

Домашнее задание №1

Оценивание коэффициентов многофакторной линейной регрессии

Пример

Регрессионный анализ фондоотдачи

На основе линейной регрессионной модели исследовать зависимость фондоотдачи в процентах на единицу основных производственных фондов (ОПФ) (y) от среднечасовой производительности вращающихся печей (x_1) и удельного веса активной части ОПФ (x_2).

В таблице 1 приводятся данные $n = 15$ цементных заводов страны.

Таблица 1

Исходные данные

№ п/п	Фондоотдача, y	Среднечасовая производительность печей, x_1	Удельный вес активной части ОПФ (%), x_2
1	26	37	39
2	33	33	40
3	24	15	35
4	29	36	48
5	42	26	53
6	24	24	42
7	52	15	54
8	56	33	54
9	26	44	50
10	45	34	53
11	27	63	46
12	54	8	50
13	34	44	43
14	48	43	55
15	45	31	51

Решение

Для определения оценок $b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$ согласно

$$b = \hat{\beta} = (X^T X)^{-1} (X^T Y)$$

найдем предварительную симметричную матрицу $X^T X$, которая имеет вид

$$X^T X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} \\ \sum_i x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_i x_{i1} x_{i2} \\ \sum_i x_{i2} & \sum_i x_{i1} x_{i2} & \sum_i x_{i2}^2 \end{pmatrix}$$

и равна для нашего примера

$$X^T X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 37 & 33 & \dots & 31 \\ 39 & 40 & \dots & 51 \end{pmatrix} \begin{pmatrix} 1 & 37 & 39 \\ 1 & 33 & 40 \\ \dots & \dots & \dots \\ 1 & 31 & 51 \end{pmatrix} = \begin{pmatrix} 15 & 486 & 713 \\ 486 & 18486 & 23132 \\ 713 & 23132 & 34455 \end{pmatrix}.$$

вектор имеет вид

$$X^T Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1} y_i \\ \sum_{i=1}^n x_{i2} y_i \end{pmatrix}$$

и для нашего примера равен

$$X^T Y = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 32 & 33 & 15 & \dots & 43 & 31 \\ 39 & 40 & 35 & \dots & 55 & 51 \end{pmatrix} \begin{pmatrix} 26 \\ 33 \\ 24 \\ \dots \\ 48 \\ 45 \end{pmatrix} = \begin{pmatrix} 565 \\ 17513 \\ 27656 \end{pmatrix}.$$

Для получения обратной матрицы $(X^T X)^{-1}$ можно воспользоваться методом полного исключения переменных Жордана—Гаусса.

Таким образом, получена обратная матрица:

$$(X^T X)^{-1} = \begin{pmatrix} 4,4075412 & -0,0111708 & -0,0837088 \\ -0,0111708 & 0,0003748 & 0,0000204 \\ -0,0837088 & -0,0000204 & 0,001775 \end{pmatrix}.$$

Подставив найденный вектор $X^T Y$ и матрицу $(X^T X)^{-1}$ в выражение

$$b = \hat{\beta} = (X^T X)^{-1} (X^T Y)$$

найдем вектор оценок:

$$b = (X^T X)^{-1} X^T Y = \begin{pmatrix} 4,4075412 & -0,0111708 & -0,0837088 \\ -0,0111708 & 0,0003748 & 0,0000204 \\ -0,0837088 & -0,0000204 & 0,001775 \end{pmatrix} \begin{pmatrix} 565 \\ 17513 \\ 27656 \end{pmatrix} =$$

$$= \begin{pmatrix} 4,4075412 \cdot 565 - 0,0111708 \cdot 17513 - 0,0837088 \cdot 27656 \\ -0,0111708 \cdot 565 + 0,0003748 \cdot 17513 + 0,0000204 \cdot 27656 \\ -0,0837088 \cdot 565 - 0,0000204 \cdot 17513 + 0,001775 \cdot 27656 \end{pmatrix} = \begin{pmatrix} -20,4102 \\ -0,3118120 \\ 1,505803 \end{pmatrix}.$$

Таким образом, $b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} -20,4102 \\ -0,3118120 \\ 1,505803 \end{pmatrix}$

И оценка уравнения регрессии имеет вид

$$\hat{y} = -20,4102 - 0,3118120x_1 + 1,505803x_2.$$

Для проверки значимости уравнения регрессии согласно выражению

$$F_{набл} = \frac{\frac{1}{k+1} Q_R}{\frac{1}{n-k-1} Q_{ост}}$$

нужно найти $Q_{ост} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ и $Q_R = \sum_{i=1}^n \hat{y}_i^2$.

Составим вспомогательную таблицу 2

Таблица 2

Вспомогательная таблица

i	y_i	x_{1i}	x_{2i}	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	26	37	39	6,779 073	0,606 954 7
2	33	33	40	29,532 124	12,026 163
3	24	15	35	27,615 725	13,073 467
4	29	36	48	40,643 112	135,562 05
5	42	26	53	51,290 247	86,308 689
6	24	24	42	35,350 038	128,823 36
7	52	15	54	56,225 982	17,858 923
8	56	33	54	50,613 366	29,015 825
9	26	44	50	41,160 222	229,832 33
10	45	34	53	48,795 751	14,407 725
11	27	63	46	29,212 582	4,895 519 1
12	54	8	50	52,385 454	2,606 758 7
13	34	44	43	30,619 601	11,427 097
14	48	43	55	49,001 049	1,002 099 1
15	45	31	51	49,719 581	2,956 958 8
Итого	565	486	713		690,403 85

ОТКУДА СЛЕДУЕТ, ЧТО $Q_{OCT} = 690,40385$.

$$\hat{s}^2 = \frac{1}{n-3} Q_{OCT} = \frac{1}{12} 690,40385 = 57,533654 ; \hat{s} = 7,5850941 \quad Q_R = \sum_{i=1}^n \hat{y}_i = 26801,319$$

Проверим на уровне значимости $\alpha = 0,05$ ЗНАЧИМОСТЬ УРАВНЕНИЯ РЕГРЕССИИ.

$$H_0: \beta = 0.$$

$$F_{набл} = \frac{\frac{1}{k+1} Q_R}{\frac{1}{n-k-1} Q_{OCT}} = \frac{\frac{1}{3} 26801,319}{\frac{1}{12} 690,40385} = 155,27907.$$

По таблице F -распределения для $\alpha = 0,05$ и чисел степеней свободы $\nu_1 = 3$ и $\nu_2 = 12$ найдем критическое значение $F_{кр}(0,05;3;12)$. Так как $F_{набл} > F_{кр}$ гипотеза $H_0: \beta = 0$ отвергается, т. е. хотя бы один элемент вектора $\beta = (\beta_0, \beta_1, \beta_2)^T$ не равен нулю.

Перед проверкой значимости отдельных коэффициентов регрессии найдем оценку ковариационной матрицы вектора b . Для этого достаточно элементы обратной матрицы $(X^T X)^{-1}$ умножить на $\hat{s}^2 = 58,533654$. Тогда будем иметь:

$$\hat{V}(b) = \hat{s}^2 (X^T X)^{-1} = \begin{pmatrix} 253,582 & -0,643 & -4,816 \\ 0,643 & 0,0216 & -0,0012 \\ -4,816 & -0,0012 & 0,1021 \end{pmatrix}.$$

Из статистического смысла ковариационной матрицы следует, что оценки дисперсии коэффициентов уравнения регрессии b_0, b_1, b_2 соответственно равны:

$$\hat{s}_{b_0}^2 = 253,582, \quad \hat{s}_{b_1}^2 = 0,0216,$$

Проверим значимость коэффициента β_1 , т. е. гипотезу $H_0: \beta = 0$.

$$t_{b_1} = \frac{b_1}{\hat{s}_{b_1}} = \frac{(-0,311812)}{(0,0216)^{1/2}} = -2,122.$$

По таблице t -распределения для значений $\alpha = 0,05$ и $\nu_2 = 12$ $t_{кр} = 2,179$.

Так как $|t_{b_1}| < t_{кр}$, гипотеза о том, что $\beta_1 = 0$, не отвергается, т. е. β_1 незначим.

Проверим теперь гипотезу $H_0: \beta_2 = 0$.

$$t_{b_2} = \frac{b_2}{\hat{s}_{b_2}} = \frac{1,515803}{(0,1021)^{1/2}} = 4,713$$

Так как $t_{b_2} > t_{кр} = 2,18$, гипотеза $H_0: \beta_2 = 0$ отвергается, т. е. β_2 не равен нулю ($\beta_2 \neq 0$).

Коэффициент регрессии при x_2 показывает, что при росте удельного веса активной части на единицу ОПФ (%) фондоотдача в среднем увеличивается на 1,416 09 единиц.

Найдем теперь с доверительной вероятностью $\gamma = 0,96$ интервальную оценку для коэффициента регрессии β_2 .

$$\beta_2 \in [1,41609 \pm 2,16 \cdot \sqrt{0,106}],$$

где $t_\gamma = 2,16$ находим по таблице t-распределения при $\alpha = 1 - \gamma = 0,05$ и $\nu = n - k - 1 = 14 - 2 = 13$. Отсюда

$$1,41609 - 0,70324 \leq \beta_1 \leq 1,41609 + 0,70324;$$

$$0,71285 \leq \beta_1 \leq 2,11933$$

Доверительную границу для условного математического ожидания y найдем в точке, определяемой вектором начальных условий: $X^0 = \begin{pmatrix} 1 \\ x_2^0 \end{pmatrix} = \begin{pmatrix} 1 \\ 50 \end{pmatrix}$ с

надежностью $\gamma = 0,95$.

Предварительно найдем матричное произведение

$$\begin{aligned} (X^0)^T (X^T X)^{-1} X^0 &= (1 \ 50) \begin{pmatrix} 4,0746515 & -0,0843188 \\ -0,0843188 & 0,0017738 \end{pmatrix} \begin{pmatrix} 1 \\ 50 \end{pmatrix} = \\ &= (-0,1413185 \ 0,0043712) \begin{pmatrix} 1 \\ 50 \end{pmatrix} = 0,0772415 \end{aligned}$$

Интервальная оценка для $\tilde{y}$ равна:

$$\tilde{y} \in [(-29,7595 + 1,41609) \pm 2,16 \cdot 7,732 \sqrt{0,0772415}].$$

Отсюда

$$\tilde{y} \in [41,0449 \pm 4,6416]$$

и окончательно

$$36,4033 \leq \tilde{y} \leq 45,6865.$$

Таким образом, с доверительной вероятностью $\gamma = 0,95$ мы можем гарантировать, что при значении $x_2^0 = 50$ удельный вес активной части ОПФ завода будет находиться в интервале от 36,403 до 45,686%.

Таблица 3

Варианты заданий

Номер предприятия	Стоимость промышленно-производственных основных фондов, тыс. руб.	Валовая продукция в оптовых ценах предприятия, тыс. руб.	Среднесписочная численность промышленно-производственного персонала, чел.	Среднесписочная численность рабочих, чел.
1	4999	5349	420	331
2	6929	6882	553	486
3	6902	7046	570	498
4	10097	7248	883	789

5	8097	5256	433	359
6	11116	14090	839	724
7	4880	3525	933	821
8	7355	5431	526	428
9	10066	7680	676	607
10	7884	8226	684	619
11	4360	4980	400	318
12	8380	6789	520	419
13	7956	5430	856	327
14	4624	8935	756	624
15	3650	1243	385	129
16	5648	6548	850	758
17	6457	1254	654	546
18	2504	4789	564	410
19	1276	2865	125	108
20	9315	8754	825	564
21	1424	2569	564	420
22	10556	9856	987	876
23	8564	5424	253	213
24	5233	2565	657	540
25	1274	2458	856	756
26	6459	6658	874	712
27	9854	8821	958	814
28	6588	5502	758	615
29	8877	8765	458	314
30	6546	5245	658	512

Выбор варианта

Студент с последней цифрой i в зачетке выбирает из таблицы 20 строк от i -й до $i+19$ -й

Если $i=0$, то от 10-й до 29-й

Многофакторная линейная регрессия(на компьютере)

*Регрессионный анализ заключается в определении аналитического выражения связи зависимой случайной величины Y (называемой также *результативным признаком*) с независимыми случайными величинами $X_1, X_2, \dots, X_m$ (называемыми также *факторами*).*

Форма связи результативного признака Y с факторами $X_1, X_2, \dots, X_m$ получила название *уравнения регрессии*. В зависимости от типа выбранного уравнения различают *линейную* и *нелинейную* регрессию (в последнем случае возможно дальнейшее уточнение: квадратичная, экспоненциальная, логарифмическая и т. д.).

В зависимости от числа взаимосвязанных признаков различают *парную* и *множественную* регрессию. Если исследуется связь между двумя признаками (результативным и факторным), то регрессия называется *парной*, если между тремя и более признаками - *множественной (многофакторной)* регрессией.

При изучении регрессии следует придерживаться определенной последовательности этапов:

1. Задание аналитической формы уравнения регрессии и определение параметров регрессии.

2. Определение в регрессии степени стохастической взаимосвязи результативного признака и факторов, проверка общего качества уравнения регрессии.

3. Проверка статистической значимости каждого коэффициента уравнения регрессии и определение их доверительных интервалов.

Основное содержание выделенных этапов рассмотрим на примере множественной линейной регрессии, реализованной в режиме «Регрессия» надстройки **Пакет анализа** Microsoft Excel.

Справочная информация по технологии работы

Режим работы «Регрессия» служит для расчета параметров уравнения *линейной* регрессии и проверки его адекватности исследуемому процессу

В диалоговом окне данного режима задаются следующие параметры:

1. *Входной интервал Y* — вводится ссылка на ячейки, содержащие данные по результативному признаку. Диапазон должен состоять из одного столбца.

2. *Входной интервал X* — вводится ссылка на ячейки, содержащие факторные признаки. Максимальное число входных диапазонов (столбцов) равно 16.

3. *Метки в первой строке/Метки в первом столбце* .

4. *Уровень надежности* — установите данный флажок в активное состояние, если в поле, расположенное напротив флажка, необходимо ввести уровень надежности, отличный от уровня 95 %, применяемого по умолчанию. Установленный уровень надежности используется для проверки значимости коэффициента детерминации R^2 и коэффициентов регрессии a_i .

Примечание. При неактивном флажке *Уровень надежности* в таблице параметров уравнения регрессии генерируются две одинаковые пары столбцов для границ доверительных интервалов.

5. *Константа-ноль* — установите данный флажок в активное состояние, если требуется, чтобы линия регрессии прошла через начало координат (т. е. $a_0 = 0$).

6. *Выходной интервал/Новый рабочий лист/Новая рабочая книга*

7. *Остатки* — установите данный флажок в активное состояние, если требуется включить в выходной диапазон столбец остатков .

8. *Стандартизованные остатки* — установите данный флажок в активное состояние, если требуется включить в выходной диапазон столбец стандартизованных остатков.

9. *График остатков* — установите данный флажок в активное состояние, если требуется вывести на рабочий лист точечные графики зависимости остатков от факторных признаков x_i .

10. *График подбора* - установите данный флажок в активное состояние, если требуется вывести на рабочий лист точечные графики зависимости теоретических результативных значений $\hat{y}$ от факторных признаков x_i .

11. *График нормальной вероятности* — установите данный флажок в активное состояние, если требуется вывести на рабочий лист точечный график

зависимости наблюдаемых значений y от автоматически формируемых интервалов персентилей. График строится на основе генерируемой таблицы «Вывод вероятности»

Пример

Данные о прибыли предприятий Y , величине оборотных средств X_1 и стоимости фондов X_2 приведены в табл. 1, сформированной на рабочем листе Microsoft Excel.

Таблица 4

Исходные данные

	В	С	Д	Е
	Номер предприятия	Прибыль Y , млн руб.	Величина оборотных средств X_1 млн руб.	Стоимость основных фондов X_2 , млн руб.
	1	188	129	510
	2	78	64	190
	3	93	69	240
	4	152	87	470
	5	55	47	110
	6	161	102	420

По представленным данным необходимо определить параметры уравнения линейной регрессии и провести его анализ.

Для решения задачи используем режим работы «Регрессия». Значения параметров, установленных в одноименном диалоговом окне, а рассчитанные в данном режиме показатели - в табл. 5- 9.

Исходные данные		
	В	С
11	ВЫВОД ИТОГОВ	
12		
13	<i>Регрессионная статистика</i>	
14	Множественный R	0,997
15	R -квадрат	0,995
16	Нормированный R - квадрат	0,991
17	Стандартная ошибка	5,050
18	Наблюдения	6

В табл. 2 сгенерированы результаты по регрессионной статистике. Эти результаты соответствуют следующим статистическим показателям:

- *Множественный R* - коэффициенту корреляции R ;
- *R -квадрат* - коэффициенту детерминации R^2 ;
- *Стандартная ошибка* — остаточному стандартному отклонению

$$\sigma_0 = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - (m + 1)}};$$

- *Наблюдения* - числу наблюдений n .

В табл. 6 сгенерированы результаты дисперсионного анализа, которые используются для проверки значимости коэффициента детерминации R^2 .

Таблица 6

	В			Е		
	Дисперсионный анализ					
		df	SS	S		<i>значимость F</i>
	Регрессия		3962,33	981,16	273,74	0,0004
	Остаток	3	6,51	25,50		
	Итого	5	14038,83			

Столбцы табл. 6 имеют следующую интерпретацию:

1. Столбец df - число степеней свободы.

Для строки *Регрессия* число степеней свободы определяется количеством факторных признаков m в уравнении регрессии $k_{\phi} = m$.

Для строки *Остаток* число степеней свободы определяется числом наблюдений n и количеством переменных в уравнении регрессии $m + 1$:
 $k_{\phi} = n - (m + 1)$.

Для строки *Итого* число степеней свободы определяется суммой $k_y = k_{\phi} + k_o$.

2. Столбец SS - сумма квадратов отклонений.

Для строки *Регрессия* — это сумма квадратов отклонений теоретических данных от среднего:

$$SS_{\phi}^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2.$$

Для строки *Остаток* - это сумма квадратов отклонений эмпирических данных от теоретических:

$$SS_O^2 = \sum_{i=1}^n (y_i - \hat{y})^2.$$

Для строки *Итого* - это сумма квадратов отклонений эмпирических данных от среднего:

$$SS_Y^2 = \sum_{i=1}^n (y_i - \bar{y})^2 \text{ или } SS_Y^2 = SS_{\phi}^2 + SS_O^2.$$

3. Столбец *MS* - дисперсии, рассчитываемые по формуле

$$MS = \frac{SS}{df}.$$

Для строки *Регрессия* - это факторная дисперсия σ_{ϕ}^2 .

Для строки *Остаток* - это остаточная дисперсия σ_O^2 .

4. Столбец *F* - расчетное значение *F*-критерия Фишера F_p , вычисляемое по формуле

$$F_p = \frac{MS(\text{Регрессия})}{MS(\text{Остатки})}.$$

5. Столбец *Значимость F* — значение уровня значимости, соответствующее вычисленному значению F_p . Определяется с помощью функции

$$= \text{FPACП}(F_p; df(\text{регрессия}); df(\text{остаток})).$$

В табл. 7 сгенерированы значения коэффициентов регрессии a_i и их статистические оценки.

Оценки

	В	С	D	Е	FF
		<i>коэффициенты</i>	<i>стандартная ошибка</i>	<i>t-статистика</i>	<i>PP-значение</i>
27	У-пересечение	1,94	,63	0,25	0,82
28	Величина оборотных средств X_1 , млн руб.	0,69	,20	3,53	0,04
29	Стоимость основных фондов X_2 , млн руб.	0,20	0,04	5,75	0,01

Столбцы табл. 7 имеют следующую интерпретацию:

1. *Коэффициенты* - значения коэффициентов a_i .
2. *Стандартная ошибка* - стандартные ошибки коэффициентов a_i .
1. *t-статистика* - расчетные значения t -критерия, вычисляемые по формуле

$$t - \text{статистика} = \frac{\text{Коэффициенты}}{\text{Стандартная ошибка}}.$$

5. *P-значение* - значения уровней значимости, соответствующие вычисленным значениям t_p . Определяются с помощью функции

$$= \text{СТЫЮДРАСП}(t_p; n - m - 1).$$

6. *Нижние 95 %* и *Верхние 95 %* — соответственно нижние и верхние границы доверительных интервалов для коэффициентов регрессии a_i . Для нахождения границ доверительных интервалов спомощью функции =СТЫДРАСПОБР (α ; $n - m - 1$) рассчитывается критическое значение t -критерия, а затем по формулам

$$\text{Нижние 95\%} = \text{Коэффициент} - \text{Стандартная ошибка} \cdot t_{кр};$$

$$\text{Верхние 95\%} = \text{Коэффициент} + \text{Стандартная ошибка} \cdot t_{кр};$$

вычисляются соответственно нижние и верхние границы доверительных интервалов.

Таблица 8

Остатки				
	В	С	Д	Е
	ВЫВОД ОСТАТКА			
	<i>Наблюдение</i>	<i>Предсказанная прибыль Y, млн руб.</i>	<i>Остатки</i>	<i>Стандартные остатки</i>
	1	190,91	-2,91	-0,74
	2	80,98	-2,98	-0,76
	3	94,57	-1,57	-0,40
	4	153,62	-1,62	-0,42

	5	52,98	2,02	0,52
	6	153,93	7,07	1,81

В табл. 8 сгенерированы теоретические значения $\hat{y}_i$ результативного признака Y и значения остатков. Последние вычисляются как разность между эмпирическими $\hat{y}_i$ и теоретическими y_i значениями результативного признака Y .

Таблица 9

Вероятности		
	В	С
	ВЫВОД ВЕРОЯТНОСТИ	
	<i>Персентиль</i>	<i>Прибыль Y, млн руб.</i>
	8,33	55
	25	78
	41,67	93
	58,33	152
	75	161
	91,67	188

В табл. 9 сгенерированы интервалы персентилей и соответствующие им эмпирические значения y .

Перейдем к анализу сгенерированных таблиц.

Рассчитанные в табл. 7 (ячейки C27 - C29) коэффициенты регрессии a_i , позволяют построить уравнение, выражающее зависимость прибыли предприятий Y от величины оборотных средств $X1$ и стоимости основных фондов $X2$:

$$\hat{y} = -1,94 + 0,69x_1 + 0,20x_2.$$

Значение множественного коэффициента детерминации $R^2 = 0,995$ (ячейка C15 в табл. 5) показывает, что 99,5 % общей вариации результативного признака объясняется вариацией факторных признаков X_1 и X_2 . Значит, выбранные факторы существенно влияют на прибыль предприятий, что подтверждает правильность их включения в построенную модель.

Рассчитанный уровень значимости $\alpha_p = 0,0004 < 0,05$ (показатель *Значимость F* в табл. 6) подтверждает значимость R^2 .

Другой подход к проверке значимости R^2 (как это делалось во всех ранее рассмотренных режимах надстройки «Пакет анализа») основан на проверке попадания F_p (показатель *F* табл. 6) в критическую область $(F_{np,\alpha}^{kp}, +\infty)$. Для рассматриваемого примера $F_{np,\alpha}^{kp} = 9,55$, которое рассчитывается по формуле

$$= \text{ФРАСПОБР}(0,05; \text{C22}; \text{C23}),$$

где в ячейке C22 вычисляется число степеней свободы $k_\phi = m = 2$, а

в ячейке C23 - число степеней свободы $k_o = n - (m + 1) = 6 - (2 + 1) = 3$.

Так как $F_p = 273,74$ попадает в критический интервал $(9,55; +\infty)$, то гипотеза $H_0: R^2 = 0$ отвергается, т. е. коэффициент детерминации R^2 является значимым.

Показатель средней ошибки аппроксимации $\bar{\varepsilon} = 2,7\%$ также подтверждает достаточно высокую адекватность построенного уравнения. Данный показатель может быть рассчитан по формуле

$$\{=\text{СУММ}(\text{ABS}(\text{D36:D41})/(\text{C3:C8}))/\text{СЧЕТ}(\text{C3:C8})*100\},$$

где в массиве D36: D41 табл. 8 рассчитаны разности между эмпирическими и теоретическими значениями результативного признака.

Следующим этапом является проверка значимости коэффициентов регрессии: a_0 , a_1 и a_2 . Сравнивая попарно элементы массивов C27:C29 и D27:D29 (см. табл. 7), видим, что абсолютное значение свободного члена a_0

меньше, чем его стандартная ошибка. Таким образом, свободный член a_0 следует исключить из уравнения регрессии.

Стандартные ошибки коэффициентов a_1 и a_2 меньше своих стандартных ошибок. К тому же эти коэффициенты являются значимыми, о чем можно судить по значениям показателя *P-значение* в табл. 7, которые меньше заданного уровня значимости $\alpha = 0,05$.

Другой распространенный способ проверки значимости коэффициентов регрессии основан на проверке попадания t_p (показатель *t-статистика* в табл. 4) в критическую область $(-\infty, t_{лев, \alpha/2}^{кр}) \cup (t_{пр, \alpha/2}^{кр}, +\infty)$. В генерируемых таблицах режима не приводится значение $t_{кр}$ но его можно легко вычислить с помощью функции СТЫЮДРАСПОБР. Для рассматриваемого примера значение $|t_{кр}| = 3,18$, которое рассчитывается по формуле

$$=СТЫЮДРАСПОБР(0,05;6-2-1),$$

где 0,05 — заданный уровень значимости; 0 — число наблюдений;
2 — число факторов в уравнении регрессии; 1 — число свободных членов в уравнении регрессии.

Так как $t_p^{a_1} = 3,53$ и $t_p^{a_2} = 5,75$ попадают в критический интервал $(-\infty; -3,18) \cup (3,18; +\infty)$, то коэффициенты регрессии a_1 и a_2 являются значимыми.