

Имеются данные 12 месяцев по району города о рынке вторичного жилья, (y – стоимость квартиры, тыс. у.е., x_1 – размер жилой площади, м², x_2 – размер кухни, м². Данные приведены в табл. 2.

Таблица 2

y	x_1	x_2
13,0	37,0	6,2
16,4	60,9	10,0
17,0	60,0	8,5
15,2	52,1	7,4
14,2	40,1	7,0
10,5	30,4	6,2
20,0	43,0	7,5
12,0	32,1	6,4
15,6	35,1	7,0
12,5	32,0	6,2
13,2	33,0	6,0
14,6	32,5	5,8

Задание:

1. Рассчитайте параметры линейного уравнения множественной регрессии.
2. Дайте оценку силы связи факторов с результатом с помощью средних коэффициентов эластичности.
3. Оцените статистическую зависимость параметров и уравнения регрессии в целом с помощью соответственно критериев Стьюдента и Фишера ($\alpha=0,01$).
4. Рассчитайте среднюю ошибку аппроксимации. Сделайте вывод.
5. Составьте матрицы парных и частных коэффициентов корреляции и укажите информативные факторы.

Решение

Результаты расчетов приведены в таблице

	A	B	C	D	E	F	G	H	I	J
1	n	y	x ₁	x ₂	yx ₁	yx ₂	x ₁ x ₂	x ₁ ²	x ₂ ²	y ²
2	1	13	37	6,2	481	80,6	229,4	1369	38,44	169
3	2	16,4	60,9	10	998,76	164	609	3708,81	100	268,96
4	3	17	60	8,5	1020	144,5	510	3600	72,25	289
5	4	15,2	52,1	7,4	791,92	112,48	385,54	2714,41	54,76	231,04
6	5	14,2	40,1	7	569,42	99,4	280,7	1608,01	49	201,64
7	6	10,5	30,4	6,2	319,2	65,1	188,48	924,16	38,44	110,25
8	7	20	43	7,5	860	150	322,5	1849	56,25	400
9	8	12	32,1	6,4	385,2	76,8	205,44	1030,41	40,96	144
10	9	15,6	35,1	7	547,56	109,2	245,7	1232,01	49	243,36
11	10	12,5	32	6,2	400	77,5	198,4	1024	38,44	156,25
12	11	13,2	33	6	435,6	79,2	198	1089	36	174,24
13	12	14,6	32,5	5,8	474,5	84,68	188,5	1056,25	33,64	213,16
14	Σ	174,2	488,2	84,2	7283,16	1243,46	3561,66	21205,06	607,18	2600,9
15	Средн.	14,51667	40,68333	7,016667	606,93	103,6217	296,805	1767,088	50,59833	216,7417
16	σ	2,451	10,581	1,168						
17	Дисперсия	6,01	111,95	1,36						

Рассматриваем уравнение вида:

$$y = a + b_1x_1 + b_2x_2 + \varepsilon.$$

Параметры уравнения можно найти из решения системы уравнений:

$$\begin{cases} \bar{y} = a + b_1\bar{x}_1 + b_2\bar{x}_2, \\ \overline{yx_1} = a\bar{x}_1 + b_1\overline{x_1^2} + b_2\overline{x_1x_2}, \\ \overline{yx_2} = a\bar{x}_2 + b_1\overline{x_1x_2} + b_2\overline{x_2^2}. \end{cases}$$

Или, перейдя к уравнению в стандартизированном масштабе:

$$y^0 = \beta_1x_1^0 + \beta_2x_2^0 + \varepsilon,$$

где $y^0 = \frac{y - \bar{y}}{\sigma_y}$, $x_j^0 = \frac{x_j - \bar{x}_j}{\sigma_{x_j}}$ – стандартизированные переменные, β_j –

стандартизированные коэффициенты:

$$b_j = \beta_j \frac{\sigma_y}{\sigma_{x_j}}, \quad j = 1, 2.$$

Коэффициенты β_j определяются из системы уравнений:

$$\begin{cases} r_{yx_1} = \beta_1 + \beta_2 r_{x_2x_1}, \\ r_{yx_2} = \beta_1 r_{x_1x_2} + \beta_2. \end{cases}$$

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{1 - r_{x_1x_2}^2},$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{1 - r_{x_1x_2}^2};$$

$$r_{x_1x_2} = \frac{\overline{x_2 x_1} - \overline{x_2} \cdot \overline{x_1}}{\sigma_{x_1} \sigma_{x_2}},$$

$$r_{yx_j} = \frac{\overline{yx_j} - \overline{y} \cdot \overline{x_j}}{\sigma_{x_j} \sigma_y};$$

$$r_{x_1x_2} = \frac{296,805 - 7,0167 \cdot 40,6833}{10,581 \cdot 1,168} = 0,918$$

$$r_{x_1x_2}^2 = 0,842;$$

$$r_{yx_1} = \frac{606,93 - 14,517 \cdot 40,683}{10,581 \cdot 2,451} = 0,63$$

$$r_{yx_1}^2 = 0,397$$

$$r_{yx_2} = \frac{103,62 - 14,517 \cdot 7,0167}{1,168 \cdot 2,451} = 0,616$$

$$r_{yx_2}^2 = 0,397$$

$$\beta_1 = \frac{0,63 - 0,616 \cdot 0,918}{1 - 0,842} = 0,413$$

$$\beta_2 = \frac{0,616 - 0,63 \cdot 0,918}{1 - 0,842} = 0,237$$

$$b_1 = 0,413 \cdot \frac{2,451}{10,581} = 0,0956,$$

$$b_2 = 0,237 \cdot \frac{2,451}{1,168} = 0,497;$$

$$a = \overline{y} - b_1 \overline{x_1} - b_2 \overline{x_2} = 7,138.$$

Стандартизированная форма уравнения регрессии имеет вид:

$$\hat{y}^0 = 0,413 x_1^0 + 0,237 x_2^0.$$

Естественная форма уравнения регрессии имеет вид:

$$\hat{y} = 0,497 + 0,0956 x_1 + 0,497 x_2.$$

Для выяснения относительной силы влияния факторов на результативный признак рассчитываются средние коэффициенты эластичности:

$$\overline{\varepsilon}_j = b_j \frac{\overline{x_j}}{y}$$

$$\overline{\varepsilon}_1 = 0,0956 \cdot \frac{40,683}{14,517} = 0,268,$$

$$\overline{\varepsilon}_2 = 0,497 \cdot \frac{7,0167}{14,517} = 0,24.$$

Следовательно, при увеличении размера жилой площади (x_1) на 1% стоимость квартиры (y) увеличивается на 0,268% от своего среднего уровня. При увеличении размера кухни на 1% стоимость квартиры повышается на 0,24%.

2. Линейные коэффициенты частной корреляции для уравнения определяются следующим образом:

$$r_{y_{x_1 x_2}} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1 x_2}^2)}} = 0,211,$$

$$r_{y_{x_2 x_1}} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1 x_2}^2)}} = 0,121.$$

Линейный коэффициент множественной корреляции рассчитывается по формуле

$$R_{y_{x_1 x_2}} = \sqrt{r_{yx_1} \beta_1 + r_{yx_2} \beta_2} = 0,637.$$

Коэффициент множественной детерминации $R^2 = 0,406$.

3. $F_{факт} = \frac{R_{yx_1 x_2}^2}{1 - R_{yx_1 x_2}^2} \cdot \frac{n - k - 1}{k}$, где n - объем выборки, k - число факторов

модели. В нашем случае

$$F_{факт} = \frac{0,406}{1 - 0,406} \cdot \frac{12 - 2 - 1}{2} = 3,075.$$

Так как $F_{табл}(\alpha = 0,05, \nu_1 = 2, \nu_2 = 9) = 4,26$, то $F_{факт} < F_{табл}$ и потому уравнение не значимо в целом а модель $\hat{y} = 0,497 + 0,0956x_1 + 0,497x_2$ плохого качества.

Выясним статистическую значимость каждого фактора в уравнении множественной регрессии.

Для этого рассчитаем частные F -статистики.

$$F_{x_1 расч} = \frac{R_{yx_1 x_2}^2 - r_{yx_2}^2}{1 - R_{yx_1 x_2}^2} \cdot \frac{n - k - 1}{k} = 0,204.$$

Так как $F_{табл}(\alpha = 0,05, \nu_1 = 1, \nu_2 = 9) = 5,12$, то $F_{факт} < F_{табл}$ и следует вывод о нецелесообразности включения в модель фактора x_1 после фактора x_2 .

$$F_{x_2 расч.} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot \frac{n - k - 1}{k} = 0,204.$$

Так как $F_{факт} < F_{табл}$, то также следует вывод о нецелесообразности включения в модель фактора x_2 после фактора x_1 .

4. Результаты расчетов позволяют сделать вывод :

- 1) о незначимости фактора x_1 и нецелесообразности включения его в уравнение регрессии;
- 2) о незначимости фактора x_2 и нецелесообразности включения его в уравнение регрессии.

В результате вся модель $\hat{y} = 0,497 + 0,0956x_1 + 0,497x_2$ оказалась плохого качества.