

Задача 4.2. По 20 предприятиям региона имеются данные (табл. 4) показателей «Выработка продукции на одного работника» (y , тыс. руб.), «Ввод в действие новых основных фондов» (x_1 , % от стоимости фондов на конец года), «Удельный вес рабочих высокой квалификации в общей численности рабочих» (x_2 , %).

Таблица 4

Исходные данные

№ предприятия	Выработка продукции на одного работника (y) тыс. руб.	Ввод в действие новых основных фондов (x_1) %	Удельный вес рабочих высокой квалификации (x_2) %
1	7,0	3,9	10
2	7,0	3,9	14
3	7,0	3,7	15
4	7,0	4,0	16
5	7,0	3,8	17
6	7,0	4,8	19
7	8,0	5,4	19
8	8,0	4,4	20
9	8,0	5,3	20
10	10,0	6,8	20
11	9,0	6,0	21
12	11,0	6,4	22
13	9,0	6,8	22
14	11,0	7,2	5
15	12,0	8,0	28
16	12,0	8,2	29
17	12,0	8,1	30
18	12,0	8,5	31
19	14,0	9,6	32
20	14,0	9,0	36

Требуется:

- 1) построить уравнение множественной линейной регрессии;
- 2) записать модель множественной линейной регрессии;
- 3) оценить качество уравнения регрессии.

Решение.

Объем выборки $n = 20$, число независимых переменных (факторов) $m = 2$.

В качестве программного средства для решения задачи воспользуемся программой «Анализ данных» в Excel, инструмент «Регрессия». Результаты расчетов представлены в табл. 2.

Таблица 2

ВЫВОД ИТОГОВ

<i>Регрессионная статистика</i>	
Множественный R	0,970128
R-квадрат	0,941148
Нормированный R-квадрат	0,934224
Стандартная ошибка	0,630417
Наблюдения	20

Дисперсионный анализ

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	2	108,0438	54,02189	135,9297	3,49E-11
Остаток	17	6,756229	0,397425		
Итого	19	114,8			

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение (b_0)	1,96175	0,488092	4,019219	0,000889	0,931965	2,991535
x_1 (b_1)	1,197856	0,113551	10,54905	7,03E-09	0,958284	1,437428
x_2 (b_2)	0,010494	0,028246	0,37152	0,71484	-0,0491	0,070087

1) *Уравнение множественной линейной регрессии* имеет вид

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 .$$

Используя результаты расчетов (см. табл. 2) в данном случае можно записать

$$\hat{y} = 1,962 + 1,198 \cdot x_1 + 0,010 \cdot x_2 .$$

Коэффициент регрессии $b_1 = 1,198$ показывает, что с увеличением величины новых основных фондов на 1 %. Выработка продукции на одного работника увеличивается на 1,198 тыс. руб., при условии, что Удельный вес рабочих высокой квалификации останется на прежнем уровне.

Коэффициент регрессии $b_2 = 0,01$ показывает, что с увеличением Удельного веса рабочих высокой квалификации на 1 %. Выработка продукции на одного работника увеличивается 0,01 тыс. руб., при условии, что Выработка продукции на одного работника не изменится.

Свободный член уравнения регрессии $b_0 = 1,962$ нельзя интерпретировать как начальный уровень Выработки продукции на одного работника, поскольку показатели «Ввод в действие новых основных фондов», «Удельный вес рабочих высокой квалификации не могут одновременно принимать нулевое значение.

2) *Множественная линейная регрессионная модель* зависимости объема реализации продукции от всех факторов имеет вид

$$y = 1,962 + 1,198 \cdot x_1 + 0,010 \cdot x_2 + \varepsilon$$

3) Для проверки качества уравнения регрессии составим расчетную таблицу (см. табл. 3) (расчеты выполнены в Excel).

Таблица 3

Расчетная таблица

% п/п	y_i	x_{i1}	x_{i2}	\hat{y}_i	$\left \frac{y_i - \hat{y}_i}{y_i} \right $
1	7	3,9	10	6,74	0,037
2	7	3,9	14	6,78	0,031
3	7	3,7	15	6,55	0,064
4	7	4	16	6,92	0,011
5	7	3,8	17	6,69	0,044
6	7	4,8	19	7,91	0,130
7	8	5,4	19	8,63	0,079
8	8	4,4	20	7,44	0,070
9	8	5,3	20	8,52	0,065
10	10	6,8	20	10,32	0,032
11	9	6	21	9,37	0,041
12	11	6,4	22	9,86	0,104
13	9	6,8	22	10,34	0,149
14	11	7,2	5	10,64	0,033
15	12	8	28	11,84	0,013
16	12	8,2	29	12,09	0,007
17	12	8,1	30	11,98	0,002
18	12	8,5	31	12,47	0,039
19	14	9,6	32	13,80	0,015
20	14	9	36	13,12	0,063
Сумма	192	123,8	426	192	1,029
Среднее	9,6	6,19	21,3	9,6	0,051

а) Найдем среднюю относительную ошибку аппроксимации

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\% = \frac{1,029}{20} \cdot 100\% = 5,1\% .$$

Так как $\bar{A} < 10\%$, **уравнение имеет высокую точность.**

б) Проверим статистическую значимость уравнения регрессии в целом с помощью критерия Фишера. Расчетные значения взяты из табл. 2.

$$F_{\text{расч}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (\hat{y}_i - y_i)^2} \cdot \frac{n - m - 1}{m} = 135,93 .$$

Табличное значение критерия Фишера с $df_1 = m = 2$ и $df_2 = n - m - 1 = 20 - 2 - 1 = 17$ степенями свободы при уровне значимости $\alpha = 0,05$ найдем с помощью встроенной функции Excel «FРАСПОБР». $F_{\text{табл}} = 3,59$.

Поскольку $F_{\text{расч}} > F_{\text{табл}}$, **уравнение регрессии в целом статистически значимо.**

в) Проверим статистическую значимость параметров уравнения регрессии с помощью критерия Стьюдента. Расчетные значения взяты из табл. 2.

Расчетные значения критерия равны

$$t_{b_0} = \left| \frac{b_0}{m_{b_0}} \right| = 0,443 , t_{b_1} = \left| \frac{b_1}{m_{b_1}} \right| = 2,346 , t_{b_2} = \left| \frac{b_2}{m_{b_2}} \right| = 4,691 , t_{b_3} = \left| \frac{b_3}{m_{b_3}} \right| = 0,423 ,$$

где средние квадратические ошибки параметров

$$m_{b_0} = 0,488 , m_{b_1} = 0,114 , m_{b_2} = 0,028$$

Табличное значение t-критерия Стьюдента при уровне значимости $\alpha = 0,05$ и числе степеней свободы $df = n - m - 1 = 20 - 2 - 1 = 17$ составляет $t_{\text{табл}} = 2,11$. (табличное значение найдено с помощью встроенной функции Excel «СТЮДРАСПОБР»).

Так как $t_{b_0} > t_{\text{табл}}$, **параметр b_0 статистически значим;**

$t_{b_1} > t_{\text{табл}}$, **параметр b_1 статистически значим.**

$t_{b_2} < t_{\text{табл}}$, **параметр b_2 статистически не значим.**

г) Интервальные оценки (95%-е доверительные интервалы) параметров уравнения регрессии приведены в табл. 2:

Интервальная оценка b_0 : (0,932 ; 2,992) .

Интервальная оценка b_1 : (0,958 ; 1,437) .

Интервальная оценка b_2 : (-0,049 ; 0,070) .

Границы доверительных интервалов для статистически незначимых параметров имеют разные знаки.