

ЛИНЕЙНАЯ РЕГРЕССИЯ: РЕШЕНИЕ ЗАДАЧ

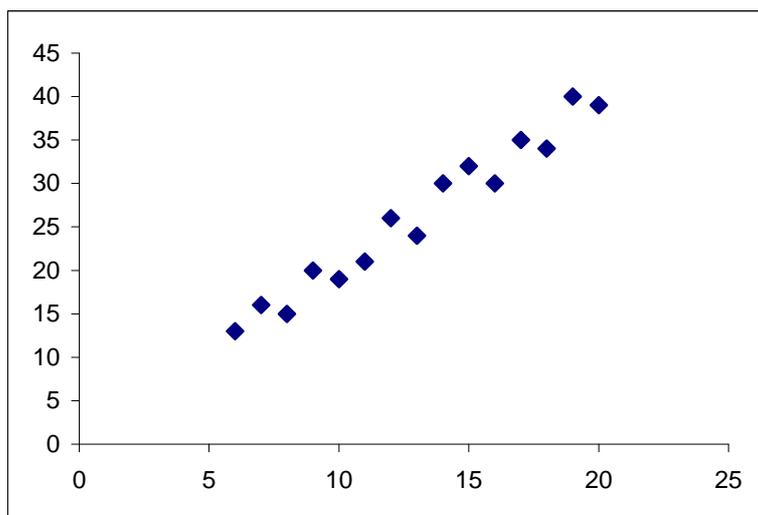
ЗАДАНИЕ.

Компанию по прокату автомобилей интересует зависимость между пробегом автомобилей X и стоимостью ежемесячного технического обслуживания Y . Для выяснения характера этой связи было отобрано 15 автомобилей. Постройте график исходных данных и определите по нему характер зависимости. Рассчитайте выборочный коэффициент линейной корреляции Пирсона, проверьте его значимость при $\alpha = 0,05$. Постройте уравнение регрессии и дайте интерпретацию полученных результатов.

X	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Y	13	16	15	20	19	21	26	24	30	32	30	35	34	40	39

РЕШЕНИЕ.

Построим график исходных данных:



По графику видно, что зависимость прямая, линейная.

Выборочный коэффициент линейной корреляции Пирсона найдем по формуле:

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

Заполним таблицу для вычисления этого коэффициента.

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{1}{15} 195 = 13,$$

$$\bar{y} = \frac{1}{n} \sum y_i = \frac{1}{15} 394 \approx 26,267.$$

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
6	13	-7	-13,267	92,867	49,000	176,004
7	16	-6	-10,267	61,600	36,000	105,404
8	15	-5	-11,267	56,333	25,000	126,938
9	20	-4	-6,267	25,067	16,000	39,271
10	19	-3	-7,267	21,800	9,000	52,804
11	21	-2	-5,267	10,533	4,000	27,738
12	26	-1	-0,267	0,267	1,000	0,071
13	24	0	-2,267	0,000	0,000	5,138
14	30	1	3,733	3,733	1,000	13,938
15	32	2	5,733	11,467	4,000	32,871
16	30	3	3,733	11,200	9,000	13,938
17	35	4	8,733	34,933	16,000	76,271
18	34	5	7,733	38,667	25,000	59,804
19	40	6	13,733	82,400	36,000	188,604
20	39	7	12,733	89,133	49,000	162,138

Сумма 195 394 0,00 0,00 540 280 1080,933

$$\text{Тогда } r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{540}{\sqrt{280 \cdot 1080,933}} \approx 0,982.$$

Введем нулевую гипотезу $H_0 : r = 0$. Проверим эту гипотезу об отсутствии корреляционной зависимости (о незначимости коэффициента корреляции). Вычислим

$$\text{значение критерия } T_{\text{набл}} = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}} = \frac{0,982 \cdot \sqrt{13}}{\sqrt{1-0,982^2}} \approx 18,75.$$

Найдем критическую точку по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = n - 2 = 13$, получаем $t_{кр} = 2,16$. Так как $|T_{\text{набл}}| = 18,75 > 2,16 = t_{кр}$, следует отвергнуть нулевую гипотезу $H_0 : r = 0$, то есть корреляционная зависимость есть (существенна), коэффициент корреляции статистически значим.

Уравнение регрессии Y на X имеет вид $\bar{y}_x - \bar{y} = r_{xy} \frac{\sigma_y}{\sigma_x} (x - \bar{x})$. Найдем средние квадратические отклонения.

$$\text{Выборочная дисперсия } \bar{D}_x = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{15} 280 \approx 18,667$$

$$\text{Выборочное квадратическое отклонение } \sigma_x = \sqrt{\bar{D}_x} \approx 4,32$$

Выборочная дисперсия $\bar{D}_y = \frac{1}{n} \sum (y_i - \bar{y})^2 = \frac{1}{15} 1080,933 \approx 72,062$

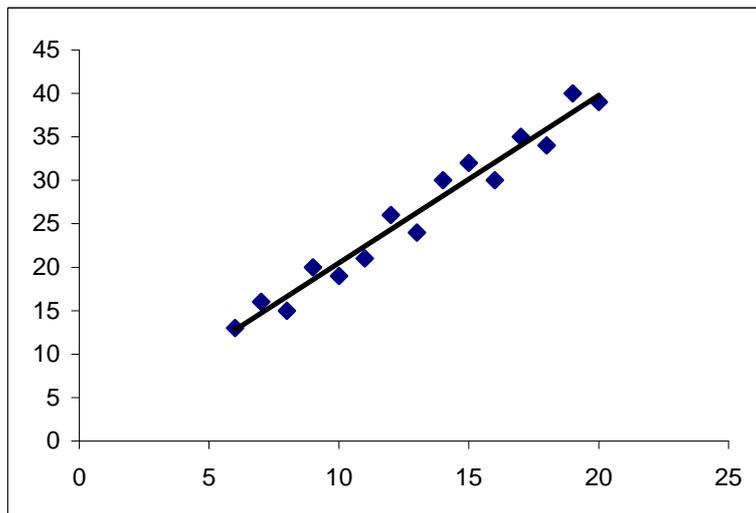
Выборочное квадратическое отклонение $\sigma_y = \sqrt{\bar{D}_y} \approx 8,489$

Подставляем все величины:

$$\bar{y}_x - 26,267 = 0,982 \frac{8,489}{4,32} (x - 13)$$

$$\bar{y}_x = 1,93x + 1,181$$

Построим линию регрессии на графике:



Таким образом, наблюдается очень тесная прямая связь между величиной пробега автомобиля и стоимостью ежемесячного технического обслуживания, которая выражается уравнением регрессии $\bar{y}_x = 1,93x + 1,181$. Через каждый год пробега стоимость обслуживания возрастает в среднем на 1,93.