

Расчет коэффициента корреляции и функции линейной регрессии

Коэффициент корреляции:

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}}$$

где

x_i - значения, принимаемые случайной переменной X,

y_i - значения, принимаемые случайной переменной Y,

\bar{x} - средняя арифметическая по X,

\bar{y} - средняя арифметическая по Y.

Функция линейной регрессии для случайных величин X и Y:

$$y = a + b \cdot x,$$

где коэффициенты a и b - числа, которые находятся по правилу:

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - \bar{x}^2} \quad a = \bar{y} - b \cdot \bar{x};$$

Пример

Найдите коэффициент корреляции и формулу регрессионной линейной зависимости продаж детских колясок, шт./тыс.чел. (величина Y) от количества рождений, чел./тыс. чел. (величина X).

	Наблюдения:					Сумма	Средняя
	i=1	i=2	i=3	i=4	i=5		
x_i	32	25	20	30	15	122	24,4
y_i	16	16	9	15	9	65	13,0
$x_i \cdot y_i$	512	400	180	450	135	1677	335,4
x_i^2	1024	625	400	900	225	3174	634,8

$$b = 0,4615$$

$$a = 1,7404$$

Уравнение регрессии:

$$Y = 1,7404 + 0,4615 \cdot X$$

Коэффициент корреляции = 0,88

Корреляционное поле

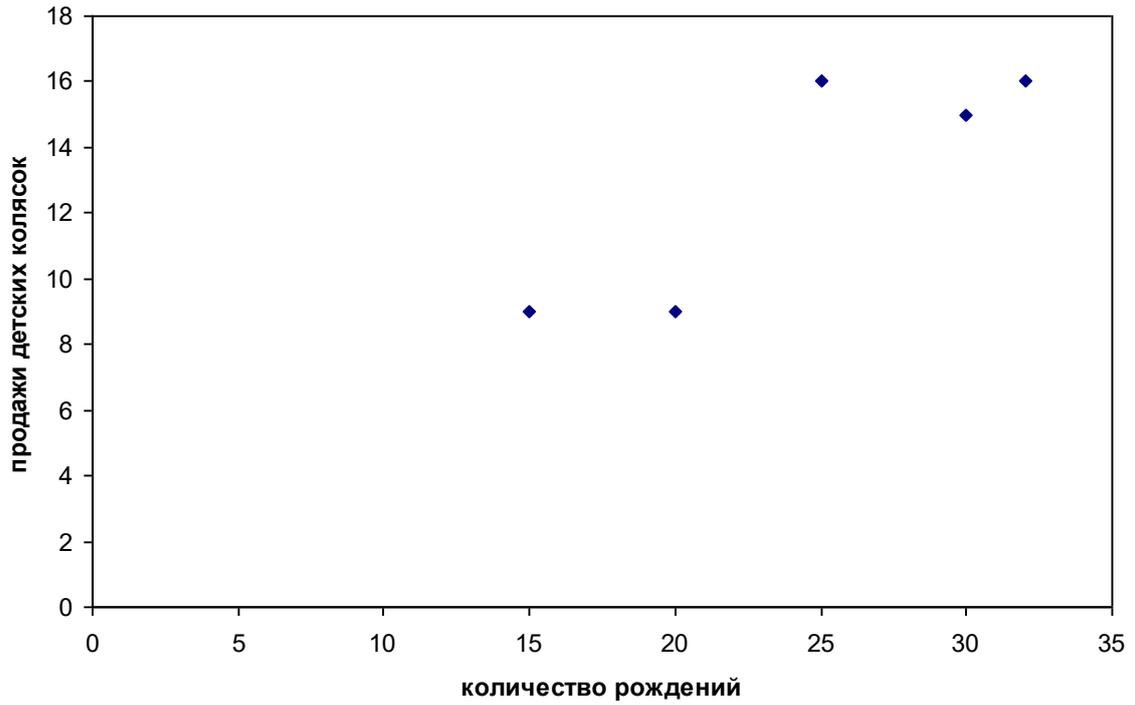
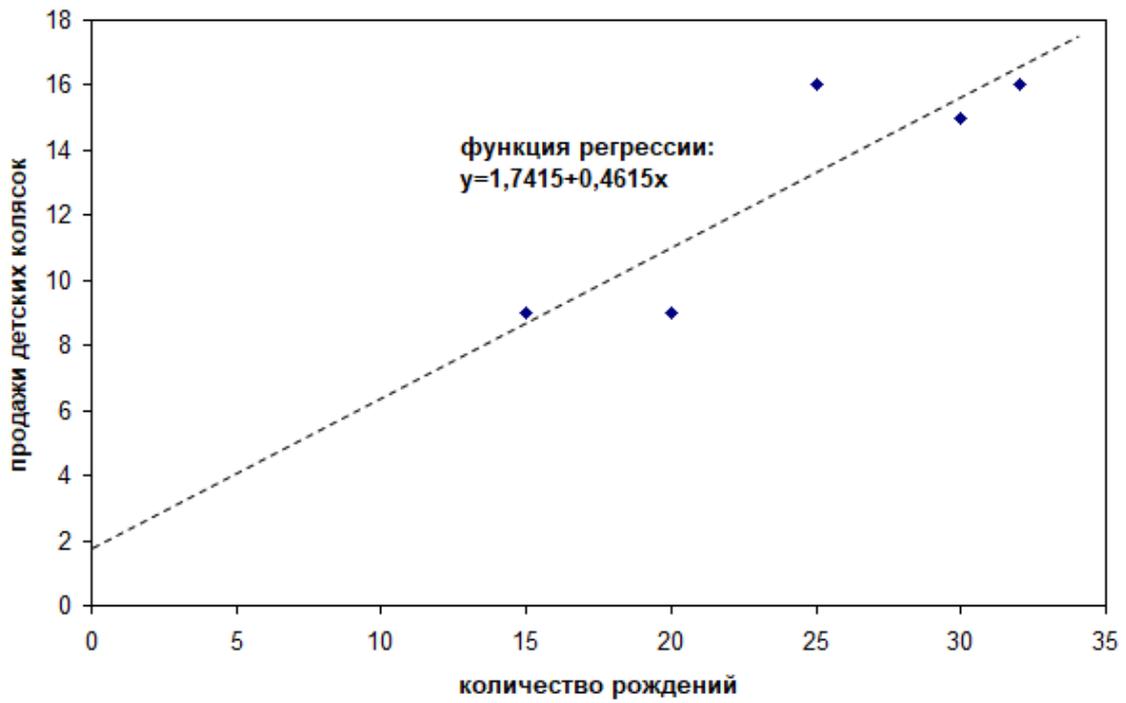
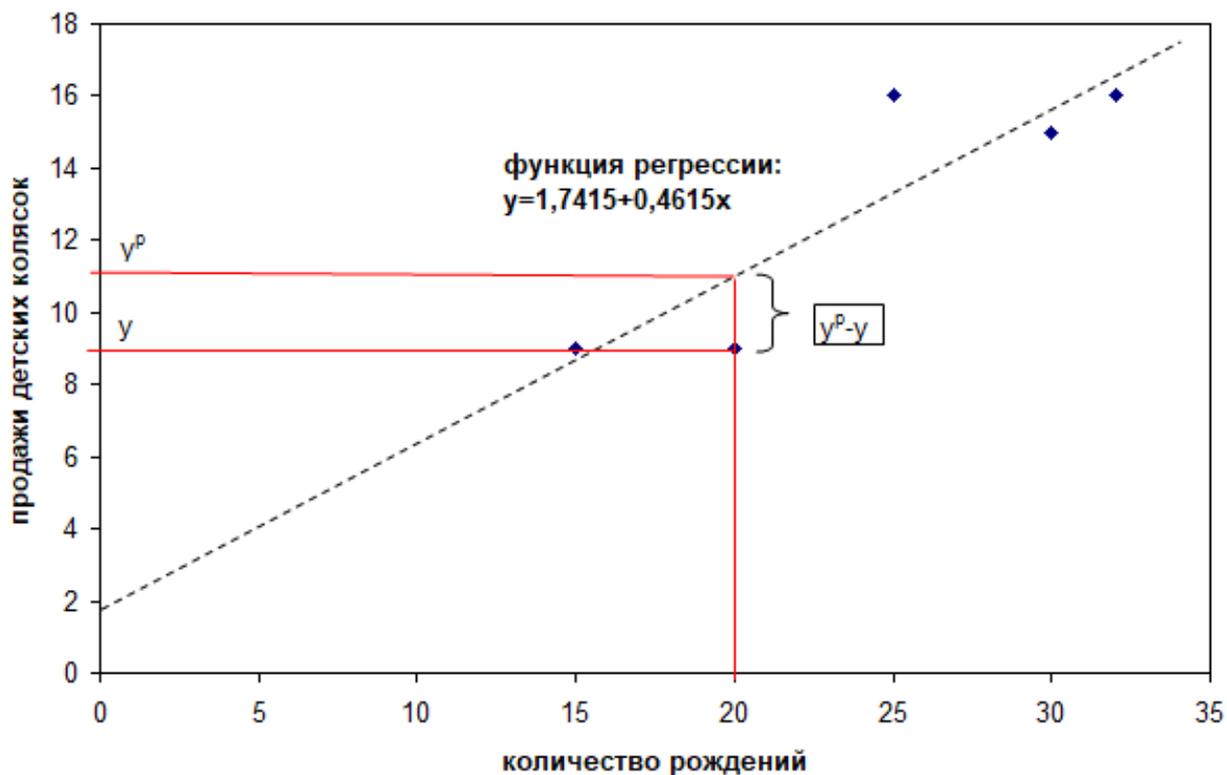


График функции регрессии



Ошибка прогноза



Расчет средней ошибки прогноза:

$$E = \sqrt{\frac{\sum_{i=1}^n (y_i^p - y_i)^2}{n}} \approx 1,55$$

то есть, в прогнозе по уравнению регрессии можно ошибиться на 1,55 коляски на 1000 жителей.