

Содержание

[Пример](#)

[Условие](#)

[Задача](#)

[Проверка значимости отдельных коэффициентов уравнения](#)

[Построение интервальных оценок коэффициентов](#)

[Интервальная оценка для уравнения регрессии \$y\$ \(размер издержек\)
при \$x_0 = 10\$](#)

Пример

→ Условие

В таблице представлены данные об издержках обращения и изменения товарооборота.

у издержки обращения, тыс. руб	х товарооборот, тыс. руб
1,1	4
1,3	5
1,4	6
1,1	3
1,9	8
1,7	9
1,4	6
1,2	4
1,9	6
1,9	9

Предполагаем, что генеральное уравнение регрессии имеет вид:

$$\hat{y} = a + bx + \varepsilon$$

Точечные оценки коэффициентов регрессии уже были получены, известно, что:

- $a = 0,68$
- $b = 0,135$

Таким образом:

$$\hat{y} = 0,68 + 0,135x$$

→ Задача

Важно знать, насколько точны эти оценки. Для этого вычислим стандартные ошибки коэффициентов и построим доверительные интервалы для коэффициентов.

Проверка значимости отдельных коэффициентов уравнения

Рассчитаем несмещённую оценку остаточной дисперсии по формуле:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum (y_i - \hat{y}_i)^2 = \frac{D[X]_{\text{ост}}}{n-2} = \frac{0,26}{8} = 0,0325$$

	y	x	x ²	xy	\hat{y}	y- \hat{y}	(y- \hat{y}) ²	(\bar{y} - \hat{y}) ²	(\bar{y} - \hat{y}) ²
	1,1	4	16	4,4	1,22	-0,12	0,01	0,15	0,07
	1,3	5	25	6,5	1,36	-0,06	0,00	0,04	0,02
	1,4	6	36	8,4	1,49	-0,09	0,01	0,01	0,00
	1,1	3	9	3,3	1,09	0,01	0,00	0,15	0,16
	1,9	8	64	15,2	1,76	0,14	0,02	0,17	0,07
	1,7	9	81	15,3	1,90	-0,19	0,04	0,04	0,16
	1,4	6	36	8,4	1,49	-0,09	0,01	0,01	0,00
	1,2	4	16	4,8	1,22	-0,02	0,00	0,08	0,07
	1,9	6	36	11,4	1,49	0,41	0,17	0,17	0,00
	1,9	9	81	17,1	1,90	0,01	0,00	0,17	0,16
Итого	14,90	60,00	400	94,8	14,90	0,00	0,26	0,99	0,73
Среднее значение	1,49	6,00	40	9,48					
СКО	0,31	2,00							
Дисперсия	0,10	4,00							

$$\hat{\sigma}^2 = 0,0325$$

Рассчитаем стандартные ошибки коэффициентов а и в по формулам:

$$\hat{\sigma}_a = \sqrt{\frac{\hat{\sigma}^2 \sum x_i^2}{n \sum_{i=1}^n x_i^2 - (\sum x_i)^2}} = \sqrt{\frac{0,0325 \times 400}{10 \times 400 - 60^2}} = 0,18$$

$$\hat{\sigma}_b = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = \sqrt{\frac{0,0325}{40}} = 0,0285$$

	y	x	x ²	xy	\hat{y}	y- \hat{y}	(y- \hat{y}) ²	(y- \bar{y}) ²	(\bar{y} - \hat{y}) ²	(x- \bar{x}) ²
	1,1	4	16	4,4	1,22	-0,12	0,01	0,15	0,07	4
	1,3	5	25	6,5	1,36	-0,06	0,00	0,04	0,02	1
	1,4	6	36	8,4	1,49	-0,09	0,01	0,01	0,00	0
	1,1	3	9	3,3	1,09	0,01	0,00	0,15	0,16	9
	1,9	8	64	15,2	1,76	0,14	0,02	0,17	0,07	4
	1,7	9	81	15,3	1,90	-0,19	0,04	0,04	0,16	9
	1,4	6	36	8,4	1,49	-0,09	0,01	0,01	0,00	0
	1,2	4	16	4,8	1,22	-0,02	0,00	0,08	0,07	4
	1,9	6	36	11,4	1,49	0,41	0,17	0,17	0,00	0
	1,9	9	81	17,1	1,90	0,01	0,00	0,17	0,16	9
Итого	14,90	60	400	94,8	14,90	0,00	0,26	0,99	0,73	40
Среднее значение	1,49	6,00	40	9,48						
СКО	0,31	2,00								
Дисперсия	0,10	4,00								

Для проверки гипотезы H₀: a=0 рассчитаем статистику:

$$\hat{y} = \underline{0,68} + 0,135x$$

$$\hat{\sigma}_a = 0,18$$

$$t_{\text{набл}} = \frac{a}{\hat{\sigma}_a} = \frac{0,68}{0,18} = 3,77$$

По таблице находим критическое значение распределения Стьюдента, где вероятность (уровень значимости) равна 0,05 и число степеней свободы $n-2=10-2=8$:

Число степеней свободы n	Уровень значимости		
	0,1	0,05	0,01
1	6,3137	12,7062	63,6559
2	2,9200	4,3027	9,9250
3	2,3534	3,1824	5,8408
4	2,1318	2,7765	4,6041
5	2,0150	2,5706	4,0321
6	1,9432	2,4469	3,7074
7	1,8946	2,3646	3,4995
8	1,8595	2,3060	3,3554
9	1,8331	2,2622	3,2498
10	1,8125	2,2281	3,1693
11	1,7959	2,2010	3,1058
12	1,7823	2,1788	3,0545

$$t_{\text{крит}} = 2,3$$

Так как $t_{\text{набл}} = 3,77 > t_{\text{кр}} = 2,31$, то гипотеза H_0 отвергается и коэффициент a считается значимым.

Для проверки гипотезы $H_0: b=0$ рассчитаем статистику:

$$\hat{y} = 0,68 + 0,135x$$

$$\hat{\sigma}_b = 0,0285$$

$$t_{\text{набл}} = \frac{b}{\hat{\sigma}_b} = \frac{0,135}{0,0285} = 4,736$$

По таблице находим критическое значение распределения Стьюдента, где вероятность (уровень значимости) равна 0,05 и число степеней свободы $n-2=10-2=8$:

Число степеней свободы n	Уровень значимости		
	0,1	0,05	0,01
1	6,3137	12,7062	63,6559
2	2,9200	4,3027	9,9250
3	2,3534	3,1824	5,8408
4	2,1318	2,7765	4,6041
5	2,0150	2,5706	4,0321
6	1,9432	2,4469	3,7074
7	1,8946	2,3646	3,4995
8	1,8595	2,3060	3,3554
9	1,8331	2,2622	3,2498
10	1,8125	2,2281	3,1693
11	1,7959	2,2010	3,1058
12	1,7823	2,1788	3,0545

$$t_{\text{крит}} = 2,3$$

Так как $t_{\text{набл}} = 4,736 > t_{\text{кр}} = 2,31$, то гипотеза H_0 отвергается и коэффициент b считается значимым, то есть товарооборот значимо влияет на издержки обращения.

Построение интервальных оценок коэффициентов

Доверительным интервалом называется такой интервал, относительно которого можно с заранее выбранной вероятностью утверждать, что он содержит значения прогнозируемого показателя.

Интервальная оценка для параметра a :

$$\hat{a} - t_{\text{кр}} \hat{\sigma}_a \leq a \leq \hat{a} + t_{\text{кр}} \hat{\sigma}_a$$

где $t_{\text{кр}}$ определяется из таблицы распределения Стьюдента для уровня значимости α и числа степеней свободы $v=n-2$.

Обычно уровень значимости берут равным 0,05 (доверительный интервал 95%) или 0,01 (доверительный интервал 99%).

Аналогично определяется интервальная оценка для коэффициента b :

$$\hat{b} - t_{\text{кр}} \hat{\sigma}_b \leq b \leq \hat{b} + t_{\text{кр}} \hat{\sigma}_b$$

Найдём 95% доверительные интервалы для каждого коэффициента регрессии.

Интервальная оценка для параметра а:

$$\hat{a} - t_{кр} \hat{\sigma}_a \leq a \leq \hat{a} + t_{кр} \hat{\sigma}_a$$

$$\hat{y} = \underline{0,68} + 0,135x$$

$$\hat{\sigma}_a = 0,18$$

$$0,68 - 2,31 \times 0,18 \leq a \leq 0,68 + 2,31 \times 0,18$$

$$0,264 \leq a \leq 1,096$$

Интервальная оценка для параметра b:

$$\hat{b} - t_{кр} \hat{\sigma}_b \leq b \leq \hat{b} + t_{кр} \hat{\sigma}_b$$

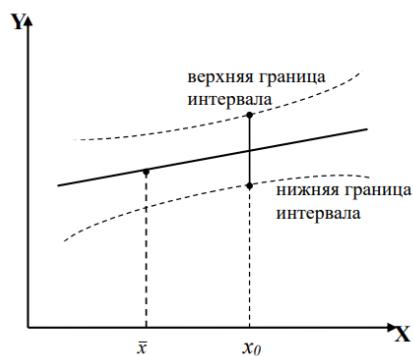
$$\hat{y} = 0,68 + \underline{0,135}x$$

$$\hat{\sigma}_b = 0,0285$$

$$0,135 - 2,31 \times 0,0285 \leq b \leq 0,135 + 2,31 \times 0,0285$$

$$0,069 \leq b \leq 0,201$$

Доверительный интервал имеет наименьшую величину, когда $x_0 = \bar{x}$, а по мере удаления x_0 от \bar{x} ширина доверительного интервала увеличивается, и точность оценки у снижается.



Интервальная оценка для уравнения регрессии y (размер издержек) при $x_0 = 10$

Точечная оценка:

$$\hat{y} = 0,68 + 0,135x = 0,68 + 0,135 \times 10 = 2,03$$

Интервальная оценка для $x_0 = 10$:

$$(a + bx) - t_{кр} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \leq y \leq (a + bx) + t_{кр} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$2,03 - 2,31 \times 0,1803 \times \sqrt{\frac{1}{10} + \frac{(10 - 6)^2}{40}} \leq y \leq 2,03 + 2,31 \times 0,1803 \times \sqrt{\frac{1}{10} + \frac{(10 - 6)^2}{40}}$$

$$1,736 \leq y \leq 2,324$$

Получено, что доверительный интервал для прогноза издержек обращения при товарообороте $x = 10$ тыс. руб. находится в пределах от 1,736 тыс. руб. до 2,354 тыс. руб. с 95% уровнем надёжности.