

МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

$$y_x = a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m + \varepsilon$$

Спецификация модели

Отбор факторов

Выбор вида уравнения

Должны быть количественно измеримы

Цена товара

Производитель

5000 руб.

16000 руб.

18500 руб

....

Отечественный

Импортный

→

0

→

1

Отбор факторов

Должны объяснять вариацию
результатирующего признака

$$R_{m+1}^2 > R_m^2$$

$$S_{ост\,m+1}^2 < S_{ост\,m}^2$$

Спецификация модели

Отбор факторов

Не должны быть взаимно коррелированы либо находится в точной функциональной зависимости

$$r_{x_i x_j} \geq 0.8$$

$$x_i = f(x_j)$$

Спецификация модели

Объем продаж, руб.	Номер квартала	Цена, руб	Цена конкурента, руб.	Реклама, руб.
Y	X1	X2	X3	X4
12 000	1	16	17	5 000
13 000	2	15	18	6 000
15 000	3	15	17	4 000
18 000	4	15	16	9 000
23 000	5	16	18	8 000
34 000	6	17	19	10 000

Отбор факторов

$$\hat{y} = f(x_1, x_2, x_3, x_4)$$

	y	x_1	x_2	x_3	x_4
y	1				
x_1	0,92	1			
x_2	0,78	0,52	1		
x_3	0,66	0,45	0,7	1	
x_4	0,81	0,81	0,51	0,32	1

Отбор факторов

$$\hat{y} = f(x_1, x_2, x_3)$$

	y	x_1	x_2	x_3
y	1			
x_1	0,92	1		
x_2	0,78	0,52	1	
x_3	0,66	0,46	0,7	1

Отбор факторов

$$\hat{y} = f(x_1, x_2)$$

	y	x_1	x_2
y	1		
x_1	0,92	1	
x_2	0,78	0,52	1

Отбор факторов

$$\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3$$

$$\begin{array}{r}
 \mathbf{r}_{x_1x_1} \quad \mathbf{r}_{x_1x_2} \quad \mathbf{r}_{x_1x_3} \quad \mathbf{1} \quad \mathbf{10} \quad \mathbf{10} \\
 \mathbf{r}_{x_2x_1} \quad \mathbf{r}_{x_2x_2} \quad \mathbf{r}_{x_2x_3} \quad \mathbf{0} \quad \mathbf{11} \quad \mathbf{10} \quad \mathbf{01} \\
 \mathbf{r}_{x_3x_1} \quad \mathbf{r}_{x_3x_2} \quad \mathbf{r}_{x_3x_3} \quad \mathbf{0} \quad \mathbf{10} \quad \mathbf{11}
 \end{array}$$

Отбор факторов

Исключение коррелированных факторов

Увеличение объема выборки

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + \varepsilon$$

Отбор факторов

Оценка параметров линейного уравнения множественной регрессии

$$y_x = a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m + \varepsilon$$

$$\sum_i \left(y_i - \hat{y}_{x_i} \right)^2 \rightarrow \min$$

Оценка параметров линейного уравнения множественной регрессии

Скалярный метод

$$b_2 = \frac{\text{cov}(x_2, y) \sigma_{x_1}^2 - \text{cov}(x_2, y) \text{cov}(x_1, x_2)}{\sigma_{x_1}^2 \sigma_{x_2}^2 - (\text{cov}(x_1, x_2))^2}$$

$a = \bar{y} - b_1 \cdot \bar{x}_1 - b_2 \cdot \bar{x}_2$

$$\hat{y} = f(x_1, x_2)$$

Оценка параметров линейного уравнения множественной регрессии

$$a = -56631,5$$

$$b_1 = 3141,732$$

$$b_2 = 4129,921$$

$$\hat{y} = -56631,5 + 3141,732 \cdot x_1 + 4129,921 \cdot x_2$$

$$\hat{y} = f(x_1, x_2)$$

Оценка параметров линейного уравнения множественной регрессии

Матричный метод

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{m1} \\ 1 & x_{12} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & \dots & x_{mn} \end{bmatrix} \quad B = \begin{bmatrix} a \\ b_1 \\ \dots \\ b_m \end{bmatrix}$$

Оценка параметров линейного уравнения множественной регрессии

Матричный метод

$$Y = XB + e$$

$$e = [e_1, e_2, \dots, e_n]'$$

$$e = Y - XB$$

Оценка параметров линейного уравнения множественной регрессии

Матричный метод

$$Q = \sum e_i^2 \Rightarrow \min$$

$$B = (X'X)^{-1} X'Y$$

Оценка параметров линейного уравнения множественной регрессии

Семья	Накопления, S	Доход, Y	Имущество, W
1	3	40	60
2	6	55	36
3	5	45	36
4	3,5	30	15
5	1,5	30	90

Оценить регрессию S на Y и W .

Матричный метод

Оценка параметров линейного уравнения множественной регрессии

$$S = \begin{bmatrix} 3 \\ 6 \\ 5 \\ 3,5 \\ 1,5 \end{bmatrix} \quad X = \begin{bmatrix} 1 & 40 & 60 \\ 1 & 55 & 36 \\ 1 & 45 & 36 \\ 1 & 30 & 15 \\ 1 & 30 & 90 \end{bmatrix} \quad B = \begin{bmatrix} a \\ b_1 \\ b_2 \end{bmatrix}$$

Матричный метод

Оценка параметров линейного уравнения множественной регрессии

$$X'X = \begin{bmatrix} 5 & 200 & 237 \\ 200 & 8450 & 9150 \\ 237 & 9150 & 14517 \end{bmatrix}$$

Матричный метод

Оценка параметров линейного уравнения множественной регрессии

$$(X'X)^{-1} = \begin{bmatrix} 5,6916 & -0,1074 & -0,0252 \\ -0,1074 & 0,0024 & 0,00024 \\ -0,0252 & 0,00024 & 0,00033 \end{bmatrix}$$

Матричный метод

Оценка параметров линейного уравнения множественной регрессии

$$B = (X'X)^{-1} X'Y =$$

$$\hat{S} = (0,27887 + 0,12229Y - 0,0294)W$$

Матричный метод

Регрессионная модель в стандартизованном масштабе

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_m t_{x_m} + \varepsilon,$$

$$t_y, t_{x_1}, \dots, t_{x_m}$$

$$t_y = \frac{y - \bar{y}}{\sigma_y}$$

$$t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$$

$$\bar{t}_y = \bar{t}_{x_i} = 0$$

$$\sigma_{t_y} = \sigma_{t_{x_i}} = 1$$

Регрессионная модель в стандартизованном масштабе

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_m t_{x_m} + \varepsilon,$$

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}$$

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - \dots - b_m \bar{x}_m$$

Регрессионная модель в стандартизованном масштабе

$$\left\{ \begin{array}{l} r_{yx_1} = \beta_1 + \beta_2 r_{x_1x_2} + \beta_3 r_{x_1x_3} + \dots + \beta_m r_{x_1x_m}, \\ r_{yx_2} = \beta_1 r_{x_1x_2} + \beta_2 + \beta_3 r_{x_1x_3} + \dots + \beta_m r_{x_1x_m}, \\ \dots \\ r_{yx_m} = \beta_1 r_{x_1x_m} + \beta_2 r_{x_2x_m} + \beta_3 r_{x_3x_m} + \dots + \beta_m, \end{array} \right.$$

Частные уравнения регрессии

$$\hat{y}_{x_i \cdot x_1, x_2, \dots, x_{i-1}, x_{i+1}, x_2, \dots, x_m} = A_i + b_i \cdot x_i$$
$$a + b_1 \cdot \bar{x}_1 + \dots + b_{i-1} \cdot \bar{x}_{i-1} + b_i \cdot x_i + b_{i+1} \cdot \bar{x}_{i+1} + \dots + b_m \cdot \bar{x}_m$$

$(i = 1, 2, \dots, m)$

$$A_i = a + b_1 \cdot \bar{x}_1 + \dots + b_{i-1} \cdot \bar{x}_{i-1} + b_{i+1} \cdot \bar{x}_{i+1} + \dots + b_m \cdot \bar{x}_m$$

$$\hat{y}_{x_i \cdot x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m} = a + b_1 \cdot \bar{x}_1 + \dots + b_{i-1} \cdot \bar{x}_{i-1} + b_i \cdot x_i + b_{i+1} \cdot \bar{x}_{i+1} + \dots + b_m \cdot \bar{x}_m$$

Частные уравнения регрессии

$$\bar{y}_{x_i} = b_i \cdot \frac{\bar{x}_i}{\hat{y}_{x_i: x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m, x_m}}$$

Частные уравнения регрессии

$$r_{yx_i \cdot x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_m} = \sqrt{\frac{1 - R_{yx_1 x_2 \dots x_i \dots x_m}^2}{1 - R_{yx_1 x_2 \dots x_{i-1} x_{i+1} \dots x_m}^2}}$$

Частные уравнения регрессии

$$r_{yx_1 \cdot x_2} = \sqrt{1 - \frac{1 - R_{yx_1 x_2}^2}{1 - r_{yx_2}^2}}$$

$$r_{yx_2 \cdot x_1} = \sqrt{1 - \frac{1 - R_{yx_1 x_2}^2}{1 - r_{yx_1}^2}}$$

Частные уравнения регрессии

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

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

Проверка существенности факторов и показатели качества регрессии

$$t_{b_i} = \frac{b_i}{m_{b_i}},$$

$$t_a = \frac{a}{m_a}$$

$$(i = 1, 2, \dots, m)$$

$$m_{b_i} = \sqrt{S_{ост}^2 \cdot [(X' \cdot X)^{-1}]_{ii}} \quad (i = 0, 1, 2, \dots, m)$$

Проверка существенности факторов и показатели качества регрессии

$$H_0 : b_i = 0$$

$$H_0 : a = 0$$

$$t_{расч} > t_{табл}(\alpha; n - m - 1)$$

Проверка существенности факторов и показатели качества регрессии

$$b_i - t_{таб} \cdot m_{b_i} \leq b_i \leq b_i + t_{таб} \cdot m_{b_i}$$

$$a - t_{таб} \cdot m_a \leq a \leq a + t_{таб} \cdot m_a$$

Проверка существенности факторов и показатели качества регрессии

$$F_{x_i} = \frac{R_{yx_1x_2\dots x_m}^2 - R_{yx_1\dots x_{i-1}x_{i+1}\dots x_m}^2}{1 - R_{yx_1x_2\dots x_m}^2} \cdot \frac{n - m - 1}{1}$$

$$F_{x_1} = \frac{R_{yx_1x_2}^2 - r_{yx_2}^2}{1 - R_{yx_1x_2}^2} \cdot (n - 3) \quad F_{x_2} = \frac{R_{yx_1x_2}^2 - r_{yx_1}^2}{1 - R_{yx_1x_2}^2} \cdot (n - 3)$$

Проверка существенности факторов и показатели
качества регрессии

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - \frac{S_{ост}^2}{S_y^2}}$$

$$R_{yx_1x_2\dots x_m} > \max(r_{yx_i}), \quad (i = \overline{1, m})$$

Проверка существенности факторов и показатели качества регрессии

$$\bar{R}^2_{yx_1x_2\dots x_m} = R^2_{yx_1x_2\dots x_m} = 1 - \frac{\sum (y - \hat{y}_{x_1x_2\dots x_m})^2}{\sum (y - \bar{y})^2} \cdot \frac{n-1}{n-m-1}$$

Проверка значимости уравнения регрессии

$$F = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}$$

$$F > F_{\text{табл}} \quad \alpha$$
$$k1 = m, \quad k2 = (n - m - 1)$$