МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Ухтинский государственный технический университет»
(УГТУ)

ПРОГРАММНЫЕ КОМПЛЕКСЫ ОБЩЕГО НАЗНАЧЕНИЯ

Методические указания

Ухта, УГТУ, 2014

УДК 681.3.016 (075.8)
ББК 32.973
Т 83

Туманова, О. Н.

Т 83 Программные комплексы общего назначения [Текст] : метод. указания / О. Н. Туманова. – Ухта : УГТУ, 2014. – 55 с.

Методические указания предназначены для студентов специальности «Менеджмент». Они соответствуют рабочей программе дисциплин «Программные комплексы общего назначения» и содержат теоретические сведения для выполнения контрольной работы, варианты заданий, пример их выполнения. Методические указания предназначены для студентов дневной и безотрывной форм обучения при выполнении контрольной работы по дисциплине «Программные комплексы общего назначения».

Содержание указаний соответствует рабочей программе дисциплины «Программные комплексы общего назначения».

УДК 681.3.016 (075.8)
ББК 32.973

Методические указания рассмотрены и одобрены кафедрой ПМИ (протокол №06 от 21.02.2014).

Рецензент и редактор: Ю. Г. Смирнов, зав. кафедрой прикладной математики и информатики УГТУ, доцент, канд. физ.-мат. наук.
Корректор и технический редактор: К. В. Коптяева.

В методических указаниях учтены замечания рецензента и редактора.

План 2014 г., позиция 90.
Подписано в печать 30.04.2014. Компьютерный набор.
Объём 55 с. Тираж 130 экз. Заказ №284.

169300, Республика Коми, г. Ухта, ул. Первомайская, д. 13.
Типография УГТУ.
169300, Республика Коми, г. Ухта, ул. Октябрьская, д. 13.

ОГЛАВЛЕНИЕ

Введение

В данных методических указаниях рассматривается компьютерное моделирование зависимостей между величинами с целью прогноза одной величины (результирующей) в зависимости от другой величины, называемой фактором. Они содержат теоретические сведения для выполнения контрольной работы, саму работу, состоящую из двух заданий. В каждое задание входит постановка задачи, варианты заданий и пример их выполнения с использованием инструментов пакета **Анализ данных** и статистических функций, встроенных в Excel. Студенты дневного отделения выполняют два задания, вариант выдаётся преподавателем, ведущим эту дисциплину. Студенты заочного отделения выполняют одно задание, вариант выбирается по сумме трёх последних цифр номера зачётной книжки, например: 2013 178 (1+7+8 =16), то есть 16 вариант.

Глава 1. Теоретические сведения к выполнению контрольной работы

Парная регрессия и корреляция

Парная регрессия – уравнение связи двух переменных *У* и *X*: $y = f(x)$, где y – зависимая переменная (результативный признак); x – независимая переменная (фактор). Различают линейные и нелинейные регрессии. Цель составления уравнения регрессии – прогноз результирующей переменной.

Линейная регрессия: $y = a_0 + a_1x + e$, где a_0, a_1 – параметры уравнения; e – погрешность.

Нелинейные регрессии делятся на два класса.

Регрессии нелинейные по независимым переменным:

1) полиномы разных степеней, например: $a_0 + a_1x + a_2x^2 + a_3x^3 + e$;
2) равносторонняя гипербола: $y = a_0 + \frac{a_1}{x} + e$.

Регрессии нелинейные по оцениваемым параметрам:

1) степенная $y = ax^b \cdot e$;
2) показательная $y = ab^x \cdot e$;
3) экспоненциальная $y = e^{a+bx} \cdot e$.

Вычисление коэффициентов линейной регрессии

Чтобы составить уравнение линейной регрессии, надо иметь таблицу наблюдений (табл. 1).

Таблица 1 – Таблица наблюдений

X	X_1	X_2	…	X_i	…	X_n
Y	Y_1	Y_2	…	Y_i		Y_n

По наблюдаемым данным можно представить диаграмму рассеяния или поле корреляции, используя инструмент Excel **Диаграммы**.

Пример. Выборочная зависимость между величиной основных производственных фондов *X* и суточной выработкой продукции *У* по данным пяти независимых наблюдений представлена в таблице 2. Требуется составить уравнение линейной парной регрессии *У* на *X*.

Таблица 2 – Выборочная зависимость между величиной основных производственных фондов *X* и суточной выработкой продукции *У*

i	1	2	3	4	5
X_i	1,20	1,50	2,50	3,00	4,50
Y_i	1,35	1,40	1,50	1,65	1,70

Используя диаграмму точечный график, построим диаграмму рассеяния **Поле корреляции** (рис. 1).

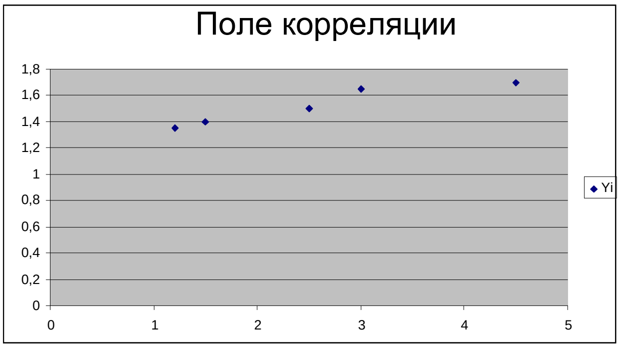


Рис. 1

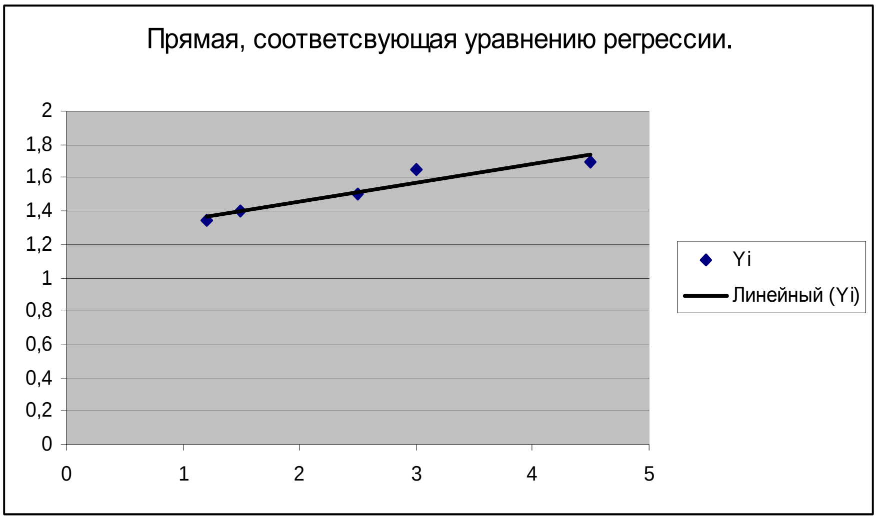


Рис. 2

Построение уравнения регрессии сводится к оценке её параметров. Для оценки регрессий линейного типа используют метод наименьших квадратов (МНК). Он позволяет получить оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических $\hat{y}$ минимальна, т. е.:

$$\sum_{i-1}^{n}(y_i - \hat{y}_i)^2 \to \min \text{ или } \sum_{i-1}^{n}(y_i - (a_0 + a_1 x_i))^2 \to \min.$$

Из этого условия находятся a_0 и a_1.

Решается система линейных уравнений:

$$\begin{cases} \sum y = a_0 \cdot n + a_1 \cdot \sum x \\ \sum y \cdot x = a_0 \cdot \sum x + a_1 \cdot \sum x^2 \end{cases}.$$

Первый способ. Система может быть решена по формулам Крамера через определители:

$$\Delta = \begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix}; \ \Delta_0 = \begin{vmatrix} \sum y & \sum x \\ \sum yx & \sum x^2 \end{vmatrix}; \ \Delta_1 = \begin{vmatrix} n & \sum y \\ \sum x & \sum yx \end{vmatrix}; \ a_0 = \frac{\Delta_0}{\Delta} \ ; \ a_1 = \frac{\Delta_1}{\Delta} .$$

Второй способ. По формулам:

$$a_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x}^2 - (\overline{x})^2} \text{ и } a_0 = \overline{y} - a_1 \cdot \overline{x},$$

где $\overline{x} = \frac{\sum x_i}{n}$, $\overline{y} = \frac{\sum y_i}{n}$, $\overline{xy} = \frac{\sum y_i x_i}{n}$, $\overline{x}^{\,2} = \frac{\sum x_i^{\,2}}{n}$, $\overline{y}^2 = \frac{\sum y_i^{\,2}}{n}$.

Далее вычисляем средние квадратические отклонения по x и по y:

$$\sigma_x = \sqrt{\overline{x}^{\,2} - \left(\overline{x}\right)^2}, \ \sigma_y = \sqrt{\overline{y}^{\,2} - \left(\overline{y}\right)^2}.$$

Коэффициент корреляции

Тесноту связи изучаемых признаков оценивает коэффициент парной корреляции r_{xy}:

$$r_{xy} = a_1 \frac{\sigma_x}{\sigma_y},$$

$$-1 <= r_{xy} <= 1.$$

Если $|r_{xy}|$ близок к единице, то связь сильная.

Если $r_{xy} < 0$, то связь обратная.

Если $r_{xy} > 0$, то связь прямая.

Коэффициент детерминации

$$R^2 = r_{xy}^{\,2}.$$

По коэффициенту детерминации также можно судить о качестве построенной модели линейной регрессии. Он показывает, на сколько процентов изменение y зависит от изменения фактора x в среднем.

Оценка значимости коэффициентов регрессии и корреляции

Проверку значимости коэффициентов уравнения и линейного коэффициента корреляции выполним с помощью t-критерия Стьюдента при уровне значимости $\alpha = 0{,}05$.

Вычислим стандартные ошибки коэффициентов уравнения регрессии и линейного коэффициента корреляции: $\sqrt{\sigma_{a0}^{\;2}}$, $\sqrt{\sigma_{a1}^{\;2}}$, $\sqrt{\sigma_{r_{xy}}^{\;2}}$.

$$\sigma_{a0}^{\;2} = \frac{\sum (y - \hat{y})^2}{n-2} \cdot \frac{\sum x^2}{n \sum (x - \overline{x})^2},$$

$$\sigma_{a1}^{\ 2} = \frac{\sum (y-\hat{y})^2}{(n-2)\times\sum (x-\overline{x})^2},$$

$$\sigma_{r_{xy}}^{\ 2} = \frac{1-r_{xy}^{\ 2}}{n-2}.$$

Для найденных коэффициентов уравнения и для линейного коэффициента корреляции определим расчётные значения t-статистик Стьюдента:

$$t_{a_0} = \frac{|a_0|}{\sqrt{\sigma_{a_0}^{\ 2}}},\ t_{a_1} = \frac{|a_1|}{\sqrt{\sigma_{a1}^{\ 2}}},\ t_{r_{xy}} = \frac{|r_{xy}|}{\sqrt{\sigma_{r_{xy}}^{\ 2}}}.$$

Расчётные значения t-статистик сравним с критическим значением $t_{крит}$, для определения которого зададим уровень значимости $\alpha = 0{,}05$ и число степеней свободы $k = n - 2$. Например, по таблице Стьюдента $t_{крит}$ (0,05;8) = 2,3. При $n = 10$, $n - 2 = 8$.

Если расчётные значения t-статистик больше $t_{крит}$, то все коэффициенты статистически значимы, в противном случае, для которых будет меньше, – не значимы.

Оценка качества построенного уравнения регрессии с помощью *F*-критерия

Пусть гипотеза H_0 предполагает, что уравнение регрессии не значимо. Оценим качество построенного уравнения регрессии с помощью F-критерия Фишера. Для этого рассчитаем фактическое значение F-критерия:

$$F_{факт} = \frac{r^2xy}{1-r^2xy}\cdot(n-2).$$

Сравним его с критическим значением.

Для определения $F_{крит}$ зададим уровень значимости $\alpha = 0{,}05$ и два числа степеней свободы. Например: $\gamma_1 = 1$(количество факторов) и $\gamma_2 = n - 2 = 10 - 2 = 8$.

По таблице Фишера определяем $F_{крит}$.

Например: $F_{крит}$(0,05; 1: 8) = 5,32.

Если $F_{факт} > F_{крит}$, то гипотеза H_0 отбрасывается и связь переменных x и y считается значимой, а построенная модель – адекватной исследуемой экономической ситуации. Если $F_{факт} < F_{крит}$, то гипотеза H_0 принимается и связь переменных x и y считается незначимой, а построенная модель – неадекватной исследуемой экономической ситуации.

Средняя ошибка аппроксимации

Рассчитаем среднюю ошибку аппроксимации:

$$A_i = \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\%,$$

$$\bar{A} = \frac{\sum A_i}{n}.$$

Если её величина находится больше 7%, то можно сделать вывод о не очень хорошем подборе модели к реальным статистическим данным.

Точечный и интервальный прогнозы

Прогнозное значение $У_{прогн}$ определяется путём подстановки в уравнение регрессии соответствующего прогнозного значения $x_{прогн}$:

$У_{прогн} = a_0 + a_1\, x_{прогн}$ – это точечный прогноз.

Дисперсия ошибки:

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

Выполним интервальный прогноз.

Найдём среднюю стандартную ошибку прогноза:

$$\sigma^2_{\hat{y}прогн} = s_e^2 \cdot \left(1 + \frac{1}{n} + \frac{(x_{прог} - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right).$$

По таблице Стьюдента находим $t_{крит}$ для $df = n - 2$ и $\alpha = 0{,}05$. Строим доверительный интервал по формуле:

$$\hat{y}_{прогн} \pm t_{крит} \cdot \sqrt{\sigma^2_{\hat{y}прогн}}\,.$$

Множественная регрессия

На любой показатель чаще всего оказывает влияние не один, а несколько факторов. Например, спрос на некоторый товар определяется не только ценой данного товара, но и ценами на замещающие и дополняющие его товары, доходом потребителя и многими другими факторами. В этом случае вместо парной регрессии рассматривается множественная регрессия. Уравнение множественной регрессии представляется в виде:

$$У = f(B, X) + e,$$

где $X = (x_1, x_2, \ldots, x_m)$ – вектор независимых переменных;

B – вектор параметров, подлежащих определению;

e – случайная ошибка (отклонение);

$У$ – зависимая переменная.

Наиболее простая из моделей множественной регрессии – это модель множественной линейной регрессии. Теоретическое уравнение линейной регрессии имеет вид:

$$У = b_0 + b_1X_1 + b_2X_2 + \ldots + b_mX_m + e$$

или для индивидуальных наблюдений $i = 1, 2, .., n$:

$$y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \ldots + b_mx_{im} + e_i.$$

Здесь $B = (b_0, b_1, \ldots, b_m)$ – вектор неизвестных параметров;

b_j, $j = 1, 2, \ldots, m$ называется j-m теоретическим коэффициентом регрессии, он характеризует чувствительность величины y к изменению X_j;

b_0 – свободный член, определяющий значение y в случае, когда все независимые переменные равны 0.

После выбора линейной функции в качестве модели необходимо оценить параметры регрессии. Самым распространённым методом оценки параметров множественной линейной регрессии является метод наименьших квадратов (МНК). Суть его состоит в минимизации суммы квадратов отклонений наблюдаемых значений зависимой переменной $У$ от значений $\hat{y}$, получаемых по уравнению регрессии. Составим функцию:

$$F(b_0, b_1, \ldots, b_m) = \sum_{i=1}^{n}(y_i - (b_0 + \sum_{j=1}^{m} b_j x_{ij}))^2 \to минимум\,.$$

Эта сумма квадратов отклонений является квадратичной относительно b_j, $j = 1, 2, \ldots, m$. Она ограничена снизу и, следовательно, имеет минимум. Необходимым условием минимума функции F является равенство нулю её частных производных по b_j, $j = 1, 2, ..., m$.

Вычислим частные производные:

$$\begin{cases} \dfrac{\partial F}{\partial b_0} = -2\sum_{i=1}^{n}(y_i - (b_0 + \sum_{j=1}^{m} b_j x_{ij})) \\ \dfrac{\partial F}{\partial b_j} = -2\sum_{i=1}^{n}(y_i - (b_0 + \sum_{j=1}^{m} b_j x_{ij})) \cdot x_j \end{cases}.$$

$$j = 1,\ 2, \ldots,\ m$$

Приравнивая их нулю, получим систему m+1 линейных уравнений с m+1 неизвестными.

$$\begin{cases} \sum_{i=1}^{n}(y_i - (b_0 + \sum_{j=1}^{m} b_j x_{ij})) = 0 \\ \sum_{i=1}^{n}(y_i - (b_0 + \sum_{j=1}^{m} b_j x_{ij})) \cdot x_j = 0 \end{cases}.$$

$$j = 1,\ 2, \ldots,\ m$$

Такая система называется системой линейных нормальных уравнений. Обычно она имеет единственное решение. Здесь y – результирующая переменная; x_1, x_2, …x_m – независимые переменные (факторы); n – число наблюдений; m – число факторов. Для решения системы может быть применён метод Крамера (метод определителей):

$$b_0 = \frac{\Delta b_0}{\Delta}, b_1 = \frac{\Delta b_1}{\Delta}, ..., \frac{\Delta b_m}{\Delta},$$

где Δ – определитель системы; Δb_0, $\Delta b_1, ...,$ Δb_{m} – частные определители.

Частные определители получаются путём замены соответствующего столбца матрицы определителя системы данными левой части системы. Для решения понадобится функция **МОПРЕД**.

Коэффициенты уравнения регрессии также могут быть найдены по формуле:

$$B = (X^T \cdot X)^{-1} \cdot (X^T Y),$$

где $B = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{bmatrix}$; $X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{m1} \\ 1 & x_{12} & \cdots & x_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & \cdots & x_{mn} \end{bmatrix}$ – матрица, первый столбец которой состоит из единиц, остальные столбцы – это наблюдаемые значения факторов x_1, x_2, … x_m.

Матрица X^T – транспонированная матрица X.

$Y = \begin{bmatrix} y_1 \\ y_2 \\ M \\ y_n \end{bmatrix}$ – наблюдаемые значения переменной Y.

В этом случае для решения в электронных таблицах будут использованы функции **ТРАНСП**, **МУМНОЖ**, **МОБР**.

Для m = 1 система имеет вид:

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases}.$$

Для m = 2 система имеет вид:

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^{n} x_{i1} + b_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i \\ b_0 \sum_{i=1}^{n} x_{i1} + b_1 \sum_{i=1}^{n} x_{i1}^2 + b_2 \sum_{i=1}^{n} x_{i1} x_{i2} = \sum_{i=1}^{n} y_i x_{i1} \\ b_0 \sum_{i=1}^{n} x_{i2} + b_1 \sum_{i=1}^{n} x_{i1} x_{i2} + b_2 \sum_{i=1}^{n} x_{i2}^2 = \sum_{i=1}^{n} y_i x_{i2} \end{cases}.$$

Рассмотрим вычисление параметров уравнения регрессии и анализ его для числа факторов $m = 2$.

Например, в электронных таблицах Excel составляется расчётная таблица следующего вида:

Таблица 3 – Пример расчётной таблицы

Таблица 1

n	x_1	x_2	y	x_1^2	x_1x_2	x_1y	x_2^2	x_2y
1	3,6	12	6	12,96	43,20	21,60	144,00	72,00
2	4,1	14	6	16,81	57,40	24,60	196,00	84,00
3	4,3	16	7	18,49	68,80	30,10	256,00	112,00
4	4,4	17	7	19,36	74,80	30,80	289,00	119,00
5	4,5	18	7	20,25	81,00	31,50	324,00	126,00
6	4,8	19	8	23,04	91,20	38,40	361,00	152,00
7	5,3	20	8	28,09	106,00	42,40	400,00	160,00
8	5,6	20	8	31,36	112,00	44,80	400,00	160,00
9	6,7	21	9	44,89	140,70	60,30	441,00	189,00
10	6,9	22	10	47,61	151,80	69,00	484,00	220,00
Итого:	50,20	179,00	76,00	262,86	926,90	393,5000	3295,00	1394,00
Среднее	5,02	17,90	7,60	26,29	92,69	39,35	329,50	139,40
σ	1,04	3,01	1,20					
σ	1,09	9,09	1,44					

y^2	$\hat{y}_x$	$\hat{y}^2_x$	y-$\hat{y}_x$	$(y$-$\hat{y}_x)^2$	Ai
36,00	5,73	32,83	0,27	0,07	4,497857004
36,00	6,38	40,68	-0,38	0,14	6,29748685
49,00	6,80	46,27	0,20	0,04	2,822075645
49,00	7,01	49,21	-0,01	0,00	0,21082074
49,00	7,23	52,23	-0,23	0,05	3,243717124
64,00	7,59	57,58	0,41	0,17	5,148670368
64,00	8,10	65,58	-0,10	0,01	1,223699591
64,00	8,32	69,24	-0,32	0,10	4,012638808
81,00	9,28	86,06	-0,28	0,08	3,078121956
100,00	9,56	91,46	0,44	0,19	4,362945646
592,00	76,00	591,14	0,00	0,86	34,90
59,20	7,60	59,11	0,00	0,09	3,49

Из неё составляется система линейных уравнений:

$$\begin{cases} 10b_0 + 50,2b_1 + 179b_2 = 76 \\ 50,2b_0 + 262,86b_1 + 926,9b_2 = 393,5. \\ 179b_0 + 926,9b_1 + 3295b_2 = 1394 \end{cases}$$

Вычисление коэффициентов уравнения

Система линейных уравнений может быть решена различными способами.

Методом Крамера по формулам:

$$b_0 = \frac{\Delta_{b0}}{\Delta}, \quad b_1 = \frac{\Delta_{b1}}{\Delta} \quad b_2 = \frac{\Delta_{b2}}{\Delta},$$

где $\Delta, \Delta_0, \Delta_1, \Delta_2$ – определители.

Вычисление коэффициентов парной линейной корреляции по формулам:

$$r_{x_1x_2} = \frac{\overline{x_1x_2} - \overline{x}_1 \cdot \overline{x}_2}{\sigma_{x_1} \cdot \sigma_{x_2}},\ r_{yx_2} = \frac{\overline{x_2y} - \overline{x}_2 \cdot \overline{y}}{\sigma_{x_2} \cdot \sigma_y},\ r_{yx_1} = \frac{\overline{x_1y} - \overline{x}_1 \cdot \overline{y}}{\sigma_{x_1} \cdot \sigma_y}.$$

Вычисление средних квадратических отклонений:

$$\sigma_{x_1} = \sqrt{\overline{x_1}^2 - (\overline{x}_1)^2},\ \sigma_{x_2} = \sqrt{\overline{x_2}^2 - (\overline{x}_2)^2},\ \sigma_y = \sqrt{\overline{y}^2 - (\overline{y})^2}.$$

Вычисление коэффициента множественной корреляции и коэффициента детерминации выполняется по формулам:

$$r_{yx} = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2}},\ R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2},\ \text{где} \quad \hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i}..$$

Или по формулам:

$$r_{x_1x_2}(y) = \frac{r_{x_1x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{1 - r^2_{yx_1}} \cdot \sqrt{1 - r^2_{yx_2}}},\ R^2 = (r_{x_1x_2}(y))^2.$$

Определение значимости коэффициентов регрессии с помощью *t*-критерия Стьюдента

Выдвинем гипотезу H_0: полученные коэффициенты уравнения регрессии не значимы.

Для каждого коэффициента определим расчётные значения t-статистик Стьюдента:

$$t_{b_i} = \frac{|b_i|}{\sigma_{b_i}},$$
$$i = 1, 2$$

где σ_{bi} – стандартные ошибки коэффициентов уравнения регрессии.

$$\sigma_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2}}{\sigma_{xi} \cdot \sqrt{1 - r^2_{x_ix_j}}} \cdot \frac{1}{\sqrt{N - m - 1}}.$$

Рассчитаем значения статистик и сравним их с критическим значением, полученным по таблице Стьюдента. n – число наблюдений, m – число факторов.

$\boldsymbol{t_{крит}}(\alpha = 0{,}05;\ df = n - m - 1)$.

Критическое значение критерия Стьюдента может быть вычислено с помощью статистической функции **=СТЬЮДРАСПОБР(*α*, *df*)**, например рис. 3.

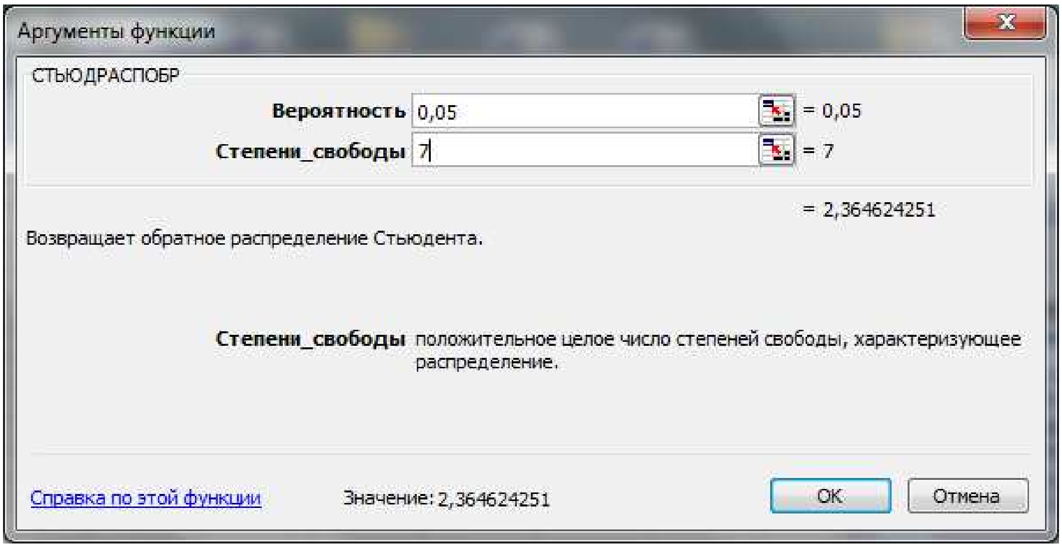


Рис. 3

$$t_{b_1} = \frac{|b_1|}{\sigma_{b_1}},\ t_{b_2} = \frac{|b_2|}{\sigma_{b_2}};$$

если:

$t_{в1} > t_{крит}$, то коэффициент b_1 статистически значим, гипотеза H_0 отвергается;

$t_{в1} < t_{крит}$, то коэффициент b_1 статистически не значим, гипотеза H_0 принимается;

$t_{в2} > t_{крит}$, то коэффициент b_2 статистически значим, гипотеза H_0 отвергается;

$t_{в2} < t_{крит}$, то коэффициент b_2 статистически не значим, гипотеза H_0 принимается.

С помощью критерия Фишера оценить статистическую надёжность построенного уравнения при уровне значимости $\alpha = 0,05$

Выдвинем гипотезу H_0: полученное уравнение регрессии не значимо.

Рассчитаем фактическое значение критерия $F_{факт}$ и сравним его с критическим значением $F_{крит}$.

$$F_{факт} = \frac{R^2}{1-R^2} \cdot \frac{N-m-1}{m},$$

где n – число наблюдений;

m – число факторов.

Для определения $F_{крит}$ задаём уровень значимости α =0,05 и два числа степеней свободы $к_1 = m$ и $к_2 = N - m - 1$: $F_{крит}$ (0,05; $к_1$; $к_2$) и по таблице Фишера определяем $F_{крит}$.

$F_{крит}$ можно определить с использованием статистической функции **=FРАСПОБР(α; $к_1$; $к_2$),**

где $к_1 = m$ (число факторов); $к_2 = n\text{-}m\text{-}1$ (n – число наблюдений).

Например:

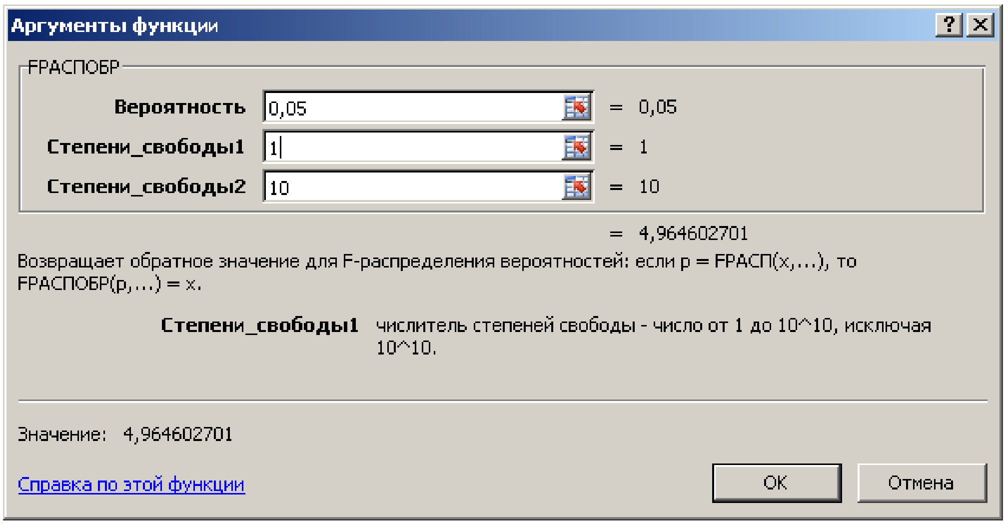


Рис. 4

Если $F_{факт} > F_{крит}$, то гипотеза H_0 отвергается, уравнение регрессии будет значимо.

Если $F_{факт} < F_{крит}$, то гипотеза H_0 принимается, уравнение регрессии будет не значимо.

Глава 2. Использование инструментов пакета Анализ данных и статистической функции Линейн для линейной парной регрессии

2.1 Задание 1. Варианты задания

По данным своего варианта, используя фактические значения независимого фактора (x) и результирующей переменной (y), провести исследование зависимости y от x.

Для этого выполнить следующие пункты.

- Построить поле корреляции.
- Выбрать и обосновать вид уравнения регрессии.
- Рассчитать коэффициенты линейной парной регрессии.
- Проанализировать коэффициенты a_0 и a_1.
- Рассчитать коэффициент парной линейной корреляции и сделать выводы о тесноте связи между переменными построенного уравнения.
- Оценить качество построенного уравнения регрессии. Проверку значимости коэффициентов уравнения и линейного коэффициента корреляции выполнить с помощью t-критерия Стьюдента. При уровне значимости $\alpha = 0{,}05$.
- Рассчитать коэффициент детерминации.
- Оценить качество построенного уравнения регрессии с помощью F-критерия Фишера.
- Рассчитать среднюю ошибку аппроксимации.
- Выполнить точечный и интервальный прогнозы для *У* при *Хпрогнозном=Хсреднее**1,1.
- Расчёты выполнить в Excel с использованием инструмента **Регрессия** из пакета **Анализ данных**, а также с помощью функции **Линейн**.

Варианты заданий

Вариант 1

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	97	161
2	73	131
3	79	135
4	99	147
5	86	139
6	91	151
7	85	135
8	77	132
9	89	161
10	95	159
11	72	120
12	115	160

Вариант 2

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	69	124
2	83	133
3	92	146
4	97	153
5	88	138
6	93	159
7	74	145
8	79	152
9	105	168
10	99	154
11	85	127
12	94	155

Вариант 3

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	77	123
2	85	152
3	79	140
4	93	142
5	89	157
6	81	181
7	79	133
8	97	163
9	73	134
10	95	155
11	84	132
12	108	165

Вариант 4

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	83	137
2	88	142
3	75	128
4	89	140
5	85	133
6	79	153
7	81	142
8	97	154
9	79	132
10	90	150
11	84	132
12	112	166

Вариант 5

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	79	134
2	91	154
3	77	128
4	87	138
5	84	133
6	76	144
7	84	160
8	94	149
9	79	125
10	98	163
11	81	120
12	115	162

Вариант 6

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	92	147
2	78	133
3	79	128
4	88	152
5	87	138
6	75	122
7	81	145
8	96	141
9	80	127
10	102	151
11	83	129
12	94	147

Вариант 7

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	75	133
2	78	125
3	81	129
4	93	153
5	86	140
6	77	135
7	83	141
8	94	152
9	88	133
10	99	156
11	80	124
12	112	156

Вариант 8

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	74	122
2	81	134
3	90	136
4	79	125
5	89	120
6	87	127
7	77	125
8	93	148
9	70	122
10	93	157
11	87	144
12	121	165

Вариант 9

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	78	133
2	94	139
3	85	141
4	73	127
5	91	154
6	88	142
7	73	122
8	82	135
9	99	142
10	113	168
11	69	124
12	83	130

Вариант 10

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	81	124
2	77	131
3	85	146
4	79	139
5	93	143
6	100	159
7	72	135
8	90	152
9	71	127
10	89	154
11	82	127
12	111	162

Вариант 11

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	87	162
2	63	133
3	69	135
4	89	147
5	76	139
6	81	151
7	75	135
8	67	132
9	79	161
10	85	159
11	62	123
12	95	162

Вариант 12

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	79	126
2	93	135
3	102	146
4	107	153
5	98	138
6	103	159
7	84	145
8	89	152
9	115	168
10	109	154
11	95	127
12	104	155

Вариант 13

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	67	125
2	75	157
3	89	141
4	93	146
5	89	159
6	81	181
7	79	133
8	98	163
9	73	134
10	95	155
11	84	138
12	109	167

Вариант 14

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	85	139
2	88	144
3	77	129
4	89	146
5	87	133
6	79	153
7	83	142
8	97	154
9	79	135
10	95	150
11	84	135
12	112	168

Вариант 15

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	78	132
2	90	154
3	78	128
4	87	138
5	86	135
6	76	144
7	87	164
8	94	149
9	79	127
10	97	163
11	86	122
12	116	164

Вариант 16

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	91	145
2	76	133
3	79	129
4	86	152
5	87	137
6	75	122
7	82	146
8	96	141
9	83	127
10	101	151
11	83	129
12	96	146

Вариант 17

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	76	135
2	79	125
3	83	129
4	92	155
5	87	140
6	78	1347
7	83	141
8	94	153
9	87	133
10	99	157
11	85	124
12	114	158

Вариант 18

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	73	125
2	81	134
3	90	138
4	78	125
5	89	120
6	86	128
7	77	125
8	94	148
9	70	123
10	95	156
11	87	147
12	120	166

Вариант 19

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	75	134
2	94	138
3	84	141
4	73	126
5	91	154
6	86	143
7	74	122
8	82	137
9	98	142
10	112	167
11	68	125
12	85	132

Вариант 20

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	83	126
2	75	132
3	85	147
4	78	138
5	93	145
6	102	159
7	73	135
8	93	153
9	75	127
10	87	152
11	83	126
12	114	165

Вариант 21

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	95	163
2	74	131
3	77	137
4	98	147
5	86	138
6	93	151
7	85	136
8	75	132
9	89	165
10	99	159
11	73	122
12	98	164

Вариант 22

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	67	126
2	85	135
3	93	146
4	98	155
5	88	138
6	94	159
7	75	147
8	79	153
9	104	168
10	99	155
11	86	127
12	95	156

Вариант 23

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	78	124
2	86	151
3	79	140
4	95	142
5	89	157
6	81	181
7	79	135
8	98	163
9	73	136
10	96	155
11	84	133
12	107	164

Вариант 24

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	85	138
2	88	142
3	75	128
4	89	142
5	87	135
6	79	155
7	83	142
8	97	154
9	79	132
10	91	155
11	84	133
12	114	167

Вариант 25

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	75	135
2	92	155
3	77	128
4	88	139
5	84	133
6	76	146
7	85	160
8	94	149
9	79	126
10	97	163
11	81	123
12	114	165

Вариант 26

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	95	145
2	79	132
3	81	128
4	88	152
5	87	138
6	75	122
7	81	145
8	97	142
9	81	127
10	102	152
11	83	129
12	96	148

Вариант 27

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	74	135
2	78	125
3	81	129
4	95	152
5	85	143
6	77	135
7	84	142
8	95	152
9	88	135
10	99	156
11	82	128
12	114	159

Вариант 28

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	73	124
2	80	136
3	93	138
4	78	124
5	88	120
6	87	127
7	77	124
8	93	149
9	74	121
10	95	158
11	87	143
12	123	165

Вариант 29

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	76	134
2	95	139
3	85	142
4	73	126
5	92	156
6	88	142
7	74	123
8	82	135
9	98	145
10	112	168
11	68	122
12	82	131

Вариант 30

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	83	126
2	78	133
3	86	146
4	79	139
5	95	143
6	100	158
7	73	135
8	90	153
9	71	128
10	88	154
11	82	126
12	114	165

2.2 Пример выполнения задания 1

Имеются данные о зависимости материалоёмкости продукции (y, кг) от размеров предприятия (выпуск продукции x, у.е.). Выполнить пункты задания, указанные выше, при этом вычислить точечный и интервальный прогнозы для *У* при *Хпрогнозном* = 50 у.е.

Выпуск продукции x	100	200	300	400	500	600	700	150	120	250
Материалоёмкость y	9	6	5	4	3,7	3,6	3,5	6	7	3,5

1. *Построить поле корреляции. При построении использовать диаграмму* ***Точечный график***

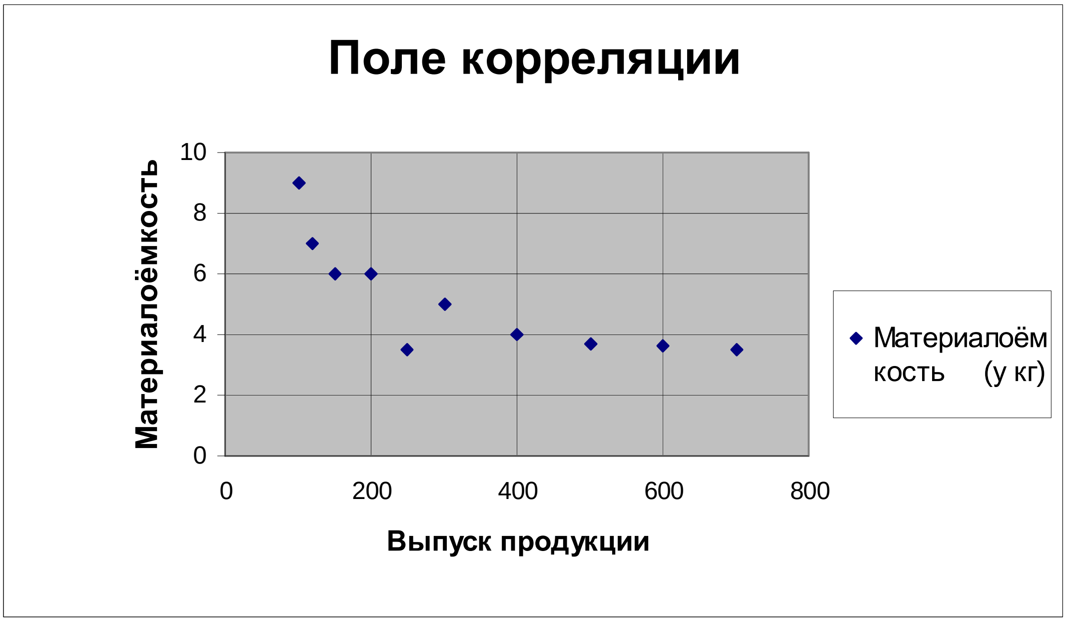


Рис. 5

2. *Выбрать и обосновать спецификацию уравнения регрессии*

Предположим, что между материалоёмкостью продукции и выпуском продукции существует линейная связь, которая описывается уравнением следующего вида:

$$y = a_0 + a_1 x + e.$$

3. *Рассчитать коэффициенты* a_0 *и* a_1 *выбранного уравнения*

Для этого составить систему нормальных уравнений и решить её методом определителей. Расчёты проведём в электронных таблицах Excel (табл. 4).

Таблица 4 – Расчёты

i	x	Y	$X \cdot Y$	X^2	Y^2	$\hat{y}_x$
1	100	9	900,00	10000,00	81,00	6,74
2	200	6	1200,00	40000,00	36,00	6,05
3	300	5	1500,00	90000,00	25,00	5,35
4	400	4	1600,00	160000,00	16,00	4,66
5	500	3,7	1850,00	250000,00	13,69	3,96
6	600	3,6	2160,00	360000,00	12,96	3,27
7	700	3,5	2450,00	490000,00	12,25	2,58
8	150	6	900,00	22500,00	36,00	6,39
9	120	7	840,00	14400,00	49,00	6,60
10	250	3,5	875,00	62500,00	12,25	5,70
Итого	3320	51,3	14275,00	1499400,00	294,15	51,3
Среднее	332	5,13	1427,50	149940,00	29,42	
σ	199,29	1,76				

Продолжение таблицы 4

$(y-\hat{y}_x)^2$	A_i	$(y\text{-}y_{cp})^2$	$(x\text{-}x_{cp})^2$
5,11	25,1%	2,593	53824,00
0,00	0,8%	0,839	17424,00
0,12	7,0%	0,049	1024,00
0,43	16,5%	0,223	4624,00
0,07	7%	1,360	28224,00
0,11	9,2%	3,460	71824,00
0,85	26,4%	6,524	135424,00
0,15	6,6%	1,596	33124,00
0,16	5,7%	2,165	44944,00
4,84	62,8%	0,324	6724,00
11,85	167%	19,133	397160,00
	16,7%		

Для вычисления коэффициентов уравнения парной линейной регрессии составим систему из двух уравнений:

$$\begin{cases} \sum y = a_0 \cdot N + a_1 \cdot \sum x \\ \sum y \cdot x = a_0 \cdot \sum x + a_1 \cdot \sum x^2 \end{cases}.$$

Найдём коэффициенты методом определителей:

$$\Delta = \begin{vmatrix} N & \sum x \\ \sum x & \sum x^2 \end{vmatrix} = \begin{vmatrix} 10 & 3320 \\ 3320 & 1499400 \end{vmatrix} = 3971600,$$

$$\Delta_0 = \begin{vmatrix} \sum \text{y} & \sum x \\ \sum yx & \sum x^2 \end{vmatrix} = \begin{vmatrix} 51{,}3 & 3320 \\ 14275 & 1499400 \end{vmatrix} = 29526220,$$

$$\Delta_1 = \begin{vmatrix} N & \sum y \\ \sum x & \sum yx \end{vmatrix} = \begin{vmatrix} 10 & 51{,}3 \\ 3320 & 14275 \end{vmatrix} = -27566,$$

$$a_0 = \frac{\Delta_0}{\Delta} = 7{,}434 \quad , \quad a_1 = \frac{\Delta_1}{\Delta} = -0{,}007.$$

4. *Модель регрессии для прогноза:* 7,434 – 0,007*x*

Проанализируем найденные коэффициенты. Коэффициент $a_1 = -$ 0,007 характеризует среднее изменение материалоёмкости продукции *y* при изменении выпуска продукции на одну у.е. Коэффициент a_0 характеризует значение *y* при $x = 0$.

5. *Рассчитать коэффициент парной линейной корреляции и сделать выводы о тесноте связи между переменными построенного уравнения*

Для расчётов используем формулы:

Среднее значение x:

$$\overline{x} = \frac{\sum x_i}{n} = \frac{3320}{10} = 332.$$

Среднее значение y:

$$\overline{y} = \frac{\sum y_i}{n} = \frac{51{,}3}{10} = 5{,}13.$$

Среднее значение произведений xy:

$$\overline{xy} = \frac{\sum y_i x_i}{n} = \frac{14275}{10} = 1427{,}5.$$

Среднее значение квадратов x:

$$\overline{x}^{\,2} = \frac{\sum x_i^{\,2}}{n} = \frac{1499400}{10} = 149940\,.$$

Среднее значение квадратов y:

$$\overline{y}^2 = \frac{\sum y_i^{\,2}}{n} = \frac{294{,}15}{10} = 29{,}415.$$

Среднее квадратическое отклонение по x:

$$\sigma_x = \sqrt{\overline{x}^2 - (\overline{x})^2} = \sqrt{149940 - 332^2} = 199,29.$$

Среднее квадратическое отклонение по y:

$$\sigma_y = \sqrt{\overline{y}^2 - (\overline{y})^2} = \sqrt{29,42 - 5,13^2} = 1,76.$$

Коэффициент корреляции:

$$r_{xy} = a_1 \frac{\sigma_x}{\sigma_y} = -0,007 \cdot \frac{199,09}{1,76} = -0,786.$$

По значению коэффициента линейной корреляции делаем вывод о наличии линейной связи между материалоёмкостью продукции и выпуском продукции. Связь обратная, средняя.

6. *Оценим качество построенного уравнения регрессии*

Проверку значимости коэффициентов уравнения и линейного коэффициента корреляции выполним с помощью *t*-критерия Стьюдента при уровне значимости $\alpha = 0,05$.

Вычислим стандартные ошибки коэффициентов уравнения регрессии и линейного коэффициента корреляции: $\sqrt{\sigma_{a0}^{\ 2}}$, $\sqrt{\sigma_{a1}^{\ 2}}$, $\sqrt{\sigma_{r_{xy}}^{\ 2}}$.

$$\sigma_{a0}^{\ 2} = \frac{\sum (y-\hat{y})^2}{N-2} \cdot \frac{\sum x^2}{N \sum (x-\overline{x})^2} = \frac{11,85}{8} \cdot \frac{3320}{10 \cdot 397160} = 0,5591,$$

$$\sqrt{\sigma_{a0}^{\ 2}} = 0,7477.$$

$$\sigma_{a1}^{\ 2} = \frac{\sum (y-\hat{y})^2}{(N-2) \cdot \sum (x-\overline{x})^2} = \frac{11,85}{8 \cdot 397160} = 0,00000037,$$

$$\sqrt{\sigma_{a1}^{\ 2}} = 0,0019.$$

$$\sigma_{r_{xy}}^{\ 2} = \frac{1 - r_{xy}^{\ 2}}{N-2} = \frac{1 - 0,786^2}{8} = 0,0478,$$

$$\sqrt{\sigma_{r_{xy}}^{\ 2}} = 0,2118.$$

Для найденных коэффициентов уравнения и для линейного коэффициента корреляции определим расчётные значения *t*-статистик Стьюдента.

$$t_{a_0} = \frac{|a_0|}{\sqrt{\sigma_{a_0}^{\ 2}}} = \frac{7,434}{0,7477} = 9,942,$$

$$t_{a_1} = \frac{|a_1|}{\sqrt{\sigma_{a1}^{\ 2}}} = \frac{0,007}{0,0019} = 3,594,$$

$$t_{r_{xy}} = \frac{|r_{xy}|}{\sqrt{\sigma_{r_{xy}}^{\ 2}}} = \frac{0,786}{0,2186} = 3,594.$$

Расчётные значения t-статистик сравним с критическим значением $t_{крит}$, для определения которого зададим уровень значимости α = 0,05 и число степеней свободы $\gamma = n - 2 = 10 - 2 = 8$. $t_{крит}$ определяем по таблице Стьюдента:

$$t_{крит}\ (0{,}05;8) = 2{,}3.$$

Так как расчётные значения t-статистик больше $t_{крит}$, то все коэффициенты статистически значимы.

7. *Для оценки качества построенного уравнения рассчитывается коэффициент детерминации:*

$$R^2 = r_{xy}{}^2 = 0{,}617.$$

То есть изменение материалоёмкости продукции y на 61,7% обусловлено изменением выпуска продукции и на 38,3% влиянием других факторов.

8. *Оценим качество построенного уравнения регрессии с помощью F-критерия Фишера*. Для этого рассчитаем фактическое значение F-критерия:

$$F_{факт} = \frac{r^2xy}{1 - r^2xy} \cdot (n-2) = \frac{0{,}786^2}{1 - 0{,}786^2} \cdot (10 - 2) = 12{,}918.$$

Сравним его с критическим значением.

Для определения $F_{крит}$ зададим уровень значимости α = 0,05 и два числа степеней свободы $\gamma_1 = 1$ (количество факторов) и $\gamma_2 = n - 2 = 10 - 2 = 8$, по таблице Фишера определяем:

$$F_{крит}(0{,}05;\ 1:\ 8) = 5{,}32.$$

У нас $F_{факт} > F_{крит}$, поэтому связь переменных x и y считается значимой, а построенная модель – адекватной исследуемой экономической ситуации.

9. *Рассчитаем среднюю ошибку аппроксимации:*

$$A_i = \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\%,$$

$$\bar{A} = \frac{\sum A_i}{n} = \frac{167\%}{10} = 16{,}7\%.$$

Поскольку её величина находится больше 7%, то можно сделать вывод о не очень хорошем подборе модели к реальным статистическим данным.

10. *Выполним точечный и интервальный прогнозы для материалоёмкости продукции при x = 50 у.е.*

Выполним точечный прогноз:

$$У_{прогн.} = 7{,}434 - 0{,}007 * 50 = 7{,}087 \text{ кг}.$$

Найдём дисперсию ошибки:

$$S_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{N} = \frac{11{,}85}{10} = 1{,}185.$$

Выполним интервальный прогноз:

$$\sigma^2_{\hat{y}прогн} = s_e^2 \cdot \left(1 + \frac{1}{n} + \frac{(x_{прог} - \overline{x})^2}{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2}\right) = 1{,}185 \cdot \left(1 + \frac{1}{10} + \frac{(50-332)^2}{397160}\right) = 1{,}825.$$

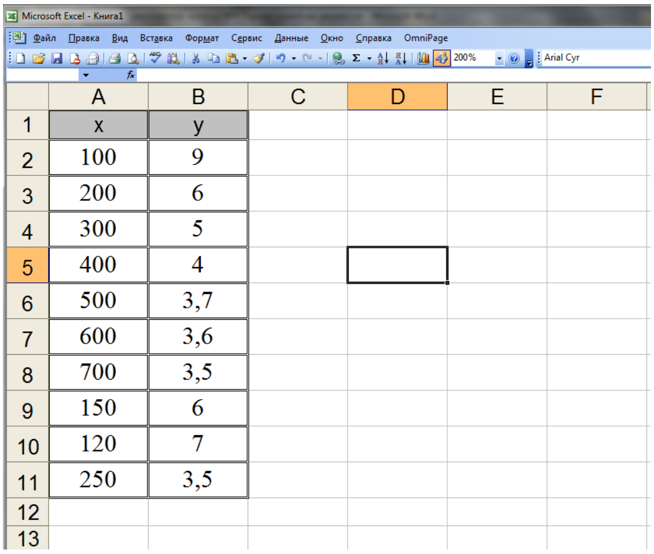
$$\hat{y}_{прогн} \pm t_{крит} \cdot \sqrt{\sigma^2_{\hat{y}прогн}} = 7{,}087 \pm 2{,}3 \cdot \sqrt{1{,}825} = 7{,}087 \pm 3{,}1073.$$

Следовательно, при выпуске продукции, равном 50 у.е., материалоёмкость продукции будет находиться в пределах: $3{,}9799 \le y \le 10{,}1946$.

Использование инструмента Регрессия из пакета Анализ данных электронных таблиц Excel

Рассмотрим последовательность действий для использования инструмента **Регрессия**.

1. Введём исходные данные на лист Excel (рис. 6):



x	y
100	9
200	6
300	5
400	4
500	3,7
600	3,6
700	3,5
150	6
120	7
250	3,5

Рис. 6

2. Войдём в меню **Данные** (версия Excel 2007 или 2010), в версии 2003 в меню **Сервис**. Выбрать **Анализ данных**. Откроется окно программы **Анализ данных** с названиями инструментов (рис. 7):

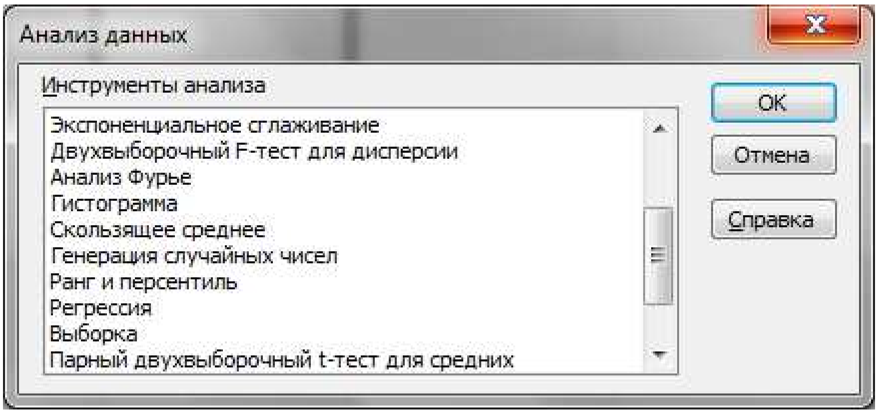


Рис. 7

3. Выберем в этом окне пункт **Регрессия**.
4. Откроется окно инструмента **Регрессия** (рис. 8):

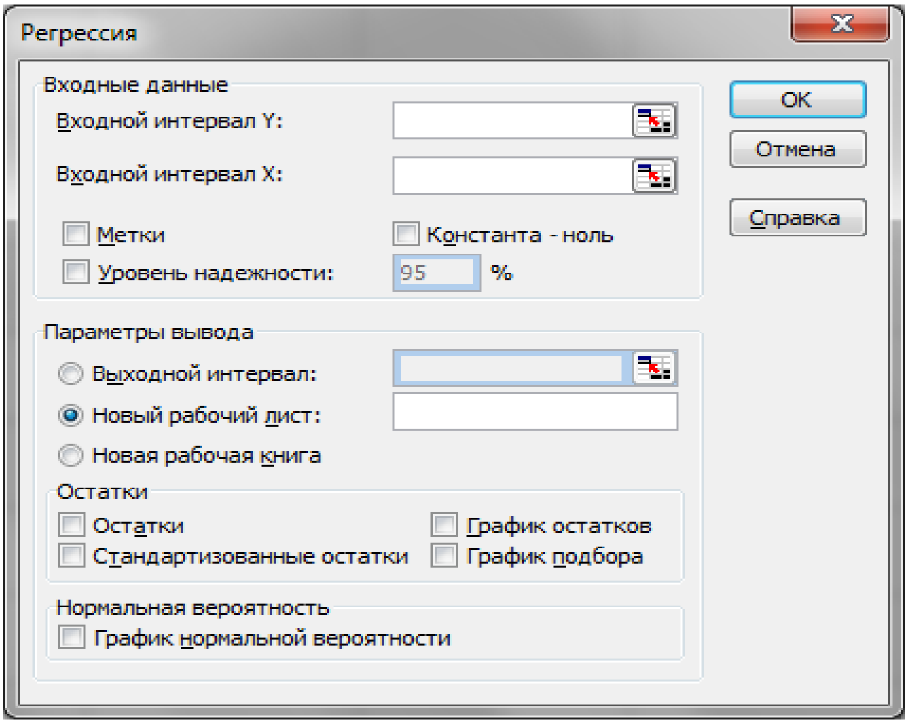


Рис. 8

5. Введём в **исходные данные** в интервал Y блок ячеек B1:B11, в интервал X – блок ячеек A1:A11; в переключатель **Метки** поставить флажок; в **Уровень надёжности** поставить флажок (уровень надёжности будет 95%), в выходной интервал поставить точку и ввести номер ячейки, начиная с которой будут выведены результаты на том же листе (рис. 9).

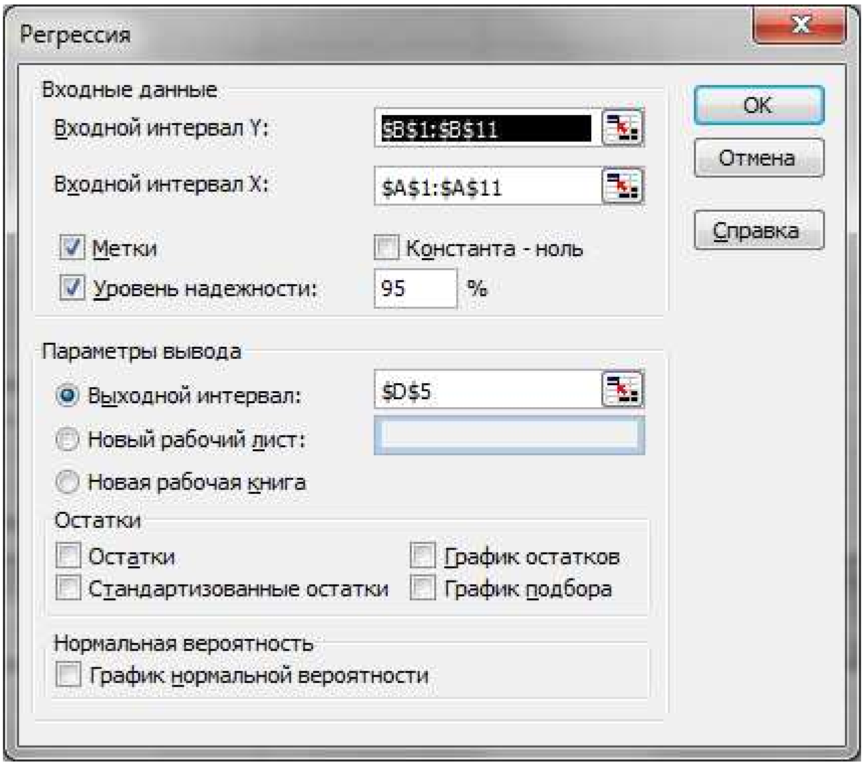


Рис. 9

6. После щелчка по **ОК** появятся результаты:

ВЫВОД ИТОГОВ	
Регрессионная статистика	
Множественный *R*	0,7859
R-квадрат	0,6176
Нормированный *R*-квадрат	0,5698
Стандартная ошибка	1,2170
Наблюдения	10,0000

Дисперсионный анализ					
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	19,1330	19,1330	12,918	0,0070
Остаток	8	11,8480	1,4810		
Итого	9	30,9810			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	7,4343	0,7477	9,9423	0,00001	5,7100	9,1586
x	–0,0069	0,0019	–3,5943	0,0070	–0,0114	–0,0025

Сравним результаты, вычисленные по формулам, с результатами, полученными с применением инструмента **Регрессия**. В первой таблице с названием ***Регрессионная статистика*** получили множественный R = 0,7859 – это коэффициент корреляции, с округлением 0,786, по формулам вычислили r_{xy} = 0,786. R-квадрат = 0,6176 – коэффициент детерминации, по формуле $R^2 = r_{xy}{}^2 = 0{,}617$. Оформим результаты в таблице (см. табл. 5).

Таблица 5 – Результаты

Результаты, полученные с применением инструмента **Регрессия**		Результаты, вычисленные по формулам	
Множественный R	0,7859	Коэффициент корреляции	0,786
R-квадрат	0,617	Коэффициент детерминации	0,617
Коэффициент Y-пересечение	7,4343	Коэффициент a_0	7,434
Коэффициент при x	–0,0069	Коэффициент a_1	-0,007
F	12,918	Фактическое значение критерия Фишера	12,918
Стандартная ошибка для a_0	0,7477	Стандартная ошибка для a_0	0,7477
Стандартная ошибка для a_1	0,0019	Стандартная ошибка для a_1	0,0019
t-статистика для a_0	9,9423	t_{a0} t-статистика Стьюдента	9,9423
t-статистика для a_1	–3,5943	t_{a1} t-статистика Стьюдента	3,5943
Значимость F	0,0070		
P-значение для a_0	0,00001		
P-значение для a_1	0,0070		

t-статистика для a_0 и *t-статистика для* a_1 для сравнения с табличными значениями берём по модулю.

Результаты совпадают с результатами, вычисленными по формулам.

Применение статистической функции Линейн для составления и анализа уравнения парной линейной регрессии

Последовательность действий

1. Введите исходные данные на лист EXCEL (рис. 10).

2. Выделите область пустых ячеек 5х2 (5 строк и 2 столбца) для вывода результатов регрессионной статистики.

3. В главном меню выберите **Вставка/Функция**.

4. В окне Категория выберите **Статистические**, в окне выберите **Функция – ЛИНЕЙН**. Щёлкните по кнопке **ОК** (рис. 11). Откроется окно функции **ЛИНЕЙН** (рис. 12).

5. В окне функции **ЛИНЕЙН** заполните аргументы функции:

– известные значения функции *У* – диапазон, содержащий данные результативного признака;

– известные значения *X* – диапазон, содержащий данные независимого признака, фактора.

- Константа – логическое значение, которое указывает на наличие или отсутствие свободного члена в уравнении. Если константа = 1, то свободный член рассчитывается обычным образом. Если константа = 0, то свободный член равен 0.
- Статистика – логическое значение, которое указывает, как выводить дополнительную информацию по регрессионному анализу. Если статистика = 1, то дополнительная информация выводится, если статистика = 0, то выводятся только коэффициенты уравнения.

6. Щёлкните по кнопке **ОК**.

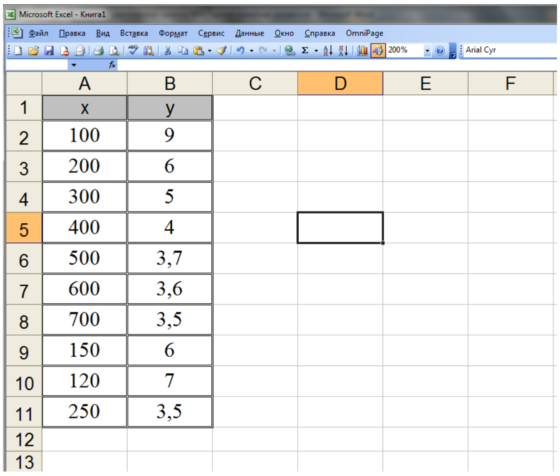


Рис. 10

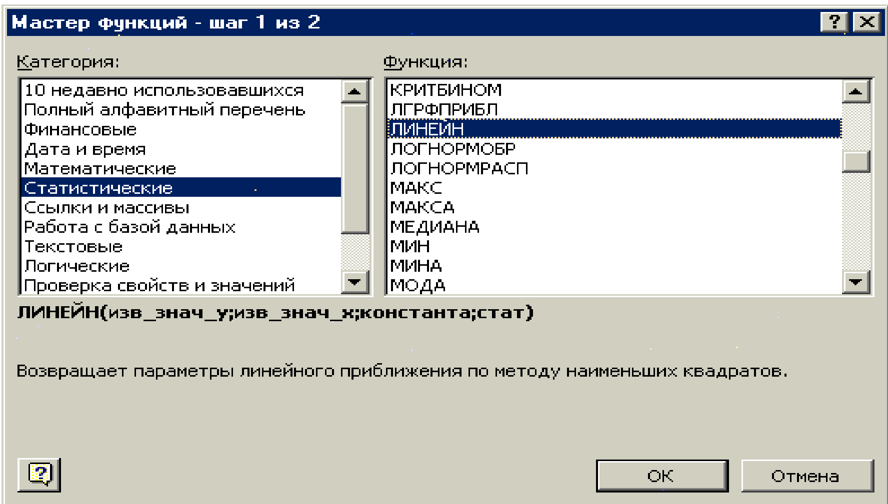


Рис. 11

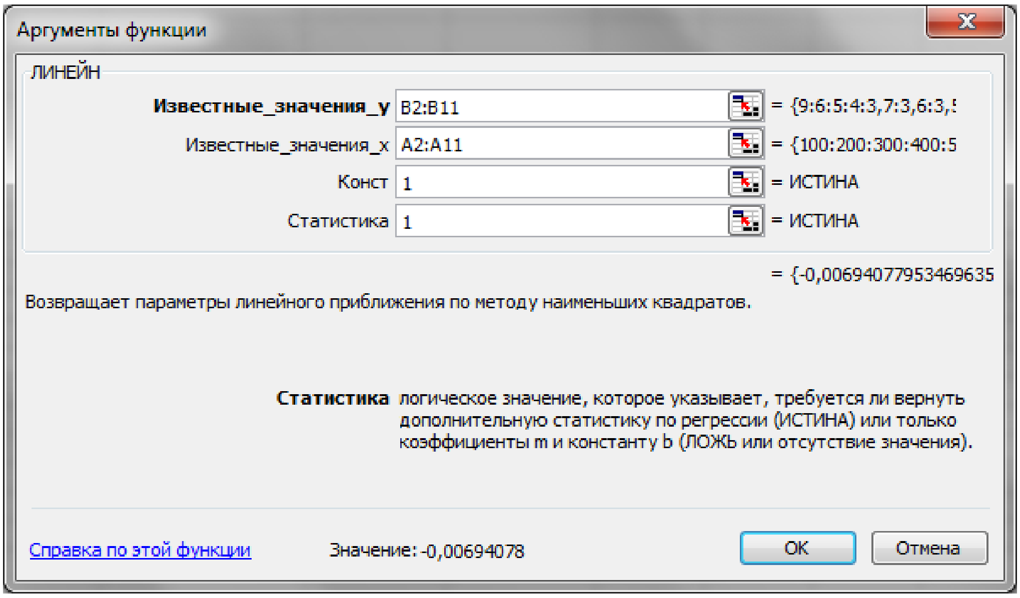


Рис. 12

7. В левой верхней ячейке выделенной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу F2, а затем – на комбинацию клавиш CTRL+SHIFT+ENTER. Появятся результаты (рис. 13).

8. Дополнительная регрессионная статистика будет выводиться в порядке, указанном в следующей схеме:

Значение коэффициента a_1	Значение коэффициента a_0
Стандартная ошибка для коэффициента a_1	Стандартная ошибка для коэффициента a_0
Коэффициент детерминации R^2	Среднеквадратическое отклонение y
F-статистика	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов

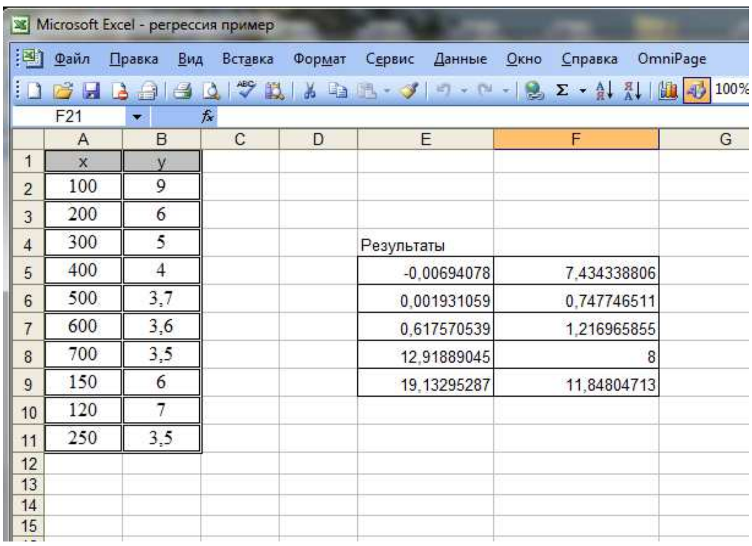


Рис. 13

- Получили уравнение у = 7,434 - 0,007x.
- Коэффициент детерминации R^2 = 0,617.
- Критерий Фишера фактическое значение = 12,918.
- Табличное критическое значение критерия Фишера $F_{крит}$(0,05; 1: 8) = 5,32.
- Уравнение значимо, так как 12,918 > 5,32.
- Стандартная ошибка для коэффициента a_1 = 0,00193.
- Стандартная ошибка для коэффициента a_0 = 0,747.
- t_{a0} t-статистика Стьюдента: $t_{a_0} = \frac{|a_0|}{\sqrt{\sigma_{a_0}{}^2}} = \frac{7,434}{0,7477} = 9,942.$
- t_{a1} t-статистика Стьюдента: $t_{a_1} = \frac{|a_1|}{\sqrt{\sigma_{a1}{}^2}} = \frac{0,007}{0,0019} = 3,594.$

Табличное критическое значение критерия Стьюдента $t_{крит}$ (0,05;8) = 2,3.

Оба коэффициента значимы, так как t-статистики Стьюдента для них больше, чем $t_{крит}$.

Результаты совпадают с результатами, вычисленными по формулам.

Глава 3. Использование инструментов пакета Анализ данных и функции Линейн для линейной множественной регрессии

3.1 Задание 2. Варианты задания

По 10 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%) (смотрите таблицу своего варианта).

Используя фактические значения независимых переменных (x_1 и x_2) и результирующего показателя (y), провести исследование зависимости y от x_1 и x_2. Для этого выполнить следующие пункты.

- Выбрать в качестве уравнения взаимосвязи переменных x_1, x_2 и y линейное регрессионное уравнение вида $y = a_0 + a_1x_1 + a_2x_2$.
- Найти коэффициенты парной линейной корреляции факторов и построить матрицу коэффициентов парной линейной корреляции. Сделать выводы о связи переменных уравнения регрессии.
- Рассчитать коэффициенты a_0, a_1, a_2 выбранного уравнения. Для этого составить систему нормальных уравнений и найти её решение методом определителей. Построить модель прогноза.
- Вычислить коэффициент множественной корреляции и коэффициент детерминации.
- Определить значимость коэффициентов регрессии с помощью t-критерия Стьюдента (при уровне значимости α = 0,05.
- С помощью критерия Фишера оценить статистическую надёжность построенного уравнения при уровне значимости α = 0,05.
- Расчёты выполнить в Excel по формулам и с использованием инструмента Регрессия из пакета Анализ данных, также с помощью функции Линейн.

Вариант 1, 2

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,6	9	1	10	6,3	21
2	7	3,6	11	2	11	6,9	23
3	7	3,7	12	3	11	7,2	24
4	8	4,1	16	4	12	7,8	25
5	8	4,3	19	5	13	8,1	27
6	8	4,5	19	6	13	8,2	29
7	9	5,4	20	7	13	8,4	31
8	9	5,5	20	8	14	8,8	33
9	10	5,8	21	9	14	9,5	35
10	10	6,1	21	10	14	9,7	34

Вариант 3, 4

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,5	9	1	10	6,3	22
2	7	3,6	10	2	10	6,5	22
3	7	3,9	12	3	11	7,2	24
4	7	4,1	17	4	12	7,5	25
5	8	4,2	18	5	12	7,9	27
6	8	4,5	19	6	13	8,2	30
7	9	5,3	19	7	13	8,4	31
8	9	5,5	20	8	14	8,6	33
9	10	5,6	21	9	14	9,5	35
10	10	6,1	21	10	15	9,6	36

Вариант 5,6

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,7	9	1	11	6,3	22
2	7	3,7	11	2	11	6,4	22
3	7	3,9	11	3	11	7,2	23
4	7	4,1	15	4	12	7,5	25
5	8	4,2	17	5	12	7,9	27
6	8	4,9	19	6	13	8,1	30
7	8	5,3	19	7	13	8,4	31
8	9	5,1	20	8	13	8,6	32
9	10	5,6	20	9	14	9,5	35
10	10	6,1	21	10	15	9,5	36

Вариант 7, 8

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,5	10	1	10	6,3	21
2	6	3,6	12	2	11	6,4	22
3	7	3,9	15	3	11	7	23
4	7	4,1	17	4	12	7,5	25
5	7	4,2	18	5	12	7,9	28
6	8	4,5	19	6	13	8,2	30
7	8	5,3	19	7	13	8,4	31
8	9	5,3	20	8	14	8,6	31
9	9	5,6	20	9	14	9,5	35
10	10	6	21	10	15	10	36

Вариант 9, 10

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,6	9	1	9	6,3	21
2	6	3,6	12	2	11	6,4	22
3	6	3,9	14	3	11	7	24
4	7	4,1	17	4	12	7,5	25
5	7	3,9	18	5	12	7,9	28
6	7	4,5	19	6	13	8,2	30
7	8	5,3	19	7	13	8	30
8	8	5,3	19	8	13	8,6	31
9	9	5,6	20	9	14	9,5	33
10	10	6,8	21	10	14	9	36

Вариант 11, 12

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,6	12	1	10	7,2	23
2	6	4,1	14	2	11	7,6	25
3	7	4,3	16	3	12	7,8	26
4	7	4,4	17	4	11	7,9	28
5	7	4,5	18	5	12	8,2	30
6	8	4,8	19	6	12	8,4	31
7	8	5,3	20	7	12	8,6	32
8	8	5,6	20	8	13	8,8	32
9	9	6,7	21	9	13	9,2	33
10	10	6,9	22	10	14	9,6	34

Вариант 13, 14

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,9	12	1	11	7,1	22
2	6	4,2	13	2	12	7,5	25
3	7	4,3	15	3	13	7,8	26
4	7	4,5	17	4	12	7,9	27
5	8	4,6	18	5	13	8,1	30
6	8	4,8	19	6	13	8,4	31
7	9	5,3	19	7	13	8,6	32
8	9	5,7	20	8	14	8,8	32
9	10	6,9	21	9	14	9,6	34
10	10	6,8	21	10	14	9,9	36

Вариант 15, 16

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,8	9	1	11	7,1	22
2	6	4,2	14	2	11	7,5	23
3	7	4,3	16	3	12	7,8	25
4	7	4,1	17	4	12	7,6	27
5	8	4,6	17	5	12	7,9	29
6	8	4,7	18	6	13	8,1	30
7	9	5,3	20	7	13	8,4	32
8	9	5,5	20	8	14	8,7	32
9	11	6,9	21	9	14	9,6	33
10	10	6,7	21	10	15	9,7	36

Вариант 17, 18

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,8	11	1	10	6,7	21
2	7	3,8	12	2	11	7,4	23
3	7	3,9	16	3	11	7,8	24
4	7	4,3	17	4	12	7,5	26
5	6	4,6	18	5	12	7,9	28
6	8	4,4	18	6	12	8,2	30
7	8	5,3	19	7	13	8,4	31
8	9	5,4	20	8	13	8,7	32
9	9	6,1	20	9	13	9,5	33
10	10	6,6	21	10	14	9,8	35

Вариант 19, 20

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,5	9	1	10	6,3	21
2	6	3,6	10	2	10	6,8	22
3	7	3,8	14	3	11	7,2	24
4	7	4,2	15	4	12	7,9	25
5	8	4,3	18	5	12	8,1	26
6	8	4,7	19	6	13	8,3	29
7	9	5,4	19	7	13	8,4	31
8	9	5,6	20	8	13	8,8	32
9	10	5,9	20	9	14	9,6	35
10	10	6,1	21	10	14	9,7	36

Вариант 21, 22

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,6	12	1	10	7,2	23
2	6	4,1	14	2	11	7,6	25
3	7	4,3	16	3	12	7,8	26
4	7	4,4	17	4	11	7,9	28
5	7	4,5	18	5	12	8,2	30
6	8	4,8	19	6	12	8,4	31
7	8	5,3	20	7	12	8,6	32
8	8	5,6	20	8	13	8,8	32
9	9	6,8	21	9	13	9,2	33
10	10	6,9	22	10	14	9,6	35

Вариант 23, 24

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,5	9	1	10	6,3	22
2	7	3,6	10	2	10	6,5	22
3	7	3,9	12	3	11	7,2	24
4	7	4,1	17	4	12	7,5	25
5	8	4,2	18	5	12	7,9	27
6	8	4,5	19	6	13	8,2	30
7	9	5,3	19	7	13	8,4	31
8	9	5,5	20	8	14	8,6	33
9	10	5,6	21	9	14	9,5	35
10	10	6,1	22	10	15	9,6	36

Вариант 25, 26

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,6	9	1	9	6,3	21
2	6	3,6	12	2	11	6,4	22
3	6	3,9	14	3	11	7	24
4	7	4,1	17	4	12	7,5	25
5	7	3,9	18	5	12	7,9	28
6	7	4,5	18	6	13	8,2	30
7	8	5,3	19	7	13	8	30
8	8	5,3	19	8	13	8,6	31
9	9	5,6	20	9	14	9,5	33
10	10	6,8	21	10	14	9	35

Вариант 27, 28

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,7	12	1	10	7,3	23
2	6	4,1	14	2	11	7,6	25
3	7	4,3	16	3	12	7,8	26
4	7	4,4	17	4	11	7,9	28
5	7	4,5	18	5	12	8,2	30
6	8	4,8	19	6	12	8,4	31
7	8	5,3	20	7	12	8,6	32
8	8	5,6	20	8	13	8,8	32
9	9	6,7	21	9	13	9,2	33
10	10	6,9	23	10	14	9,6	35

Вариант 29, 30

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,8	12	1	11	7,2	22
2	6	4,2	13	2	12	7,5	25
3	7	4,3	15	3	13	7,8	26
4	7	4,4	17	4	12	7,9	27
5	8	4,6	18	5	13	8,2	30
6	8	4,8	19	6	13	8,4	31
7	9	5,3	19	7	13	8,6	32
8	9	5,7	20	8	14	8,8	33
9	10	6,9	21	9	14	9,6	34
10	10	6,8	22	10	14	9,9	36

3.2 Пример выполнения задания 2

Имеются данные (в у.е.) о стоимости основных фондов (x_1), численности рабочих (x_2) и выпуске продукции (y).

Выпуск продукции $У$	7	7	8	8	10	11	11	12	13	14
Стоимость основных фондов X_1	3,9	4	4,8	4,8	6,8	6,4	7,2	8,2	8,5	9
Численность рабочих X_2	14	16	19	20	21	22	24	28	33	36

1. Выбрать в качестве уравнения взаимосвязи переменных x_1, x_2 и y линейное регрессионное уравнение вида $y = a_0 + a_1x_1 + a_2x_2$.

Предположим, что на этапе спецификации выявлено, что между выпуском продукции y, стоимостью основных фондов x_1 и численностью рабочих x_2 существует линейная зависимость, которая описывается уравнением $y = a_0 + a_1x_1 + a_2x_2$. Для расчёта коэффициентов такого уравнения и проведения его дальнейшего анализа составим расчётную таблицу в электронных таблицах Excel. Все расчёты проведём по формулам, используя возможности Excel.

Таблица 5

n	x_1	x_2	y	x_1^2	x_1x_2	x_1y	x_2^2	x_2y
1	3,9	14	7	15,21	54,60	27,30	196,00	98,00
2	4	16	7	16,00	64,00	28,00	256,00	112,00
3	4,8	19	8	23,04	91,20	38,40	361,00	152,00
4	4,8	20	8	23,04	96,00	38,40	400,00	160,00
5	6,8	21	10	46,24	142,80	68,00	441,00	210,00
6	6,4	22	11	40,96	140,80	70,40	484,00	242,00
7	7,2	24	11	51,84	172,80	79,20	576,00	264,00
8	8,2	28	12	67,24	229,60	98,40	784,00	336,00
9	8,5	33	13	72,25	280,50	110,50	1089,00	429,00
10	9	36	14	81,00	324,00	126,00	1296,00	504,00
Итого:	63,60	233,00	101,00	436,82	1596,30	684,60	5883,00	2507,00
Среднее	6,36	23,30	10,10	43,68	159,63	68,46	588,30	250,70
σ	1,80	6,74	2,39					
$σ^2$	3,23	45,41	5,69					

Продолжение таблицы

y^2	$\hat{y}_x$	$\hat{y}^2{}_x$	y-$\hat{y}_x$	$(y$-$\hat{y}_x)^2$	A_i
49,00	6,83	46,72	0,17	0,03	2,36
49,00	7,11	50,49	-0,11	0,01	1,51
64,00	8,17	66,68	-0,17	0,03	2,07
64,00	8,25	68,07	-0,25	0,06	3,13
100,00	10,35	107,05	-0,35	0,12	3,46
121,00	10,03	100,59	0,97	0,94	8,82
121,00	11,00	121,09	0,00	0,00	0,04
144,00	12,35	152,52	-0,35	0,12	2,92
169,00	13,08	171,02	-0,08	0,01	0,60
196,00	13,84	191,42	0,16	0,03	1,17
1077,00	101,00	1075,65	0,00	1,35	26,08325697
107,70					

2. Найти коэффициенты парной линейной корреляции факторов и построить матрицу коэффициентов парной линейной корреляции. Сделать выводы о связи переменных уравнения регрессии.

Вычислим коэффициенты парной линейной корреляции по формулам:

$$r_{yx_1} = \frac{\overline{x_1 y} - \overline{x}_1 \cdot \overline{y}}{\sigma_{x_1} \cdot \sigma_y} = \frac{68,46 - 6,36 \cdot 10,1}{1,8 \cdot 2,39} = 0,98,$$

$$r_{yx_2} = \frac{\overline{x_2 y} - \overline{x}_2 \cdot \overline{y}}{\sigma_{x_2} \cdot \sigma_y} = \frac{250,7 - 23,3 \cdot 10,1}{6,14 \cdot 2,39} = 0,96,$$

$$r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \overline{x}_1 \cdot \overline{x}_2}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{159,63 - 6,36 \cdot 23,3}{1,8 \cdot 6,74} = 0,94.$$

Составим матрицу коэффициентов парной линейной корреляции:

$$\begin{bmatrix} 1 & 0,98 & 0,96 \\ 0,98 & 1 & 0,94 \\ 0,96 & 0,94 & 1 \end{bmatrix}.$$

Коэффициент матрицы $r_{x1x2} = 0,94 > 0,7$. Значит, переменные x_1 и x_2 коллениарны, факторы x_1 и x_2 сильно связаны между собой.

3. Рассчитать коэффициенты a_0, a_1, a_2 выбранного уравнения. Для этого составить систему нормальных уравнений и найти её решение методом определителей. Построить модель прогноза.

Используя данные таблицы расчёта, подставим их в систему линейных уравнений:

$$\begin{cases} Na_0 + a_1 \sum x_1 + a_2 \sum x_2 = \sum y_i \\ a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 = \sum (y_i x_1) \\ a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 = \sum (y_i x_2) \end{cases},$$

$$\begin{cases} 10a_0 + 63{,}6a_1 + 233a_2 = 101 \\ 63{,}6a_0 + 436{,}82a_1 + 1596{,}3a_2 = 684{,}6. \\ 233a_0 + 1596{,}3a_1 + 5883a_2 = 2507 \end{cases}$$

Решим систему методом определителей:

$$D = \begin{vmatrix} 10 & 63{,}6 & 233 \\ 63{,}6 & 435{,}82 & 1596{,}3 \\ 233 & 1596{,}3 & 5883 \end{vmatrix} = 15863{,}92, \quad D_1 = \begin{vmatrix} 101 & 63{,}6 & 233 \\ 684{,}6 & 435{,}82 & 1596{,}3 \\ 2507 & 1596{,}3 & 5883 \end{vmatrix} = 27316{,}57,$$

$$D_2 = \begin{vmatrix} 10 & 101 & 233 \\ 63{,}6 & 684{,}6 & 1596{,}3 \\ 233 & 2507 & 5883 \end{vmatrix} = 15948{,}3, \quad D_3 = \begin{vmatrix} 10 & 63{,}6 & 101 \\ 63{,}6 & 435{,}82 & 684{,}6 \\ 233 & 1596{,}3 & 2507 \end{vmatrix} = 1350{,}98.$$

a_0 = 1,722;

a_1 = 1,005;

a_2 = 0,085.

Получили уравнение зависимости, модель прогноза:

$$Y = 1{,}722 + 1{,}005X_1 + 0{,}085X_2.$$

4. Вычислить индекс множественной корреляции и коэффициент детерминации:

$$\rho_{yx} = \sqrt{1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}} = \sqrt{1 - \frac{1{,}35}{56{,}9}} = 0{,}988,$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{1{,}35}{56{,}9} = 0{,}976.$$

По индексу множественной корреляции делаем вывод, что связь между y и факторами x_1 и x_2 очень сильная.

По значению коэффициента детерминации делаем вывод: изменение выпуска продукции на 97,6% объясняется влиянием факторов стоимости основных фондов и численности рабочих и на 2,4% – влиянием каких-то других факторов.

5. Оценить качество построенного уравнения:

а) определить значимость коэффициентов регрессии с помощью t-критерия Стьюдента (при уровне значимости α = 0,05).

Для каждого коэффициента определим расчётные значения t-статистик Стьюдента:

$$t_{a_i} = \frac{|a_i|}{\sigma_{a_i}},$$

где σ_{ai} – стандартные ошибки коэффициентов уравнения регрессии.

$$\sigma_{a_i} = \frac{\sigma_y \cdot \sqrt{1-R^2}}{\sigma_{xi} \cdot \sqrt{1-r^2_{x_i x_j}}} \cdot \frac{1}{\sqrt{N-m-1}},$$

$$\sigma_{a_1} = \frac{2,39 \cdot \sqrt{1-0,976}}{1,80 \cdot \sqrt{1-0,94^2}} \cdot \frac{1}{\sqrt{10-2-1}} = 0,234,$$

$$\sigma_{a_2} = \frac{2,39 \cdot \sqrt{1-0,976}}{6,74 \cdot \sqrt{1-0,94^2}} \cdot \frac{1}{\sqrt{10-2-1}} = 0,0628.$$

Рассчитаем значения статистик и сравним их с критическим значением $t_{крит.}(\alpha = 0,05;\ df = 10 - 2 - 1 = 7) = 2,36$**.**

Критическое значение критерия Стьюдента может быть вычислено с помощью статистической функции =СТЬЮДРАСПОБР(***α, df***):

$$t_{a_1} = \frac{|a_1|}{\sigma_{a_1}} = \frac{1,005}{0,2346} = 4,28,$$

$$t_{a_2} = \frac{|a_2|}{\sigma_{a_2}} = \frac{0,085}{0,062} = 1,36,$$

$t_{a1} > t_{крит}$, следовательно, коэффициент a_1 статистически значим;

$t_{a2} < t_{крит}$, следовательно, коэффициент a_2 статистически не значим;

б) с помощью критерия Фишера оценить статистическую надёжность построенного уравнения (при уровне значимости α = 0,05).

Рассчитаем фактическое значение критерия, $F_{факт}$, и сравним его с критическим значением, $F_{крит}$.

$$F_{факт} = \frac{R^2}{1-R^2} \cdot \frac{N-m-1}{m},$$

ult N – число наблюдений;

m – число факторов.

Для определения $F_{крит}$ задаём уровень значимости α =0,05 и два числа степеней свободы $к_1 = m = 2$ и $к_2 = N - m - 1 = 10 - 2 - 1 = 7$. $\boldsymbol{F_{крит}}$ **(0,05;2;7) = 4,74.**

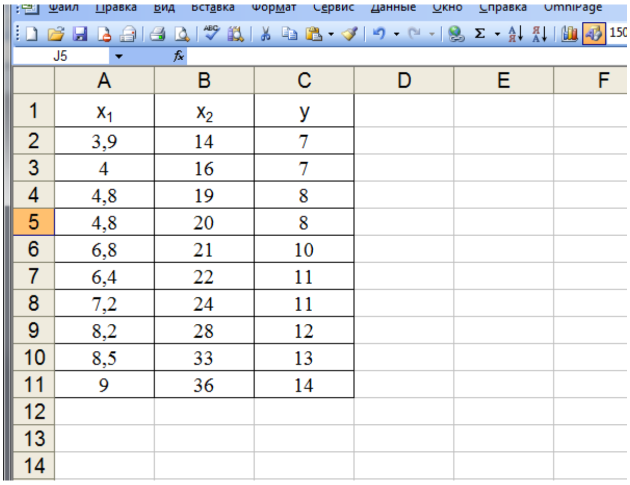
$$F_{факт} = \frac{0,976}{1-0,976} \cdot \frac{10-2-1}{2} = 144,44.$$

$F_{факт} > F_{крит}$. Следовательно, можно сделать вывод о не случайности выявленной зависимости выпуска продукции от стоимости основных фондов и численности рабочих, а также об адекватности построенного уравнения взаимосвязи этих показателей реальным данным.

Использование инструмента Регрессия из пакета «Анализ данных» электронных таблиц Excel для множественной регрессии.

Рассмотрим последовательность действий для использования инструмента Регрессия.

1. Введём исходные данные на лист Excel (рис. 14).



	A	B	C	D	E	F
1	x_1	x_2	y			
2	3,9	14	7			
3	4	16	7			
4	4,8	19	8			
5	4,8	20	8			
6	6,8	21	10			
7	6,4	22	11			
8	7,2	24	11			
9	8,2	28	12			
10	8,5	33	13			
11	9	36	14			
12						
13						
14						

Рис. 14

2. Войти в меню Данные (версия Excel-2007 или 2010) , в версии 2003 в меню Сервис. Выбрать Анализ данных. Откроется окно программы Анализ данных с названиями инструментов (рис. 15).

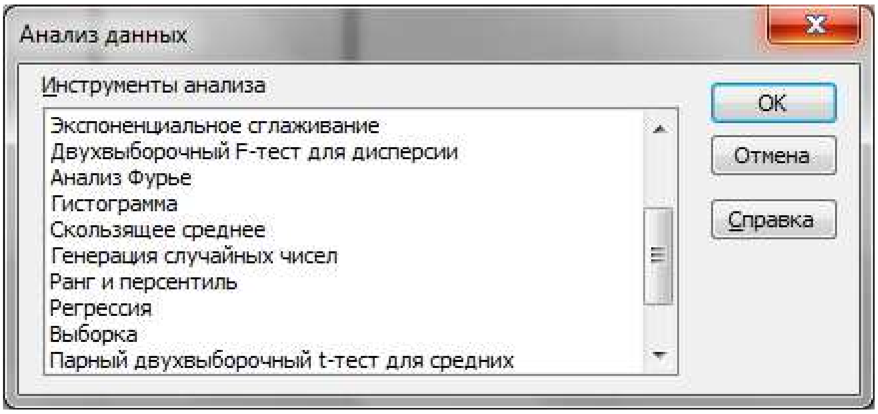


Рис. 15

3. Выбрать в этом окне пункт Регрессия.

4. Откроется окно инструмента Регрессия (рис. 16).

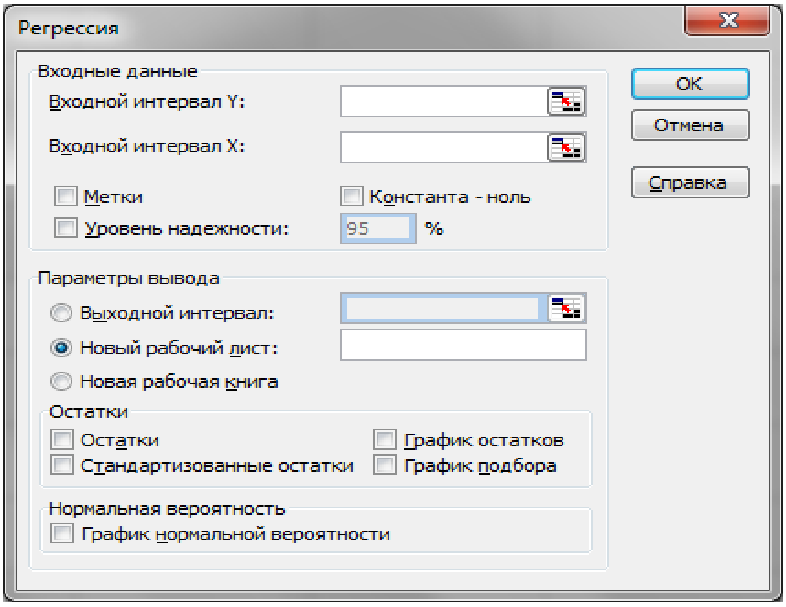


Рис. 16

5. Ввести исходные данные в интервал У блок ячеек C1:C11 и в интервал X блок ячеек A1:B11, в переключатель метки поставить флажок, в уровень надёжности поставить флажок (уровень надёжности будет 95%), в выходной интервал поставить точку и ввести номер ячейки, начиная с которой будут выведены результаты на том же листе (рис. 17).

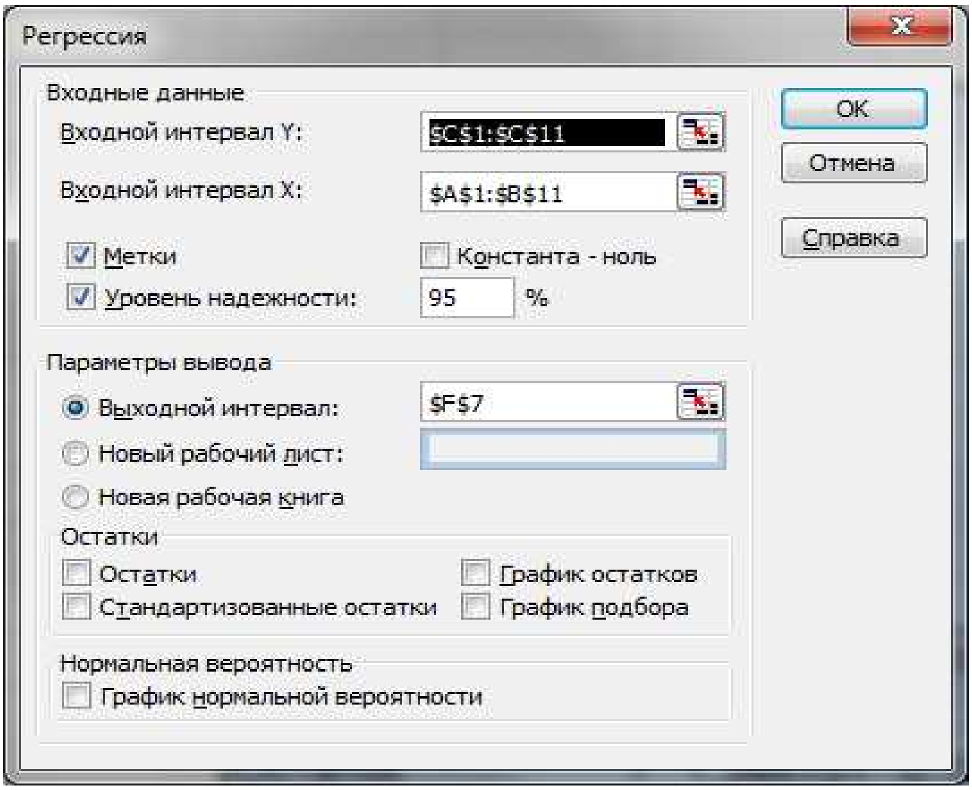


Рис. 17

6. После щелчка по ОК появятся результаты.

Таблица 6

ВЫВОД ИТОГОВ						
Регрессионная статистика						
Множественный R	0,9881					
R-квадрат	0,9763					
Нормированный R-квадрат	0,9696					
Стандартная ошибка	0,4385					
Наблюдения	10,0000					
Дисперсионный анализ						
	df	*SS*	*MS*	*F*	*Значимость F*	
Регрессия	2	55,5538	27,7769	144,4403	0,000002	
Остаток	7	1,3462	0,1923			
Итого	9	56,9000				
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	1,7219	0,5122	3,3621	0,0120	0,5109	2,9330
x_1	1,0053	0,2346	4,2848	0,0036	0,4505	1,5601
x_2	0,0852	0,0626	1,3605	0,2159	-0,0629	0,2332

Сравним результаты, вычисленные по формулам, с результатами, полученными с применением инструмента Регрессия. В первой таблице с названием *Регрессионная статистика* получили множественный R = 0,9881 –это коэффициент корреляции, R-квадрат = 0,9763– коэффициент детерминации. Оформим результаты в таблице.

t-статистика для a_0 и *t-статистика для* a_1 для сравнения с табличными значениями берём по модулю.

Таблица 7

Результаты, полученные с применением инструмента Регрессия		Результаты, вычисленные по формулам	
Множественный R	0,9881	Коэффициент корреляции	0,988
R-квадрат	0,9763	Коэффициент детерминации	0,976
Коэффициент Y-пересечение	1,7219	Коэффициент a_0	1,722
Коэффициент при x_1	1,0053	Коэффициент a_1	1,005
Коэффициент при x_2	0,0852	Коэффициент a_2	0,085
F	144,4403	Фактическое значение критерия Фишера	144,44

Окончание таблицы 7

Стандартная ошибка для a_0	0,5122	Стандартная ошибка для a_0	0,7477
Стандартная ошибка для a_1	0,2346	Стандартная ошибка для a_1	0,234
Стандартная ошибка для a_2	0,0626	Стандартная ошибка для a_2	0,062
t-статистика для a_0	3,3621	t_{a0} *t*-статистика Стьюдента	3,36
t-статистика для a_1	4,2848	t_{a1} *t*-статистика Стьюдента	4,28
t-статистика для a_2	1,3605	t_{a2} *t*-статистика Стьюдента	1,36
Значимость F	0,0070		
Р-значение для a_0	0,0120		
Р-значение для a_1	0,0036		

Результаты совпадают с вычисленными по формулам.

Применение статистической функции Линейн для составления и анализа уравнения парной линейной регрессии

Последовательность действий:

1. Введите исходные данные на лист EXCEL (рис. 18).
2. Выделите область пустых ячеек 5х3 (5 строк и 3 столбца) для вывода результатов регрессионной статистики.
3. В главном меню выберите Вставка/Функция.
4. В окне Категория выберите Статистические, в окне выберите Функция – **ЛИНЕЙН**. Щёлкните по кнопке ОК (рис. 19). Откроется окно функции **ЛИНЕЙН** (рис. 20).
5. В окне функции **ЛИНЕЙН** заполните аргументы функции:

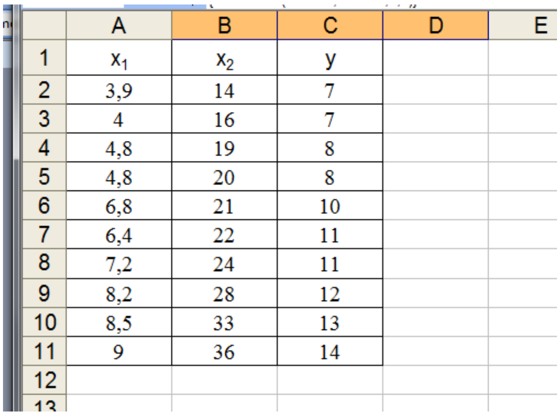
– известные значения функции *У* – диапазон, содержащий данные результативного признака;

– известные значения *X* – диапазон, содержащий данные независимого признака, фактора.

– Константа – логическое значение, которое указывает на наличие или отсутствие свободного члена в уравнении. Если константа = 1, то свободный член рассчитывается обычным образом. Если константа = 0, то свободный член равен 0.

– Статистика – логическое значение, которое указывает, как выводить дополнительную информацию по регрессионному анализу. Если статистика = 1, то дополнительная информация выводится, если статистика = 0, то выводятся только коэффициенты уравнения.

6. Щёлкните по кнопке **ОК**.



	A	B	C	D	E
1	x_1	x_2	y		
2	3,9	14	7		
3	4	16	7		
4	4,8	19	8		
5	4,8	20	8		
6	6,8	21	10		
7	6,4	22	11		
8	7,2	24	11		
9	8,2	28	12		
10	8,5	33	13		
11	9	36	14		
12					
13					

Рис. 18

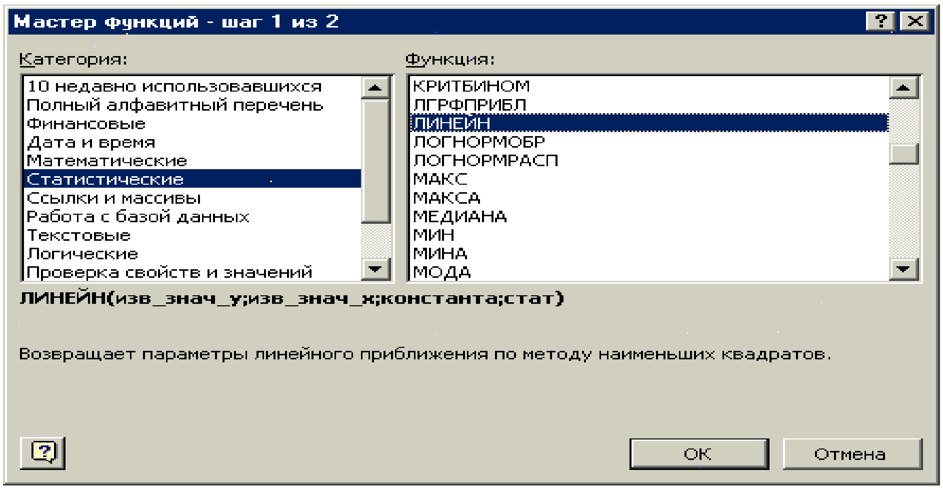


Рис. 19

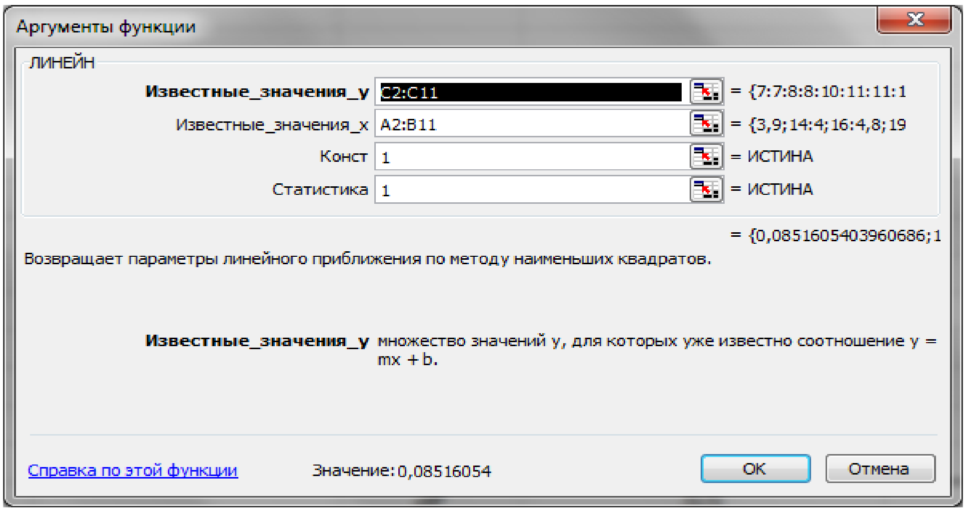
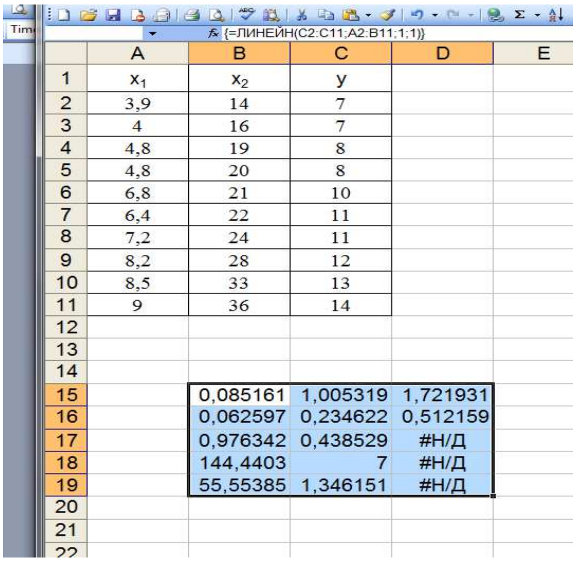


Рис. 20

- В левой верхней ячейке выделенной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу F2, а затем – на комбинацию клавиш CTRL+SHIFT+ENTER. Появятся результаты (рис. 21).
- Дополнительная регрессионная статистика будет выводиться в порядке, указанном в следующей схеме:

Значение коэффициента a_2	Значение коэффициента a_1	Значение коэффициента a_0
Стандартная ошибка для коэффициента a_2	Стандартная ошибка для коэффициента a_1	Стандартная ошибка для коэффициента a_1
Коэффициент детерминации R^2	Среднеквадратическое отклонение y	Нет данных
F-статистика	Число степеней свободы	Нет данных
Регрессионная сумма квадратов	Остаточная сумма квадратов	Нет данных


{=ЛИНЕЙН(C2:C11;A2:B11;1;1)}

	A	B	C	D
1	x_1	x_2	y	
2	3,9	14	7	
3	4	16	7	
4	4,8	19	8	
5	4,8	20	8	
6	6,8	21	10	
7	6,4	22	11	
8	7,2	24	11	
9	8,2	28	12	
10	8,5	33	13	
11	9	36	14	
15		0,085161	1,005319	1,721931
16		0,062597	0,234622	0,512159
17		0,976342	0,438529	#Н/Д
18		144,4403	7	#Н/Д
19		55,55385	1,346151	#Н/Д

Рис. 21

- Получили уравнение $y = 1{,}7219 + 1{,}0053x_1+0{,}0851x_2$.
- Коэффициент детерминации $R^2 = 0{,}976$.
- Критерий Фишера: фактическое значение = 144,44.
- Табличное критическое значение критерия Фишера $F_{крит}(0{,}05;\ 2:\ 7) = 4{,}74$.
- Уравнение значимо, так как $144{,}44 > 4{,}74$.
- Стандартная ошибка для коэффициента $a_0 = 0{,}512$.
- Стандартная ошибка для коэффициента $a_1 = 0{,}234$.
- Стандартная ошибка для коэффициента $a_1 = 0{,}0625$.
- t_{a0} t-статистика Стьюдента: $t_{a_0} = \dfrac{|a_0|}{\sqrt{\sigma_{a_0}^{\ 2}}} = \dfrac{1{,}7219}{0{,}512} = 3{,}362$.
- t_{a1} t-статистика Стьюдента: $t_{a_1} = \dfrac{|a_1|}{\sqrt{\sigma_{a1}^{\ 2}}} = \dfrac{1{,}0053}{0{,}234} = 4{,}28$.
- t_{a2} t-статистика Стьюдента: $t_{a2} = \dfrac{|a_2|}{\sqrt{\sigma_{a2}^{\ 2}}} = \dfrac{0{,}0851}{0{,}0625} = 1{,}36$.

Табличное критическое значение критерия Стьюдента $t_{крит}$ (0,05;7) = 2,36.

Коэффициенты a_0, a_1 значимы, так t-статистики Стьюдента для них больше, чем $t_{крит}$. Коэффициент a_2 не значим, так как $t_{a2} < t_{крит}$.

Библиографический список

1. Орлова, И. В. Экономико-математические методы и модели: компьютерное моделирование : учеб. пособие /И. В. Орлова, В. А. Половников. – 3-е изд., перераб. и доп. – М. : Вузовский учебник: ИНФРА-М, 2013. – 389 с.
2. Экономико-математические методы и прикладные модели : учеб. пособие для вузов / В. В. Федосеев [и др.] – М .: ЮНИТИ, 2001.
3. Туманова, О. Н. Использование компьютерных технологий для решения экономических задач : учеб. пособие / О. Н. Туманова. – Ухта : УГТУ, 2010. – 55 с.