,25	2275	358,23	67,74	3025
4	400	6	160000	36	2400	381,91	327,21	25
5	440	5	193600	25	2200	429,27	115,07	1225
6	380	5,6	144400	31,36	2128	400,86	434,97	625
7	450	4,5	202500	20,25	2025	452,95	8,73	2025
8	420	5	176400	25	2100	429,27	85,99	225
Сумма	3240	44,1	1320200	246,11	17718	3240	1250,93	8000
Среднее	405	5,51	165025	30,76	2214,75	405	156,37	1000

Определим ряд характеристик моделируемых рядов:

$$\sigma_x = \sqrt{x^2 - (\bar{x})^2} = \sqrt{30,76 - 5,51^2} = 0,61$$

$$\sigma_y = \sqrt{y^2 - (\bar{y})^2} = \sqrt{165025 - 405^2} = 31,62$$

$$\sum (x - \bar{x})^2 = \sigma_x^2 * n = 0,61^2 * 8 = 3$$

Рассчитаем коэффициенты регрессии:

$$b = \frac{2214,74 - 5,51 * 405}{30,76 - 5,51^2} = -47,36$$

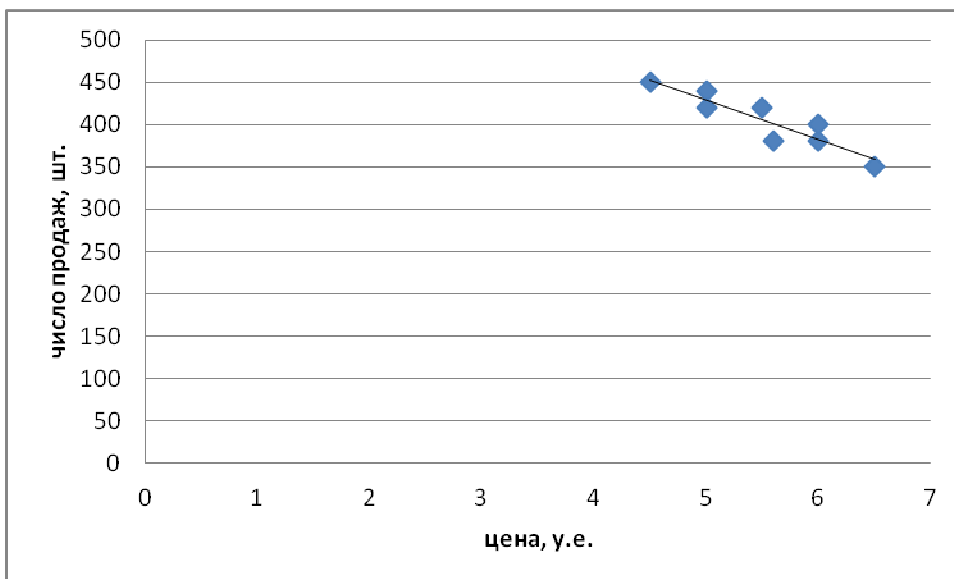
$$a = 405 + 47,36 * 5,51 = 666,08$$

Таким образом, регрессионное уравнение имеет следующий вид:

$$\hat{y}_x = 666,08 - 47,36x$$

Можно сделать вывод, что при росте цены на 1 у.е. число продаж видеомагнитофонов сокращается в среднем на $47,36 \approx 47$ шт. Условный объём продаж при нулевой цене равен 666 шт.

3) Построим теоретическую линию регрессии:



Так как эмпирические точки находятся вблизи теоретической прямой, то уравнение регрессии хорошо аппроксимирует данные.

4) Определим линейный коэффициент корреляции:

$$r = \frac{\overline{xy} - \bar{x} * \bar{y}}{\sigma_x * \sigma_y} = \frac{2214,74 - 5,51 * 405}{0,61 * 31,62} = -0,919$$

Коэффициент корреляции указывает на наличие связи и находится в интервале [-1;1]. В данном случае его отрицательное значение говорит об обратной связи между ценой видеомagniтофона и объёмом продаж. Связь весьма высокая по тесноте, так как $|r| > 0,9$.

Рассчитаем коэффициент детерминации:

$$R^2 = 1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2} = 1 - \frac{1250,93}{8000} = 0,844$$

Коэффициент детерминации показывает, что 84,4% вариации числа продаж видеомagniтофонов объясняется вариацией их цен, а остальные 15,6% - вариацией прочих факторов.

5) На основании данных расчётной таблицы рассчитаем остаточное среднее квадратическое отклонение:

$$S_{\text{ост}} = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{n - p - 1}}, \text{ где}$$

p – число факторов в модели.

$$S_{\text{ост}} = \sqrt{\frac{1250,93}{8 - 1 - 1}} = 14,44$$

Определим стандартную ошибку для коэффициента регрессии:

$$m_b = S_{\text{ост}} * \sqrt{\frac{1}{\sum (x - \bar{x})^2}} = 14,44 * \sqrt{\frac{1}{3}} = 8,32$$

Определим расчётное значение t-критерия Стьюдента:

$$t_b = \frac{b}{m_b} = \frac{-47,36}{8,32} = -5,69$$

Табличное значение t-критерия Стьюдента при $\alpha = 0,05$ и $k = 8 - 1 - 1 = 6$ будет равно $t_{\text{табл}} = 2,45$. Таким образом, коэффициент регрессии оказался значим, так как расчётное значение t-критерия по модулю выше табличного значения.

Для определения значимости полученного уравнения регрессии необходимо рассчитать эмпирическую величину F-критерия Фишера:

$$F_p = \frac{R^2}{1 - R^2} * \frac{n - p - 1}{p} = \frac{0,844}{1 - 0,844} * \frac{8 - 1 - 1}{1} = 32,37$$

По таблице F-распределения Снедекора-Фишера при $\alpha = 0,05$ и $K_1 = 1$, $K_2 = 8 - 2 = 6$ величина $F_T = 5,99$. Это означает, что гипотеза H_0 о несущественности связи между y и x с вероятностью ошибочности суждения $\alpha = 0,05$ отклоняется, то есть связь между этими переменными может быть признана существенной.

б) Спрогнозируем число продаж видеомagnитофонов при цене 5,75 у.е. за 1 шт.

$$\hat{y}_{xp} = 666,08 - 47,36 * 5,75 = 393,75$$

Построим доверительные интервалы для индивидуальной прогнозной величины \hat{y}_{xp} в точке x^* по следующим формулам:

$$\hat{y}_{xp} - dy \leq \hat{y}_{xp} \leq \hat{y}_{xp} + dy, \text{ где}$$

$$dy = t_{\text{табл}} * S_{\text{ост}} * \sqrt{1 + h}, \text{ где}$$

$$h = \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x - \bar{x})^2} = \frac{1}{8} + \frac{(5,75 - 5,51)^2}{3} = 0,14$$

Тогда:

$$dy = 2,45 * 14,44 * \sqrt{1 + 0,14} = 37,79$$

$$393,75 - 37,79 \leq \hat{y}_{xp} \leq 393,75 + 37,79$$

$$355,97 \leq \hat{y}_{xp} \leq 431,54$$

Следовательно, с 95% уверенностью можно утверждать, что число продаж видеомagnитофонов компанией при цене 5,75 у.е. за 1 шт., находится в пределах от 356 до 431 шт.

Если нужно сделать вывод относительно числа продаж во всех компаниях со средней ценой видеомagnитофона 5,75 у.е., то необходимо построить доверительный интервал для генерального среднего значения $\hat{y}_{ген.}$ при заданном x^* . Вариация в этом случае будет

меньше, поскольку имеет место не отдельное, а со среднее значение y . Следовательно, интервал будет уже:

$$dy = t_{\text{табл}} * S_{\text{ост}} * \sqrt{h} = 2,45 * 14,44 * \sqrt{0,14} = 13,4$$

$$393,75 - 13,4 \leq \hat{y}_{\text{ген}} \leq 393,75 + 13,4$$

$$380,36 \leq \hat{y}_{\text{ген}} \leq 407,15$$

Таким образом, с 95% уверенностью можно утверждать, что число продаж видеоманитофонов во всех компаниях по средней цене 5,75 у.е. находится в пределах от 380 до 407 шт.

7) Коэффициент эластичности определим по формуле:

$$\mathcal{E}_{yx} = b * \frac{\bar{x}}{y} = (-47,36) * \frac{0,61}{31,62} = -0,64\%$$

Итак, в относительном выражении при увеличении цены видеоманитофона на 1% возможно сокращение продаж в анализируемой компании на 0,64%.