ПАРНАЯ РЕГРЕССИЯ

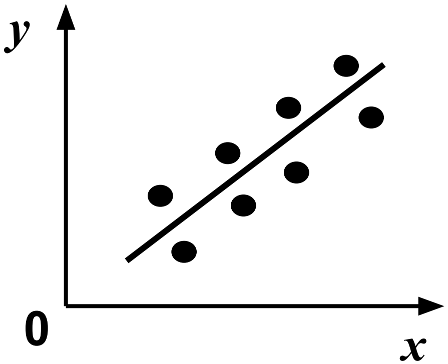
Регрессия – зависимость между независимыми (объясняющими) переменными и условным математическим ожиданием зависимой (объясняемой) переменной

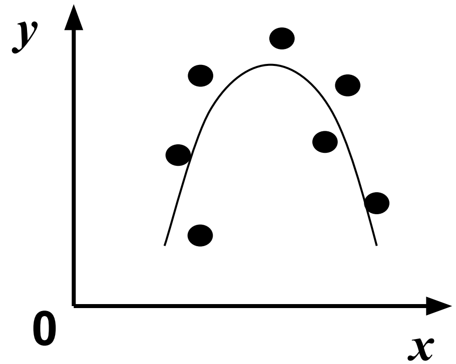
$$f(x) = M(Y / X)$$

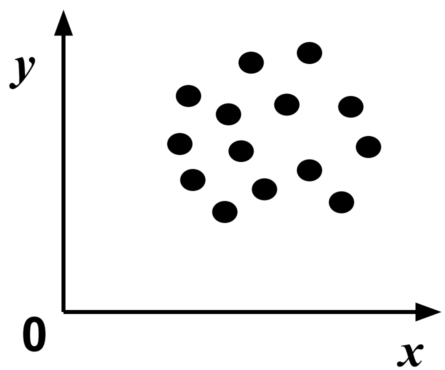
Парная регрессия -- регрессия между двумя переменными y **и** x

$$y = f(a, x) + \varepsilon$$

$$y_i = \hat{y}_{x_i} + \varepsilon_i$$





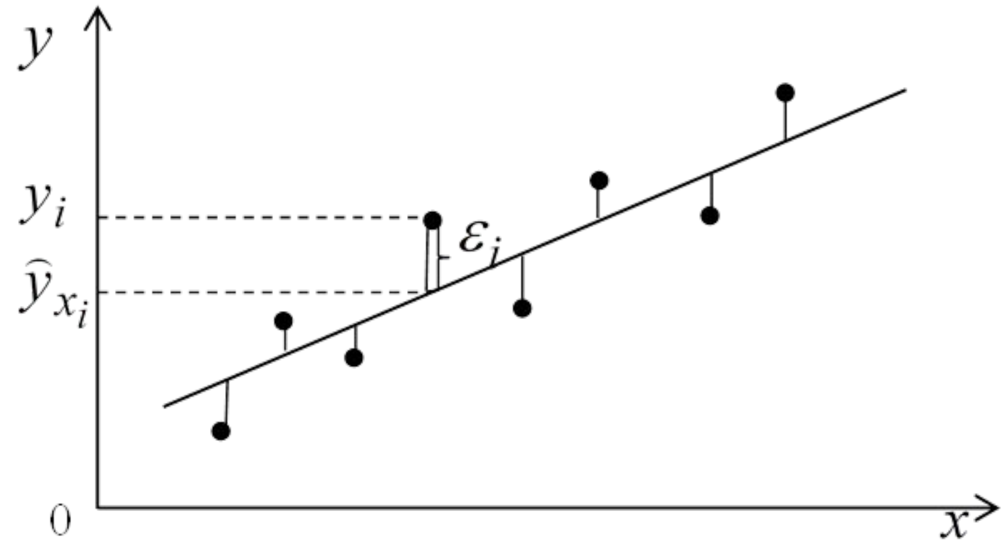


$$\sigma^2_{ocm} = \frac{1}{n}\sum(y-\hat{y}_x)^2$$

ЛИНЕЙНАЯ МОДЕЛЬ ПАРНОЙ РЕГРЕССИИ

$$\hat{y}_x = a + b \cdot x$$

$$y_x = a + b \cdot x + \varepsilon$$



Оценка параметров. Метод наименьших квадратов (МНК)

$$\sum_{i=1}^{n} (y_i - \hat{y}_{x_i})^2 = \sum_{i=1}^{n} \varepsilon_i^2 \to \min$$

Оценка параметров. МНК

Обозначим $\sum_i \varepsilon_i^2$ **через** $S(a,\ b)$

$$S(a,\ b) = \sum (y - a - b \cdot x)^2$$

,

$$\begin{cases} \dfrac{\partial S}{\partial a} = -2\sum (y - a - b \cdot x) = 0; \\ \dfrac{\partial S}{\partial b} = -2\sum x(y - a - b \cdot x) = 0. \end{cases}$$

$$\begin{cases} a \cdot n \quad + b \cdot \sum x \ = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 = \sum x \cdot y. \end{cases}$$

Оценка параметров. МНК

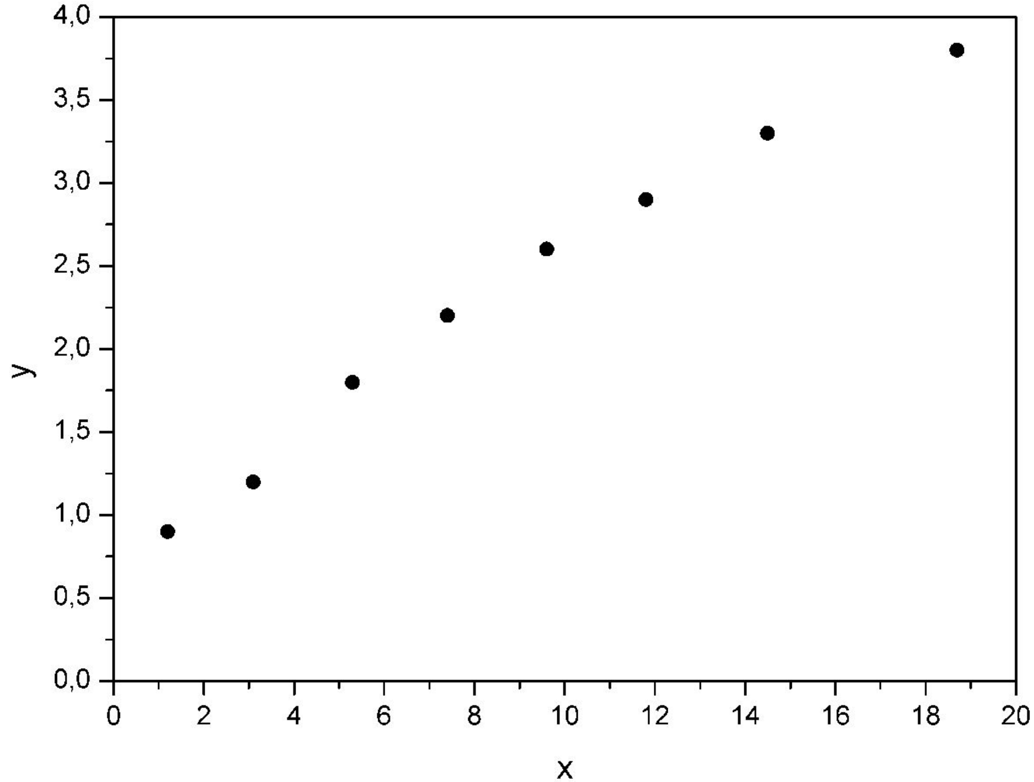
$$b = \frac{\operatorname{cov}(x,\ y)}{\sigma_x^2} \qquad a = \overline{y} - b \cdot \overline{x}$$

$$\operatorname{cov}(x, y) = \overline{y \cdot x} - \overline{y} \cdot \overline{x} \qquad \sigma_x^2 = \overline{x^2} - \overline{x}^2$$

$$\overline{x} = \frac{1}{n}\sum x \qquad \overline{y \cdot x} = \frac{1}{n}\sum y \cdot x$$

$$\overline{y} = \frac{1}{n}\sum y \qquad \overline{x^2} = \frac{1}{n}\sum x^2$$

Расходы на продукты питания, *y*	0,9	1,2	1,8	2,2	2,6	2,9	3,3	3,8
Доходы семьи, *x*	1,2	3,1	5,3	7,4	9,6	11,8	14,5	18,7



	x	y	$x \cdot y$	x^2	y^2
1	**2**	**3**	**4**	**5**	**6**
1	1,2	0,9	1,08	1,44	0,81
2	3,1	1,2	3,72	9,61	1,44
3	5,3	1,8	9,54	28,09	3,24
4	7,4	2,2	16,28	54,76	4,84
5	9,6	2,6	24,96	92,16	6,76
6	11,8	2,9	34,22	139,24	8,41
7	14,5	3,3	47,85	210,25	10,89
8	18,7	3,8	71,06	349,69	14,44
Итого	71,6	18,7	208,71	885,24	50,83
Среднее значение	8,95	2,34	26,09	110,66	6,35
σ	5,53	0,935	–	–	–
σ^2	30,56	0,874	–	–	–

$$b = \frac{\mathrm{cov}(x,\ y)}{\sigma_x^2} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{26,09 - 8,95 \cdot 2,34}{30,56} = 0,168$$

$$a = \bar{y} - b \cdot \bar{x} = 2,34 - 0,168 \cdot 8,95 = 0,836$$

$$y_x = 0,836 + 0,168 \cdot x$$

Исследование уравнения регрессии

$$\boxed{r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = \frac{\operatorname{cov}(x,\ y)}{\sigma_x \cdot \sigma_y}}$$

$$-1 \le r_{xy} \le 1$$

$$\boxed{r_{xy}^2 = 1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}}$$

$$\sigma_{\text{ост}}^2 = \frac{1}{n} \sum \left(y - \hat{y}_x \right)^2$$

$$\sigma_y^2 = \frac{1}{n} \sum \left(y - \overline{y} \right)^2 = \overline{y^2} - \overline{y}^2$$

Оценка качества модели из относительных отклонениях по каждому наблюдению

$$\overline{A} = \frac{1}{n}\sum\left|\frac{y - \hat{y}_x}{y}\right| \cdot 100\%$$

Оценка значимости модели

Схема дисперсионного анализа

$$\sum\left(y-\overline{y}\right)^2=\sum\left(\hat{y}_x-\overline{y}\right)^2+\sum\left(y-\hat{y}_x\right)^2$$

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на одну степень свободы
Общая	$\sum\left(y-\overline{y}\right)^2$	$n-1$	$S^2_{\text{общ}}=\dfrac{\sum\left(y-\overline{y}\right)^2}{n-1}$
Факторная	$\sum\left(\hat{y}_x-\overline{y}\right)^2$	m	$S^2_{\text{факт}}=\dfrac{\sum\left(\hat{y}_x-\overline{y}\right)^2}{m}$
Остаточная	$\sum\left(y-\hat{y}_x\right)^2$	$n-m-1$	$S^2_{\text{ост}}=\dfrac{\sum\left(y-\hat{y}_x\right)^2}{n-m-1}$

$$k_1=m$$

$$F=\frac{S^2_{\text{факт}}}{S^2_{\text{ост}}}$$

$$F_{\text{табл}}\left(\alpha;\,k_1;\,k_2\right)$$

$$\alpha$$

$$k_2=n-m-1$$

Оценка значимости параметров модели

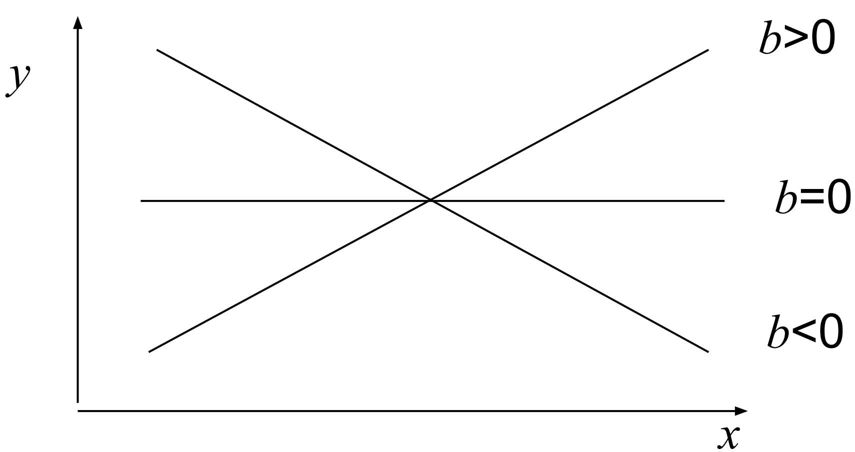
$$m_b = \sqrt{\frac{S_{\text{ост}}^2}{\sum\left(x-\bar{x}\right)^2}} = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}}$$

$$S_{\text{ост}}^2 = \frac{\sum\left(y-\hat{y}_x\right)^2}{n-2}$$

$$t_b = \frac{b}{m_b}$$

$$\alpha$$

$$\left(n-2\right)$$



$$b \pm t_{\text{табл}} \cdot m_b$$

$$\gamma = 1-\alpha$$

Оценка значимости параметров модели

$$m_a = \sqrt{S_{\text{ост}}^2 \cdot \frac{\sum x^2}{n \cdot \sum \left(x - \bar{x}\right)^2}} = S_{\text{ост}} \cdot \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n}$$

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

$$\alpha$$

$$(n-2)$$

$$a \pm t_{табл} \cdot m_a$$

$$\gamma = 1 - \alpha$$

Построение интервальных оценок для уравнения парной регрессии

$$\hat{y}_i \pm t_{табл} \cdot m_{\hat{y}_i}$$

$$m_{\hat{y}_i} = S_{ост} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{n \cdot \sigma_x^2}}$$

	x	y	$x \cdot y$	x^2	y^2	$\hat{y}_x$	$y - \hat{y}_x$	$\left(y - \hat{y}_x\right)^2$	A_i , %
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1,2	0,9	1,08	1,44	0,81	1,038	–0,138	0,0190	15,33
2	3,1	1,2	3,72	9,61	1,44	1,357	–0,157	0,0246	13,08
3	5,3	1,8	9,54	28,09	3,24	1,726	0,074	0,0055	4,11
4	7,4	2,2	16,28	54,76	4,84	2,079	0,121	0,0146	5,50
5	9,6	2,6	24,96	92,16	6,76	2,449	0,151	0,0228	5,81
6	11,8	2,9	34,22	139,24	8,41	2,818	0,082	0,0067	2,83
7	14,5	3,3	47,85	210,25	10,89	3,272	0,028	0,0008	0,85
8	18,7	3,8	71,06	349,69	14,44	3,978	–0,178	0,0317	4,68
Итого	71,6	18,7	208,71	885,24	50,83	18,717	–0,017	0,1257	52,19
Среднее значение	8,95	2,34	26,09	110,66	6,35	2,34	–	0,0157	6,52
σ	5,53	0,935	–	–	–	–	–	–	–
σ^2	30,56	0,874	–	–	–	–	–	–	–

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = 0,168 \cdot \frac{5,53}{0,935} = 0,994 \qquad r_{xy}^2 = 0,987$$

$$A_i = \left| \frac{y_i - \hat{y}_{x_i}}{y_i} \right| \cdot 100\% \qquad \overline{A} = 6,52\%$$

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n-2) = \frac{0,987}{1-0,987} \cdot 6 = 455,54$$

$$F_{\text{табл}} = 5,99 \qquad k_1 = 1 \qquad F_{\text{факт}} > F_{\text{табл}}$$

$$k_2 = n - 2 = 6$$

$$\alpha = 0,05$$

$$\left(S_{\text{ост}}^2 = \frac{\sum\left(y - \hat{y}_x\right)^2}{n-2} = \frac{0,1257}{8-2} = 0,021 \right)$$

$$m_b = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}} = \frac{\sqrt{0,021}}{5,53 \cdot \sqrt{8}} = 0,0093 \qquad t_b = \frac{0,168}{0,0093} = 18,065$$

$$t_{\text{табл}} = 1,943 \qquad \alpha = 0,05$$

$$\nu = n - 2 = 6$$

$$m_a = S_{\text{ост}} \cdot \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n} = \frac{\sqrt{0,021 \cdot 885,24}}{5,53 \cdot 8} = 0,0975 \qquad t_a = \frac{0,836}{0,0975} = 8,574$$

$$x_p = 1,1 \cdot \bar{x} = 1,1 \cdot 8,95 = 9,845$$

$$y_p = 0,836 + 0,168 \cdot 9,845 = 2,490$$

$$m_{y_p} = S_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{n \cdot \sigma_x^2}} = \sqrt{0,021 \cdot \left(1 + \frac{1}{8} + \frac{(9,845 - 8,95)^2}{8 \cdot 30,56}\right)} = 0,154$$

$$y_p - \Delta_{y_p} \leq \hat{y}_p \leq y_p + \Delta_{y_p}$$

$$2,113 < \hat{y}_p < 2,867$$