



А.А. Чалганова

**Построение нелинейных моделей парной
регрессии с использованием табличного
процессора Excel
для студентов всех форм обучения**

**Учебное пособие по дисциплине
«Эконометрика»**

УДК 33

ББК 65.05

Ч-16

Чалганова, Алла Анатольевна.

Построение нелинейных моделей парной регрессии с использованием табличного процессора Excel. Учебное пособие по дисциплине «Эконометрика» / А.А. Чалганова. – [Текст: электронный]. – Санкт-Петербург : РГГМУ, 2022. – 90 с.

В учебном пособии особое внимание уделяется реализации алгоритма решения задачи построения нелинейных моделей парной регрессии путем обращения к функции «Регрессия» модуля «Анализ данных» (надстройка табличного процессора Excel).

Учебное пособие предназначено для студентов всех форм обучения по направлению подготовки 38.03.01 «Экономика».

Учебное пособие «Построение нелинейных моделей парной регрессии с использованием табличного процессора Excel» по дисциплине «Эконометрика» для студентов всех форм обучения одобрены на заседании кафедры экономики предприятия природопользования и учетных систем от 15 июня 2022 г., протокол № 12.

© Чалганова А.А., 2022

© «Российский государственный гидрометеорологический университет» (РГГМУ), 2022

Введение

Построение эконометрических моделей обуславливает существенный объем вычислений, что требует использования вычислительной техники и соответствующего программного обеспечения, особенно при большом объеме исходных данных. Удобным инструментом для решения эконометрических задач является табличный процессор Excel. Будучи удобной универсальной вычислительной средой, Excel позволяет реализовать алгоритм решения путем программирования арифметических или логических выражений в ячейках электронной таблицы, либо путем обращения к стандартным функциям и модулям.

Использование современных алгоритмов решения эконометрических задач на практике предполагает знакомство с возможностями использования ЭВМ при обработке массовой статистической информации, моделировании и прогнозировании явлений.

Поэтому основной целью данного методических указаний является изложение численной методики решения основных задач линейного регрессионного анализа в вычислительной среде табличного процессора Excel.

1. Общие методические указания для решения задач по дисциплине «Эконометрика»

Целью данного пособия является рассмотрение различных форм нелинейных моделей, изучение и применение метода линеаризации нелинейных уравнений регрессии, итерационных методов подбора нелинейных моделей, методов проверки качества таких моделей.

Решение задач предполагает наличие предварительной теоретической подготовки по той теме дисциплины, для усвоения которой предлагается определенная задача. Чтобы очертить круг вопросов, которые надо прояснить для решения задачи, необходимо, прежде всего, внимательно разобраться в условии задачи и в вопросах, на которые требуется дать ответы. Это позволит целенаправленно проводить работу с конспектом лекций и эффективнее готовиться к аттестации по предмету, используя теоретические материалы, которые помогут в решении задачи. Поскольку решение задач по дисциплине «Эконометрика» связано с расчетом различных показателей, необходимо хорошо ориентироваться в них, что достигается выписыванием терминов, понятий, названий показателей и формул их расчета, а также систематизацией подобных записей. Использование современных программных средств упрощает выполнение расчетов, сокращает время решения задач, расширяет их практическое применение для планирования и составления прогнозов. Определив необходимые для решения задачи показатели и формулы их расчета, выбрав средства автоматизации расчетов, студенты обычно легко справляются с вычислениями, однако полученные цифры не являются конечной целью решения задачи. Задача исследователя, в роли которого выступает решающий учебную задачу студент, не ограничивается получением какого-то числового значения, хотя это важно и полезно само по себе.

Главная цель исследования – дать оценку и интерпретацию полученным значениям, т.е. разъяснить их смысл, оценить статистическую значимость эконометрической модели и ее параметров, сделать вывод о возможности

использования данной модели для анализа и прогнозирования. Поэтому решение задачи по эконометрике должно заканчиваться выводами, полученными в результате корреляционно-регрессионного анализа.

2. Нелинейные модели парной регрессии и корреляции

Если между экономическими явлениями существуют нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных функций.

Различают два класса нелинейных регрессий:

1. Регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, например:

– полиномы различных степеней:

$$y_x = a + b \cdot x + c \cdot x^2,$$

$$y_x = a + b \cdot x + c \cdot x^2 + d \cdot x^3;$$

– равносторонняя гиперболола:

$$y_x = a + b/x;$$

– полулогарифмическая функция:

$$y_x = a + b \cdot \ln x.$$

2. Регрессии, нелинейные по оцениваемым параметрам, например:

– степенная:

$$y_x = a \cdot x^b;$$

– показательная:

$$y_x = a \cdot b^x;$$

– экспоненциальная:

$$y_x = e^{a+b \cdot x}.$$

Регрессии, нелинейные по включенным переменным, приводятся к линейному виду простой заменой переменных, а дальнейшая оценка параметров производится с помощью метода наименьших квадратов. Рассмотрим некоторые функции.

Парабола второй степени $y_x = a + b \cdot x + c \cdot x^2$ приводится к линейному виду с помощью замены: $x = x_1$, $x^2 = x_2$. В результате приходим к двухфакторному уравнению $y_x = a + b \cdot x_1 + c \cdot x_2$, оценка параметров которого при помощи МНК, как можно показать, приводит к системе следующих нормальных уравнений:

$$\begin{cases} a \cdot n + b \cdot \sum x_1 + c \cdot \sum x_2 = \sum y; \\ a \cdot \sum x_1 + b \cdot \sum x_1^2 + c \cdot \sum x_1 \cdot x_2 = \sum x_1 \cdot y; \\ a \cdot \sum x_2 + b \cdot \sum x_1 \cdot x_2 + c \cdot \sum x_2^2 = \sum x_2 \cdot y. \end{cases}$$

А после обратной замены переменных получим:

$$\begin{cases} a \cdot n + b \cdot \sum x + c \cdot \sum x^2 = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 + c \cdot \sum x^3 = \sum x \cdot y; \\ a \cdot \sum x^2 + b \cdot \sum x^3 + c \cdot \sum x^4 = \sum x^2 \cdot y. \end{cases} \quad (2.1)$$

Парабола второй степени обычно применяется в случаях, когда для определенного интервала значений фактора меняется характер связи рассматриваемых признаков: прямая связь меняется на обратную или обратная на прямую.

Равносторонняя гипербола $y_x = a + b/x$ может быть использована для характеристики связи удельных расходов сырья, материалов, топлива с объемами выпускаемой продукции, зависимости времени обращения товаров от величины товарооборота, процента прироста заработной платы от уровня безработицы (например, кривая А.В. Филлипса), расходов на непродовольственные товары от доходов или общей суммы расходов (например, кривые Э. Энгеля) и в других случаях. Вид кривой Филлипса представлен на рисунке 1.

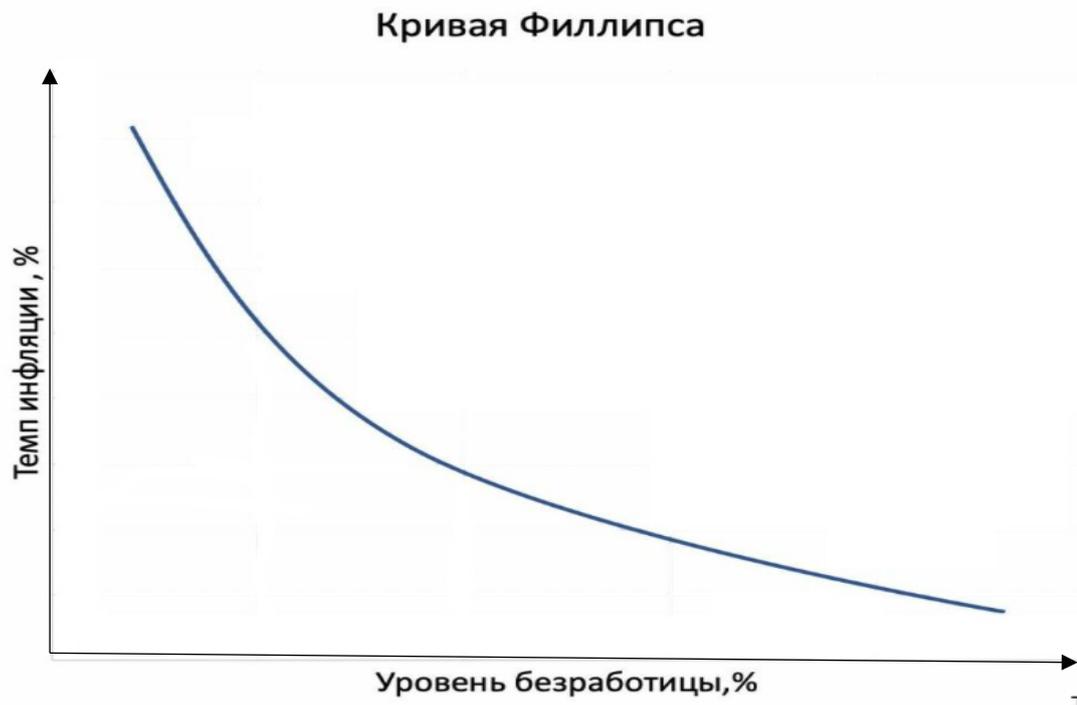


Рис. 1. Вид кривой Филлипса

Вид зависимостей расходов на все блага, кроме блага X, от объема потребления блага X, а также кривой Энгеля представлен на рисунке 2.

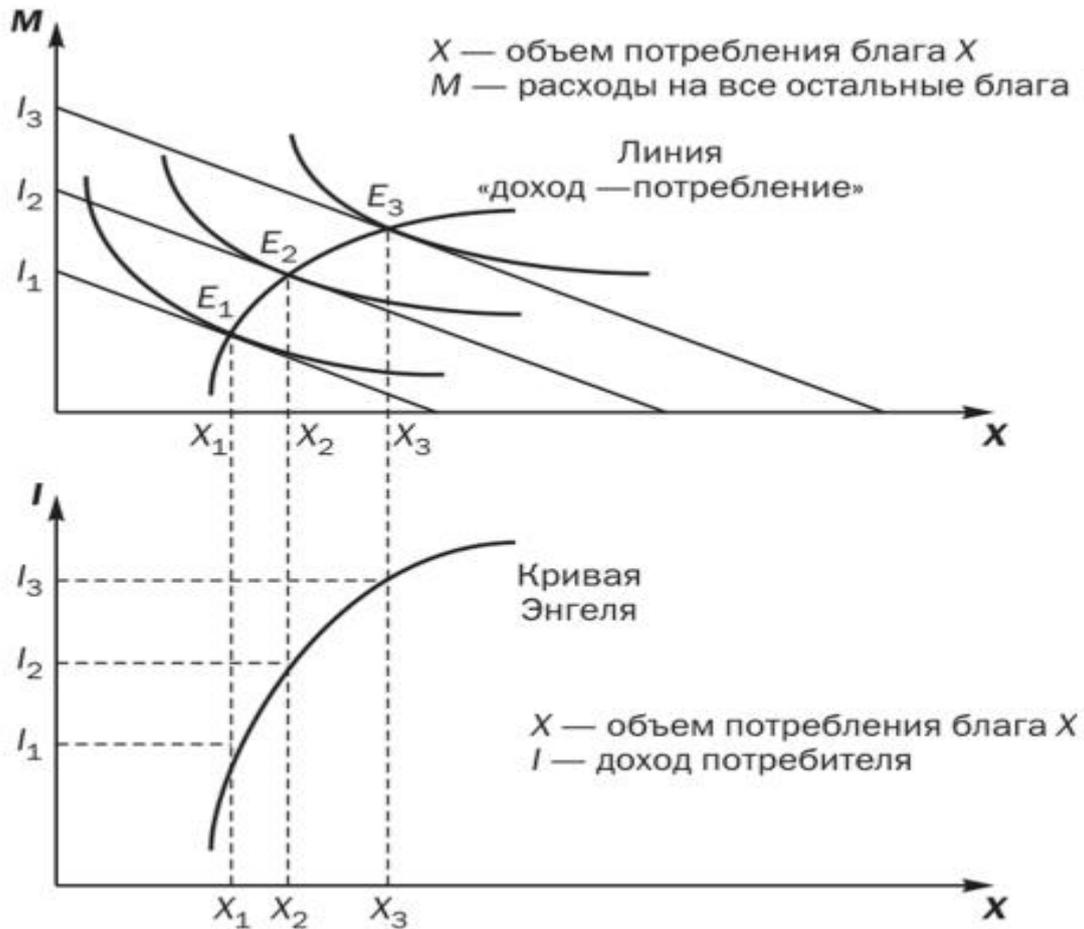


Рис. 2. Кривая «доход-потребление» и кривая Энгеля

Гипербола $y_x = a + b/x$ приводится к линейному уравнению простой заменой: $z = 1/x$. Система линейных уравнений при применении МНК будет выглядеть следующим образом:

$$\begin{cases} a \cdot n + b \cdot \sum \frac{1}{x} = \sum y; \\ a \cdot \sum \frac{1}{x} + b \cdot \sum \frac{1}{x^2} = \sum \frac{1}{x} \cdot y. \end{cases} \quad (2.2)$$

Аналогичным образом приводятся к линейному виду зависимости $y_x = a + b \cdot \ln x$, $y_x = a + b \cdot \sqrt{x}$ и другие.

Несколько иначе обстоит дело с регрессиями нелинейными по оцениваемым параметрам, которые делятся на два типа: нелинейные модели внутренне линейные (приводятся к линейному виду с помощью соответствующих преобразований, например, логарифмированием) и нелинейные модели внутренне нелинейные (к линейному виду не приводятся).

К внутренне линейным моделям относятся, например:

Степенная функция:

$$y_x = a \cdot x^b,$$

Показательная функция:

$$y_x = a \cdot b^x,$$

Экспоненциальная функция:

$$y_x = e^{a+b \cdot x},$$

Логистическая:

$$y_x = \frac{a}{1 + b \cdot e^{-c \cdot x}},$$

Обратная:

$$y_x = \frac{1}{a + b \cdot x}.$$

К внутренне нелинейным моделям можно, например, отнести следующие модели:

$$y_x = a + b \cdot x^c,$$

$$y_x = a \cdot \left(1 - \frac{1}{1 - x^b} \right).$$

Среди нелинейных моделей наиболее часто используется степенная функция $y = a \cdot x^b \cdot \varepsilon$, которая приводится к линейному виду логарифмированием:

$$\ln y = \ln(a \cdot x^b \cdot \varepsilon);$$

$$\ln y = \ln a + b \cdot \ln x + \ln \varepsilon;$$

Введя обозначения:

$$Y = \ln y, X = \ln x, A = \ln a, E = \ln \varepsilon,$$

получим:

$$Y = A + b \cdot X + E,$$

Таким образом, путем логарифмирования получаем линейное уравнение, применение метода наименьших квадратов (МНК) к которому используется для получения оценок параметров данного уравнения.

Т.е. МНК мы применяем для преобразованных данных:

$$\begin{cases} A \cdot n + b \cdot \sum X = \sum Y, \\ A \cdot \sum X + b \cdot \sum X^2 = \sum X \cdot Y, \end{cases}$$

а затем потенцированием находим искомое уравнение.

Широкое использование степенной функции связано с тем, что параметр b в ней имеет четкое экономическое истолкование – он является коэффициентом эластичности. (Коэффициент эластичности показывает, на сколько процентов измениться в среднем результат, если фактор изменится на 1%.) Формула для расчета коэффициента эластичности имеет вид:

$$\varepsilon = f'(x) \cdot \frac{x}{y}. \quad (2.3)$$

Так как для остальных функций коэффициент эластичности не является постоянной величиной, а зависит от соответствующего значения фактора x , то обычно рассчитывается средний коэффициент эластичности:

$$\bar{\varepsilon} = f'(\bar{x}) \cdot \frac{\bar{x}}{\bar{y}}. \quad (2.4)$$

Формулы для расчета средних коэффициентов эластичности для наиболее часто используемых типов уравнений регрессии приведены в таблице 2.1.

Таблица 2.1

Формулы для расчета средних коэффициентов эластичности

Вид функции, y	Первая производная, y'	Средний коэффициент эластичности, $\bar{\varepsilon}$
1	2	3
$y = a + b \cdot x + \varepsilon$	b	$\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$
$y = a + b \cdot x + c \cdot x^2 + \varepsilon$	$b + 2c \cdot x$	$\frac{(b + 2c \cdot \bar{x}) \cdot \bar{x}}{a + b \cdot \bar{x} + c \cdot \bar{x}^2}$
$y = a + \frac{b}{x} + \varepsilon$	$-\frac{b}{x^2}$	$-\frac{b}{a \cdot \bar{x} + b}$
$y = a \cdot x^b \cdot \varepsilon$	$a \cdot b \cdot x^{b-1}$	b
$y = a \cdot b^x \cdot \varepsilon$	$a \cdot \ln b \cdot b^x$	$\bar{x} \cdot \ln b$
$y = a + b \cdot \ln x + \varepsilon$	$\frac{b}{x}$	$\frac{b}{a + b \cdot \ln \bar{x}}$
$y = \frac{a}{1 + b \cdot e^{-c \cdot x + \varepsilon}}$	$\frac{a \cdot b \cdot c \cdot e^{-c \cdot x}}{(1 + b \cdot e^{-c \cdot x})^2}$	$\frac{b \cdot c \cdot \bar{x}}{b + e^{c \cdot \bar{x}}}$
$y = \frac{1}{a + b \cdot x + \varepsilon}$	$-\frac{b}{(a + b \cdot x)^2}$	$-\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$

Возможны случаи, когда расчет коэффициента эластичности не имеет смысла. Это происходит тогда, когда для рассматриваемых признаков бессмысленно определение изменения в процентах.

Уравнение нелинейной регрессии, так же, как и в случае линейной зависимости, дополняется показателем тесноты связи. В данном случае это индекс корреляции:

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}}, \quad (2.5)$$

где $\sigma_y^2 = \frac{1}{n} \sum (y - \bar{y})^2$ – общая дисперсия результативного признака y ,

$\sigma_{\text{ост}}^2 = \frac{1}{n} \sum (y - y_x)^2$ – остаточная дисперсия.

Величина данного показателя находится в пределах: $0 \leq \rho_{xy} \leq 1$. Чем ближе значение индекса корреляции к единице, тем теснее связь рассматриваемых признаков, тем более надежно уравнение регрессии.

Квадрат индекса корреляции носит название индекса детерминации и характеризует долю дисперсии результативного признака y , объясняемую регрессией, в общей дисперсии результативного признака:

$$\rho_{xy}^2 = 1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2} = \frac{\sigma_{\text{объясн}}^2}{\sigma_y^2}, \quad (2.6)$$

т.е. имеет тот же смысл, что и в линейной регрессии; $\sigma_{\text{объясн}}^2 = \frac{1}{n} \sum (y_x - \bar{y})^2$.

Индекс детерминации ρ_{xy}^2 можно сравнивать с коэффициентом детерминации r_{xy}^2 для обоснования возможности применения линейной функции. Чем больше кривизна линии регрессии, тем величина r_{xy}^2 меньше ρ_{xy}^2 . А близость этих показателей указывает на то, что нет необходимости усложнять форму уравнения регрессии и можно использовать линейную функцию.

Если нелинейное относительно объясняющей переменной уравнение регрессии при линеаризации принимает форму линейного уравнения парной регрессии, то для оценки тесноты связи может быть использован линейный коэффициент корреляции, величина которого в этом случае совпадает с индексом корреляции:

$$\rho_{yx} = r_{yz}, \quad (2.7)$$

где z – преобразованная величина признака-фактора, например:

$$z = \ln x \text{ (или } z = \sqrt{x}, \text{ либо } z = \frac{1}{x}\text{)}.$$

В качестве примера можно привести равностороннюю гиперболу:

$$\hat{y}_x = a + \frac{b}{x}.$$

Заменив $\frac{1}{x}$ на z , имеем линейное уравнение $\hat{y}_z = a + b \cdot z$, для которого

может быть определен линейный коэффициент корреляции: $r_{yz} = b \cdot \frac{\sigma_z}{\sigma_y}$. Возведя

данное выражение в квадрат, получим:

$$r_{yz}^2 = b^2 \cdot \frac{\sigma_z^2}{\sigma_y^2},$$

$$\text{где } \sigma_z^2 = \frac{\sum (z - \bar{z})^2}{n} \text{ и } \sigma_y^2 = \frac{\sum (y - \bar{y})^2}{n}.$$

Отсюда:

$$r_{yz}^2 = b^2 \cdot \frac{\sum (z - \bar{z})^2}{\sum (y - \bar{y})^2}. \quad (2.8)$$

Поскольку по определению коэффициент детерминации характеризует долю дисперсии результативного признака y , объясняемую регрессией, в общей дисперсии результативного признака:

$$r_{yz}^2 = \frac{\sigma_{y \text{ объясн.}}^2}{\sigma_{y \text{ общ.}}^2} = \frac{\sum (\hat{y}_z - \bar{y})^2}{\sum (y - \bar{y})^2}. \quad (2.9)$$

Сравнивая формулы 2.8 и 2.9, мы видим, что знаменатель в обеих формулах совпадает, значит, равны и числители дробей. Следовательно, для линейной регрессии по z :

$$\sigma_{y \text{ объясн.}}^2 = b^2 \cdot \sum (z - \bar{z})^2, \quad (2.10)$$

или

$$\sum (\hat{y}_z - \bar{y})^2 = b^2 \cdot \sum (z - \bar{z})^2. \quad (2.11)$$

К этому же выводу придем, если рассмотрим содержательную сторону расчетного (теоретического) значения признака y , т.е. \hat{y}_z .

Величина \hat{y}_z определяется по уравнению линейной регрессии:

$$\hat{y}_z = a + b \cdot z.$$

Параметр a можно найти как $a = \bar{y} - b \cdot \bar{z}$. Подставив выражение параметра a в линейную модель, получим:

$$\hat{y}_z = a + b \cdot \bar{z} = \bar{y} - b \cdot \bar{z} + b \cdot z = \bar{y} + b \cdot (z - \bar{z})$$

или

$$\hat{y}_z - \bar{y} = b \cdot (z - \bar{z})$$

Возведя в квадрат, получим:

$$(\hat{y}_z - \bar{y})^2 = b^2 \cdot (z - \bar{z})^2 \quad (2.12)$$

Подставим в числитель формулы 2.8 левую часть полученного выражения

2.11:

$$r_{yz}^2 = b^2 \cdot \frac{\sum (z - \bar{z})^2}{\sum (y - \bar{y})^2} = \frac{\sum (\hat{y}_z - \bar{y})^2}{\sum (y - \bar{y})^2}.$$

Но так как

$$\sum (y - \bar{y})^2 = \sum (\hat{y}_z - \bar{y})^2 + \sum (y - \hat{y}_z)^2,$$

$$\sum (\hat{y}_z - \bar{y})^2 = \sum (y - \bar{y})^2 - \sum (y - \hat{y}_z)^2,$$

$$r_{yz}^2 = \frac{\sum (\hat{y}_z - \bar{y})^2}{\sum (y - \bar{y})^2} = \frac{\sum (y - \bar{y})^2 - \sum (y - \hat{y}_z)^2}{\sum (y - \bar{y})^2},$$

$$r_{yz}^2 = 1 - \frac{\sum (y - \hat{y}_z)^2}{\sum (y - \bar{y})^2}$$

Т.е. пришли к формуле индекса корреляции:

$$r_{yz} = \sqrt{1 - \frac{\sum (y - \hat{y}_z)^2}{\sum (y - \bar{y})^2}}.$$

Заменяя z обратно на $\frac{1}{x}$, получим: $\hat{y}_z = \hat{y}_x$, соответственно:

$$r_{yz} = \rho_{yx} \tag{2.13}$$

Аналогично можно показать, что это выражение будет справедливо для полулогарифмической функции $\hat{y}_z = a + b \cdot \ln x$ (а также \sqrt{x}), поскольку здесь преобразования в линейный вид $z = \ln x$ (или $z = \sqrt{x}$) не затрагивают зависимую переменную и требование МНК $\sum (y - \hat{y}_x)^2 \rightarrow \min$ выполнимо.

Индекс детерминации используется для проверки существенности в целом уравнения регрессии по F -критерию Фишера:

$$F = \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} \cdot \frac{n - m - 1}{m}, \quad (2.14)$$

где ρ_{xy}^2 – индекс детерминации, n – число наблюдений, m – число параметров при переменной x . Фактическое значение F -критерия (2.23) сравнивается с табличным при уровне значимости α и числе степеней свободы $k_2 = n - m - 1$ (для остаточной суммы квадратов) и $k_1 = m$ (для факторной суммы квадратов).

Если нелинейное относительно объясняющей переменной уравнение регрессии при линеаризации принимает форму линейного уравнения парной регрессии, то величина линейного коэффициента детерминации в этом случае совпадает с величиной индекса детерминации:

$$r_{yz}^2 = \rho_{yx}^2$$

Следовательно, значения F -критерия Фишера будут также совпадать для нелинейного уравнения регрессии и линеаризованного.

О качестве нелинейного уравнения регрессии можно также судить и по средней ошибке аппроксимации, которая, так же, как и в линейном случае, вычисляется по следующей формуле:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - y_x}{y} \right| \cdot 100\% \quad (2.15)$$

Значение средней ошибки аппроксимации не должно превышать 8-10%, чтобы можно было считать качество аппроксимации исходных данных рассматриваемой моделью хорошим.

3. Контрольные вопросы

1. Нелинейная регрессия. Классы нелинейных регрессий.
2. Регрессии нелинейные относительно включенных в анализ объясняющих переменных.
3. Регрессии нелинейные по оцениваемым параметрам.
4. Коэффициенты эластичности для разных видов регрессионных моделей.
5. Какой должна быть функция регрессии, чтобы применить к ней МНК?
6. Что означает понятие «линеаризация модели»? Как она может быть произведена?
7. Показатели корреляции и детерминации для нелинейной регрессии.
8. F -критерий Фишера для нелинейной регрессии.
9. Почему конкретные значения параметров уравнения регрессии называют их оценками?

4. Построение моделей нелинейной регрессии.

Рассмотрим **пример 1**. По данным проведенного опроса восьми групп семей известны данные связи расходов населения на продукты питания с уровнем доходов семьи.

Таблица 4.1.

Расходы на продукты питания, y , тыс. руб.	9	12	8	22	26	29	33	38
Доходы семьи, x , тыс. руб.	12	31	53	74	96	118	145	187

Предположим, что связь между признаками носит нелинейный характер, и найдем параметры следующих нелинейных уравнений:

$$1) y = a + b \cdot \ln x + \varepsilon,$$

$$2) y = a + b \cdot \sqrt{x} + \varepsilon$$

$$3) y = a \cdot x^b \cdot \varepsilon.$$

2.1. Для нахождения параметров регрессии $y_x = a + b \cdot \ln x$ делаем замену $z = \ln x$ и получаем линейное уравнение вида:

$$\hat{y}_x = a + b \cdot z$$

Определить параметры линейного уравнения данного вида можно с использованием функции «Регрессия» надстройки «Анализ данных» табличного процессора Excel.

Прежде всего, надо сформировать таблицу с исходными данными для анализа на листе Excel. Скриншот листа Excel с расположенной на нем таблицей исходных данных представлен на рис. 1.

Исходные данные задачи			
№п.п.	x	y	
1	2	3	
1	12	9	
2	31	12	
3	53	18	
4	74	22	
5	96	26	
6	118	29	
7	145	33	
8	187	38	

Рис. 1. Скриншот листа Excel с расположенной на нем таблицей исходных данных для решения задачи.

Следует обратить внимание на расположение числовых данных: они должны быть введены в виде столбцов, а не строк, как было представлено условие задачи в табл. 4.1.

Далее нужно составить вспомогательную таблицу, куда заносятся рассчитанные значения переменной z , которая была введена с целью линеаризации логарифмической функции. Вспомогательная таблица может быть, например, такого вида, как таблица 4.2.

Таблица 4.2.

№ п.п.	x	z	y
1	2	3	4
1	12	2,485	9
2	31	3,434	12
3	53	3,970	18
4	74	4,304	22
5	96	4,564	26
6	118	4,771	29
7	145	4,977	33
8	187	5,231	38

При расчетах вручную удобнее расположить столбец значений переменной z перед (слева) от столбца значений y , при расчетах в Excel это значения не имеет, поскольку столбцы значений результатного признака и факторного вводятся отдельно (по очереди) путем задания диапазона ячеек со значениями, например, выделением соответствующих столбцов.

Для того, чтобы рассчитать натуральный логарифм, можно воспользоваться математической функцией LN (число), которая является обратной функцией к EXP, чем можно будет воспользоваться в дальнейшем. Аргументом функции LN должно быть положительное вещественное число, для которого вычисляется натуральный логарифм. Вид окна выбора функции LN представлен на рисунке 2.

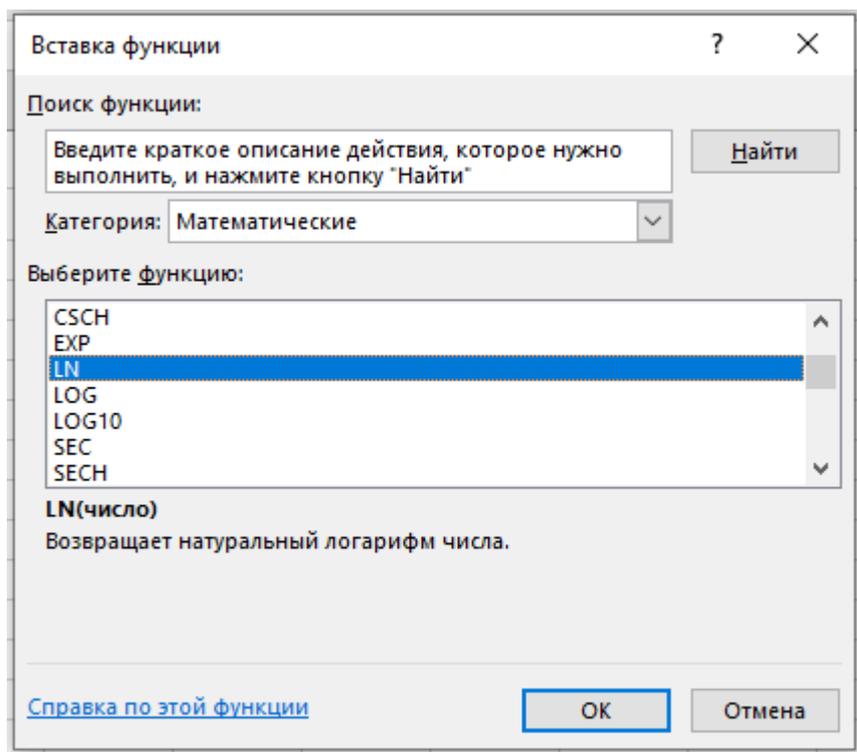


Рис. 2. Скриншот окна выбора функции.

Для того, чтобы рассчитать в ячейке таблицы $\ln(x)$, необходимо сделать данную ячейку активной (установив курсор), ввести знак равенства «=» и, обратившись к функциям, выбрать необходимую. Стрелка на рисунке 3 указывает на знак вызова функций Excel (f_x), кликнув который, получим окно выбора для вставки функции.

Поскольку нужно вычислить $\ln(x)$, аргументами функции будут значения из столбца x . Задать аргумент в окне функции можно, установив курсор в поле **Число**, а затем кликнув на необходимой ячейке из столбца x . Задав функцию, надо нажать кнопку **ОК**. Введя формулу, нужно нажать **Enter** на клавиатуре, после чего в ячейке появится результат расчета по заданной формуле.

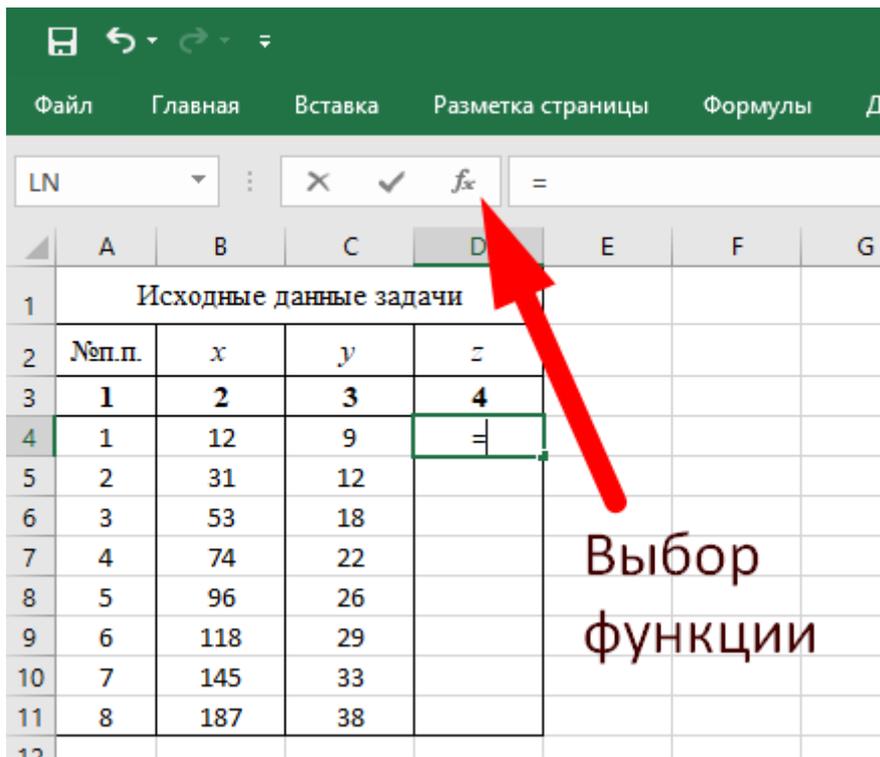


Рис. 3. Скриншот листа Excel при выборе функции.

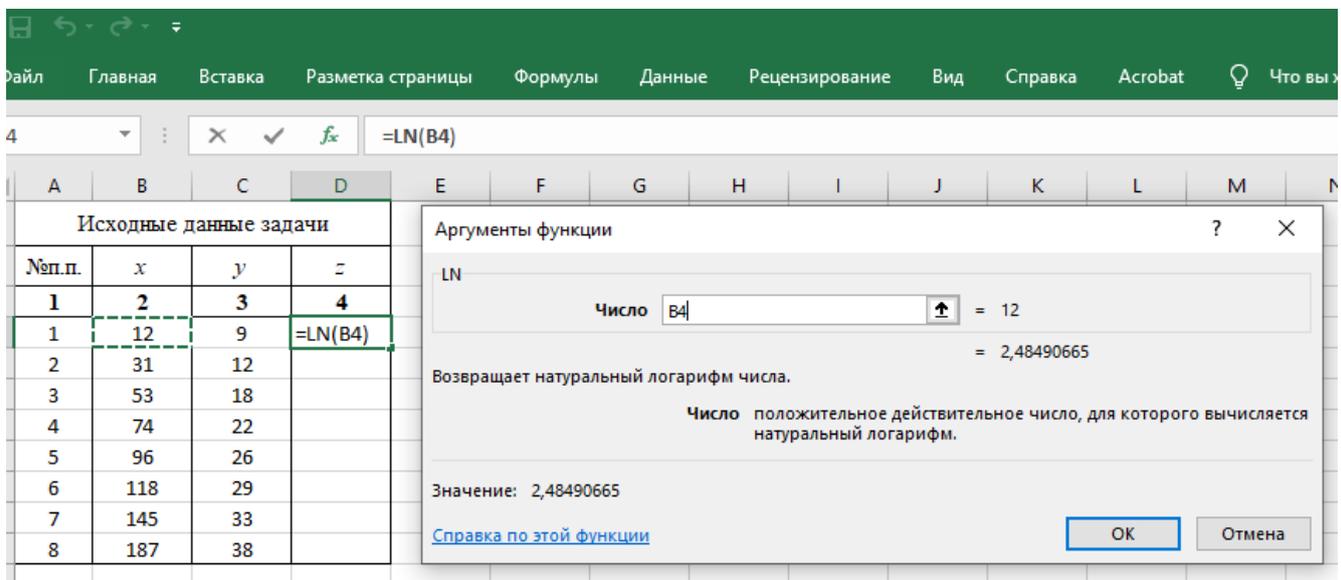


Рис. 4. Скриншот листа Excel при вводе аргумента функции.

Нажав на черный квадратик в правом нижнем углу ячейки, левой кнопкой мыши можно протащить черный крест вниз, заполнив все ячейки столбца z. Результат заполнения представлен на рисунке 5.

The screenshot shows an Excel spreadsheet with the following data:

Исходные данные задачи				
№п.п.	x	y	z	
1	2	3	4	
1	12	9	2,484907	
2	31	12	3,433987	
3	53	18	3,970292	
4	74	22	4,304065	
5	96	26	4,564348	
6	118	29	4,770685	
7	145	33	4,976734	
8	187	38	5,231109	

The formula bar shows the formula $=LN(B11)$ applied to cell D11.

Рис. 5. Скриншот листа Excel с результатами вычислений значений введенной переменной z.

На листе Excel теперь есть значения для независимой переменной z и зависимой переменной y. Для оценки параметров уравнения линейной регрессии $\hat{y}_x = a + b \cdot z$ можно воспользоваться методом наименьших квадратов, который реализуется функцией «Регрессия» надстройки «Анализ данных» табличного процессора Excel. На рисунке 6 стрелками указаны пункты меню, к которым нужно обратиться, чтобы воспользоваться функцией **Регрессия** табличного процессора Excel.

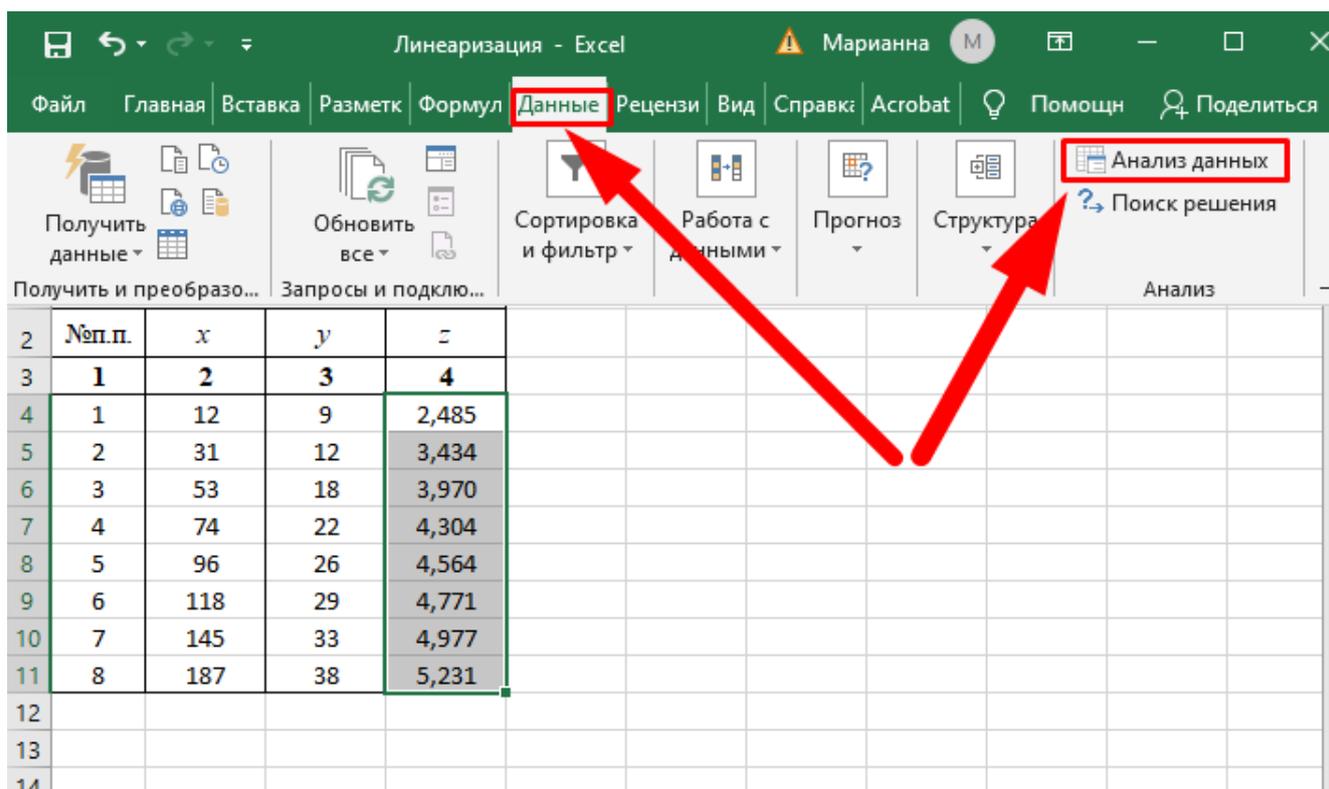


Рис. 6. Скриншот листа Excel с выделенными пунктами меню.

Следует заметить, что пакет анализа должен быть установлен на компьютере. Если же это не было сделано ранее, нужно его установить. Для этого можно зайти во вкладку **Надстройки** (**Файл – Параметры – Надстройки – Надстройки Excel – Перейти**), после чего требуется установить флажок для **Пакета анализа**.

После вызова **Анализа данных** на экране появляется окно с перечнем инструментов анализа. Передвигая бегунок, следует найти требуемую функцию, выделить ее и выбрать, нажав кнопку **ОК**.

Скриншот листа Excel с окном **Анализ данных** и выбранной функцией **Регрессия** представлен на рисунке 7.

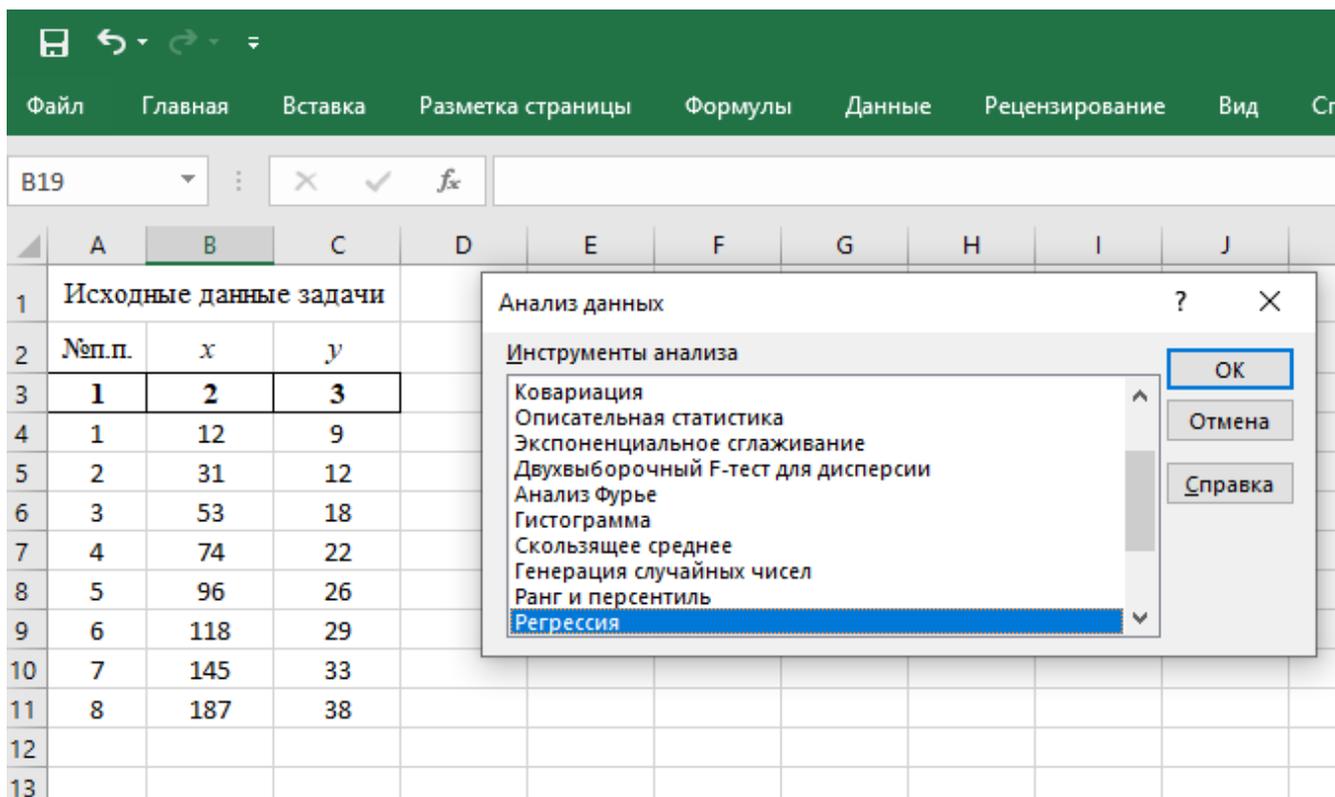


Рис. 7. Скриншот листа Excel с окном «Анализ данных».

После вызова режима **Регрессия** на экране появляется диалоговое окно. Вид диалогового окна **Регрессия** представлен на рис. 8.

В диалоговом окне задаются следующие параметры:

1. *Входной интервал Y* – вводится диапазон адресов ячеек, содержащих значения y_i , т.е. зависимой переменной или признака-результата (ячейки должны составлять один столбец).
2. *Входной интервал X* – вводится диапазон адресов ячеек, содержащих значения независимых переменных или признаков-факторов. Значения каждой переменной представляются одним столбцом. В режиме **Регрессия** можно построить не только линейную парную, но и множественную регрессию. Количество переменных может быть не более 16.

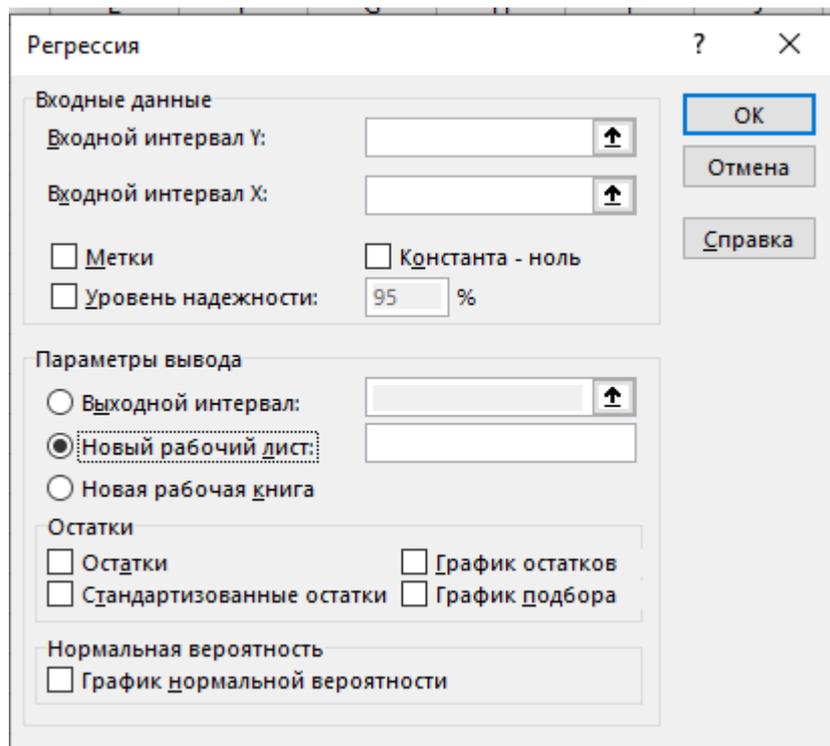


Рис. 8. Скриншот диалогового окна **Регрессия**.

3. *Метки* – включается, если первая строка во входном диапазоне содержит заголовок, чтобы программа не трактовала его как числовые данные. В этом случае автоматически будут созданы стандартные названия.
4. *Уровень надежности* – при включении этого параметра задается надежность при построении доверительных интервалов.
5. *Константа-ноль* – при включении этого параметра коэффициент a (свободный член регрессии) равен 0.
6. *Выходной интервал* – при включении активизируется поле, куда необходимо ввести адрес левой верхней ячейки выходного диапазона, который будет содержать ячейки с результатами вычислений режима **Регрессия**.
7. *Новый рабочий лист* - при включении этого параметра открывается новый лист, в который, начиная с ячейки A1, вставляются результаты работы режима **Регрессия**.

8. *Новая рабочая книга* – при включении этого параметра открывается новая книга, на первом листе которой, начиная с ячейки А1, вставляются результаты работы режима **Регрессия**.
9. *Остатки* – при включении вычисляется столбец, содержащий остатки для всех точек наблюдений (исходных данных) $y_i - \hat{y}_i, i = 1, \dots, n$.
10. *Стандартизованные остатки* – при включении вычисляется столбец, содержащий стандартизованные остатки.
11. *График остатков* – при включении выводятся точечные графики остатков $y_i - \hat{y}_i, i = 1, \dots, n$, в зависимости от значений переменных $x_j, j = 1, \dots, m$. Количество графиков равно числу m переменных x_j .
12. *График подбора* – при включении выводятся точечные графики предсказанных по построенной регрессии значений \hat{y}_i от значений переменных x_j , при $j = 1, \dots, m$. Количество графиков равно числу m переменных x_j .
Для парной регрессии $m = 1$, поскольку признак-фактор в модели (уравнении регрессии) только один. Будет построен один график подбора и один график остатков.

В диалоговом окне **Регрессия** нужно сделать следующее:

- ввести в окне редактирования **Входной интервал Y** диапазон зависимой переменной, что можно сделать, выделив нужный столбец y_i таблицы исходных данных на листе Excel;

Скриншот диалогового окна **Регрессия** после ввода диапазона ячеек со значениями зависимой переменной y путем выделения соответствующего столбца таблицы исходных данных представлен на рисунке 9.

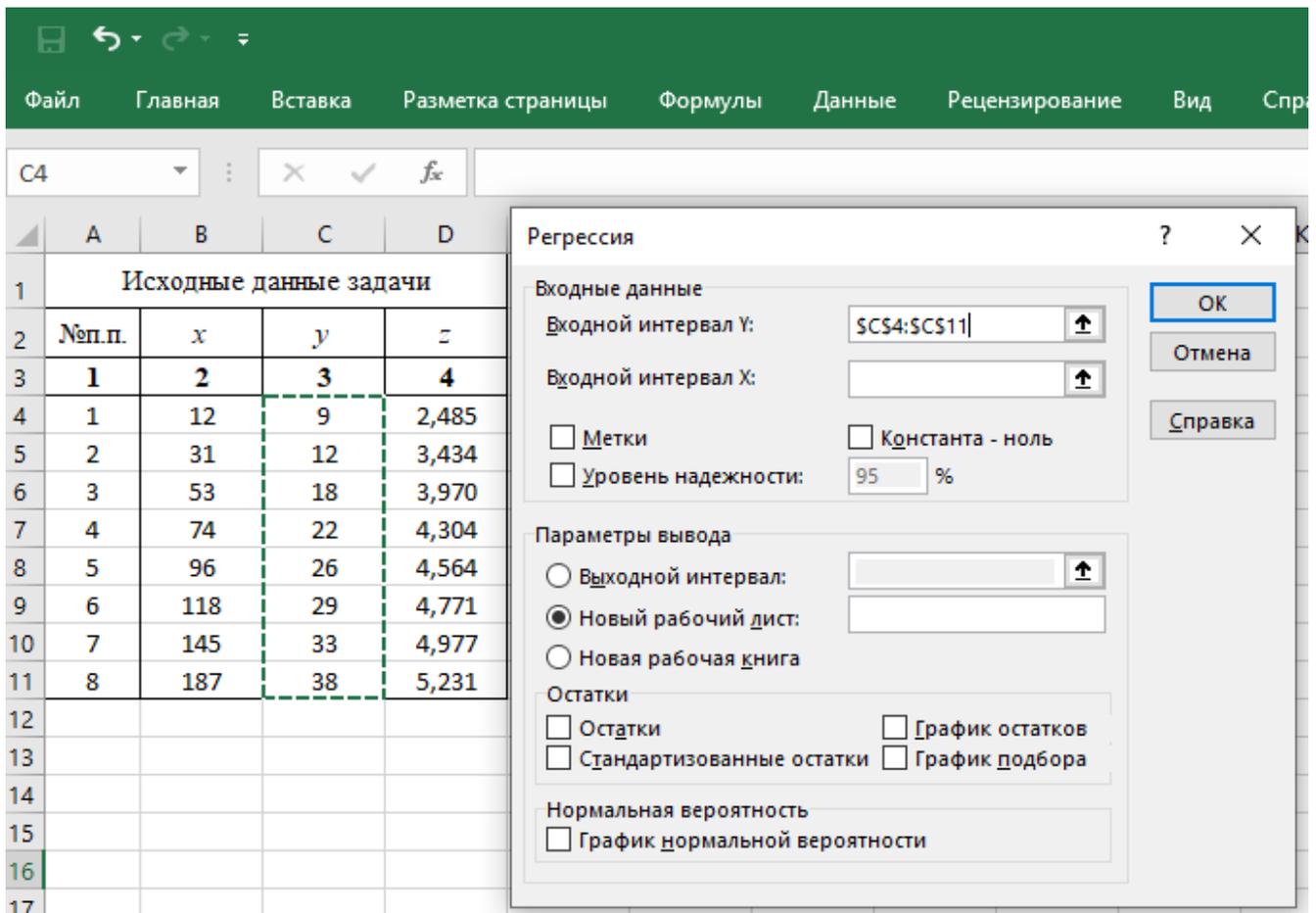


Рис. 9. Скриншот диалогового окна **Регрессия** после ввода диапазона ячеек со значениями зависимой переменной y .

- ввести в окне редактирования **Входной интервал X** диапазон факторной переменной x_i ;

Следует обратить внимание на то, что в данной задаче после введения новой независимой переменной z для определения оценок значений параметров для регрессии вида:

$$\hat{y}_x = a + b \cdot z ,$$

следует вводить в качестве признака-фактора именно z (значения из столбца z таблицы исходных данных).

На рисунке 10 представлен скриншот вида листа Excel с диалоговым окном **Регрессия** после ввода диапазона ячеек со значениями независимой переменной z (выделен столбец z в таблице исходных данных).

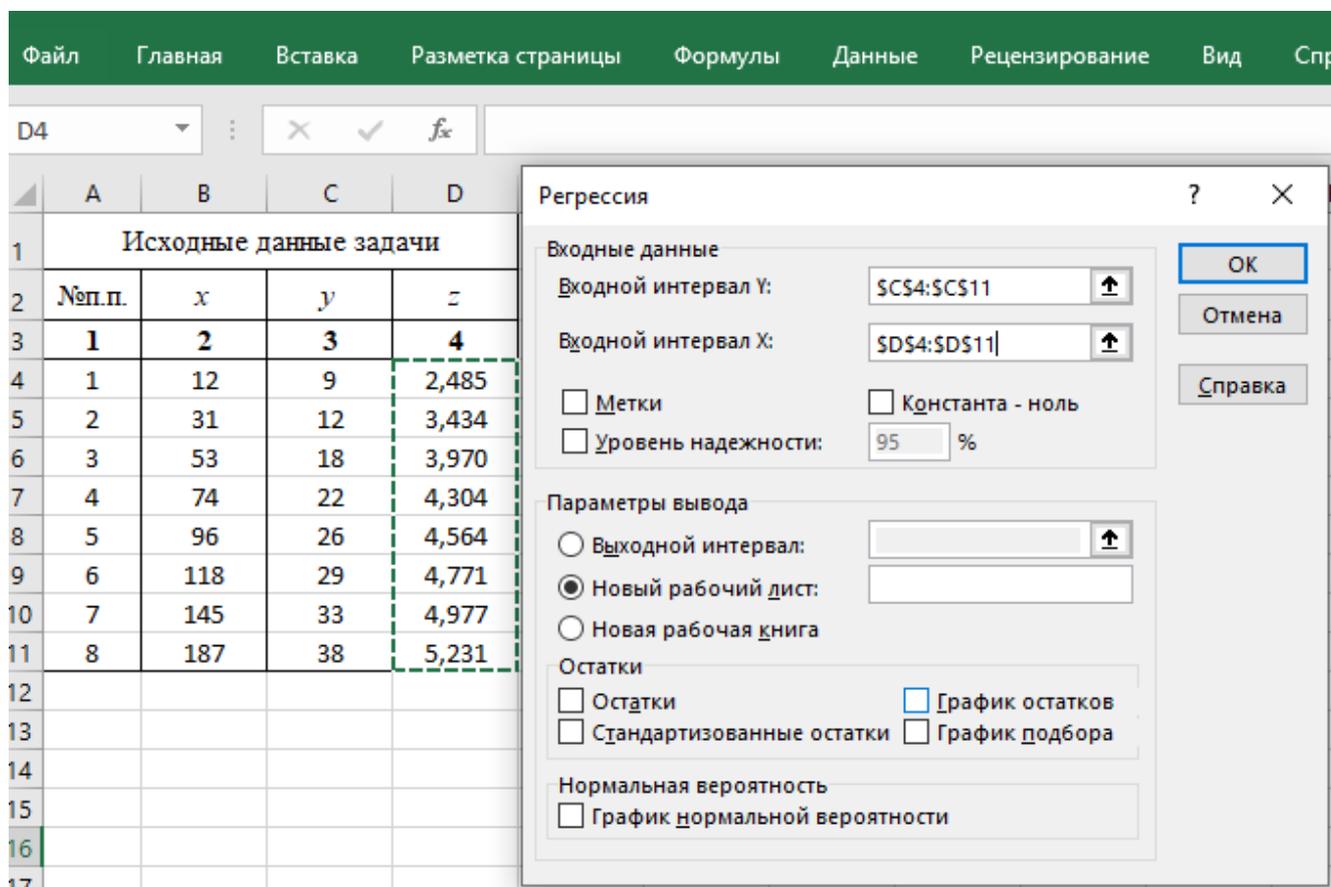


Рис. 10. Скриншот диалогового окна **Регрессия** после ввода диапазона ячеек со значениями независимой переменной z .

-установить флажок **Метки**, если первая строка содержит название столбцов;

Если в диалоговом окне введены только указания на ячейки, содержащие данные, как в иллюстрации решения данной задачи, флажок **Метки** не должен быть установлен.

-установить флажок **Константа-ноль**, если в уравнении регрессии отсутствует свободный член a ; В данной задаче не предполагается равенство нулю

свободного члена регрессии, поэтому флажок **Константа-ноль** не должен быть установлен.

- ввести в окне редактирования **Выходной интервал** номер свободной ячейки на рабочем столе (или выбрать новый лист);

На рис. 10 приведен скриншот диалогового окна **Регрессия**, где в параметрах вывода установлен флажок у поля **Новый рабочий лист**, следовательно вывод результатов будет выполнен на новый лист, т.е. Excel создаст новый рабочий лист и поместит туда таблицы итогов работы и другие результаты, которые были заданы в диалоговом окне **Регрессия**. Если же флажок установлен у поля **Выходной интервал**, то необходимо ввести номер ячейки на текущем листе (левую верхнюю диапозона), начиная с которой будут размещены итоги работы.

- нажать **ОК**.

Таблицы итогов работы режима **Регрессия** приведены для данной задачи на рис. 11.

Вывод итогов								
<i>Регрессионная статистика</i>								
Множественный R	0,959249268							
R-квадрат	0,920159159							
Нормированный R-квадрат	0,906852352							
Стандартная ошибка	3,077788677							
Наблюдения	8							
<i>Дисперсионный анализ</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
Регрессия	1	655,0383	655,0383	69,14951	0,000164			
Остаток	6	56,836699	9,472783					
Итого	7	711,875						
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	-21,77401562	5,5373947	-3,93218	0,007693	-35,32353	-8,2245	-35,3235	-8,2245
Переменная X 1	10,70639008	1,2875038	8,315618	0,000164	7,555982	13,8568	7,555982	13,8568

Рис. 11. Скриншот результатов работы режима **Регрессия**.

Заметим, что таблицы, которые являются результатом работы режима **Регрессия**, являются такими же таблицами, как и все другие таблицы Excel. Это значит, что работать с ними можно так же, как и с другими, меняя шрифт, цвет, размер и многое другое, например, приведя их в более удобный и привычный для восприятия вид в соответствии с потребностями или желаниями.

Дадим краткую интерпретацию показателям, значения которых вычисляются в режиме **Регрессия**.

а) *Регрессионная статистика*:

Сначала рассмотрим показатели, объединенные названием «Регрессионная статистика». Для удобства эта таблица представлена отдельно на рис. 12.

<i>Регрессионная статистика</i>	
Множественный R	0,959249268
R-квадрат	0,920159159
Нормированный R-квадрат	0,906852352
Стандартная ошибка	3,077788677
Наблюдения	8

Рис. 12. Таблица «Регрессионная статистика»

Значения показателей, необходимых решения задачи, округлим до трех знаков после запятой.

- Множественный R – для данной задачи – это коэффициент парной линейной корреляции $r_{zy} = 0,959$;
- R-квадрат- коэффициент детерминации $r^2_{zy} = 0,920$;
- Нормированный R-квадрат - приведенный коэффициент детерминации

$$R^2 = 0,907$$

- Стандартная ошибка – оценка Se для среднеквадратического отклонения B .

- наблюдения – число наблюдений $n = 8$;

Значение коэффициента парной линейной корреляции говорит о высокой тесноте связи между результатным показателем и факторным.

Значение коэффициента детерминации можно выразить в процентах. Он показывает, что на 92% вариация результативного признака определяется вариацией факторного признака (z) в этой модели.

б) *Дисперсионный анализ:*

Показатели, объединенные названием «Дисперсионный анализ» для удобства представлены на рис. 13.

Дисперсионный анализ					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	Значимость <i>F</i>
Регрессия	1	655,0383	655,0383	69,14951	0,000164
Остаток	6	56,836699	9,4727831		
Итого	7	711,875			

Рис. 13. Таблица «Дисперсионный анализ»

- столбец *df* – число степеней свободы для уравнения регрессии (строка **Регрессия**), для остаточной вариации (строка **Остаток**), и общая вариация (строка **Итого**). Для строки *Регрессия* показатель равен t – числу параметров при переменной x (z в данной задаче, т.е. 1); для строки *Остаток* – равен $n-t-1$ (т.е. 6); для строки *Итого* (общая дисперсия) – $n-1$ или числу наблюдений без единицы, т.е. 7;

- столбец *SS* содержит суммы квадратов отклонений: сумму квадратов отклонений теоретических данных от среднего значения (строка **Регрессия**), сумму

квадратов отклонений фактических данных от теоретических (строка **Остаток**) и сумму квадратов отклонений фактических данных от среднего значения (строка **Итого**);

- в столбце MS показаны дисперсии на одну степень свободы: объясненная (факторная) дисперсия (для строки **Регрессия**) и остаточная дисперсия (для строки **Остаток**);

- в столбце F показано расчетное значение F -критерия Фишера, которое сравнивают с табличным $F_{\text{табл}}(\alpha; k_1; k_2)$ при уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n - m - 1$. При этом, если фактическое значение F -критерия больше табличного, то признается статистическая значимость уравнения в целом. Таблица значений F -критерия Фишера при уровне значимости $\alpha = 0,05$ приведена в Приложении 1;

Для нашей задачи, т.е. при $\alpha=0,05$, $k_1=1$, $k_2=8-1-1=6$ $F_{\text{табл}}$ равно 5,99. Расчетное значение F -критерия равно 69,15, что значительно превышает табличное, т.е. следует признать статистическую значимость уравнения регрессии в целом.

- в столбце **значимость F** показано значение уровня значимости, соответствующее вычисленной величине F -критерия и равное вероятности того, что расчетное значение F -критерия меньше или равно табличному. Если вероятность меньше уровня значимости α (обычно $\alpha = 0,05$), то построенная регрессия является значимой;

Значимость F рассчитанная на два порядка меньше уровня значимости $\alpha = 0,05$, следовательно построенная регрессия является значимой.

в) *Перейдем к следующей группе показателей, объединенных в таблице, показанной на рис. 14.*

	<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	-21,77401562	5,5373947	-3,9321769	0,007693	-35,32353	-8,2245
Переменная X 1	10,70639008	1,2875038	8,3156184	0,000164	7,555982	13,8568

Рис. 14. Продолжение результатов работы режима *Регрессия*

Приведенная на рис. 14 таблица включает, кроме оценок параметров, также их среднеквадратические ошибки, вероятности ошибочного решения (P-значение), нижние и верхние интервальные оценки параметров с вероятностью 95%.

В столбце *Коэффициенты* показаны значения коэффициентов уравнения регрессии. В строке *Y-пересечение* – представлено значение параметра *a*, в строке *Переменная x1* – значение параметра *b*.

В данном случае это свободный член и коэффициент регрессии для следующего уравнения:

$$\hat{y}_z = a + b \cdot z$$

Подставим найденные МНК значения для параметров в уравнение регрессии:

$$\hat{y}_z = -21,774 + 10,706 \cdot z$$

Полученное методом линеаризации искомое уравнение регрессии имеет вид:

$$\hat{y}_x = -21,774 + 10,706 \cdot \ln x$$

Продолжим рассмотрение показателей из группы, представленной в таблице на рис. 14.

В столбце *Стандартная ошибка* приведены значения стандартных ошибок для параметров регрессии;

В столбце *t-статистика* – значения статистик Стьюдента, рассчитанные для соответствующих параметров регрессии. Фактические (расчетные) значения сравнивают с табличными, если фактическое значение больше табличного, то признается статистическая значимость параметра;

В столбце *P-значение* содержатся вероятности случайных событий непревышения расчетной статистикой Стьюдента для соответствующего параметра регрессии табличного значения. Если эта вероятность меньше уровня значимости α (обычно $\alpha = 0,05$), то принимается гипотеза о значимости соответствующего коэффициента регрессии.

Табличное значение *t-критерия* Стьюдента находят по таблице, при $\alpha = 0,05$ и числе степеней свободы $n - 2$. Таблица критических значений *t* -критерия Стьюдента при уровне значимости 0,10; 0,05; 0,01 приведена в Приложении 1;

Для рассматриваемой задачи $n - 2 = 6$, $t_{крит} = 2,4469$, т.е. расчетные значения *t* – статистики для обоих параметров больше критического (табличного) значения, следовательно, принимается гипотеза о значимости обоих параметров регрессии.

Столбцы *Нижние 95%* и *Верхние 95%* - содержат соответственно нижние и верхние границы интервалов для оцениваемых параметров *a* и *b*.

Для оценки аппроксимации построенной моделью исходных данных рассчитывают среднюю ошибку аппроксимации по формуле 2.8:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - y_x}{y} \right| \cdot 100\%$$

Прежде чем выполнять расчет данного показателя, обратимся к группе показателей, объединенных в таблице ВЫВОД ОСТАТКА, представленной на рисунке 15.

з) ВЫВОД ОСТАТКА

ВЫВОД ОСТАТКА		
<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>
1	4,830364271	4,1696357
2	14,9915909	-2,991591
3	20,73347832	-2,733478
4	24,30698418	-2,306984
5	27,09367656	-1,093677
6	29,30279489	-0,302795
7	31,50883713	1,4911629
8	34,23227376	3,7677262

Рис. 15. Таблица ВЫВОД ОСТАТКА

Данная таблица выводится, если в диалоговом окне режима Регрессия был задан параметр *Остатки*.

Столбец *Наблюдение* – содержит номера наблюдений;

Столбец *Предсказанное Y* – содержит значения \hat{y}_i , вычисленные по построенному уравнению регрессии;

Столбец *Остатки* – включает значения остатков $y_i - \hat{y}_i$;

Если в диалоговом окне режима *Регрессия* был задан параметр *График подбора*, будет построен точечный график предсказанных по построенной регрессии значений \hat{y}_i от значений переменных $x_j, j=1, \dots, m$. На этом же графике есть точки исходных наблюдений из условия задачи.

Поскольку для решения задачи была произведена замена переменной (введение новой переменной), то будет построен график зависимости от переменной z , а не от x , т.е. не той зависимости, которая анализируется.

График изучаемой зависимости по имеющимся данным можно построить, используя средства построения диаграмм Excel. Такой график зависимости для модели логарифмической функции, с точками исходных данных приведен на рисунке 16.

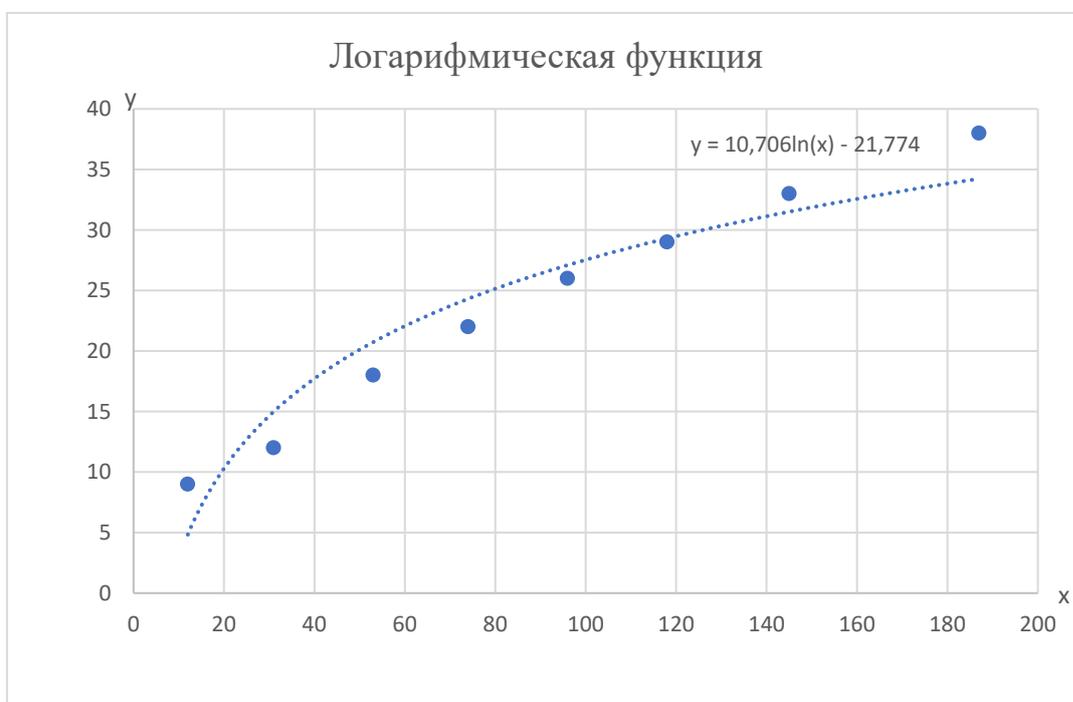


Рис. 16. График модели логарифмической функции и точечной диаграммы исходных данных.

Уравнение нелинейной регрессии, так же, как и линейной зависимости, дополняется показателем корреляции, а именно индексом корреляции. Как было показано в разделе 2 (2.13), если нелинейное относительно объясняющей переменной уравнение регрессии при линеаризации принимает форму линейного уравнения парной регрессии, то для оценки тесноты связи может быть использован

линейный коэффициент корреляции, величина которого в этом случае совпадает с индексом корреляции и приведена в таблице на рис. 12:

$$r_{yz} = \rho_{yx} = 0,959,$$

Коэффициент (и индекс) детерминации: $r_{yz}^2 = \rho_{yx}^2 = 0,920$.

F -критерий Фишера для нелинейной модели:

$$F = \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} \cdot \frac{n - m - 1}{m} = \frac{0,920}{1 - 0,920} \cdot \frac{8 - 1 - 1}{1} = 69,1495$$

Поскольку индекс детерминации и линейный коэффициент детерминации, рассчитанный в Excel и приведенный в таблице на рис.12, для данной модели совпадают $r_{yz}^2 = \rho_{yx}^2$, то и расчетные значения F -критерий Фишера тоже будут совпадать для обеих функций

$$\hat{y}_z = -21,774 + 10,706 \cdot z$$

и

$$\hat{y}_x = -21,774 + 10,706 \cdot \ln x.$$

Соответственно, F -критерий Фишера, рассчитанный в Excel и приведенный в таблице дисперсионного анализа на рис.13, будет совпадать со значением F -критерия Фишера для нелинейной модели:

$$F = 69,1495.$$

Поскольку для каждого конкретного i предсказанные по регрессии теоретические значения зависимой переменной y от x_i и от z_i будут одинаковыми, можно использовать это в дальнейших расчетах:

$$\hat{y}_i = \hat{y}_i(x_i) = \hat{y}_i(z_i).$$

В столбце *Остатки* таблицы ВЫВОД ОСТАТКА уже рассчитаны значения $y_i - \hat{y}_i$. Можно этим воспользоваться для расчета показателя средней относительной аппроксимации по формуле 2.8.

Далее необходимо каждое значение i -го остатка разделить на y_i и взять по модулю, воспользовавшись функцией ABS. Скриншот обращения к математической функции ABS приведен на рисунке 17.

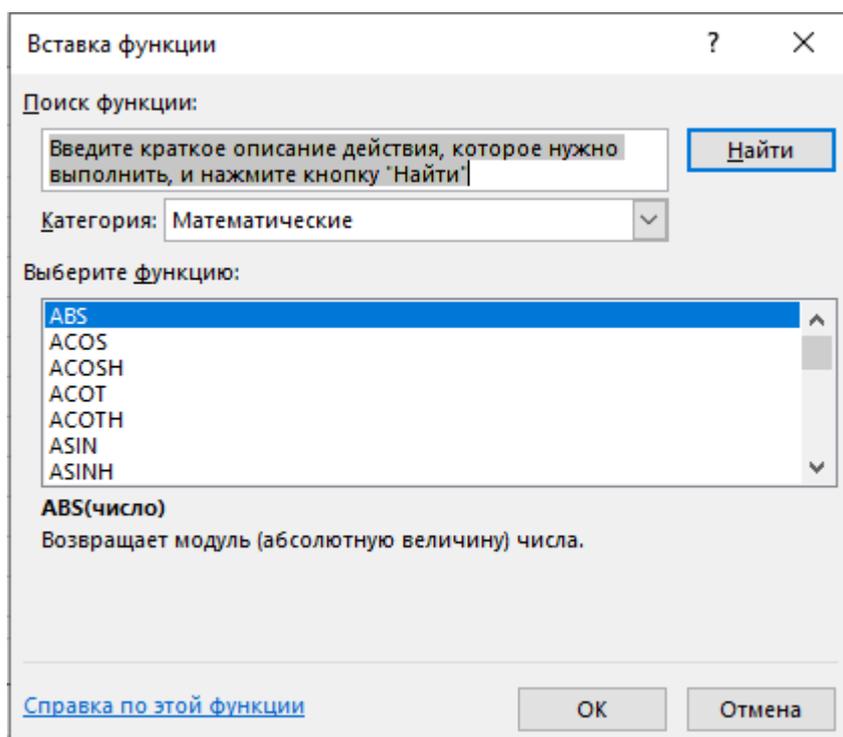


Рис. 17. Скриншот окна **Вставка функции** при обращении к математической функции ABS.

Далее необходимо найти сумму модулей и разделить ее на число наблюдений (строк исходных данных).

Такие расчеты легко выполняются в таблицах Excel.

Построим вспомогательную таблицу 4.3 для расчета средней относительной аппроксимации по формуле 2.8.

Таблица 4.3.

Расчет средней ошибки аппроксимации

№п. п. (i)	y_i	\hat{y}_i	$y_i - \hat{y}_i$	$\frac{y_i - \hat{y}_i}{y_i}$	$\left \frac{y_i - \hat{y}_i}{y_i} \right $
1	9	4,83036	4,169636	4,169636	0,463293
2	12	14,99159	-2,99159	-2,99159	0,249299
3	18	20,73348	-2,73348	-2,73348	0,15186
4	22	24,30698	-2,30698	-2,30698	0,10486
5	26	27,09368	-1,09368	-1,09368	0,04206
6	29	29,30279	-0,30279	-0,30279	0,01044
7	33	31,50884	1,49116	1,491163	0,045187
8	38	34,23227	3,76773	3,767726	0,09915
Итого					1,166158

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - y_x}{y} \right| \cdot 100\% = \frac{1,166158}{8} = 0,1458 \cdot 100\% = 14,58\%$$

Средняя ошибка аппроксимации равна 14,58 %, что недопустимо велико.

2.2. Рассмотрим следующую модель нелинейной регрессии:

$$y_x = a + b \cdot \sqrt{x}.$$

Для нахождения параметров регрессии выполняем замену переменной $z = \sqrt{x}$ и получаем линейное уравнение вида:

$$\hat{y}_z = a + b \cdot z$$

Для построения данной нелинейной модели с применением функции «Регрессия» надстройки «Анализ данных» удобно создать новый лист табличного процессора Excel, куда можно скопировать таблицу со значениями измерений (исходными данными).

Исходные данные x и y нужно расположить по столбцам. Столбец новой переменной z можно получить, обратившись к математической функции КОРЕНЬ, которая возвращает значение квадратного корня. Скриншот окна Вставка функции при обращении к функции КОРЕНЬ, представлен на рисунке 18.

Далее необходимо установить маркер на черный квадратик и, нажав левую кнопку мыши, протащить черным крестом вниз, чтобы заполнить столбец z таблицы данных для корреляционно-регрессионного анализа. После этого на листе будет сформирована таблица, где будут необходимые для построения регрессии значения независимой переменной z и зависимой (результатной) переменной y . Для построения парной линейной регрессии необходимо будет, как и в предыдущем примере, обратиться к функции «Регрессия» надстройки «Анализ данных», введя в окне редактирования *Входной интервал* Y диапазон зависимой переменной, что можно сделать, выделив нужный столбец y_i таблицы исходных данных на листе Excel.

Затем ввести в окне редактирования *Входной интервал* X диапазон факторной переменной z_i .

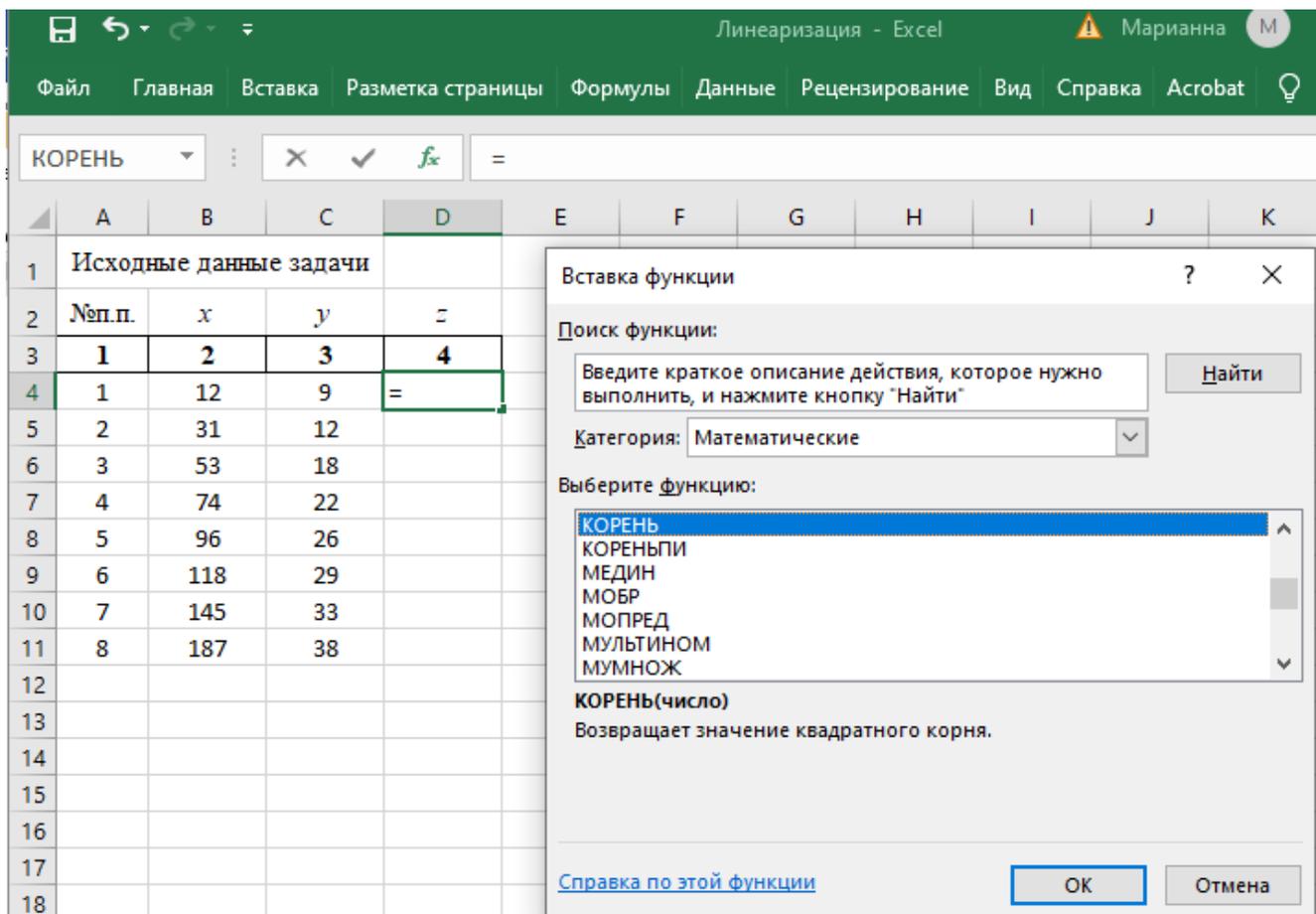


Рис. 18. Скриншот листа Excel с окном **Вставка функции** при обращении к математической функции КОРЕНЬ.

Для определения оценок значений параметров для регрессии вида:

$$\hat{y}_z = a + b \cdot z ,$$

следует вводить в качестве признака-фактора именно z (значения из столбца z таблицы исходных данных).

Скриншот таблицы исходных данных на листе Excel представлен на рисунке 19.

	A	B	C	D	E	F
1	Исходные данные задачи					
2	№п.п.	x	y	z		
3	1	2	3	4		
4	1	12	9	3,464102		
5	2	31	12	5,567764		
6	3	53	18	7,28011		
7	4	74	22	8,602325		
8	5	96	26	9,797959		
9	6	118	29	10,86278		
10	7	145	33	12,04159		
11	8	187	38	13,67479		
12						

Рис. 19. Скриншот листа Excel с таблицей исходных данных для корреляционно-регрессионного анализа.

Скриншот листа Excel с окном функции **Регрессия** для корреляционно-регрессионного анализа, где заданы входные интервалы данных для зависимой переменной y и независимой переменной z представлены на рисунке 20.

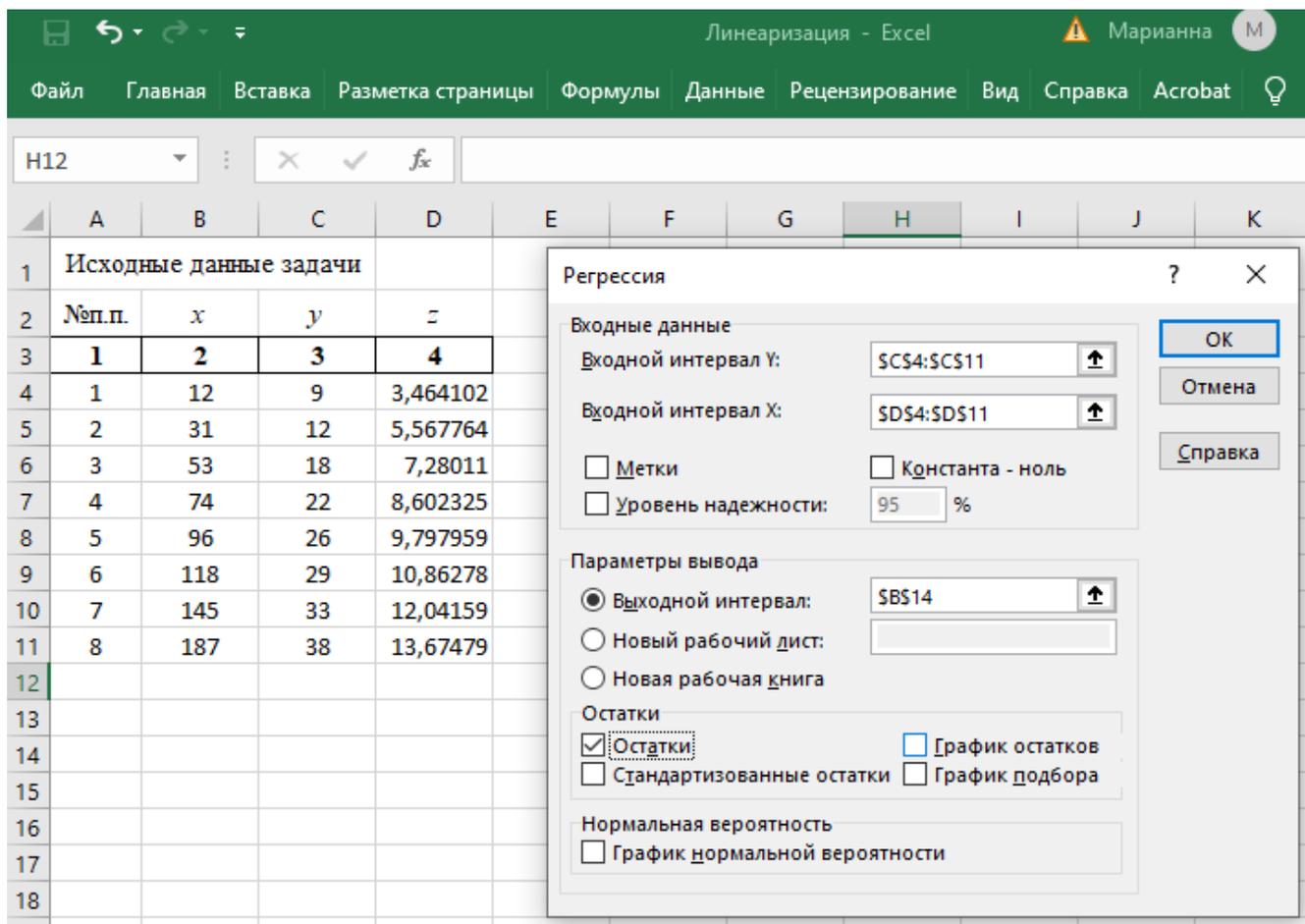


Рис. 20. Скриншот листа Excel с таблицей исходных данных и окном функции **Регрессия** для корреляционно-регрессионного анализа.

В диалоговом окне введены только указания на ячейки, содержащие данные для решения данной задачи, поэтому флажок **Метки** не должен быть установлен.

Был задан параметр *Остатки* для печати таблицы **ВЫВОД ОСТАТКА**.

После нажатия клавиши **ОК** на выбранном месте листа (или на новом листе) появятся таблицы с результатами работы режима **Регрессия**. Таблицы итогов работы режима **Регрессия** приведены для данной задачи на рис. 21.

ВЫВОД ИТОГОВ						
<i>Регрессионная статистика</i>						
Множественный R	0,995649296					
R-квадрат	0,99131752					
Нормированный R-квадрат	0,98987044					
Стандартная ошибка	1,014958161					
Наблюдения	8					
<i>Дисперсионный анализ</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	1	705,6941596	705,6942	685,0468	2,05211E-07	
Остаток	6	6,18084041	1,03014			
Итого	7	711,875				
	<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	-2,978704792	1,068921313	-2,78665	0,031718	-5,594261021	-0,36315
Переменная X 1	2,9572929	0,112988494	26,1734	2,05E-07	2,680820016	3,233766
ВЫВОД ОСТАТКА						
<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>				
1	7,265658321	1,734341679				
2	13,48680523	-1,486805229				
3	18,5507125	-0,550712497				
4	22,46089065	-0,460890647				
5	25,99672971	0,003270289				
6	29,14571883	-0,145718833				
7	32,63181736	0,368182635				
8	37,4616674	0,538332603				

Рис. 21. Скриншот результатов работы режима **Регрессия**.

Сначала рассмотрим показатели, объединенные названием «Регрессионная статистика». Значения показателей, необходимых решения задачи, округлим до трех знаков после запятой.

- Множественный R – для данной задачи – это коэффициент парной линейной корреляции $r_{zy} = 0,996$;

- R -квадрат- коэффициент детерминации $r^2_{zy} = 0,991$;

Значение коэффициента парной линейной корреляции говорит о высокой тесноте связи между результатным показателем и факторным.

Значение коэффициента детерминации можно выразить в процентах. Он показывает, что на 99,1% вариация результативного признака определяется вариацией факторного признака (z) в этой модели.

Рассмотрим показатели, объединенные названием «Дисперсионный анализ», представленные на рис. 21.

- в столбце F показано расчетное значение F -критерия Фишера, которое сравнивают с табличным $F_{\text{табл}}(\alpha; k_1; k_2)$ при уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n - m - 1$. При этом, если фактическое значение F -критерия больше табличного, то признается статистическая значимость уравнения в целом. Таблица значений F -критерия Фишера при уровне значимости $\alpha = 0,05$ приведена в Приложении 1;

Для нашей задачи, т.е. при $\alpha=0,05$, $k_1=1$, $k_2=8-1-1=6$ $F_{\text{табл}}$ равно 5,99. Расчетное значение F -критерия равно 685,05, что значительно превышает табличное, т.е. признается статистическая значимость уравнения регрессии в целом.

- в столбце **значимость F** показано значение уровня значимости, соответствующее вычисленной величине F -критерия и равное вероятности того,

что расчетное значение F -критерия меньше или равно табличному. Если вероятность меньше уровня значимости α (обычно $\alpha = 0,05$), то построенная регрессия является значимой;

Значимость F рассчитанная на несколько порядков меньше уровня значимости $\alpha = 0,05$, следовательно построенная регрессия является значимой.

Обратимся к следующей группе показателей, объединенных ниже в таблице, показанной также на рис. 21.

Эта таблица включает, кроме оценок параметров, также их среднеквадратические ошибки, вероятности ошибочного решения (Р-значение), нижние и верхние интервальные оценки параметров с вероятностью 95%.

В столбце *Коэффициенты* показаны значения коэффициентов уравнения регрессии. В строке *Y-пересечение* – представлено значение параметра a , в строке *Переменная x1* – значение параметра b .

В данном случае это свободный член и коэффициент регрессии для следующего уравнения:

$$\hat{y}_z = a + b \cdot z$$

Подставим найденные МНК значения для параметров в уравнение регрессии:

$$\hat{y}_z = -2,979 + 2,957 \cdot z$$

Полученное методом линеаризации искомое уравнение регрессии имеет вид:

$$\hat{y}_x = -2,979 + 2,957 \cdot \sqrt{x}$$

В столбце *Стандартная ошибка* приведены значения стандартных ошибок для параметров регрессии;

В столбце *t-статистика* – значения статистик Стьюдента, рассчитанные для соответствующих параметров регрессии. Фактические (расчетные) значения

сравнивают с табличными, если фактическое значение больше табличного, то признается статистическая значимость параметра;

Для рассматриваемой задачи $n - 2 = 6$, $t_{крит} = 2,4469$, т.е. расчетные значения t – статистики для обоих параметров больше критического (табличного) значения, следовательно, принимается гипотеза о значимости обоих параметров регрессии.

В столбце *P-значение* содержатся вероятности случайных событий непревышения расчетной статистикой Стьюдента для соответствующего параметра регрессии табличного значения. Если эта вероятность меньше уровня значимости α (обычно $\alpha = 0,05$), то принимается гипотеза о значимости соответствующего коэффициента регрессии. Значения вероятности, представленные в рассматриваемой таблице, ниже уровня значимости $\alpha = 0,05$, следовательно коэффициенты регрессии признаются статистически значимыми.

Столбцы *Нижние 95%* и *Верхние 95%* - содержат соответственно нижние и верхние границы интервалов для оцениваемых параметров a и b .

Если в диалоговом окне режима *Регрессия* задать параметр *График подбора*, будет построен точечный график предсказанных по построенной регрессии значений \hat{y}_i от значений переменных $x_j, j=1, \dots, m$. На этом же графике есть точки исходных наблюдений из условия задачи.

Поскольку для решения задачи была произведена замена переменной (введение новой переменной), то (в наших обозначениях) будет построен график зависимости от переменной z , а не от x , т.е. не той зависимости, которая анализируется. График изучаемой зависимости по имеющимся данным можно построить, используя средства построения диаграмм Excel.

Такой график зависимости для модели функции $\hat{y}_x = -2,979 + 2,957 \cdot \sqrt{x}$, с точками исходных данных приведен на рисунке 22.

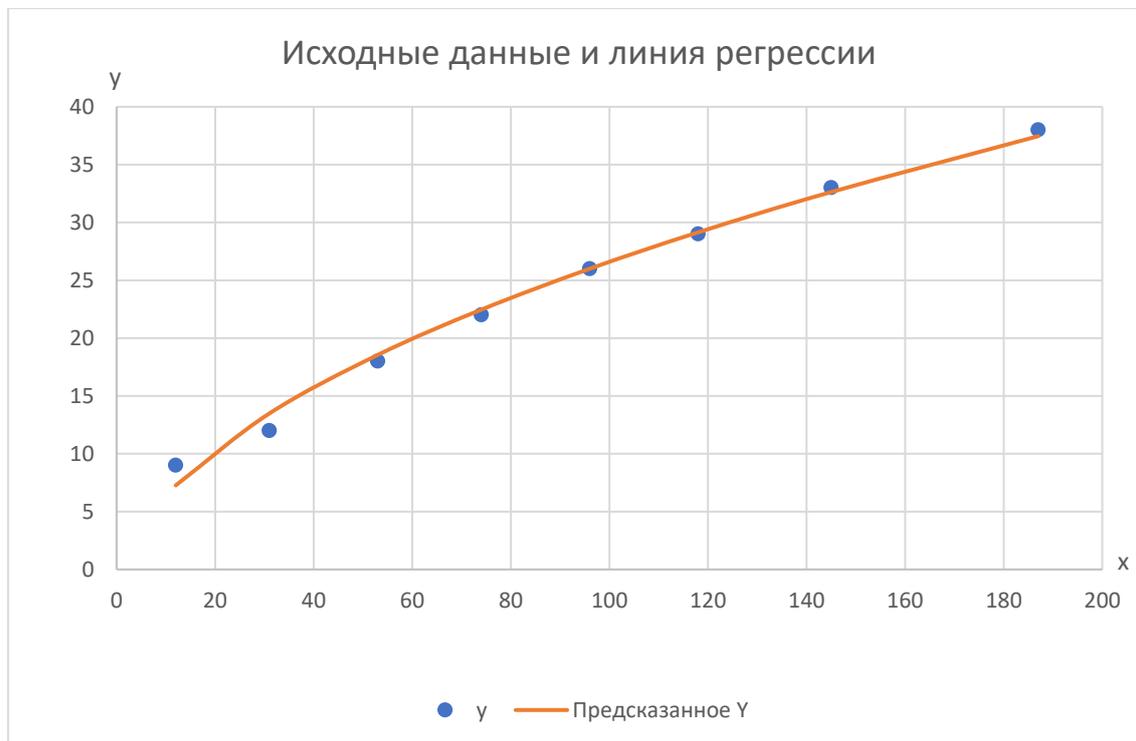


Рис. 22. График нелинейной регрессии $\hat{y}_x = -2,979 + 2,957 \cdot \sqrt{x}$ и точечной диаграммы исходных данных.

Как было показано в разделе 2 (2.13), если нелинейное относительно объясняющей переменной уравнение регрессии при линейаризации принимает форму линейного уравнения парной регрессии, то для оценки тесноты связи может быть использован линейный коэффициент корреляции, величина которого в этом случае совпадает с индексом корреляции и приведена в таблице на рис. 19:

$$r_{yz} = \rho_{yx} = 0,996,$$

Коэффициент (и индекс) детерминации: $r_{yz}^2 = \rho_{yx}^2 = 0,991$.

Уравнение нелинейной регрессии, так же, как и линейной зависимости, дополняется показателем корреляции, а именно индексом корреляции. Как было показано в разделе 2 (2.13), если нелинейное относительно объясняющей переменной уравнение регрессии при линеаризации принимает форму линейного уравнения парной регрессии, то для оценки тесноты связи может быть использован линейный коэффициент корреляции, величина которого в этом случае совпадает с индексом корреляции и приведена в таблице на рис. 21:

$$r_{yz} = \rho_{yx} = 0,996,$$

$$\text{Коэффициент (и индекс) детерминации: } r_{yz}^2 = \rho_{yx}^2 = 0,991.$$

F -критерий Фишера для нелинейной модели:

$$F = \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} \cdot \frac{n - m - 1}{m} = \frac{0,991}{1 - 0,991} \cdot \frac{8 - 1 - 1}{1} = 685,05$$

Поскольку индекс детерминации и линейный коэффициент детерминации, рассчитанный в Excel и приведенный в таблице на рис.21, для данной модели совпадают $r_{yz}^2 = \rho_{yx}^2$, то и расчетные значения F -критерий Фишера тоже будут совпадать для обеих функций

$$\hat{y}_z = -2,979 + 2,957 \cdot z$$

и

$$\hat{y}_x = -2,979 + 2,957 \cdot \sqrt{x}.$$

Соответственно, F -критерий Фишера, рассчитанный в Excel и приведенный в таблице дисперсионного анализа на рис.21, будет совпадать со значением F -критерия Фишера для нелинейной модели:

$$F = 685,05.$$

Поскольку для каждого конкретного i предсказанные по регрессии теоретические значения зависимой переменной y от x_i и от z_i будут одинаковыми, можно использовать это в дальнейших расчетах:

$$\hat{y}_i = \hat{y}_i(x_i) = \hat{y}_i(z_i).$$

В столбце *Остатки* таблицы ВЫВОД ОСТАТКА уже рассчитаны значения $y_i - \hat{y}_i$. Скриншот таблицы ВЫВОД ОСТАТКА представлен на рисунке 21.

Вспользуемся этим для расчета показателя средней относительной аппроксимации по формуле 2.8:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - y_x}{y} \right| \cdot 100\%$$

Каждое значение i -го остатка необходимо разделить на y_i и взять по модулю, воспользовавшись функцией ABS. Скриншот обращения к математической функции ABS приведен на рисунке 17.

Потом нужно найти сумму модулей и разделить ее на число наблюдений (строк исходных данных).

Такие расчеты легко выполняются в таблицах Excel.

В качестве аргумента функции ABS можно задать выражение, в частности, разделив остаток $y_i - \hat{y}_i$ на y_i . Сократив таким образом объем вычислений, можно ограничиться таблицей из трех столбцов.

Построим вспомогательную таблицу 4.4 для расчета средней относительной аппроксимации по формуле 2.8.

Таблица 4.4.

Расчет средней относительной ошибки аппроксимации

№п.п. (i)	y_i	$y_i - \hat{y}_i$	$\left \frac{y_i - \hat{y}_i}{y_i} \right $
1	9	1,73434	0,19270
2	12	-1,48681	0,12390
3	18	-0,55071	0,030595
4	22	-0,46089	0,02095
5	26	0,00327	0,000126
6	29	-0,14572	0,00502
7	33	0,36818	0,01116
8	38	0,53833	0,01417
Итого			0,3986

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - y_x}{y} \right| \cdot 100\% = \frac{0,3986}{8} = 0,049828 \cdot 100\% = 4,98\%$$

Средняя ошибка аппроксимации равна 4,98 %, что говорит о хорошей аппроксимации исходных данных построенной моделью.

2.3. Рассмотрим следующую модель нелинейной регрессии:

$$y = a \cdot x^b \cdot \varepsilon.$$

Для нахождения параметров регрессии $y = a \cdot x^b \cdot \varepsilon$ необходимо провести ее линеаризацию, как было показано выше:

$$Y = A + b \cdot X + E,$$

где $Y = \ln y$, $X = \ln x$, $A = \ln a$, $E = \ln \varepsilon$.

Для того, чтобы найти параметры линеаризованной регрессии A и b , можно воспользоваться функцией «Регрессия» надстройки «Анализ данных» табличного процессора Excel.

Прежде всего, сформируем таблицу с исходными данными для анализа на листе Excel. Для того, чтобы определить параметры A и b необходимо ввести диапазон значений зависимой переменной (результатного признака), в качестве которой выступает $Y = \ln y$, и независимой переменной (признака-фактора), $X = \ln x$.

Для построения вспомогательной таблицы можно воспользоваться математической функцией **LN**, которую уже использовали при построении первой нелинейной модели. Аргументом функции **LN** должно быть положительное вещественное число, для которого вычисляется натуральный логарифм. Вид окна выбора функции **LN** был представлен на рисунке 2.

Поскольку нужно вычислить $X = \text{LN}(x)$ и $Y = \text{LN}(y)$, аргументами функции будут значения из столбцов x и y . Задать аргумент в окне функции можно, установив курсор в поле **Число**, а затем кликнув на необходимой ячейке из столбца x . Задав функцию, надо нажать кнопку **ОК**. Введя формулу, нужно нажать **Enter** на клавиатуре, после чего в ячейке появится результат расчета по заданной формуле.

Скриншот листа Excel при построении вспомогательной таблицы представлен на рисунке 23.

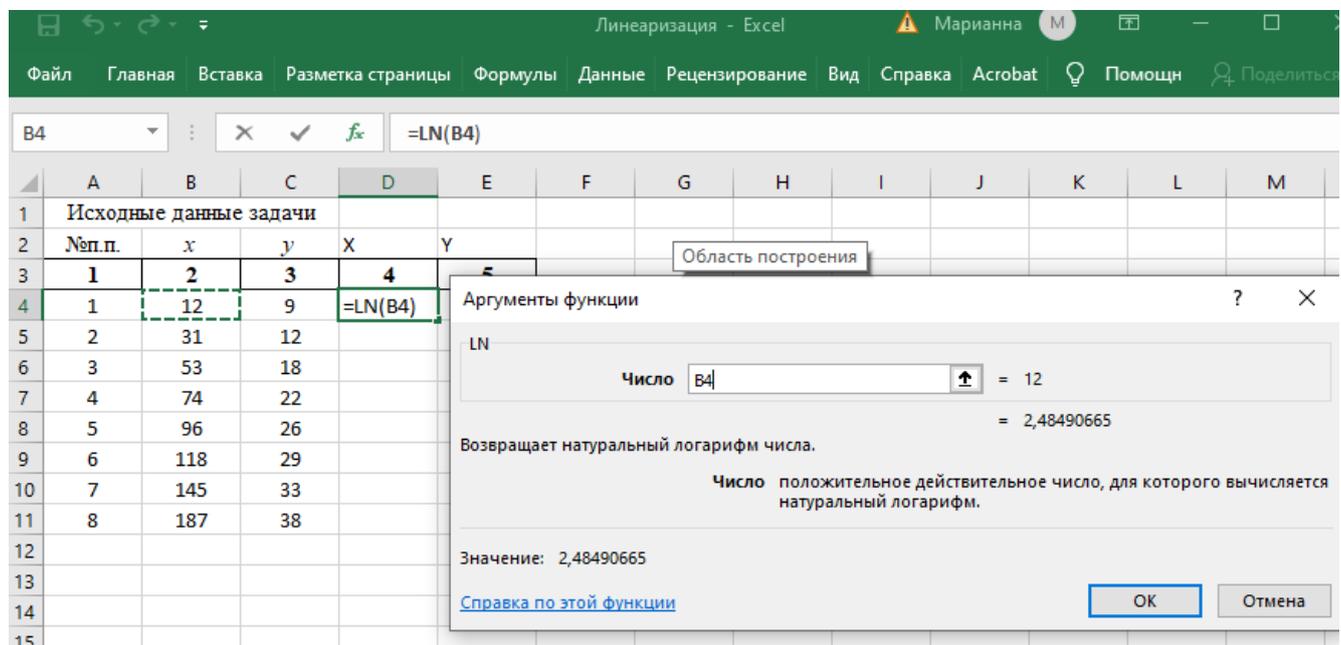


Рис. 23. Скриншот листа Excel при построении вспомогательной таблицы.

Нажав на черный квадратик в правом нижнем углу ячейки, левой кнопкой мыши можно протащить черный крест вниз или вправо, заполнив все ячейки столбцов X и Y. Результат представлен на рисунке 24.

Исходные данные задачи				
№п.п.	x	y	X	Y
1	2	3	4	5
1	12	9	2,484907	2,197225
2	31	12	3,433987	2,484907
3	53	18	3,970292	2,890372
4	74	22	4,304065	3,091042
5	96	26	4,564348	3,258097
6	118	29	4,770685	3,367296
7	145	33	4,976734	3,496508
8	187	38	5,231109	3,637586

Рис. 24. Скриншот листа Excel со вспомогательной таблицей.

Найдем уравнение регрессии. Итоги работы функции **Регрессия** представлены на рисунке 25.

Вывод итогов						
<i>Регрессионная статистика</i>						
Множественный R	0,991233					
R-квадрат	0,982542					
Нормированный коэффициент	0,979632					
Стандартная ошибка	0,071607					
Наблюдено	8					
<i>Дисперсионный анализ</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	1	1,731497	1,731497	337,6804	1,67E-06	
Остаток	6	0,030766	0,005128			
Итого	7	1,762263				
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
У-пересечение	0,731608	0,128832	5,678762	0,001285	0,416367	1,046849
Переменная X 1	0,550453	0,029955	18,37608	1,67E-06	0,477157	0,62375

Рис. 25. Скриншот листа Excel с результатами работы функции **Регрессия**.

Параметры функции $Y = A + b \cdot X + E$:

$A=0,731608$;

$b=0.550453$.

Т.е. получаем следующее уравнение регрессии:

$$\hat{Y} = 0,731608 + 0,550453 \cdot X :$$

После потенцирования найдем искомое уравнение регрессии. Потенцирование означает возведение основания логарифма в степень, равную значению логарифма.

Если $A = \ln a$, то $a = e^A$.

Для того, чтобы выполнить действие, обратное логарифмированию (потенцирование) можно воспользоваться математической функцией EXP. Вид окна **Вставка функции** приведен на рисунке 25.

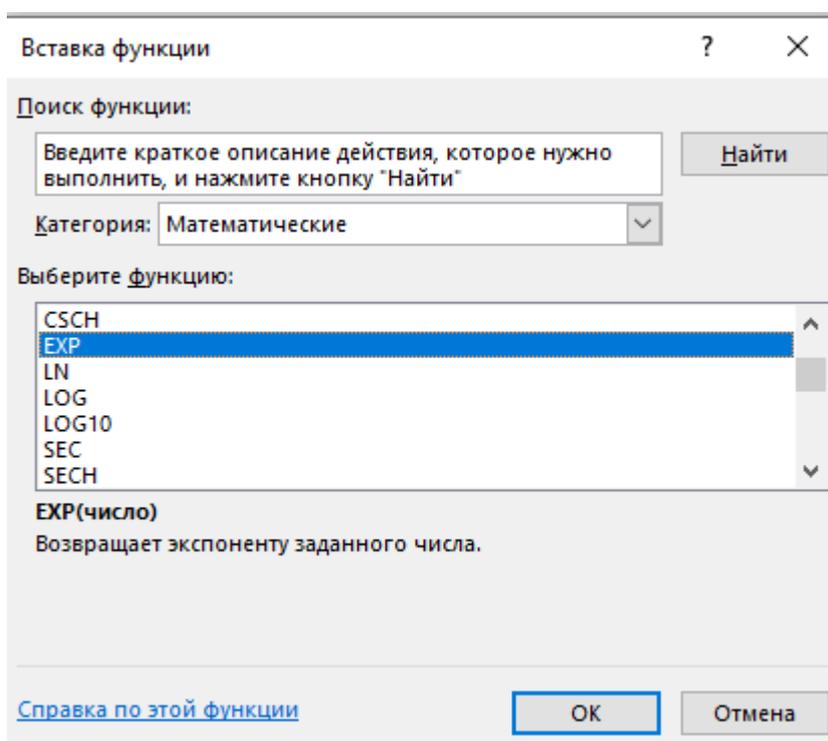


Рис. 25. Скриншот окна **Вставка функции** Excel при обращении к EXP.

Получим, выполнив потенцирование, значение параметра регрессии:

$$a = e^A = e^{0,7316} = 2.07842$$

Искомое уравнение регрессии: $\hat{y}_x = a \cdot x^b = 2.07842 \cdot x^{0.550453}$

При построении модели преобразования в линейный вид затрагивали не только не только независимую переменную, но и зависимую. В отличие от предыдущих рассмотренных функций, индекс корреляции не будет равен коэффициенту корреляции линеаризованной функции, который был рассчитан в результате работы функции **Регрессия** надстройки Excel пакета **Анализ данных**.

Расчет предсказанных по полученному уравнению регрессии (теоретических, выровненных) значений результатного показателя выполним в Excel.

$$\hat{y}_x = 2.07842 \cdot x^{0.550453}$$

Выполнить вычисление теоретических значений результата можно, воспользовавшись математической функцией **СТЕПЕНЬ**. Скриншот окна **Вставка функции** Excel при обращении к математической функции **СТЕПЕНЬ** приведен на рисунке 26.

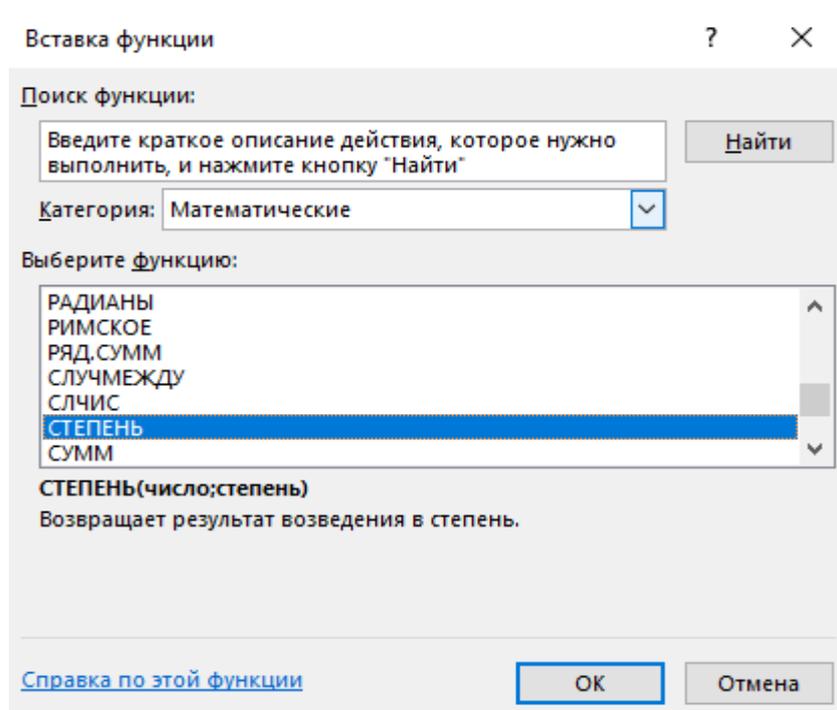


Рис. 26. Скриншот окна **Вставка функции** Excel при обращении к математической функции **СТЕПЕНЬ**.

Для функции СТЕПЕНЬ (число;степень) требуется задать два обязательных аргумента.

Число — это основание, которое возводится в степень, им может быть любое вещественное число.

Степень — показатель степени, в которую возводится основание.

Вместо функции СТЕПЕНЬ для возведения числа в степень можно использовать оператор ^, например: 5^2.

Скриншот листа Excel с таблицей необходимых расчетов для оценок нелинейной модели приведен на рисунке 27.

В столбце 6 верхней таблицы рассчитаны теоретические значения зависимой переменной для степенной функции. Теперь можно заполнить столбцы 7-10 данной таблицы.

Для расчета общей и остаточной дисперсий можно воспользоваться статистической функцией СРЗНАЧ (число1;[число2];...), у которой обязательным является первый аргумент, он может быть числом (первым, в случае перечня), а также ссылкой на ячейку (первую, в случае перечня) или диапазон, для которого требуется вычислить среднее значение. Скриншот листа Excel с окном функции СРЗНАЧ приведен на рисунке 28.

В нижней строке верхней таблицы, приведенной на рисунке 28, рассчитаны средние значения по столбцам 8 и 9, которые являются значениями остаточной и общей дисперсии для данной функции, и необходимы для расчета индексов корреляции и детерминации:

$$\sigma_{ост.}^2 = \frac{1}{n} \sum (y - \hat{y}_x)^2 = 0,7464814$$

$$\sigma_{общ.}^2 = \frac{1}{n} \sum (y - \bar{y}_x)^2 = 88,984375$$

F4 X ✓ fx = $\$B\$16*\text{СТЕПЕНЬ}(B4;\$B\$17)$										
	A	B	C	D	E	F	G	H	I	J
1	Исходные данные задачи			X	Y	\hat{y}_x	$y - \hat{y}_x$	$(y - \hat{y}_x)^2$	$(y - \bar{y}_x)^2$	$\frac{ y - \hat{y}_x }{y}$
2	№п.п.	x	y							
3	1	2	3	4	5	6	7	8	9	10
4	1	12	9	2,4849066	2,1972246	8,161545	0,838455	0,703007	206,640625	0,093162
5	2	31	12	3,4339872	2,4849066	13,76127	-1,76127	3,102073	129,390625	0,146773
6	3	53	18	3,9702919	2,8903718	18,48702	-0,48702	0,237190	28,890625	0,027057
7	4	74	22	4,3040651	3,0910425	22,21562	-0,21562	0,046491	1,890625	0,009801
8	5	96	26	4,5643482	3,2580965	25,63784	0,362162	0,131162	6,890625	0,013929
9	6	118	29	4,7706846	3,3672958	28,72156	0,278445	0,077531	31,640625	0,009602
10	7	145	33	4,9767337	3,4965076	32,17109	0,828906	0,687085	92,640625	0,025118
11	8	187	38	5,2311086	3,6375862	37,00636	0,993636	0,987313	213,890625	0,026148
12										
13	Сумма		187							
	Среднее значение		23,375					0,746481	88,984375	0,043949
14										
15										
16	a=	2,07842		ВЫВОД ИТОГОВ						
17	b=	0,55045								
18				<i>Регрессионная статистика</i>						
19				Множественный R	0,9912325					
20				R-квадрат	0,9825419					
21				Нормированный коэффициент детерминации	0,9796322					
22				Стандартная ошибка	0,0716074					
23				Наблюдения в выборке	8					
24										
25				<i>Дисперсионный анализ</i>						
26					df	SS	MS	F	Значимость F	
27				Регрессия	1	1,731497	1,731497	337,6803821	1,6738E-06	
28				Остаток	6	0,030766	0,005128			
29				Итого	7	1,762263				
30										
31					Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Верхние 95%	Нижние 95%
32				Y-пересечение	0,7316079	0,128832	5,678762	0,001284798	0,416366673	1,046849
33				Переменная X	0,5504535	0,029955	18,37608	1,6738E-06	0,477156518	0,62375

Рис. 27. Скриншот листа Excel с таблицей необходимых расчетов для оценок нелинейной модели.

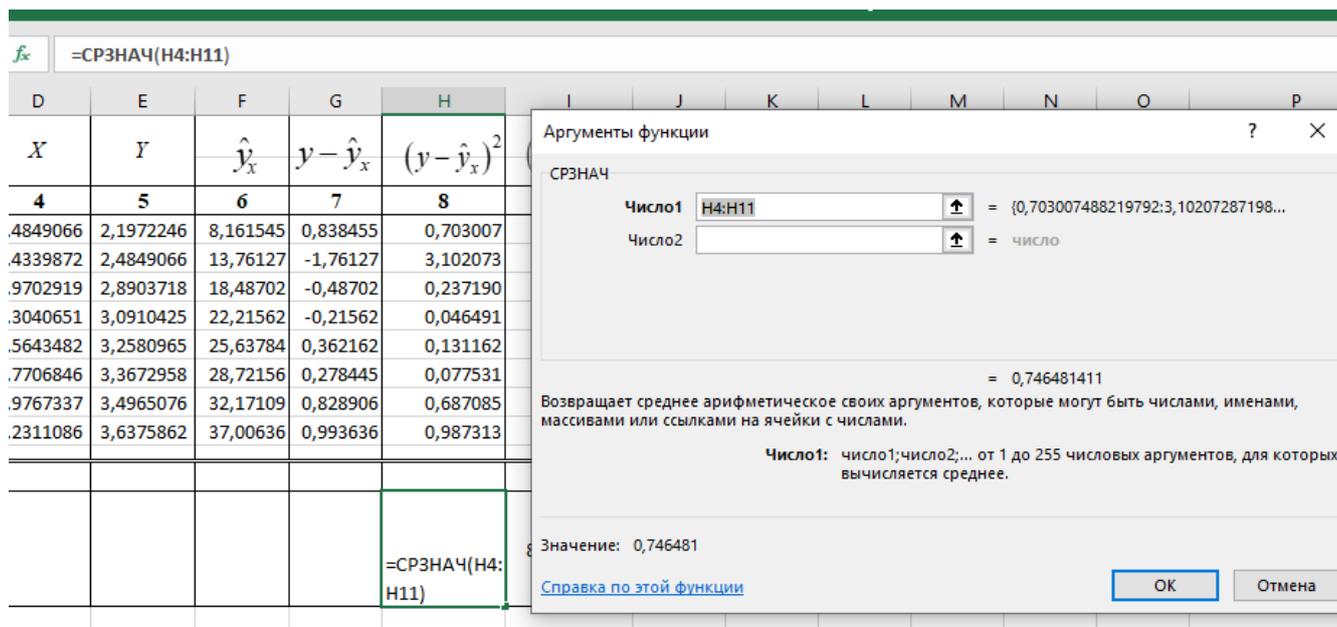


Рис. 28. Скриншот листа Excel с окном функции СРЗНАЧ.

Индекс корреляции находим по формуле 2.5:

$$\rho_{yx} = \sqrt{1 - \frac{\sigma_{ост.}^2}{\sigma_{общ.}^2}} = \sqrt{1 - \frac{0,746481}{88,984375}} = 0,9958$$

Индекс детерминации: $\rho_{yx}^2 = 0,9916$ показывает, что 99,16% вариации результативного признака объясняется вариацией признака-фактора, а 0,84% приходится на долю прочих факторов.

Критерий Фишера рассчитывается по формуле 2.14:

$$F = \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} \cdot \frac{n - m - 1}{m} = \frac{0,9916}{0,84} \cdot \frac{8 - 1 - 1}{1} = 709,23$$

Расчетное значение F -критерия значительно превышает табличное $F_{табл}=5,99$, что говорит о статистической значимости уравнения регрессии в целом.

В столбце 10 рассчитана средняя ошибка аппроксимации: $\bar{A} = 4.39\%$, что говорит о хорошей аппроксимации исходных данных линией регрессии, построенной в форме степенной функции.

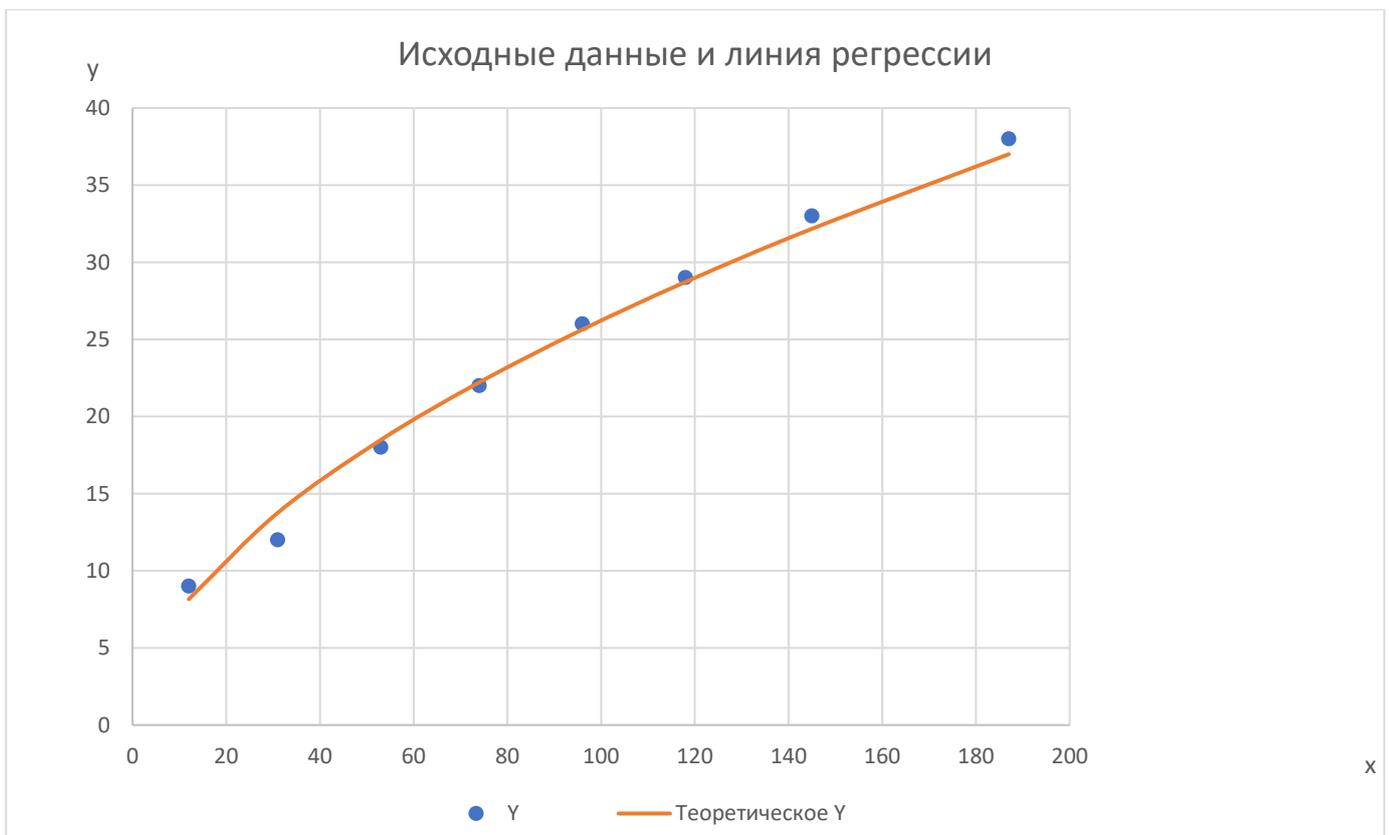


Рис. 29. График нелинейной регрессии $\hat{y}_x = 2.07842 \cdot x^{0.550453}$ и точечная диаграмма исходных данных.

Построим также парную линейную регрессию для сравнения моделей. Скриншот листа Excel с результатами работы функции **Регрессия** представлен на рисунке 30.

	A	B	C	D	E	F	G	H	
1	Исходные данные задачи			\hat{y}_x	$y - \hat{y}_x$	$\frac{ y - \hat{y}_x }{y}$			
2	№п.п.	x	y						
3	1	2	3	4	5	5			
4	1	12	9	10,26544	-1,26544	0,140605			
5	2	31	12	13,4794	-1,4794	0,123283			
6	3	53	18	17,20082	0,799178	0,044399			
7	4	74	22	20,75309	1,246911	0,056678			
8	5	96	26	24,47451	1,525489	0,058673			
9	6	118	29	28,19593	0,804067	0,027726			
10	7	145	33	32,76313	0,236867	0,007178			
11	8	187	38	39,86767	-1,86767	0,049149			
12									
13					$\bar{A} =$	6,35%			
14	ВЫВОД ИТОГОВ								
15									
16	<i>Регрессионная статистика</i>								
17		Множест	0,991181						
18		R-квадрат	0,982439						
19		Нормиро	0,979512						
20		Стандарт	1,443455						
21		Наблюдени	8						
22									
23	<i>Дисперсионный анализ</i>								
24			df	SS	MS	F	значимость F		
25		Регрессия	1	699,3736	699,3736	335,6625	1,7E-06		
26		Остаток	6	12,50137	2,083562				
27		Итого	7	711,875					
28									
29	<i>Коэффициент корреляции</i>							<i>Значение</i>	<i>нижние 95%</i>
30	Y-пересеч	8,235578	0,971226	8,479568	0,000147	5,859073	10,61208		
31	Переменн	0,169156	0,009233	18,3211	1,7E-06	0,146564	0,191747		

Рис. 30. Скриншот листа Excel с результатами работы функции **Регрессия**.

Сравнение построенных моделей по индексу детерминации и средней ошибке аппроксимации выполним в таблице 4.5.

Таблица 4.5

Модель	Индекс детерминации	Средняя ошибка аппроксимации
Линейная модель $y = a + b \cdot x + \varepsilon$	0,9824	6,35
Полулогарифмическая модель $y = a + b \cdot \ln x + \varepsilon$	0,9202	14,58
Модель с квадратным корнем $y = a + b \cdot \sqrt{x} + \varepsilon$	0,9913	4,98
Степенная модель $y = a \cdot x^b \cdot \varepsilon$	0,9916	4,39

Наиболее хорошо исходные данные аппроксимирует степенная модель, чуть хуже – модель с квадратным корнем. Но в данном случае, так как индексы детерминации линейной модели и модели степенной отличаются всего на 0,009, то вполне можно обойтись более простой линейной функцией.

5. Задания для самостоятельной работы

Задача 1.

Для определения зависимости между сменной добычей угля на одного рабочего (обозначим переменной y , измеряемой в тоннах) и мощностью угольного пласта (переменная x , измеряемая в метрах) на 10 шахтах были проведены исследования, результаты которых представлены в таблице 5.1.

Таблица 5.1.

i	1	2	3	4	5	6	7	8	9	10
x_i	8	11	12	9	8	8	9	10	6	12
y_i	5	10	10	7	5	6	6	8	4	11

1. Подбором нелинейных преобразований исходных переменных в варианте задания добиться улучшения представления данных с помощью нелинейной функции регрессии. Предположив, что связь между признаками носит нелинейный характер, нужно найти параметры следующих нелинейных уравнений:

$$y = a + b \cdot \ln x + \varepsilon,$$

$$y = a + b \cdot \sqrt{x} + \varepsilon,$$

$$y = a + b \cdot \frac{1}{x} + \varepsilon$$

$$y = a \cdot x^b \cdot \varepsilon.$$

2. Записать оцененную нелинейную функцию регрессии и построить её график на диаграмме рассеяния.

3. Оцените тесноту связи с помощью показателей корреляции и детерминации.

4. Оцените с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования (оцените статистическую надежность уравнения регрессии в целом).

5. Оцените с помощью средней ошибки аппроксимации качество полученной модели.

6. Выполните сравнительный анализ построенных нелинейных регрессионных моделей и линейной регрессии на основе показателей детерминации и средней относительной ошибки аппроксимации.

Задача 2.

Изучается зависимость доли расходов на товары длительного пользования в общих расходах семьи в процентах (y) от среднемесячного дохода семьи в тысячах долларов США (x). Данные представлены в таблице 5.2.

Таблица 5.2.

Доля расходов на товары длительного пользования в зависимости от дохода

	семьи							
x	1	2	3	4	5	6	7	8
y	10,0	13.4	15.4	16.5	18.6	19.1	20,0	20,5

1. Подбором нелинейных преобразований исходных переменных в варианте задания добиться улучшения представления данных с помощью нелинейной функции регрессии. Предположив, что связь между признаками носит нелинейный характер, нужно найти параметры следующих нелинейных уравнений:

$$y = a + b \cdot \ln x + \varepsilon,$$

$$y = a + b \cdot \sqrt{x} + \varepsilon,$$

$$y = a + b \cdot \frac{1}{x} + \varepsilon$$

$$y = a \cdot x^b \cdot \varepsilon.$$

2. Записать оцененную нелинейную функцию регрессии и построить её график на диаграмме рассеяния.

3. Оцените тесноту связи с помощью показателей корреляции и детерминации.

4. Оцените с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования (оцените статистическую надежность уравнения регрессии в целом).

5. Оцените с помощью средней ошибки аппроксимации качество полученной модели.

6. Выполните сравнительный анализ построенных нелинейных регрессионных моделей и линейной регрессии на основе показателей детерминации и средней относительной ошибки аппроксимации.

Задача 3.

Изучается зависимость производительности труда y (т/час) от уровня механизации работ x (%) по данным 14 промышленных предприятий. Данные для решения задачи представлены по вариантам 1-12 в таблицах 5.4 – 5.15.

1. Подбором нелинейных преобразований исходных переменных в варианте задания добиться улучшения представления данных с помощью нелинейной функции регрессии. Предположив, что связь между признаками носит нелинейный характер, нужно найти параметры следующих нелинейных уравнений:

$$y = a + b \cdot \ln x + \varepsilon,$$

$$y = a + b \cdot \sqrt{x} + \varepsilon,$$

$$y = a + b \cdot \frac{1}{x} + \varepsilon$$

$$y = a \cdot x^b \cdot \varepsilon.$$

Пункты 2-6 задания нужно выполнить для каждой нелинейной функции.

2. Записать оцененную нелинейную функцию регрессии и построить её график на диаграмме рассеяния.

3. Найти средний коэффициент эластичности зависимой переменной по независимой.

4. Оценить тесноту связи с помощью показателей корреляции и детерминации.

5. Оценить с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования (оцените статистическую надежность уравнения регрессии в целом).

6. Оценить с помощью средней ошибки аппроксимации качество полученной модели.

7. Выполните сравнительный анализ построенных нелинейных регрессионных моделей и линейной регрессии на основе показателей детерминации и средней относительной ошибки аппроксимации.

Анализ можно выполнить в таблице 5.3.

Таблица 5.3

Модель	Индекс детерминации	Средняя ошибка аппроксимации
Линейная модель: $y = a + b \cdot x + \varepsilon$		
Полулогарифмическая модель: $y = a + b \cdot \ln x + \varepsilon$		
Модель с квадратным корнем: $y = a + b \cdot \sqrt{x} + \varepsilon$		
Степенная модель: $y = a \cdot x^b \cdot \varepsilon$		

Равносторонняя гипербола:		
$y = a + b \cdot \frac{1}{x} + \varepsilon$		

Результатом анализа является выбор наиболее подходящей модели исследуемой зависимости и обоснование данного выбора с использованием значений показателей, характеризующих качество построенной регрессионной модели.

Ниже в таблицах 5.4 – 5.15 приведены данные наблюдений для решения поставленной задачи по вариантам 1-12.

Таблица 5.4.

Вар. 1

<i>x</i>	32	30	36	40	41	47	56	54	60	55	61	67	69	76
<i>y</i>	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Таблица 5.5.

Вар. 2

<i>x</i>	55	46	40	39	35	29	31	75	68	66	60	54	59	53
<i>y</i>	33	32	30	29	27	23	19	47	44	42	40	39	37	36

Таблица 5.6.

Вар. 3

<i>x</i>	48	57	55	61	56	62	68	70	77	42	41	37	31	33
<i>y</i>	34	35	38	39	41	42	44	46	49	32	31	29	25	21

Таблица 5.7.

Вар. 4

x	52	54	45	39	38	34	28	30	74	67	65	59	53	58
y	35	32	31	29	28	26	22	18	46	43	41	39	38	36

Таблица 5.8.

Вар. 5

x	43	49	58	56	62	57	63	69	71	78	34	32	38	42
y	33	35	36	39	40	42	43	45	47	50	22	26	30	32

Таблица 5.9.

Вар. 6

x	52	57	51	53	44	38	37	33	27	29	73	66	64	58
y	37	35	34	31	30	28	27	25	21	17	45	42	40	38

Таблица 5.10.

Вар. 7

x	39	43	44	50	59	57	63	58	64	70	72	79	35	33
y	31	33	34	36	37	40	41	43	44	46	48	51	23	27

Таблица 5.11.

Вар. 8

x	63	57	51	56	50	52	43	37	36	32	26	28	72	65
y	39	37	36	34	33	30	29	27	26	24	20	16	44	41

Таблица 5.12.

Вар. 9

x	64	59	65	71	73	80	36	34	40	44	45	51	60	58
y	42	44	45	47	49	52	24	28	32	34	35	37	38	41

Таблица 5.13.

Вар. 10

x	46	52	61	59	65	60	66	72	74	81	37	35	41	45
y	36	38	39	42	43	45	46	48	50	53	25	29	33	35

Таблица 5.14.

Вар. 11

x	62	30	36	50	41	47	56	54	60	55	61	67	69	66
y	50	20	25	40	30	36	40	38	42	43	47	48	50	46

Таблица 5.15.

Вар. 12

x	45	46	40	39	35	29	61	75	68	66	60	54	59	53
y	7	6	5	6	5	3	9	10	9	8	9	8	7	6

Задача 4.

Изучается зависимость материалоемкости продукции от объема выпуска продукции предприятием по 12 однородным заводам. Данные для решения задачи представлены в таблице 5.16.

Таблица 5.16

№	Потреблено материалов на единицу продукции, кг. (y)	Выпуск продукции, тыс. ед. (x)
1	9,0	100
2	6,0	200
3	5,0	300
4	4,0	400
5	3,7	500
6	3,5	600
7	3,5	700
8	6,0	150
9	7,0	120
10	3,5	250
11	4,5	300
12	5,2	380

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.
2. Рассчитайте параметры уравнений линейной и гиперболической парной регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.

4. Дайте с помощью общего коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оцените с помощью средней ошибки аппроксимации качество уравнений.
6. Оцените с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования. По значениям характеристик, рассчитанных в пунктах 3, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте обоснование выбора.
7. Оцените с помощью t-критерия Стьюдента статистическую надежность коэффициентов регрессии.
8. По линейной функции постройте доверительный интервал для коэффициента регрессии.
9. По линейной функции рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 5 процентов от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.
10. Оцените полученные результаты, выводы оформите в аналитической записке.

Задача 5.

Изучается зависимость цены квартиры от ее общей площади. Имеются данные о цене однокомнатной квартиры и величине ее общей площади по 12 сделкам одного района города. Данные для решения задачи представлены в таблице 5.17.

Таблица 5.17

№	Площадь, м ² (x)	Цена квартиры, млн. руб. (y)
1	35	2,9
2	35	3,1
3	33	3,5
4	34	3,5
5	38	4,5
6	40	4,6
7	40	4,5
8	39	4,4
9	37	3,8
10	36	3,7
11	42	4,8
12	45	5,0

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.
2. Рассчитайте параметры уравнений линейной и степенной парной регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. Дайте с помощью общего коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оцените с помощью средней ошибки аппроксимации качество уравнений.
6. Оцените с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования. По значениям характеристик,

рассчитанных в пунктах 3, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте обоснование выбора.

7. Оцените с помощью t-критерия Стьюдента статистическую надежность коэффициентов регрессии.

8. По линейной функции постройте доверительный интервал для коэффициента регрессии.

9. По линейной функции рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 3 процента от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.

10. Оцените полученные результаты, выводы оформите в аналитической записке.

Задача 6.

Изучается зависимость набора слушателей на образовательную программу «Мастер делового администрирования» (МВА) от ее стоимости. Имеются данные о годовой цене программы «Мастер делового администрирования» (МВА) и числе слушателей в образовательном учреждении. Данные для решения задачи представлены в таблице 5.18.

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.
2. Рассчитайте параметры уравнений линейной и показательной парной регрессии.

3. Оцените тесноту связи с помощью показателей корреляции и детерминации.

4. Дайте с помощью общего коэффициента эластичности сравнительную оценку силы связи фактора с результатом.

Таблица 5.18

№	Цена программы, тыс. руб., (y)	Число слушателей, чел., (x)
1	80	5
2	50	10
3	49	12
4	40	15
5	38	20
6	35	22
7	38	25
8	37	30
9	36	35
10	35	36
11	34	40
12	30	50
13	30	60

5. Оцените с помощью средней ошибки аппроксимации качество уравнений.

6. Оцените с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования. По значениям характеристик, рассчитанных в пунктах 3, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.

7. Оцените с помощью t-критерия Стьюдента статистическую надежность коэффициентов регрессии.

8. По линейной функции постройте доверительный интервал для коэффициента регрессии.

9. По линейной функции рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 4 процента от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.

10. Оцените полученные результаты, выводы оформите в аналитической записке.

Задача 7.

Изучается зависимость инвестиций в основной капитал от размера валового регионального продукта (ВРП). Имеются данные по 13 регионам о размере ВРП и объеме инвестиций в основной капитал. Данные для решения задачи представлены в таблице 5.19.

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.
2. Рассчитайте параметры уравнений линейной и степенной парной регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. Дайте с помощью общего коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оцените с помощью средней ошибки аппроксимации качество уравнений.
6. Оцените с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования. По значениям характеристик, рассчитанных в пунктах 3, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.

Таблица 5.19

№ региона	ВРП, млрд. руб., (x)	Инвестиции в основной капитал, млрд. руб., (y)
1	24,6	5,0
2	41,1	9,0
3	29,5	4,8
4	27,6	5,4
5	31,9	7,4
6	38,8	6,6
7	39,2	7,8
8	40,2	9,3
9	41,6	9,6
10	41,3	8,0
11	47,0	10,8
12	54,7	9,9
13	53,3	10,0

7. Оцените с помощью t-критерия Стьюдента статистическую надежность коэффициентов регрессии.
8. По линейной функции постройте доверительный интервал для коэффициента регрессии.
9. По линейной функции рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 5 процентов от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.
10. Оцените полученные результаты, выводы оформите в аналитической записке.

Задача 8.

Имеются данные о размере прибыли предыдущего года и суммах, выделяемых на приобретение ценных бумаг, по 17 предприятиям отрасли. Данные для решения задачи представлены в таблице 5.20.

Таблица 5.20

№	Нераспределенная прибыль предыдущего года, млн. руб. (x)	Приобретено ценных бумаг, млн. руб. (y)
1	68,4	15,3
2	67,3	9,2
3	57,3	5,4
4	55,3	7,7
5	2,9	4,4
6	51,5	3,1
7	48,8	3,6
8	42,1	3,8
9	44,8	12,0
10	33,1	4,0
11	30,1	3,7
12	27,1	1,8
13	24,3	2,7
14	21,9	1,9
15	19,7	2,4
16	19,2	2,3
17	65,0	4,5

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.

2. Рассчитайте параметры уравнений линейной и степенной парной регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. Дайте с помощью общего коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оцените с помощью средней ошибки аппроксимации качество уравнений.
6. Оцените с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования. По значениям характеристик, рассчитанных в пунктах 3, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.
7. Оцените с помощью t-критерия Стьюдента статистическую надежность коэффициентов регрессии.
8. По линейной функции постройте доверительный интервал для коэффициента регрессии.
9. По линейной функции рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 4 процента от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.
10. Оцените полученные результаты, выводы оформите в аналитической записке.

Тесты

1. Какое уравнение регрессии нельзя свести к линейному виду:

а) $y_x = a + b \cdot \ln x$;

б) $y_x = a \cdot x^b$;

в) $y_x = a + b \cdot x^c$.

2. Какое из уравнений является степенным:

а) $y_x = a + b \cdot \ln x$;

б) $y_x = a \cdot x^b$;

в) $y_x = a + b \cdot x^c$.

3. Параметр b в степенной модели является:

а) коэффициентом детерминации;

б) коэффициентом эластичности;

в) коэффициентом корреляции.

4. Коэффициент корреляции r_{xy} может принимать значения:

а) от -1 до 1 ;

б) от 0 до 1 ;

в) любые.

5. Для функции $y = a + \frac{b}{x} + \varepsilon$ средний коэффициент эластичности

имеет вид:

а) $\bar{\varepsilon} = \frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$;

б) $\bar{\varepsilon} = -\frac{b}{a \cdot \bar{x} + b}$;

$$\text{в) } \bar{\Theta} = -\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}.$$

6. Какое из следующих уравнений нелинейно по оцениваемым параметрам:

а) $y = a + b \cdot x + \varepsilon;$

б) $y = a + b \cdot \ln x + \varepsilon;$

в) $y = a \cdot x^b \cdot \varepsilon.$

Рекомендуемая литература

1. Методы эконометрики: Учебник / С.А. Айвазян; Московская школа экономики МГУ им. М.В. Ломоносова (МШЭ). - М.: Магистр: ИНФРА-М, 2010. - 512 с.: Режим доступа: <http://znanium.com/bookread2.php?book=196548>
2. Эконометрика и эконометрическое моделирование : учебник / Л.О. Бабешко, М.Г. Бич, И.В. Орлова. - М. : Вузовский учебник : ИНФРА-М, 2018. - 385 с. : ил. — (Высшее образование: Бакалавриат). Режим доступа: <http://znanium.com/bookread2.php?book=>
3. Эконометрика / Уткин В.Б., - 2-е изд. - М.: Дашков и К, 2017. - 564 с.: ISBN 978-5-394-02145-9 Режим доступа: <http://znanium.com/bookread2.php?book=415317>
4. Артамонов, Н. В. Введение в эконометрику: Учебник / Артамонов Н.В., - 2-е изд., испр. и доп. - Москва : МЦНМО, 2014. - 222 с.: ISBN 978-5-4439-2010-8. - Текст : электронный. - URL: <https://znanium.com/catalog/product/969269>. - Режим доступа: по подписке.
5. Эконометрика: учеб. / под ред. И.И. Елисейевой. – М.: Проспект. 2009. – 288с.
6. Эконометрика: учеб. / И.И. Елисейева, С.В. Курышева, Т.В. Костеева и др.; под ред. И.И. Елисейевой. – 2-е изд., перераб. и доп. - М.: Финансы и статистика, 2006. – 576 с.: ил.
7. Кремер Н.Ш. Эконометрика: учебник для студентов вузов / Н.Ш. Кремер, Б.А. Путко; под ред. Н.Ш. Кремера. — 3-е изд., перераб. и доп. — М.: ЮНИТИ-ДАНА, 2010. — 328 с. — (Серия «Золотой фонд российских учебников»).
8. Колпаков, В. Ф. Экономико-математическое и эконометрическое моделирование: компьютерный практикум: учеб. пособие / В.Ф. Колпаков. - Москва : ИНФРА-М, 2018. – 396 с. - (Высшее образование: Бакалавриат). - www.dx.doi.org/10.12737/24417. - ISBN 978-5-16-010967-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/975797>. - Режим доступа: по подписке.
9. Бородич, С. А. Эконометрика. Практикум : учебное пособие / С.А. Бородич. — Москва : ИНФРА-М, 2021. — 329 с. : ил. — (Высшее образование: Бакалавриат). - ISBN 978-5-16-009429-8. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1228789>. - Режим доступа: по подписке.
10. Орлова, И. В. Эконометрика. Компьютерный практикум для студентов третьего курса, обучающихся по специальностям 080105.65 "Финансы и кредит", 080109.65 "Бухгалтерский учет, анализ и аудит" / И. В.

- Орлова, Е. С. Филонова, А. В. Агеев. - Москва : ВЗФЭИ, 2011. - 96 с. - Текст : электронный. - URL: <https://znanium.com/catalog/product/453458>. – Режим доступа: по подписке.
11. Хайяши, Ф. Эконометрика / Ф. Хайяши ; пер. с англ. под науч. ред. В.П. Носко. — Москва : Издательский дом «Дело» РАНХиГС, 2017. — 728 с. — (Академический учебник). - ISBN 978-5-7749-1197-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1043302> (дата обращения: 11.01.2023). – Режим доступа: по подписке.
 12. Экономико-математические методы в примерах и задачах : учебное пособие / И.В. Орлова, Н.В. Концевая, Е.Н. Горбатенко, В.А. Большаков ; под ред. А.Н. Гармаша. — Москва : Вузовский учебник : ИНФРА-М, 2021. — 416 с. + Доп. материалы [Электронный ресурс]. - ISBN 978-5-9558-0322-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1659052> (дата обращения: 11.01.2023). – Режим доступа: по подписке.
 13. Эконометрика: краткий курс лекций для обучающихся направления подготовки «Менеджмент» / Сост.: В. А. Шибайкин // ФГБОУ ВО Саратовский ГАУ. – Саратов, 2017. – с.60
 14. Доугерти К. Введение в эконометрику: Учебник. 3-е изд. / Пер. с англ. – М.: ИНФРА-М, 2009. – XIV, 465 с/ - (Университетский учебник).
 15. Кузнецова О.А. Эконометрика: практикум / О.А. Кузнецова, О.Н. Мазурмович. – Самара:Изд-во Самарского университета, 2019. – 72 с.
 16. Максимова Т.Г., Попова И.Н. Эконометрика: учебно-методическое пособие / Т.Г. Максимова, И.Н. Попова. – СПб.: Университет ИТМО, 2018. – 70 с.

Приложение 1

**1.1. Таблица значений F -критерия Фишера при уровне значимости
 $\alpha = 0,05$**

$k_1 \backslash k_2$	1	2	3	4	5	6	8	12	24	∞
1	2	3	4	5	6	7	8	9	10	11
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48

1	2	3	4	5	6	7	8	9	10	11
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1

1.2. Критические значения t -критерия Стьюдента при уровне значимости 0,10, 0,05, 0,01 (двухсторонний)

Число степеней свободы d.f.	α			Число степеней свободы d.f.	α		
	00,10	0,05	0,01		00,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,5041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

СОДЕРЖАНИЕ

Введение	3
1. Общие методические указания для решения задач по дисциплине «Эконометрика».....	4
2. Нелинейные модели парной регрессии и корреляции	6
Таблица 2.1	12
Формулы для расчета средних коэффициентов эластичности.....	12
3. Контрольные вопросы.....	19
4. Построение моделей нелинейной регрессии.	20
5. Задания для самостоятельной работы	66
Тесты.....	83
Рекомендуемая литература	85
Приложение 1. Математико-статистические таблицы.....	87

Учебное издание

А.А. Чалганова

Построение нелинейных моделей парной регрессии с
использованием табличного процессора Excel

Публикуется в авторской редакции.

Подписано к публикации 30.12.2022. Формат 60×90 1/16.

Гарнитура Times New Roman. Усл. печ. л. 5,625. Заказ № 1343.

РГГМУ, 192007, Санкт-Петербург, Воронежская ул., д. 79.