

Задание 1.

Компанию по продаже автомобилей интересует зависимость между пробегом автомобилей X и стоимостью ежемесячного обслуживания Y . Для выяснения связи отобрано 10 автомобилей:

1. Вычислить точечные оценки для математического ожидания, дисперсии и среднеквадратического отклонения показателей X и Y .
2. Найти уравнение линейной регрессии $\hat{y}_x = a + bx$.
3. Найти среднюю ошибку аппроксимации. Сделать вывод о качестве модели.
4. Найти парный коэффициент линейной корреляции и с доверительной вероятностью 0,95 проверить его значимость.
5. Сделать точечный и интервальный прогноз для случая расходов на рекламу, равных 5 млн. руб.
6. Построить график линии регрессии с нанесением на него опытных данных.

X	Y
1,50	12,00
2,00	17,40
2,50	18,60
3,00	18,00
3,50	21,30
4,00	21,30
4,50	24,40
5,00	24,10
5,50	27,20
6,00	28,70

Решение:

1). Точечная оценка математического ожидания – это выборочная средняя \bar{x} :

$$\bar{x} = \frac{\sum x_i}{n},$$

где n - число автомобилей.

Составим вспомогательную таблицу.

Наблюдения	X _i	Y _i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1,0000	1,5000	12,0000	5,0625	86,4900
2,0000	2,0000	17,4000	3,0625	15,2100
3,0000	2,5000	18,6000	1,5625	7,2900
4,0000	3,0000	18,0000	0,5625	10,8900
5,0000	3,5000	21,3000	0,0625	0,0000
6,0000	4,0000	21,3000	0,0625	0,0000
7,0000	4,5000	24,4000	0,5625	9,6100
8,0000	5,0000	24,1000	1,5625	7,8400
9,0000	5,5000	27,2000	3,0625	34,8100
10,0000	6,0000	28,7000	5,0625	54,7600
Итого	37,5000	213,0000	20,6250	226,9000

$$\bar{x} = 37,5/10 = 3.75$$

$$\bar{y} = 213.00/10 = 21.30$$

б). Точечная оценка дисперсии:

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$s_x^2 = 20,625/10 = 2,0625$$

$$s_y^2 = 226.90/10 = 22.69$$

в). Точечная оценка СКО:

$$s_y = \sqrt{22.69} = 4.763$$

$$s_x = \sqrt{2,0625} = 1,436,$$

2). Найдем уравнение линейной регрессии:

$$\tilde{y} = ax + b.$$

Для этого нужно решить систему нормальных уравнений:

$$\begin{cases} nb + a \sum x_i = \sum y_i \\ b \sum x_i + a \sum (x_i^2) = \sum (x_i y_i) \end{cases}$$

где n - объем выборки.

Составим расчетную таблицу:

i	X _i	Y _i	X _i *Y _i	x _i ²
1	1,5000	12,0000	18,0000	2,2500
2	2,0000	17,4000	34,8000	4,0000
3	2,5000	18,6000	46,5000	6,2500
4	3,0000	18,0000	54,0000	9,0000
5	3,5000	21,3000	74,5500	12,2500
6	4,0000	21,3000	85,2000	16,0000
7	4,5000	24,4000	109,8000	20,2500
8	5,0000	24,1000	120,5000	25,0000
9	5,5000	27,2000	149,6000	30,2500
10	6,0000	28,7000	172,2000	36,0000
Итого	37,5000	213,0000	865,1500	161,2500

$$\begin{cases} 10b + 37.5a = 213 \\ 22.5b + 161.25 = 865.15 \end{cases}$$

Решая эту систему, находим a и b :

$$a = 3.2134, \quad b = 9.2273$$

Тогда уравнение регрессии:

$$\tilde{y} = 3.2134x + 9.2273.$$

3). Средняя ошибка аппроксимации \bar{A} .

Y _i	Y _x	$\left \frac{y - \hat{y}_x}{y} \right $
12,0000	14,0564	0,1714
17,4000	15,6661	0,0997
18,6000	17,2758	0,0712
18,0000	18,8855	0,0492
21,3000	20,4952	0,0378
21,3000	22,1048	0,0378
24,4000	23,7145	0,0281
24,1000	25,3242	0,0508
27,2000	26,9339	0,0098
28,7000	28,5436	0,0054
Итого		0,5611

$$\bar{A} = 0,5611 * 100\% = 56.11\%.$$

4). Парный коэффициент линейной корреляции:

$$r = a \frac{s_x}{s_y}.$$

Отсюда:

$$r = 3.2134 \cdot 1,436 / 4.763 = 0,969.$$

С доверительной вероятностью 0,95 проверим его значимость.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$t = \frac{0,969\sqrt{10-2}}{\sqrt{1-0,969^2}} = 15.563.$$

Табличное значение $t_{0,95;8} = 2,31$. Поскольку $t > t_{0,95;8}$, коэффициент корреляции значим с доверительной вероятностью 0,95.

5). Точечный прогноз для случая расхода на рекламу $x_0 = 5$ млн. руб.:

$$\tilde{y}(x_0) = 3.2134x + 9.2273 = 3.2134 \cdot 5 + 9.2273 = 25.2943$$

Интервальный прогноз:

$$\tilde{y}(x_0) - t_{1-\alpha;n-2} s_{y_0} \leq y_0^* \leq \tilde{y}(x_0) + t_{1-\alpha;n-2} s_{y_0},$$

$$\text{где } s_{y_0}^2 = s^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right],$$

а остаточная дисперсия s^2 равна:

$$s^2 = \frac{\sum (\tilde{y}_i - y_i)^2}{n-2}.$$

Для нахождения нужных значений составим таблицу:

наблюдения	X _i	Y _i	Y _x	($\tilde{y}_i - y_i$) ²	(x _i - \bar{x}) ²
1	1,5000	12,0000	14,0564	4,2286	5,0625
2	2,0000	17,4000	15,6661	0,0997	3,0625
3	2,5000	18,6000	17,2758	0,0712	1,5625
4	3,0000	18,0000	18,8855	0,0492	0,5625
5	3,5000	21,3000	20,4952	0,0378	0,0625
6	4,0000	21,3000	22,1048	0,0378	0,0625
7	4,5000	24,4000	23,7145	0,0281	0,5625
8	5,0000	24,1000	25,3242	0,0508	1,5625
9	5,5000	27,2000	26,9339	0,0098	3,0625
10	6,0000	28,7000	28,5436	0,0054	5,0625
Итого	37,5000	213,0000		4,6184	20,6250

$$s^2 = 4.6184/8 = 0,5773$$

$$s_{y_0}^2 = 0,5773 \left[1 + 1/10 + \frac{(5 - 3.75)^2}{20.625} \right] = 0,6788$$

$$s_{y_0} = \sqrt{0,6788} = 0,8239$$

$$\tilde{y}(x_0) - t_{1-\alpha;n-2} s_{y_0} \leq y_0^* \leq \tilde{y}(x_0) + t_{1-\alpha;n-2} s_{y_0}$$

$$23.3911 \leq y_0^* \leq 227.1975$$

б). График линейной регрессии:

График линейной регрессии

