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

О. Н. Туманова, В. И. Серкова

Прикладные программные продукты

Учебное пособие

Ухта, УГТУ, 2016

УДК 004.4(075.8)
ББК 32.(972 я7)
Т 83

Туманова, О. Н.

Т 83 Прикладные программные продукты [Текст] : учеб. пособие / О. Н. Туманова, В. И. Серкова. – Ухта : УГТУ, 2016. – 80 с.

ISBN 978-5-88179-924-3

Настоящее пособие предназначено для студентов дневной и заочной форм обучения по специальности 131000 Нефтегазовое дело.

Пособие может быть использовано студентами дневного обучения при выполнении лабораторных работ по дисциплинам «Пакеты прикладных программ» и «Экономико-математические методы и модели». В пособии раскрыто применение возможностей Microsoft Excel и пакета Mathcad для решения экономических задач, таких как определение оптимального ассортимента продукции, определение оптимальных затрат при перевозках, составление и анализ уравнений регрессии с целью прогноза, оптимальное распределение производственных ресурсов и инвестиций и др.

Учебное пособие отвечает требованиям рабочей учебной программы дисциплин «Пакеты прикладных программ» и «Экономико-математические методы и модели» и рекомендуется к использованию в учебном процессе. Пособие также может быть применено в курсовых и дипломных работах.

УДК 004.4(075.8)
ББК 32.(972 я7)

Учебное пособие рекомендовано к изданию Редакционно-издательским советом Ухтинского государственного технического университета.

Рецензенты: С. М. Мильков, доцент МИИТ, к.ф.-м.н.; В. П. Дробышевский, директор ООО «АТИКС».

ISBN 978-5-88179-924-3

Оглавление

Введение

К прикладному программному обеспечению относятся средства, позволяющие пользователям решать конкретные задачи: набор текста, расчёты, построение диаграмм, создание графических изображений, обработка звука и видеоизображения и др. Крупные разработчики программного обеспечения объединяют самые популярные прикладные программы в комплекты – пакеты, включающие в себя обычно текстовый редактор, электронные таблицы, графический редактор, систему управления базами данных. Компоненты пакетов хорошо интегрированы и позволяют переносить информацию из одного в другой.

Наиболее популярным среди пользователей персональных компьютеров является пакет Microsoft Excel, программа презентации и слайдов Microsoft PowerPoint, клиент электронной почты и органайзер Microsoft Outlook, средство объединения документов Microsoft Binder, графический редактор Microsoft Graph, редактор математических формул Microsoft Equation, Microsoft Access – система управления базами данных.

В данном методическом пособии рассматриваются задания по использованию электронных таблиц Excel:

1. для решения задач оптимизации (модуль поиска решения);
2. для решения задач регрессионного анализа и прогнозирования.

Кроме того, рассматривается прикладная программа Mathcad, которая относится к универсальным программам, сочетая в себе возможности проведения расчётов и подготовки форматированных научных и технических документов.

1. Надстройка Excel Solver (Поиск решения)

1.1 Общие сведения

Оптимизация значений таблицы Excel, удовлетворяющих определённым критериям, может быть сложным процессом. К счастью, Microsoft предлагает надстройку для численной оптимизации. Хотя данный сервис не может решить всех проблем, он может быть полезным в качестве инструмента.

Мощным средством анализа данных Excel является надстройка Solver (Поиск решения). С её помощью можно определить, при каких значениях указанных влияющих ячеек формула в целевой ячейке принимает нужное значение (минимальное, максимальное или равное какой-либо величине). Для процедуры поиска решения можно задать ограничения, причём не обязательно, чтобы при этом использовались те же влияющие ячейки. Для расчёта заданного значения применяются различные математические методы поиска.

Вы можете установить режим, в котором полученные значения переменных автоматически заносятся в таблицу. Кроме того, результаты работы программы могут быть оформлены в виде отчёта.

Программа Поиск решений (в оригинале Excel Solver) – дополнительная надстройка табличного процессора MS Excel, которая предназначена для решения определённых систем уравнений, линейных и нелинейных задач оптимизации, используется с 1991 года.

Надстройка **Поиск решения** используется при решении задач оптимизации, а также для решения уравнений и систем уравнений. Надстройка доступна во всех версиях Excel. Обратите внимание, что скриншоты могут не соответствовать вашей версии. Несмотря на то, что некоторые функции могут менять своё местоположение в зависимости от версии надстройки, функционал остаётся практически неизменным.

Процедура поиска решения позволяет найти оптимальное значение формулы, содержащейся в ячейке, которая называется целевой. Эта процедура работает с группой ячеек, прямо или косвенно связанных с формулой в целевой ячейке. Чтобы получить по формуле, содержащейся в целевой ячейке, заданный результат, процедура изменяет значения во влияющих ячейках. Чтобы сузить множество значений, используемых в модели, применяются ограничения. Эти ограничения могут ссылаться на другие влияющие ячейки.

Процедуру поиска решения можно использовать для определения значения влияющей ячейки, которое соответствует экстремуму зависимой ячейки, например, можно изменить объём планируемого бюджета и увидеть, как это повлияет на проектируемую сумму расходов.

Размер задачи, которую можно решить с помощью базовой версии программы, ограничивается такими предельными показателями:

- Количество неизвестных (decision variable) – 200;
- количество формульных ограничений (explicit constraint) на неизвестные – 100;
- количество предельных условий (simple constraint) на неизвестные – 400.

Надстройка поставляется вместе с Excel, но по умолчанию отключена.

Чтобы включить его, перейдите по вкладке Файл/Параметры/Надстройки (рис. 1.1), в появившемся диалоговом окне выберите надстройки Excel.

В окне устанавливаем галочку напротив поля, жмём Поиск решения (рис. 1.2).

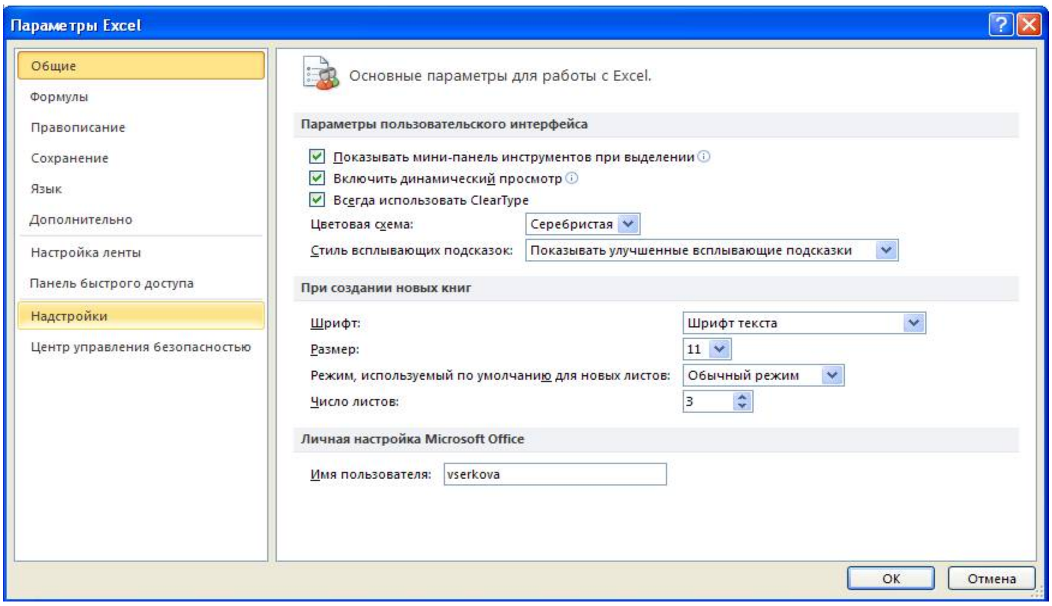


Рисунок 1.1 – Параметры Excel

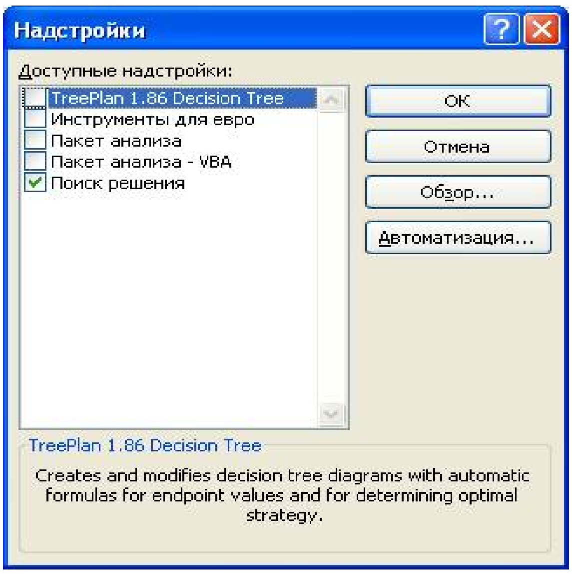


Рисунок 1.2 – Надстройки

Чтобы открыть окно поиска решения, надо войти в меню Данные, пункт Поиск решения (рис. 1.3).

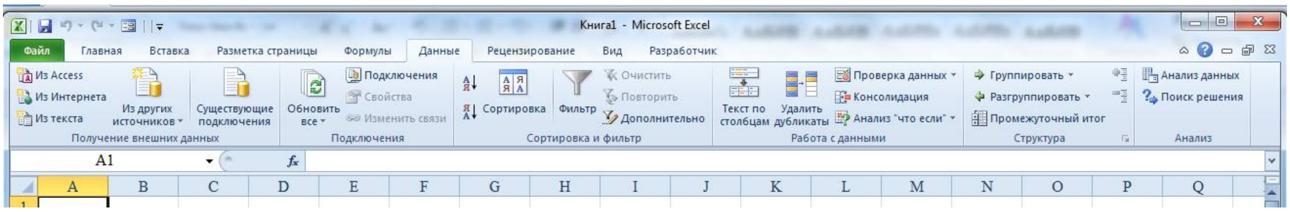


Рисунок 1.3 –Лента окна Excel

Откроется окно Поиск решения (рис. 1.4).

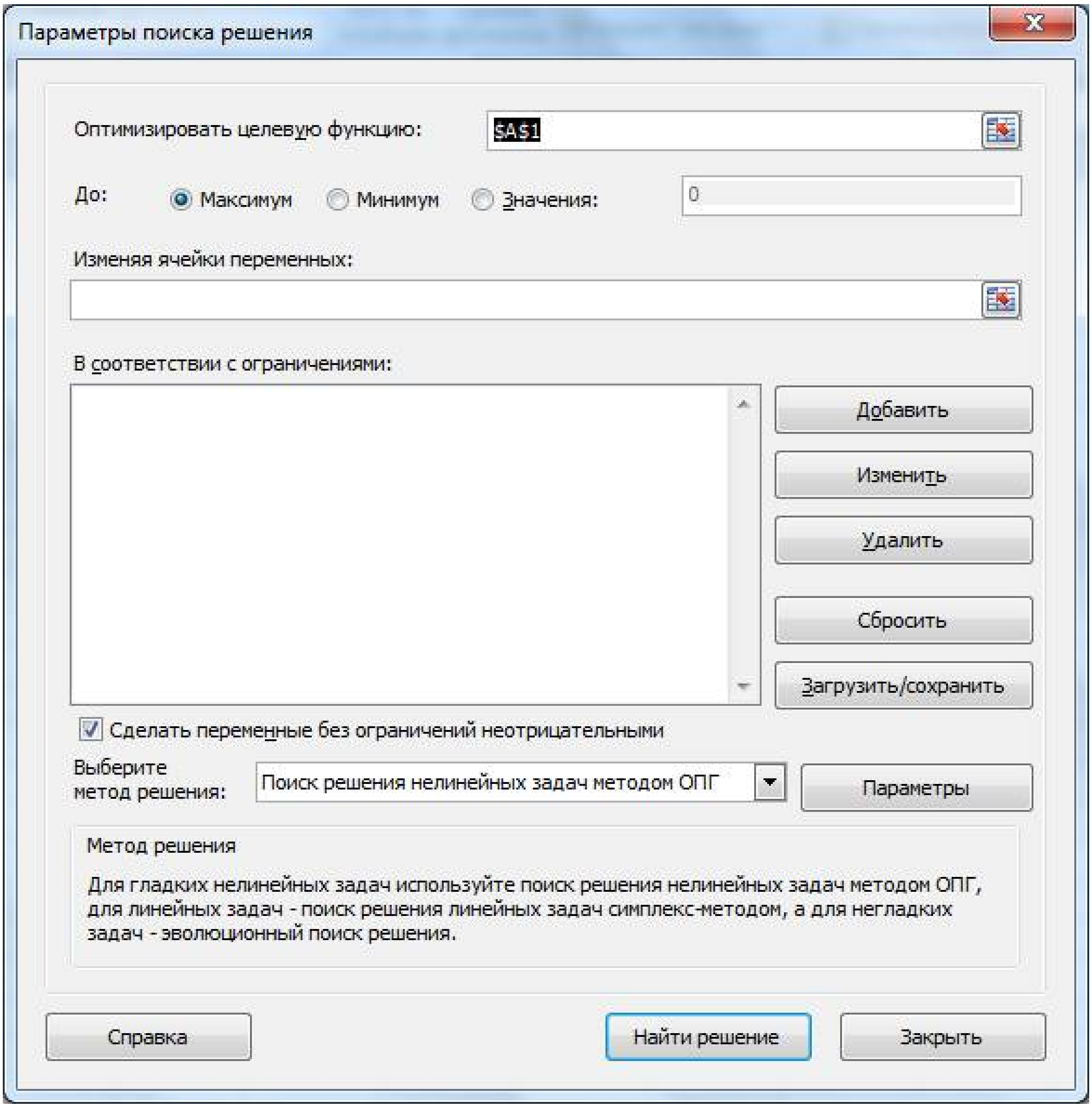


Рисунок 1.4. – Параметры поиска решения

Элементы окна Поиск решения

Оптимизировать целевую ячейку

Служит для указания целевой ячейки, значение которой необходимо максимизировать, минимизировать или установить равным заданному числу. Эта ячейка должна содержать формулу.

Равно

Служит для выбора варианта оптимизации значения целевой ячейки (максимизация, минимизация или подбор заданного числа).

Изменяя ячейки

Служит для указания ячеек, значения которых изменяются в процессе поиска решения до тех пор, пока не будут выполнены наложенные ограничения и условие оптимизации значения ячейки, указанной в поле **Установить целевую ячейку**.

Ограничения

Служит для отображения списка граничных условий поставленной задачи.

Добавить

Служит для отображения диалогового окна **Добавить ограничение**.

Изменить

Служит для отображения диалогового окна **Изменить ограничение**.

Удалить

Служит для снятия указанного ограничения.

Выполнить

Служит для запуска поиска решения поставленной задачи.

Закрыть

Служит для выхода из окна диалога без запуска поиска решения поставленной задачи. При этом сохраняются установки, сделанные в окнах диалога, появлявшихся после нажатий на кнопки **Параметры**, **Добавить**, **Изменить** или **Удалить**.

Параметры

Служит для отображения диалогового окна **Параметры поиска решения**, в котором можно загрузить или сохранить оптимизируемую модель и указать предусмотренные варианты поиска решения.

Восстановить

Служит для очистки полей окна диалога и восстановления значений параметров поиска решения, используемых по умолчанию.

1.2 Использование Поиска решения для решения задач оптимизации в нефтегазовом деле

Математическая модель задачи:

$$F(x_1, x_2) = 3x_1 + 2x_2 \to \max$$

$$\begin{cases} 4x_1 + 2x_2 \le 80 \\ 2x_1 + 3x_2 \le 60 \\ 0 \cdot x_1 + x_2 \le 15 \end{cases}$$

$$x_1 \ge 0 \quad x_2 \ge 0$$

Требуется найти x_1, x_2.

Далее запишем последовательность действий при использовании модуля Поиск решения.

1. Введём на лист Excel исходные данные

Таблица 1.2.1 – Исходные данные

	A	B	C	D	E	F
1	Затраты		Запасы		Расход ресурсов	
2	4	2	80			
3	2	3	60			
4	0	1	15			
5	X_1	X_2		Прибыль от единицы продукции		
6					3	2
7	Целевая функция					
8						

2. Ввод формул на лист Excel

Для расчёта целевой функции в ячейку A8 вводим формулу:

=E6 · A6 + F6 · B6.

Далее вводим формулы ограничения:

в ячейку E2

=A2 · A6 + B2 · B6;

в ячейку E3

=A3 · A6 + B3 · B6;

в ячейку E4

=A4 · A6 + B4 · B6.

3. Использование модуля Поиск решения

Входим в меню Сервис и выбираем Поиск решения в версии 2003, а в версии 2010 войти в меню Данные и выбрать Поиск решения.

Открывается окно Поиск решения, его заполняем (рис. 1.2.1)

Целевая ячейка – A8.

Устанавливаем поиск максимального значения целевой ячейки.
Задаём изменяемые ячейки – A6 : B6.

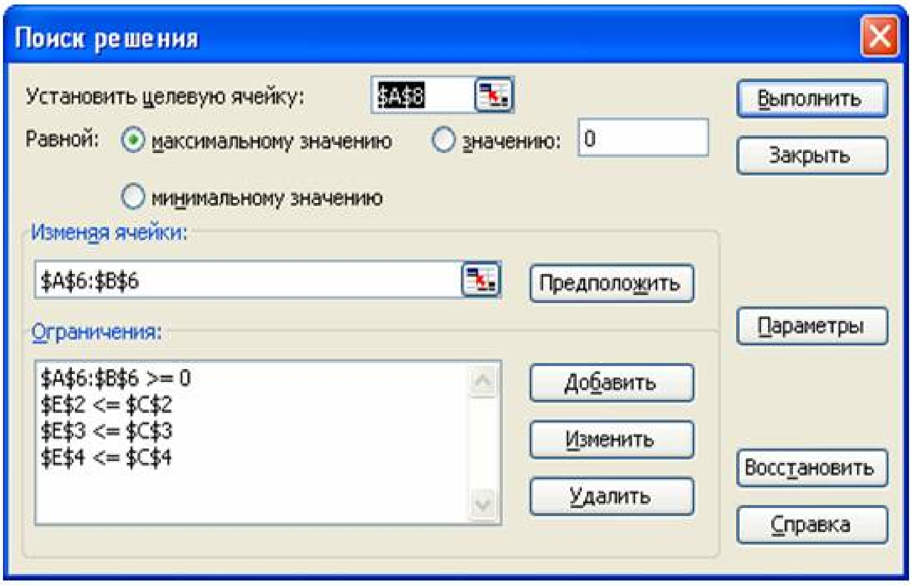


Рисунок 1.2.1 – Поиск решения

Вводим систему ограничений, щёлкнем по кнопке Добавить: Откроется окно ограничений. Вводим по очереди ограничения.

E2 ≤ C2; E3 ≤ C3; E4 ≤ C4; A6; B6 ≥ 0 (рис. 1.2.2).

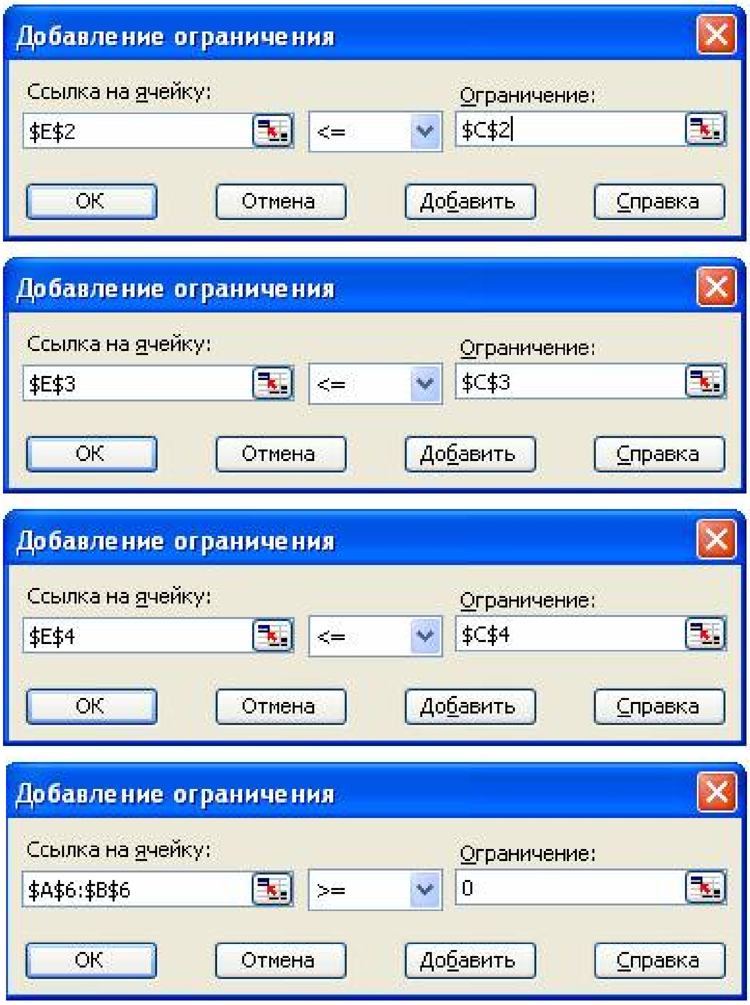


Рисунок 1.2.2 – Ограничения

Щёлкнув по кнопке ОК, выйдем из окна ограничений.

В окне Поиск решения щёлкнуть по кнопке Параметры. В окне Параметры поставим флажок напротив надписи Линейная модель. В окне Поиск решения щёлкнуть по кнопке Выполнить. Получаем необходимое максимальное значение целевой функции и изменяемых параметров.

Таблица 1.2.2 – Результаты работы

	A	B	C	D	E	F
1	Затраты		Запасы		Расход ресурсов	
2	4	2	80		80	
3	2	3	60		60	
4	0	1	15		10	
5	X1	X2		Прибыль от единицы продукции		
6	15	10			3	2
7	Целевая функция					
8	65					

Получаем: $x_1 = 15$, $x_2 = 10$ $f_{\max} = 65$.

Решить задачу линейного программирования, используя модуль Поиск решения электронных таблиц Excel.

Таблица 1.2.3 – Варианты заданий

Номер варианта	Условие	Номер варианта	Условие
1	$z = 6x_1 + 4x_2 \to max$ $\begin{cases} 16x_1 + 4x_2 \le 784 \\ 8x_1 + 7x_2 \le 552 \\ 5x_1 + 9x_2 \le 567 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	2	$z = 6x_1 + 2x_2 \to \max$ $\begin{cases} 12x_1 + 3x_2 \le 684 \\ 10x_1 + 5x_2 \le 690 \\ 3x_1 + 6x_2 \le 558 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$
3	$z = 2x_1 + 3x_2 \to \max$ $\begin{cases} 8x_1 + 3x_2 \le 864 \\ 7x_1 + 6x_2 \le 864 \\ 4x_1 + 9x_2 \le 945 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	4	$z = 5x_1 + 2x_2 \to \max$ $\begin{cases} 11x_1 + 3x_2 \le 671 \\ 8x_1 + 4x_2 \le 588 \\ 5x_1 + 3x_2 \le 423 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$
5	$z = 3x_1 + 2x_2 \to \max$ $\begin{cases} 15x_1 + 4x_2 \le 1095 \\ 11x_1 + 5x_2 \le 865 \\ 9x_1 + 10x_2 \le 1080 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	6	$z = 3x_1 + 9x_2 \to \max$ $\begin{cases} 9x_1 + 5x_2 \le 1431 \\ 7x_1 + 8x_2 \le 1224 \\ 4x_1 + 16x_2 \le 1328 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$

Номер варианта	Условие	Номер варианта	Условие
7	$z = 3x_1 + 9x_2 \to \max$ $\begin{cases} 6x_1 + 3x_2 \le 714 \\ 5x_1 + 10x_2 \le 910 \\ 3x_1 + 12x_2 \le 948 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	8	$z = 3x_1 + 2x_2 \to \max$ $\begin{cases} 9x_1 + 4x_2 \le 801 \\ 6x_1 + 7x_2 \le 807 \\ 3x_1 + 8x_2 \le 768 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0$
9	$z = 2x_1 + 3x_2 \to \max$ $\begin{cases} 3x_1 + 5x_2 \le 453 \\ 4x_1 + 8x_2 \le 616 \\ 3x_1 + 11x_2 \le 627 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	10	$z = 7x_1 + 9x_2 \to \max$ $\begin{cases} 10x_1 + 9x_2 \le 1870 \\ 5x_1 + 11x_2 \le 1455 \\ 4x_1 + 15x_2 \le 1815 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$
11	$z = 7x_1 + 4x_2 \to \max$ $\begin{cases} 5x_1 + 2x_2 \le 505 \\ 3x_1 + 3x_2 \le 393 \\ 2x_1 + 3x_2 \le 348 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	12	$z = 6x_1 + 5x_2 \to \max$ $\begin{cases} 7x_1 + 3x_2 \le 1365 \\ 6x_1 + 3x_2 \le 1245 \\ x_1 + 2x_2 \le 650 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$
13	$z = 6x_1 + 3x_2 \to \max$ $\begin{cases} 6x_1 + 2x_2 \le 600 \\ 4x_1 + 3x_2 \le 520 \\ 3x_1 + 4x_2 \le 600 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	14	$z = 5x_1 + 6x_2 \to \max$ $\begin{cases} 5x_1 + 3x_2 \le 750 \\ 4x_1 + 3x_2 \le 630 \\ 3x_1 + 4x_2 \le 700 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$
15	$z = 6x_1 + 2x_2 \to \max$ $\begin{cases} 8x_1 + 2x_2 \le 840 \\ 6x_1 + 3x_2 \le 870 \\ 3x_1 + 2x_2 \le 560 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	16	$z = 4x_1 + 5x_2 \to \max$ $\begin{cases} 3x_1 + 2x_2 \le 273 \\ 3x_1 + 3x_2 \le 300 \\ 2x_1 + 5x_2 \le 380 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$
17	$z = 7x_1 + 5x_2 \to \max$ $\begin{cases} 2x_1 + x_2 \le 438 \\ 3x_1 + 6x_2 \le 747 \\ 3x_1 + 7x_2 \le 812 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	18	$z = 2x_1 + 4x_2 \to \max$ $\begin{cases} 4x_1 + 3x_2 \le 480 \\ 3x_1 + 4x_2 \le 444 \\ 2x_1 + 6x_2 \le 546 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$
19	$z = 6x_1 + 5x_2 \to \max$ $\begin{cases} 4x_1 + 3x_2 \le 440 \\ 3x_1 + 4x_2 \le 393 \\ 3x_1 + 5x_2 \le 450 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	20	$z = 3x_1 + 8x_2 \to \max$ $\begin{cases} 2x_1 + 3x_2 \le 428 \\ 3x_1 + 6x_2 \le 672 \\ 2x_1 + 8x_2 \le 672 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$

Номер варианта	Условие	Номер варианта	Условие
21	$z = 7x_1 + 4x_2 \to \max$ $\begin{cases} 5x_1 + 2x_2 \le 505 \\ 3x_1 + 3x_2 \le 393 \\ 2x_1 + 3x_2 \le 348 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	22	$z = 6x_1 + 5x_2 \to \max$ $\begin{cases} 7x_1 + 3x_2 \le 1365 \\ 6x_1 + 3x_2 \le 1245 \\ x_1 + 2x_2 \le 650 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$
23	$z = 6x_1 + 3x_2 \to max$ $\begin{cases} 6x_1 + 2x_2 \le 600 \\ 4x_1 + 3x_2 \le 520 \\ 3x_1 + 4x_2 \le 600 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	24	$z = 5x_1 + 6x_2 \to \max$ $\begin{cases} 5x_1 + 3x_2 \le 750 \\ 4x_1 + 3x_2 \le 630 \\ 3x_1 + 4x_2 \le 700 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$
25	$z = 6x_1 + 2x_2 \to \max$ $\begin{cases} 8x_1 + 2x_2 \le 840 \\ 6x_1 + 3x_2 \le 870 \\ 3x_1 + 2x_2 \le 560 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	26	$z = 4x_1 + 5x_2 \to \max$ $\begin{cases} 3x_1 + 2x_2 \le 273 \\ 3x_1 + 3x_2 \le 300 \\ 2x_1 + 5x_2 \le 380 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$
27	$z = 6x_1 + 2x_2 \to \max$ $\begin{cases} 12x_1 + 3x_2 \le 684 \\ 10x_1 + 5x_2 \le 620 \\ 3x_1 + 6x_2 \le 558 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	28	$z = 2x_1 + 3x_2 \to \max$ $\begin{cases} 8x_1 + 3x_2 \le 864 \\ 7x_1 + 6x_2 \le 864 \\ 4x_1 + 9x_2 \le 945 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$
29	$z = 5x_1 + 2x_2 \to \max$ $\begin{cases} 11x_1 + 3x_2 \le 671 \\ 8x_1 + 4x_2 \le 588 \\ 5x_1 + 3x_2 \le 423 \end{cases}$ $x_1 \ge 0;\ x_2 \ge 0.$	30	$z = 10x_1 + 12x_2 \to \max$ $\begin{cases} 16x_1 + 2x_2 \le 304 \\ 3x_1 + 2x_2 \le 83 \\ 6x_1 + 15x_2 \le 375 \end{cases}$ $x_1 \ge 0; x_2 \ge 0$

1.3 Обоснование плана интенсификации прироста скважинной добычи нефти

Постановка задачи

В НГДУ проводятся различные виды мероприятий по интенсификации добычи, например, обработки призабойных зон скважин (ПЗС): кислотные (КО), глинокислотные (ГКО), соляно-кислотные (СКО) и ванны (СКВ), гидроразрывы пластов (ГРП) и т. д. Для проведения обработок требуются различные виды производственных ресурсов: рабочая сила, оборудование, реагенты и т. д.

Затраты ресурсов различаются в зависимости от вида обработок ПЗС, принятой технологии проведения.

Результатом осуществления мероприятий является прирост добычи нефти, который также различен при разных видах обработок ПЗС.

Задача заключается в определении такого плана проведения мероприятий по интенсификации, который, исходя из рационального использования имеющихся ресурсов, обеспечивает максимальный прирост добычи нефти.

Составим математическую модель задачи.

Обозначим b_1 фонд времени оборудования группы А,

b_2 – фонд времени оборудования группы Б,

b_3 – фонд времени оборудования группы В,

b_4 – трудозатраты бригады рабочих,

b_5 – материалы, реагенты,

i – индекс производственного ресурса,

j – индекс вида обработки ПЗС,

X_1 – количество обработок вида СКО,

X_2 – количество обработок вида КО,

X_3 – количество обработок вида ГКО,

X_4 – количество обработок вида СКВ.

Известны:

b_i – запас i-го производственного ресурса на планируемый период;

a_{ij} – нормы затрат i-го вида ресурса на проведение одной скважинной операции j-го вида обработки ПЗС;

c_j – прирост добычи нефти на одну скважинную операцию при проведении определённого вида обработки.

Таблица 1.3.1 – Исходные данные

Виды ресурсов	Фонд ресурсов	Единица измерения	Нормы затрат ресурсов на одну скважинную операцию			
1. Оборудование А	b_1	Маш.-смен	a_{11}	a_{12}	a_{13}	a_{14}
2. Оборудование Б	b_2	Маш.-смен	a_{21}	a_{22}	a_{23j}	a_{24}
3. Оборудование В	b_3	Маш.-смен	a_{31}	a_{32}	a_{33}	a_{34}
4. Трудозатраты бригад	b_4	Чел. час	a_{41}	a_{42}	a_{43}	a_{44}
5. Материалы и реагенты	b_5	Тыс. руб.	a_{51}	a_{52}	a_{53}	a_{54}
Прирост добычи на 1 скважинную операцию		Тонны	C_1	C_2	C_3	C_4

Математическая модель имеет вид:

$F(x_1, x_2, x_3, x_4) = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 \to \max$ – целевая функция.

Система ограничений:

$$
\begin{cases}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 \le b_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 \le b_2 \\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 \le b_3 \\
a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 \le b_4 \\
a_{51}x_1 + a_{52}x_2 + a_{53}x_3 + a_{54}x_4 \le b_5
\end{cases}
$$

$$x_{ij} \ge 0$$

Составить математическую модель для своего варианта и, используя модуль Поиск решения, рассчитать количество операций обработок каждого вида для получения максимального прироста добычи нефти.

Задания для выполнения

Таблица 1.3.2 – Варианты задания

Вариант	Вид ресурса	Фонд ресурса	Норма затрат ресурсов на одну скважинную операцию по видам				Вариант	Вид ресурса	Фонд ресурса	Норма затрат ресурсов на одну скважинную операцию по видам			
1	1	480	0	1	0	1	2	1	280	0	0	1	1
	2	960	1	0	0	1		2	360	1	1	3	2
	3	1000	2	1	1	1		3	640	1	3	2	0
	4	800	1	1	4	2		4	300	1	2	1	1
	5	300	3	2	1	1		5	580	1	1	2	1
		C_j	40	20	30	10			C_j	16	18	25	20
3	1	560	0	2	1	0	4	1	840	2	0	3	1
	2	500	1	0	0	1		2	300	0	1	2	2
	3	400	1	1	0	1		3	760	3	0	0	4
	4	690	1	1	1	2		4	560	1	1	1	1
	5	420	3	2	2	1		5	900	2	1	3	1
		C_j	20	24	30	20			C_j	25	15	35	40
5	1	680	0	1	2	0	6	1	1000	1	0	1	0
	2	600	1	0	0	1		2	460	0	1	1	1
	3	380	1	0	1	0		3	1200	1	0	0	3
	4	800	2	1	1	1		4	600	2	1	1	2
	5	500	1	2	2	3		5	900	2	2	3	1
		C_j	10	15	12	10			C_j	20	18	19	28
7	1	400	0	0	1	2	8	1	680	2	0	1	3
	2	350	1	1	0	0		2	700	0	1	2	1
	3	450	1	0	0	1		3	800	1	0	4	0
	4	900	2	1	1	1		4	500	1	1	1	1
	5	950	2	1	1	2		5	700	1	1	2	2
		C_j	20	18	16	10			C_j	40	20	70	30
9	1	200	0	0	1	1	10	1	960	1	1	1	0
	2	580	1	1	1	2		2	720	0	1	0	1
	3	960	1	4	2	0		3	1220	4	2	1	0
	4	500	1	1	1	1		4	600	1	1	1	2
	5	880	2	1	1	2		5	550	1	1	1	1
		C_j	40	30	15	15			C_j	20	17	15	27

Вариант	Вид ресурса	Фонд ресурса	Норма затрат ресурсов на одну скважинную операцию по видам				Вариант	Вид ресурса	Фонд ресурса	Норма затрат ресурсов на одну скважинную операцию по видам			
11	1	300	0	0	1	1	12	1	500	1	1	0	1
	2	480	2	1	1	2		2	800	1	1	0	1
	3	920	1	4	2	0		3	600	0	1	0	1
	4	520	1	1	1	1		4	200	4	1	1	1
	5	800	1	2	2	1		5	900	1	1	1	1
		C_j	20	30	12	10			C_j	20	15	15	28
13	1	500	0	1	1	0	14	1	600	1	0	0	1
	2	200	1	2	1	0		2	500	0	1	1	1
	3	800	2	1	2	1		3	250	2	1	1	1
	4	360	1	1	1	1		4	480	1	2	2	1
	5	820	2	1	1	2		5	620	2	1	1	1
		C_j	10	15	18	20			C_j	15	10	24	20
15	1	600	0	1	2	1	16	1	300	1	1	0	1
	2	300	1	2	0	1		2	350	2	1	1	1
	3	800	1	2	1	1		3	480	1	1	1	1
	4	400	2	1	1	1		4	240	2	1	2	1
	5	550	2	2	1	1		5	660	2	2	1	1
		C_j	20	15	12	24			C_j	12	18	24	20
17	1	400	1	2	1	0	18	1	420	1	2	0	1
	2	420	0	1	2	0		2	580	1	0	0	1
	3	580	2	1	0	2		3	600	1	1	1	1
	4	660	2	4	1	2		4	820	2	1	2	1
	5	880	1	2	2	1		5	940	2	2	1	1
		C_j	20	15	12	24			C_j	40	30	12	10
19	1	240	1	0	1	0	20	1	420	1	0	1	1
	2	220	2	1	1	1		2	380	0	1	1	1
	3	520	1	2	1	1		3	640	0	1	2	1
	4	400	2	1	2	1		4	880	2	1	1	1
	5	980	2	2	1	1		5	920	2	2	1	3
		C_j	20	12	14	22			C_j	18	20	28	22
21	1	500	1	1	2	1	22	1	600	1	2	1	2
	2	300	4	3	4	1		2	750	3	4	2	1
	3	240	2	2	3	1		3	270	1	0	3	4
	4	360	3	4	2	1		4	240	1	1	2	2
	5	250	1	1	0	1		5	360	4	3	2	0
		C_j	20	25	32	40			C_j	15	20	18	24
23	1	480	1	2	0	1	24	1	550	1	1	2	3
	2	360	4	3	0	1		2	450	4	1	1	3
	3	240	2	3	1	0		3	320	3	5	2	1
	4	560	3	4	1	2		4	340	4	2	1	0
	5	350	5	4	2	1		5	280	2	5	3	1
		C_j	30	25	34	18			C_j	16	24	28	32
25	1	270	1	0	1	0	26	1	600	2	0	1	1
	2	200	4	3	2	2		2	280	4	2	1	1
	3	350	2	4	2	3		3	350	3	4	2	1
	4	420	3	2	1	4		4	300	1	1	0	1
	5	1000	1	0	3	2		5	240	1	2	1	2
		C_j	20	28	32	40			C_j	35	28	20	15

Вариант	Вид ресурса	Фонд ресурса	Норма затрат ресурсов на одну скважинную операцию по видам				Вариант	Вид ресурса	Фонд ресурса	Норма затрат ресурсов на одну скважинную операцию по видам			
27	1	160	1	0	2	0	28	1	360	2	1	1	1
	2	230	1	1	1	1		2	450	3	2	1	0
	3	240	3	4	1	2		3	240	4	3	1	1
	4	380	1	2	1	2		4	450	3	2	2	2
	5	400	4	2	1	1		5	320	1	2	1	2
		C_j	30	40	45	18			C_j	32	26	40	18
29	1	380	1	1	2	1	30	1	500	2	0	1	1
	2	240	2	3	3	1		2	280	4	2	0	1
	3	320	1	1	1	1		3	350	3	4	2	1
	4	400	4	2	2	0		4	400	1	1	0	1
	5	560	5	1	1	2		5	240	1	2	1	2
		C_j	20	32	40	22			C_j	24	32	40	18

1.4 Использование Поиска решения для решения задач в нефтегазовом деле

1.4.1 Применение надстройки для решения транспортной задачи

Постановка задачи. Имеются три пункта поставки однородного груза A_1, A_2 и A_3 и пять пунктов B_1, B_2, B_3, B_4, B_5 потребления этого груза. На пунктах A_1, A_2 и A_3 находится груз, соответственно, в количестве a_1, a_2 и a_3 тонн. В пункты B_1, B_2, B_3, B_4, B_5 требуется доставить, соответственно, b_1, b_2, b_3, b_4, b_5 тонн груза. Стоимость перевозки единицы груза от пункта поставки до пункта потребления приведена в следующей матрице – таблице 1.4.1:

Таблица 1.4.1 – Транспортная задача

Пункты поставки	Пункты потребления				
	B_1	B_2	B_3	B_4	B_5
A_1	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
A_2	d_{21}	d_{22}	d_{23}	d_{24}	d_{25}
A_3	d_{31}	d_{32}	d_{33}	d_{34}	d_{35}

Найти план закрепления потребителей за поставщиками однородного груза, чтобы общие затраты по перевозкам были минимальными. Составить математическую модель задачи и решить её, используя Поиск решения таблиц Excel.

$a_1 = 400$

$a_2 = 250$

$a_3 = 350$

$b_1 = 200$; $b_2 = 170$; $b_3 = 230$; $b_4 = 225$; $b_5 = 175$.

$$D = \begin{pmatrix} 13 & 9 & 5 & 11 & 17 \\ 14 & 5 & 12 & 14 & 22 \\ 20 & 17 & 13 & 18 & 21 \end{pmatrix}.$$

Решение

Найдём сумму запасов и сумму потребностей:

$a_1 + a_2 + a_3 = 400 + 250 + 350 = 1000;$

$b_1 + b_2 + b_3 + b_4 + b_5 = 200 + 170 + 230 + 225 + 175 = 1000$. Если сумма запасов равна сумме потребностей, то модель транспортной задачи называется закрытой, в противном случае модель будет открытой. Открытую модель можно привести к закрытой по правилу: если сумма запасов больше суммы потребностей, то добавляется фиктивный потребитель, потребности которого равны разности между суммой запасов и суммой потребностей, а тарифы перевозок к нему равны нулям, задача превращается в закрытую форму. Если сумма потребностей больше суммы запасов, то добавляется фиктивный поставщик, запасы которого равны разности между суммой потребностей и суммой запасов, а тарифы перевозок от него равны нулям, задача превращается в закрытую форму.

Рассмотрим сначала закрытую модель. Сумма запасов равна сумме потребностей, то есть все запасы должны быть вывезены, и все потребности удовлетворены.

Обозначим $x_{i,j}$ – количество груза, перевозимого от A_i к B_j, и составим математическую модель задачи.

Целевая функция – общие затраты на перевозки:

$$\begin{aligned} Z &= 13x_{11} + 9x_{12} + 5x_{13} + 11x_{14} + 17x_{15} + \\ &+ 14x_{21} + 5x_{22} + 12x_{23} + 14x_{24} + 22x_{25} + \\ &+ 20x_{31} + 17x_{32} + 13x_{33} + 18x_{34} + 21x_{35} \to \min. \end{aligned}$$

Ограничения:

$$\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} + x_{15} = 400 \\ x_{21} + x_{22} + x_{23} + x_{24} + x_{25} = 250 \\ x_{31} + x_{32} + x_{33} + x_{34} + x_{35} = 350 \\ x_{11} + x_{21} + x_{31} = 200 \\ x_{12} + x_{22} + x_{32} = 170 \\ x_{13} + x_{23} + x_{33} = 230 \\ x_{14} + x_{24} + x_{34} = 225 \\ x_{15} + x_{25} + x_{35} = 175 \end{cases}$$

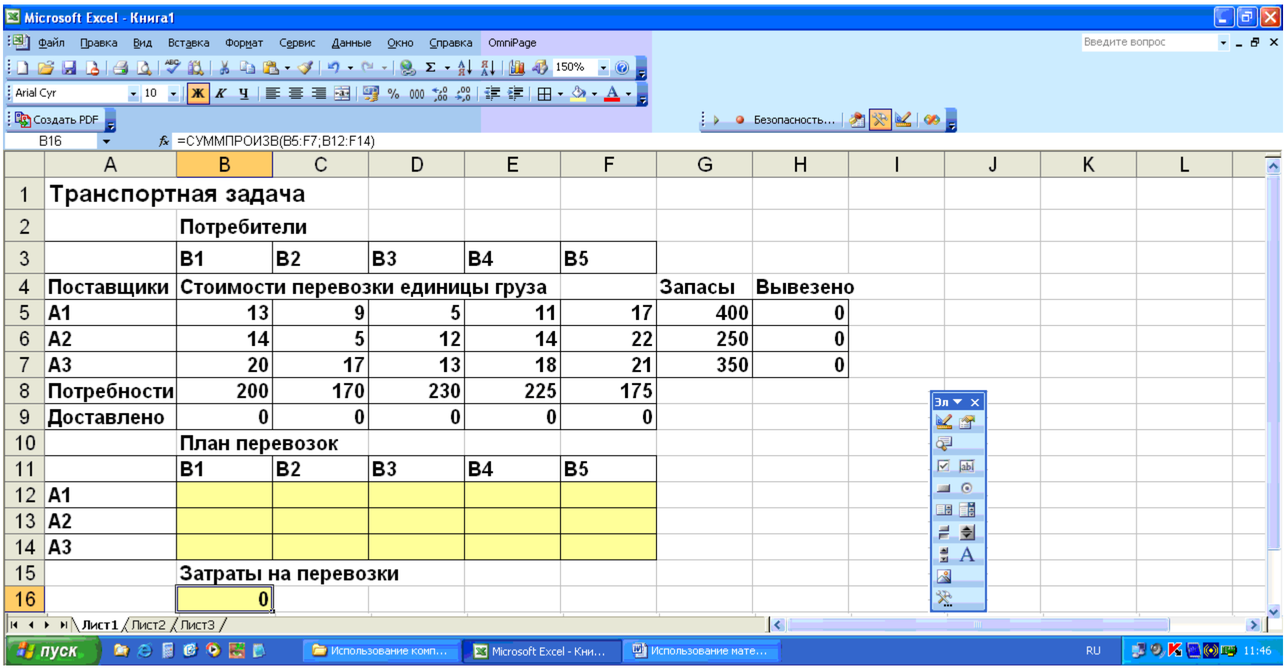
$$X_{ij} \ge 0, i = 1..3, j = 1..5.$$

Найти неотрицательные значения x_{ij}, удовлетворяющие системе ограничений и минимизирующие функцию z – затраты на перевозки.

Рассмотрим последовательность действий для решения этой задачи, используя модуль Поиск решения.

1. **Подготовка исходных данных на листе Excel**. В ячейку A1 ввести текст «**Транспортная задача**». В ячейку B2 текст «**Потребители**». В ячейки B3 по F3 названия потребителей. В ячейку A4 ввести текст «**Поставщики**». В ячейку G4 текст «**Запасы**». В ячейки A5 по A7 названия поставщиков. В ячейку A8 текст «**Потребности**». Значения запасов ввести в блок ячеек G5 : G7. Значения потребностей ввести в блок ячеек B8 : F8. Стоимости перевозки единицы груза от Ai к Bj ввести в блок ячеек B5 : F7. Для плана перевозок отведём блок ячеек B12 : F14. В ячейку A9 ввести текст «**Доставлено**». В ячейку H4 ввести текст «**Вывезено**». В ячейку B10 ввести текст «**План перевозок**». В ячейки B11 по F11 названия потребителей. В ячейки A12 по A14 названия поставщиков. В B15 ввести текст «**Затраты на перевозки**» (рис. 1.4.1).

2. **Ввод формул и функций**. В ячейку B16 вставим функцию Суммпроизв. В окне этой функции указать первый массив B5 : F7, второй массив B12 : F14. В ячейки H5, H6, H7 ввести функции: в H5 – функцию СУММ(B12 : F12), в H6 – функцию СУММ(B13 : F13), в H7 – функцию СУММ(B14 : F14). В ячейки B9, C9, D9, E9, F9 ввести функции: в B9 – функцию СУММ(B12 : B14), в C9 – функцию СУММ(C12 : C14), в D9 – функцию СУММ(D12 : D14), в E9 – функцию СУММ(E12 : E14), в F9 – функцию СУММ(F12 : F14).



	A	B	C	D	E	F	G	H
1	Транспортная задача							
2		Потребители						
3		B1	B2	B3	B4	B5		
4	Поставщики	Стоимости перевозки единицы груза					Запасы	Вывезено
5	A1	13	9	5	11	17	400	0
6	A2	14	5	12	14	22	250	0
7	A3	20	17	13	18	21	350	0
8	Потребности	200	170	230	225	175		
9	Доставлено	0	0	0	0	0		
10		План перевозок						
11		B1	B2	B3	B4	B5		
12	A1							
13	A2							
14	A3							
15		Затраты на перевозки						
16		0						

Рисунок 1.4.1 – Исходные данные

3. **Войти в меню Данные, Поиск решения** или в поздних версиях Офиса в меню Данные, Поиск решения. В окне Поиска решения установим целевую ячейку A16, переключатель установим на min, в поле Изменяя ячейки укажем мышкой блок B12 : F14. Далее ввести ограничения. Щёлкнуть по кнопке Добавить.

4. В окне ограничений ввести одно за другим 9 ограничений:

первое ограничение: H5 = G5; второе ограничение: H6 = G6;
третье ограничение: H7 = G7; четвёртое ограничение: B9 = B8;
пятое ограничение: C9 = C8; шестое ограничение: D9 = D8;
седьмое ограничение: E9 = E8; восьмое ограничение: F9 = F8;
девятое ограничение: B12 : F14 ≥ 0 (рис. 1.4.2).

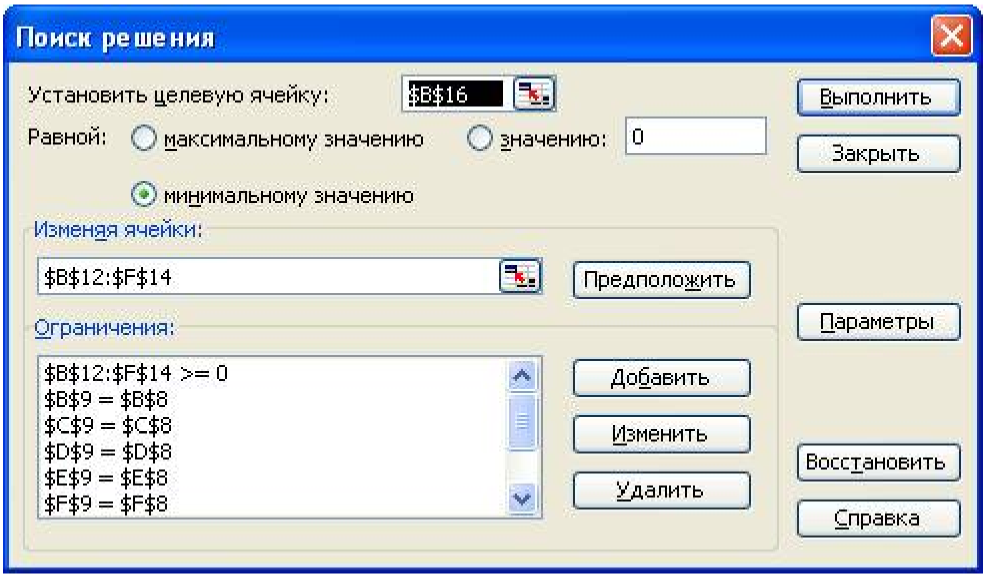
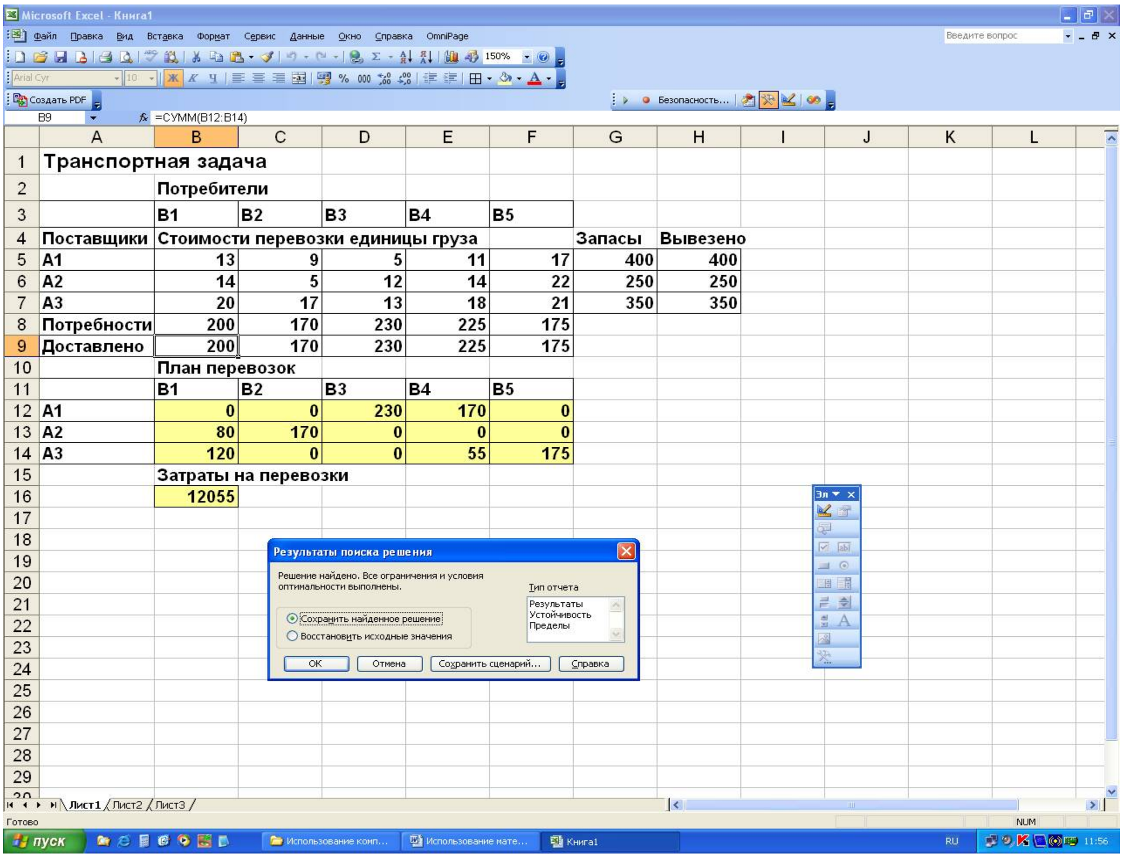


Рисунок 1.4.2 – Поиск решения

5. После ввода ограничений снова войти в окно Поиск решения и проверить правильность ввода ограничений. Если есть ошибки, использовать кнопки **Изменить** или **Удалить**. Затем войти в окно **Параметры** и поставить флажок в поле напротив надписи **Линейная функция**. Далее в окне Поиск решения щёлкнуть по кнопке **Выполнить**.

6. В результате получим оптимальный план перевозок и значение функции минимальных затрат: Z_{min} = 12 055 единиц (рис. 1.4.3).

$x_{11} = 0$ $x_{12} = 0$ $x_{13} = 230$ $x_{14} = 170$ $x_{15} = 0$;
$x_{21} = 80$ $x_{22} = 170$ $x_{23} = 0$ $x_{24} = 0$ $x_{25} = 0$;
$x_{31} = 120$ $x_{32} = 0$ $x_{33} = 0$ $x_{34} = 55$ $x_{35} = 175$.



	A	B	C	D	E	F	G	H
1	Транспортная задача							
2		Потребители						
3		B1	B2	B3	B4	B5		
4	Поставщики	Стоимости перевозки единицы груза					Запасы	Вывезено
5	A1	13	9	5	11	17	400	400
6	A2	14	5	12	14	22	250	250
7	A3	20	17	13	18	21	350	350
8	Потребности	200	170	230	225	175		
9	Доставлено	200	170	230	225	175		
10		План перевозок						
11		B1	B2	B3	B4	B5		
12	A1	0	0	230	170	0		
13	A2	80	170	0	0	0		
14	A3	120	0	0	55	175		
15		Затраты на перевозки						
16		12055						

Рисунок 1.4.3 – Результат работы

1.4.2 Транспортная задача открытого типа

В регионе расположено несколько НГДУ, обеспечивающих определённые объёмы добычи нефти, которая поступает на НПЗ, расположенные в различных регионах страны и имеющие различные производственные мощности. В силу разноудалённости потребителей от НГДУ затраты на транспортировку нефти различаются.

В задаче необходимо составить план закрепления поставщиков за потребителями, который учитывает, по возможности, наиболее полное удовлетворение потребителей НПЗ и при этом обеспечивает минимальные затраты на транспортировку нефти.

***Введём условные обозначения*:**

i – индекс НГДУ, i = 1,m;

m – общее число НГДУ в регионе;

j – Индекс НПЗ, j = 1,n;

n – общее число НПЗ.

Известно:

a_i – объёмы добычи нефти в i-ом НГДУ, тыс. т,

b_j – потребность j-го НПЗ в нефти, тыс. т,

c_{ij} – издержки на транспортировку 1000 т нефти, тыс. руб.

Таблица 1.4.2 – Варианты задания

№ вар.	b_j \ a_i							№ вар.	b_j \ a_i						
№ вар.	b_j \ a_i	180	170	200	210	160	130	№ вар.	b_j \ a_i	150	130	120	200	240	170
1	500	3	9	7	5	8	6	16	350	5	6	7	5	8	9
	320	6	4	10	8	4	9		450	4	7	7	9	10	6
	300	8	7	5	12	10	6		350	9	3	11	8	10	6
№ вар.	b_j \ a_i	110	190	220	240	100	170	№ вар.	b_j \ a_i	130	100	110	120	150	140
2	400	9	6	5	3	4	8	17	390	3	8	6	5	7	4
	500	8	3	2	8	7	7		370	6	4	7	4	8	5
	200	3	4	6	9	11	5		210	8	5	8	3	6	9
№ вар.	b_j \ a_i	150	160	130	110	150	170	№ вар.	b_j \ a_i	210	200	190	180	160	140
3	400	7	6	5	9	4	9	18	520	5	7	9	5	8	9
	450	8	6	2	5	3	8		410	7	5	7	3	6	5
	200	3	4	6	8	4	3		300	8	5	6	7	2	4
№ вар.	b_j \ a_i	130	120	160	140	180	100	№ вар.	b_j \ a_i	180	190	110	210	200	120
4	320	7	8	5	8	4	8	19	490	5	7	8	4	6	9
	340	9	5	7	2	6	7		270	7	2	5	8	6	7
	370	9	4	6	7	9	5		380	5	4	7	6	9	8
№ вар.	b_j \ a_i	180	200	220	210	160	130	№ вар.	b_j \ a_i	120	240	230	220	200	170
5	340	9	8	3	7	3	4	20	520	5	4	5	6	8	3
	390	4	5	7	7	2	8		440	7	2	8	4	5	7
	380	4	6	6	8	3	6		410	4	3	6	5	2	6
№ вар.	b_j \ a_i	50	100	150	100	150	50	№ вар.	b_j \ a_i	110	140	150	100	150	120
6	200	3	5	4	2	1	4	21	150	4	3	7	8	1	3
	300	8	7	3	6	5	7		250	5	9	3	10	5	2
	150	2	4	9	8	1	3		100	7	2	1	2	4	1
№ вар.	b_j \ a_i	40	110	200	300	250	100	№ вар.	b_j \ a_i	100	50	150	200	110	100
7	300	2	1	4	8	7	1	22	210	1	4	3	2	1	7
	350	5	9	6	3	2	2		300	7	8	5	4	2	3
	400	7	8	10	1	4	5		390	3	2	5	1	7	1
№ вар.	b_j \ a_i	220	200	150	170	130	100	№ вар.	b_j \ a_i	100	150	70	130	180	100
8	300	1	4	3	5	2	1	23	220	8	4	3	2	1	2
	350	7	9	13	4	1	3		380	7	9	8	3	2	3
	350	2	3	5	8	1	4		350	4	5	9	3	1	6
№ вар.	b_j \ a_i	70	130	240	160	100	100	№ вар.	b_j \ a_i	200	150	130	170	150	100
9	320	2	4	3	5	1	2	24	350	1	4	3	2	1	2
	280	7	8	9	6	3	4		350	5	8	9	7	3	3
	250	1	4	5	9	7	1		280	8	1	2	4	1	6
№ вар.	b_j \ a_i	150	250	300	100	180	100	№ вар.	b_j \ a_i	180	250	120	210	190	110
10	400	2	4	5	8	9	7	25	300	10	12	3	6	11	4
	320	1	5	7	6	3	4		350	4	5	9	7	8	7
	380	2	1	11	9	10	2		450	8	7	3	2	1	2

№ вар.	b_j \ a_i	150	250	300	100	180	100	№ вар.	b_j \ a_i	80	250	120	210	190	110
11	300	2	5	3	8	9	7	26	200	15	12	3	6	10	4
	320	1	4	7	6	3	4		350	4	5	9	3	18	7
	380	2	1	12	9	10	2		450	8	7	3	2	1	2
№ вар.	b_j \ a_i	50	200	300	100	180	100	№ вар.	b_j \ a_i	180	250	220	210	190	110
12	200	3	4	5	8	9	7	27	400	12	15	3	6	11	4
	320	1	5	7	6	3	4		350	4	3	9	7	8	7
	380	2	1	11	9	12	2		450	8	7	3	2	1	2
№ вар.	b_j \ a_i	250	250	300	100	180	100	№ вар.	b_j \ a_i	180	350	120	210	190	110
13	500	7	4	5	8	9	2	28	300	10	11	3	6	11	4
	320	1	5	7	6	3	4		450	14	5	8	7	8	7
	380	2	1	11	9	10	2		450	8	7	3	2	7	3
№ вар.	b_j \ a_i	150	250	200	100	280	100	№ вар.	b_j \ a_i	180	250	120	310	190	110
14	400	2	8	6	3	9	7	29	300	16	12	1	6	11	4
	220	1	5	7	5	3	4		350	14	5	9	3	8	7
	480	2	1	11	9	15	2		550	8	7	3	2	1	2
№ вар.	b_j \ a_i	150	250	300	200	180	100	№ вар.	b_j \ a_i	280	250	120	210	190	110
15	400	2	7	5	8	6	4	30	400	10	16	5	8	11	4
	420	1	5	7	9	3	4		350	4	5	9	7	8	7
	380	2	1	11	9	10	2		450	8	7	3	5	2	4

Указания.

Модель задачи.

1. В качестве неизвестных задачи принимаются переменные x_{ij}, означающие объём перевозок нефти от i-го НГДУ к j-му НПЗ.

2. В качестве коэффициентов целевой функции выступают издержки на перевозку 1000 т нефти. Целевая функция минимизируется.

2. Надстройка Excel Анализ данных

В состав Microsoft Excel входит набор средств анализа данных (так называемый пакет анализа), предназначенный для решения сложных статистических и инженерных задач. Для анализа данных с помощью этих инструментов следует указать входные данные и выбрать параметры; анализ будет выполнен с помощью подходящей статистической или инженерной макрофункции, а результат будет помещён в выходной диапазон. Другие средства позволяют представить результаты анализа в графическом виде.

Обращение к средствам анализа данных. Они доступны через команду **Анализ данных** в меню **Сервис**. Если этой команды нет в меню, необходимо загрузить надстройку **Пакет анализа** (рис. 2.1).

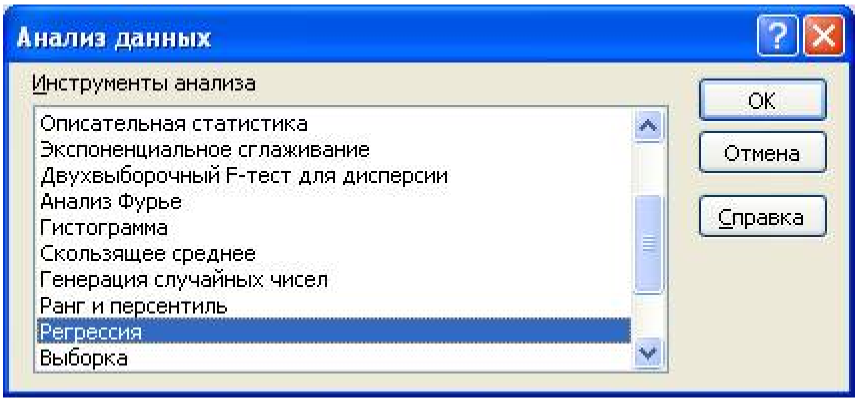


Рисунок 2.1 – Инструменты анализа

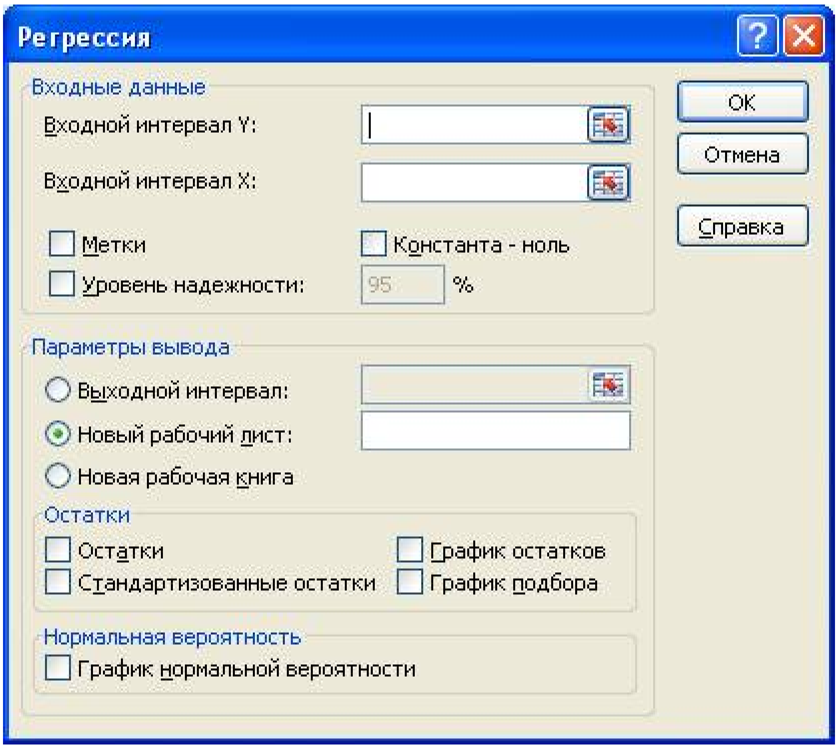


Рисунок 2.2 – Регрессия

Линейный регрессионный анализ заключается в подборе графика для набора наблюдений с помощью метода наименьших квадратов. Регрессия используется для анализа воздействия на отдельную зависимую переменную значений одной или более независимых переменных.

Элементы диалогового окна Регрессия (рис. 2.2).

Входной интервал Y

Введите ссылку на диапазон анализируемых зависимых данных. Диапазон должен состоять из одного столбца.

Входной интервал X

Введите ссылку на диапазон независимых данных, подлежащих анализу. Microsoft Excel располагает независимые переменные этого диапазона слева направо в порядке возрастания. Максимальное число входных диапазонов равно 16.

Заголовки

Установите флажок, если первая строка или первый столбец входного интервала содержит заголовки. Снимите флажок, если заголовки отсутствуют; в этом случае подходящие названия для данных выходного диапазона будут созданы автоматически.

Уровень надёжности

Установите флажок, чтобы включить в выходной диапазон дополнительный уровень. В соответствующее поле введите уровень надёжности, который будет использован дополнительно к уровню 95 %, применяемому по умолчанию.

Константа – ноль

Установите флажок, чтобы линия регрессии прошла через начало координат.

Выходной диапазон

Введите ссылку на левую верхнюю ячейку выходного диапазона. Отведите, по крайней мере, семь столбцов для итогового диапазона, который будет включать в себя: результаты дисперсионного анализа, коэффициенты регрессии, стандартную погрешность вычисления Y, среднеквадратичные отклонения, число наблюдений, стандартные погрешности для коэффициентов.

Новый лист

Установите переключатель, чтобы открыть новый лист в книге и вставить результаты анализа, начиная с ячейки A1. Если в этом есть необходимость, введите имя нового листа в поле, расположенном напротив соответствующего положения переключателя.

Остатки

Установите флажок, чтобы включить остатки в выходной диапазон.

Стандартизированные остатки

Установите флажок, чтобы включить стандартизированные остатки в выходной диапазон.

График остатков

Установите флажок, чтобы построить диаграмму остатков для каждой независимой переменной.

График подбора

Установите флажок, чтобы построить диаграммы наблюдаемых и предсказанных значений для каждой независимой переменной.

График нормальной вероятности

Установите флажок, чтобы построить диаграмму нормальной вероятности.

Задача. В таблице приведены данные о курсе евро X_1, фондовом индексе X_2 и котировке акций Y за последние 10 дней. Требуется с помощью линейной регрессии спрогнозировать котировку акций, если курс евро составит 32 руб., а значение фондового индекса окажется равным 5.

X_1	24,4	24,8	25,8	26,4	26,5	27	27,5	27,7	27,9	28
X_2	4,2	4,2	4,3	4,4	4,5	4,7	4,8	4,9	5,0	5,1
Y	44,4	44,9	45,1	45,5	47,7	48,4	48,8	49,7	49,9	50

Замечание. Уравнение регрессии имеет вид: $Y = a_0 + a_1X_1 + a_2X_2$.

Найти a_0, a_1, a_2, использовав команду **регрессия** из пакета **анализ данных** электронных таблиц **Excel**. Найти также интервальные оценки параметров a_1, a_2 и показать значимость уравнения регрессии.

Решение:

1. Присваиваем листу имя Анализ данных.

2. В ячейке A1 введём текст «Курс евро». В ячейках A2 : A11 введём данные X_1.

3. В ячейку B1 введём текст «Фондовый индекс». В ячейках B2 : B11 введём данные X_2.

4. В ячейку C1 введём текст «Котировки акций». В ячейках C2 : C11 введём данные Y (рис. 2.3).

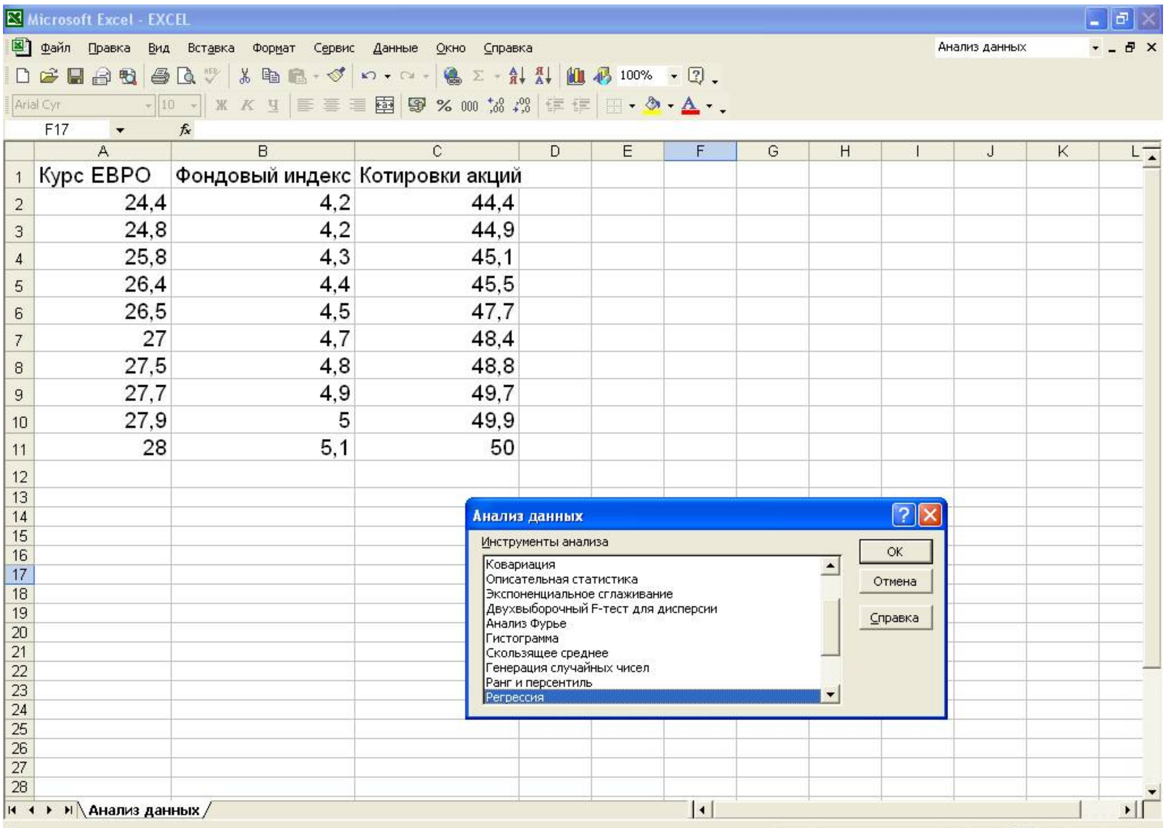


Рисунок 2.3 – Исходные данные

5. Войти в меню – Сервис – Анализ данных – регрессия.

6. В поле Входной интервал Y введём диапазон C2 : C11.

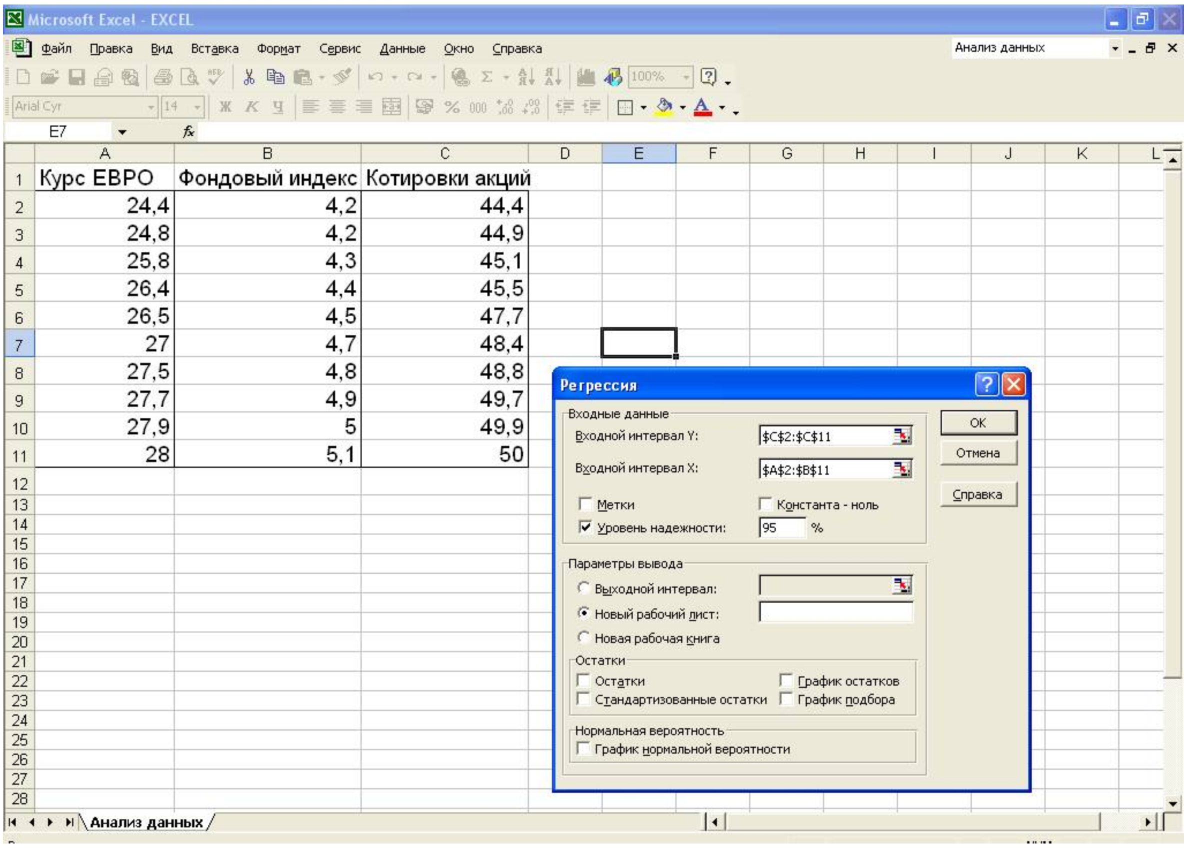
7. В поле Входной интервал X введём диапазон A2 : B11.

8. Поставим флажок на поле Уровень надёжности (95 %) и укажем ячейку, с которой будет начинаться вывод итогов (рис. 2.4).

9. Щёлкнем на кнопке ОК и получим таблицу Вывод итогов (рис. 3.5).

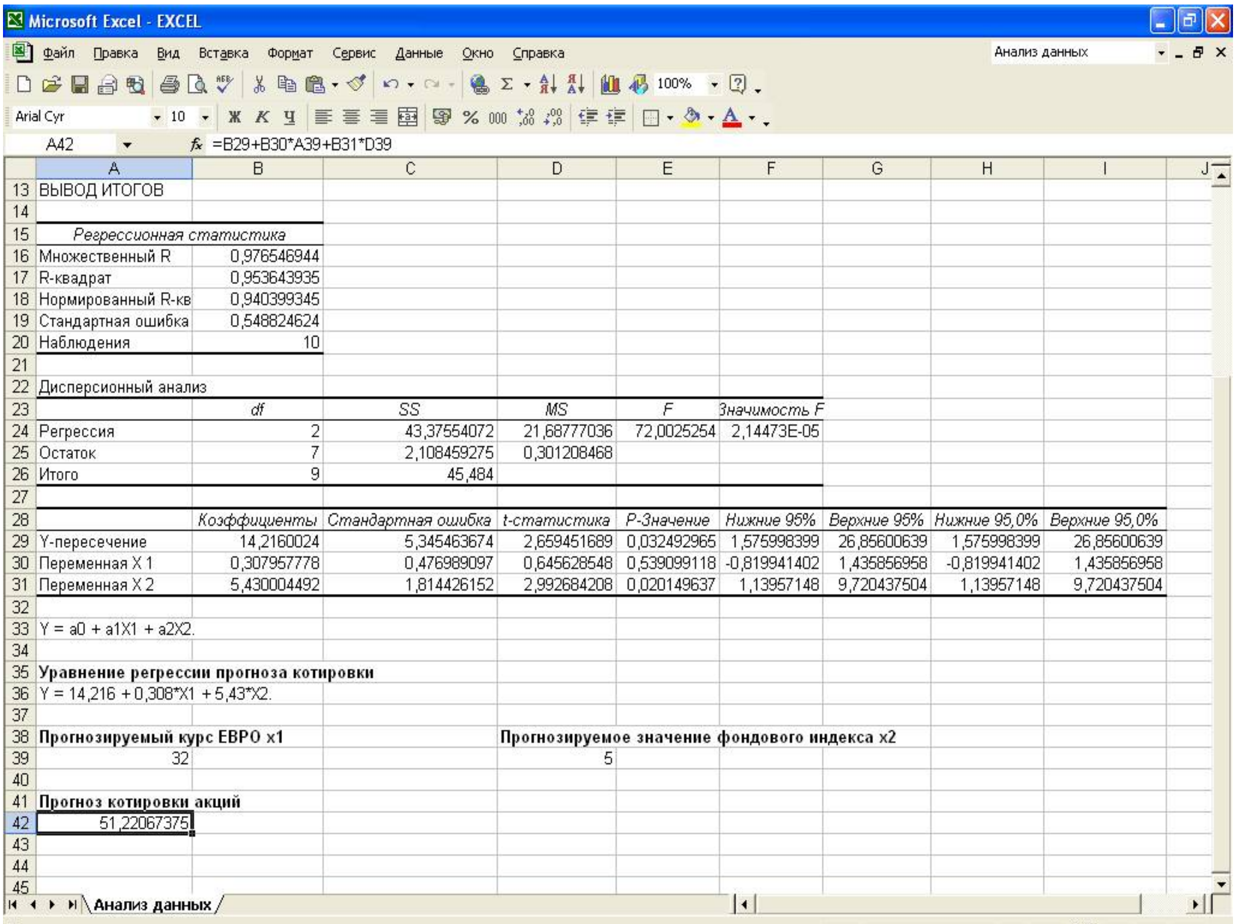
Результаты выведены, начиная с ячейки A13. Поясним смысл регрессионной статистики. Здесь множественный R – это коэффициент корреляции. R = 0,976. Он показывает, что связь между Y и факторами X1, X2 высокая, то есть степень зависимости котировки акций Y от двух факторов курса доллара и фондового индекса достаточно велика. R-квадрат или R^2 – коэффициент детерминации. $R^2 = 0,96$. Это означает, что 96 % изменений котировок акций связано с линейным влиянием курса доллара и фондового индекса. Стандартная ошибка – 0,54. Стандартная ошибка s-ошибка, возникающая при замене фактических наблюдений y_i оценками y_i, рассчитываемых по линейному уравнению регрессии, вычисляется по формуле:

$$s = \sqrt{\frac{\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}{n - m - 1}},$$ здесь n – число наблюдений, m – число факторов.



Курс ЕВРО	Фондовый индекс	Котировки акций
24,4	4,2	44,4
24,8	4,2	44,9
25,8	4,3	45,1
26,4	4,4	45,5
26,5	4,5	47,7
27	4,7	48,4
27,5	4,8	48,8
27,7	4,9	49,7
27,9	5	49,9
28	5,1	50

Рисунок 2.4 – Параметры регрессии



ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,976546944
R-квадрат	0,953643935
Нормированный R-кв	0,940399345
Стандартная ошибка	0,548824624
Наблюдения	10

Дисперсионный анализ

	df	SS	MS	F	Значимость F
Регрессия	2	43,37554072	21,68777036	72,0025254	2,14473E-05
Остаток	7	2,108459275	0,301208468		
Итого	9	45,484			

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	14,2160024	5,345463674	2,659451689	0,032492965	1,575998399	26,85600639	1,575998399	26,85600639
Переменная X 1	0,307957778	0,476989097	0,645628548	0,539099118	-0,819941402	1,435856958	-0,819941402	1,435856958
Переменная X 2	5,430004492	1,814426152	2,992684208	0,020149637	1,13957148	9,720437504	1,13957148	9,720437504

Рисунок 2.5 – Результат

Вторая таблица в выводе итогов – это дисперсионный анализ. В этой таблице рассмотрим F и значимость F. Выдвигается гипотеза H_0, состоящая в том, что уравнение регрессии не значимо. Уровень значимости α – вероятность совершить ошибку при принятии гипотезы. Примем $\alpha = 0{,}05$.

Если значимость F > α, то гипотезу H_0 принимают, и уравнение регрессии будет не значимо. В противном случае H_0 отбрасывается, и уравнение регрессии будет значимо. В примере значимость F = 0,0000214 < 0,05. Поэтому уравнение регрессии будет значимо.

В третьей таблице получим оценки коэффициентов уравнения регрессии. В столбце коэффициентов получим $a_0 = 14{,}216$; $a_1 = 0{,}308$; $a_2 = 5{,}430$. Окончательно уравнение регрессии примет вид: $14{,}216 + 0{,}308 \cdot x1 + 5{,}430 \cdot x2$.

Во второй и в третьей строках этой таблицы приведены 95 %-е интервальные оценки параметров a_1 и a_2: $-0{,}820 < a_1 < 1{,}436$ и $1{,}139 < a_2 < 9{,}720$. В интервале для a1 содержится нуль, поэтому a_1 не значим. В интервале для a_2 нуль не содержится, a2 будет значим. По полученным результатам сделаем прогноз. Продолжим последовательность действий.

10. В ячейку A33 введём текст «Уравнение регрессии:
$Y = a_0 + a_1X_1 + a_2X_2$».

11. В ячейку A35 введём текст «Уравнение регрессии прогноза котировки».

12. В ячейку A36 введём текст «$Y = 14{,}216 + 0{,}308{*}X_1 + 5{,}43{*}X_2$».

13. В ячейку A38 введём текст «Прогнозируемый курс евро X_1».

14. В ячейку A39 введём = 32.

15. В ячейку D38 введём текст «Прогнозируемое значение фондового индекса X_2».

16. В ячейку D39 введём = 5.

17. В ячейку A41 введём текст «Прогноз котировки акций».

18. В ячейке A42 произведём расчёт. Для этого введём в ячейку A42 формулу = B29 + B30 · A39 + B31 · D39 и нажмём клавишу <Enter>.

Таблица 2.1 – Варианты заданий

Варианты	Значения										
1	X_1	24,3	24,7	24,9	25,9	26,3	26,9	27,3	27,9	28,5	28,9
	X_2	3,9	4,1	4,2	4,9	5,1	5,1	5,0	4,8	4.9	4,5
	Y	45,9	47,3	48,2	45,2	49,7	50,2	49,1	47,2	40,3	41,5
2	X_1	23,1	24,1	28,4	28,9	29,5	30,1	30,8	31,2	31,8	29
	X_2	3,7	4,3	4,8	4,9	5,0	5,2	5,1	4,9	4,5	4,3
	Y	42,3	43,5	44,1	45,2	49,2	48,2	50,2	48,4	40,3	48,5

Вари-анты	Значения										
3	X_1	21,5	22,8	23,5	24,1	25,9	22,4	24,1	23,2	22,3	23,8
	X_2	3,9	4,1	4,3	4,5	4,1	4,8	4,2	4,7	4,2	4,5
	Y	43,4	44,8	47,2	45,3	49,2	42,3	43,1	44,5	42,3	45,1
4	X_1	29,5	30,2	31,4	30,8	31,8	31,9	32,3	31,1	30,1	29,2
	X_2	4,2	4,5	5,1	5,0	5,2	5,1	5,2	5,0	5,2	4,8
	Y	44,1	45,8	48,4	49,2	49,8	50,1	50,2	48,1	48,7	46,2
5	X_1	22,1	23,2	24,2	23,8	24,1	24,2	25,1	25,3	25,7	26,2
	X_2	4,1	4,5	4,7	4,6	4,7	4,8	4,9	4,9	4,8	4,9
	Y	45,2	46,1	46,8	47,4	48,1	48,7	48,9	49,1	49,7	50,2
6	X_1	27,2	28,1	28,7	28,9	29,1	29,3	28,2	29,4	29,5	29,7
	X_2	4,3	4,4	4,6	4,8	4,8	4,9	4,7	4,9	5,0	5,1
	Y	45,2	46,1	47,2	47,4	48,1	49,7	50,1	50,3	50,5	50,9
7	X_1	24,3	25,1	25,8	23,9	25,8	26,2	27,4	28,3	29,5	29,5
	X_2	4,3	4,6	4,7	4,2	4,8	4,8	4,9	4,9	5,1	5,2
	Y	45,8	46,7	47,2	48,4	48,2	47,1	48,4	48,9	49	50,2
8	X_1	23,1	24,2	25,1	25,7	26,2	27,1	27,8	29,1	27,3	29,8
	X_2	4,7	4,8	4,8	4,9	4,8	4,9	5,0	5,1	4,7	5,2
	Y	45,8	46,7	47,2	48,4	48,2	47,1	48,4	48,9	49	50,2
9	X_1	22,3	24,1	27,8	25,3	26,8	27,9	29,4	30,1	31,0	31,5
	X_2	4,3	4,7	4,7	4,8	4,9	5,0	5,1	5,1	5,2	5,2
	Y	45,4	45,6	45,7	45,8	45,9	46,2	46,8	46,9	46,9	47,2
10	X_1	23,2	25,4	26,2	28,1	29,2	29,8	30,1	31,2	31,4	32,1
	X_2	4,5	4,8	4,8	4,9	4,9	5,0	5,1	5,2	5,2	5,1
	Y	45,4	45,8	45,9	46,1	46,8	46,9	47,2	47,3	48,1	48,5
11	X_1	24,1	25,3	26,1	26,8	27,3	28,1	28,4	29,1	29,3	29,6
	X_2	4,3	4,8	4,9	4,9	5,0	5,1	5,0	5,1	5,2	5,2
	Y	44,1	45,8	45,9	45,9	46,2	46,7	46,8	47,1	47,8	47,9
12	X_1	25,2	26,1	26,8	27,2	27,7	27,3	28,4	28,3	28,5	28,9
	X_2	4,3	4,2	4,7	4,8	4,9	5,0	5,1	5,2	5,2	5,3
	Y	45,1	45,2	46,2	46,8	46,9	46,7	46,9	47,2	47,8	47,9
13	X_1	24,1	24,7	24,9	25,2	25,8	26,2	27,1	27,8	27,3	28,3
	X_2	4,1	4,2	4,4	4,7	4,6	4,8	4,8	4,9	5,0	5,1
	Y	44,1	44,2	44,9	45,2	45,8	45,9	46,2	46,8	46,9	47,1
14	X_1	24,2	24,5	24,9	25,2	25,7	26,1	26,7	27,8	27,4	28,9
	X_2	4,2	4,3	4,1	4,6	4,7	4,8	4,9	4,8	5,1	5,0
	Y	44,1	44,6	44,9	45,1	45,7	45,9	46,1	46,7	46,8	46,9
15	X_1	25,1	25,4	25,8	25,9	25,9	26,2	26,8	26,9	27	27,3
	X_2	4,2	4,3	4,5	4,7	4,8	4,9	4,9	5,1	5,2	5,1
	Y	44,2	44,8	45,0	45,6	45,9	46,2	46,7	46,8	48	48,4
16	X_1	24,1	25,9	26,1	26,2	27,3	27,5	27,8	28,2	29,2	29,3
	X_2	4,2	4,3	4,5	4,6	4,6	4,7	4,9	5,0	5,1	5,2
	Y	44,3	45,1	46,8	47,2	48,3	49,4	49,6	49,8	50,1	50,3
17	X_1	24,4	24,8	25,8	26,4	26,5	27	27,5	27,7	27,9	28
	X_2	4,2	4,2	4,3	4,4	4,5	4,7	4,8	4,9	5,0	5,1
	Y	44,4	44,9	45,1	45,5	47,7	48,4	48,8	49,7	49,9	50

Варианты	Значения										
18	X_1	24,1	25,3	26,1	26,8	27,3	28,1	28,4	29,1	29,3	29,6
	X_2	4,3	4,4	4,7	4,8	4,9	5,0	5,1	5,1	5,2	5,2
	Y	42,3	43,5	44,1	45,2	47,6	49,2	49,5	49,7	50,3	50,9
19	X_1	21,1	21,4	21,5	21,7	24,4	26,3	27,8	28,4	28,8	29,1
	X_2	4,1	4,3	4,5	4,6	4,6	4,8	4,9	5,0	5,1	5,1
	Y	42,1	42,5	43,1	43,6	45,7	49,2	49,7	50,0	50,1	50,1
20	X_1	27,2	28,1	28,7	28,8	28,9	29,3	29,7	29,8	30,1	30,5
	X_2	4,1	4,2	4,4	4,7	4,6	4,8	4,8	4,9	5,0	5,1
	Y	44,2	44,6	45,9	46,1	46,9	47,2	48,5	49,6	49,9	49,9
21	X_1	24,7	25,1	26,4	26,7	27,5	27,9	28,4	29,5	29,7	30,1
	X_2	4,1	4,1	4,2	4,3	4,3	4,7	4,8	4,8	5,0	5,0
	Y	44,4	45,1	45,3	46,1	46,4	46,8	49,4	49,6	50,0	50,1
22	X_1	23,1	23,5	23,6	23,8	24,1	26,4	28,4	29,5	29,7	29,8
	X_2	4,2	4,5	4,5	4,7	4,8	4,8	4,9	4,9	5,0	5,0
	Y	43,9	44,1	44,5	44,8	45,2	45,8	46,3	46,5	47,4	48,4
23	X_1	23,2	23,4	23,6	23,8	24,1	24,6	25,7	26,8	27,9	28,6
	X_2	3,9	4,1	4,2	4,5	5,2	4,9	4,9	5,1	5,0	5,1
	Y	44,1	44,5	44,8	44,7	47,4	48,7	49,5	49,6	49,9	50,1
24	X_1	22,3	24,5	25,6	26,7	28,7	28,8	29,1	29,4	29,6	29,9
	X_2	4,3	4,5	4,8	5,0	5,1	5,0	5,1	5,3	5,3	5,1
	Y	43,1	43,2	43,6	43,8	44,1	44,4	44,8	44,7	45,2	47,2
25	X_1	21,7	21,9	22,7	22,9	23,4	24,4	25,1	27,4	28,9	30,1
	X_2	4,1	4,1	4,2	4,3	4,8	4,9	5,0	5,1	5,1	5,3
	Y	43,8	44,2	44,5	44,6	44,9	45,3	45,8	46,7	47,8	49,1
26	X_1	23,7	25,1	26,2	27,1	26,8	26,9	27,1	28,1	28,3	28,5
	X_2	4,2	4,4	4,7	4,8	4,5	4,7	4,8	5,1	5,0	5,3
	Y	43,5	44,1	44,7	44,9	44,7	44,9	45	45,3	46,1	47,1
27	X_1	21,3	21,9	21,7	22,1	23,4	23,8	23,9	24,1	27,2	27,4
	X_2	4,1	4,3	4,6	4,5	4,9	4,9	5,1	5,2	5,2	5,4
	Y	44,1	44,5	44,7	44,8	45,0	45,1	45,3	46,2	46,4	47,0
28	X_1	24,7	24,9	25,1	25,8	26,1	26,9	27,2	28,1	28,3	29,1
	X_2	4,1	4,3	4,5	4,9	4,9	5,1	5,3	5,8	5,6	5,8
	Y	44,5	45,1	45,8	45,9	46,1	47,2	47,4	47,5	47,8	48
29	X_1	21,7	22,8	23,1	24,5	24,9	24,8	25,3	26,1	27,3	28,7
	X_2	4,5	4,7	4,9	5,1	5,3	5,3	5,8	5,9	5,9	6,0
	Y	42,1	42,5	42,7	43,1	44,1	44,5	45,1	45,8	45,9	46
30	X_1	21,9	22,7	23,5	24,1	25,1	26,5	26,9	27,1	27,8	27,9
	X_2	4,3	4,5	4,9	5,1	5,3	5,4	5,5	5,9	6,1	6,3
	Y	45,1	45,8	45,9	46,0	47,2	48,1	48,3	48,9	49,0	50,0

2.2.Использование Анализа данных инструмента Регрессия для решения задач прогнозирования в нефтегазовом деле

2.2.1 Парная регрессия и корреляция

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

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

Нелинейные регрессии делятся на два класса.

Регрессии, нелинейные по независимым переменным:

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

Регрессии, нелинейные по оцениваемым параметрам:

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

Вычисление коэффициентов линейной регрессии.

Чтобы составить уравнение линейной регрессии, надо иметь таблицу наблюдений.

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

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

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

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

Используя диаграмму точечного графика, построим диаграмму рассеяния (поле корреляции) (рис. 2.2.1).

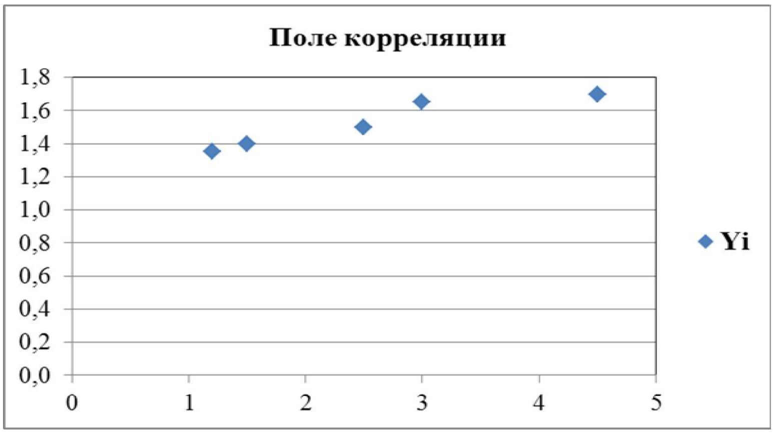


Рисунок 2.2.1 – Поле корреляции

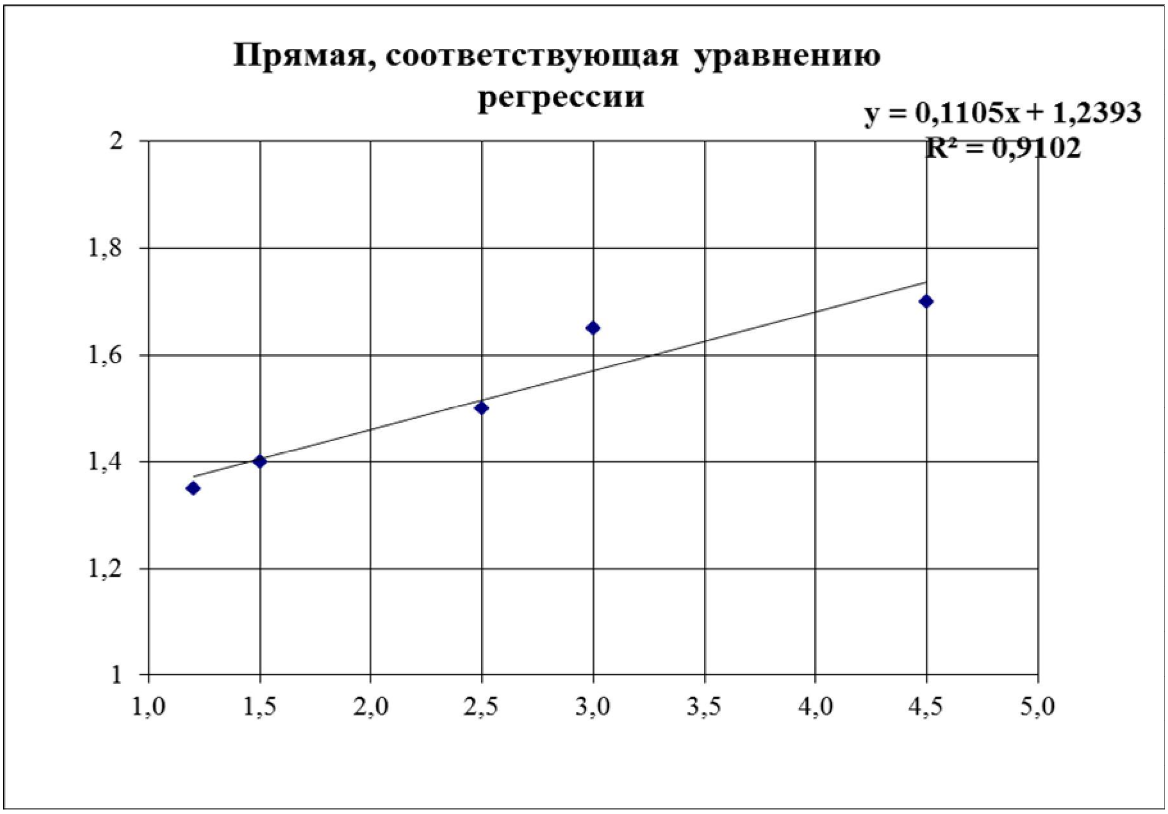


Рисунок 2.2.2 – Линия регрессии

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

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

Из этого условия находятся a_0 и a_1.

Решается система линейных уравнений:

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

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

$$\Delta = \begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix} \qquad \Delta_0 = \begin{vmatrix} \sum \text{y} & \sum x \\ \sum yx & \sum x^2 \end{vmatrix} \qquad \Delta_1 = \begin{vmatrix} n & \sum y \\ \sum x & \sum yx \end{vmatrix}$$

$$a_0 = \frac{\Delta_0}{\Delta} \qquad a_1 = \frac{\Delta_1}{\Delta}$$

Второй способ. По формулам.

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

$$\overline{x} = \frac{\sum x_i}{n}; \quad \overline{y} = \frac{\sum y_i}{n}; \quad \overline{xy} = \frac{\sum y_i x_i}{n};$$

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

Далее вычисляем средние квадратические отклонения по x и по y

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

Коэффициент корреляции.

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

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

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

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

Если $r_{xy} < 0$, то связь обратная.

Если $r_{xy} > 0$, то связь прямая.

Коэффициент детерминации:

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

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

Оценка значимости коэффициентов регрессии и корреляции.

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

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

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

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

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

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

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

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

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

Оценка качества построенного уравнения регрессии с помощью F-критерия.

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

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

Сравним его с критическим значением.

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

По таблице Фишера определяем $F_{крит}$.

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

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

Средняя ошибка аппроксимации

Рассчитаем среднюю ошибку аппроксимации:

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

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

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

Точечный и интервальный прогнозы

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

$Y_{прогн} = a_0 + a_1 X_{прогн}$, это точечный прогноз.

Дисперсия ошибки:

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

Выполним интервальный прогноз.

Найдём среднюю стандартную ошибку прогноза

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

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

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

2.2.2 Множественная регрессия

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

$Y = f(B, X) + e$, где $X = (x_1, x_2, \ldots, x_m)$ – вектор независимых переменных; B – вектор параметров, подлежащих определению; e – случайная ошибка (отклонение); Y – зависимая переменная.

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

$Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_mX_m + e$ или для индивидуальных наблюдений $i = 1, 2, \ldots, m$. $y_i = b_0 + b_1x_{i1} + b_2x_{i2} + \ldots + b_mx_{im} + e_i$.

Здесь $B = (b_0, b_1, \ldots, b_m)$ – вектор неизвестных параметров. b_j, $j = 1, 2, \ldots, m$ называется j-ым теоретическим коэффициентом регрессии. Он характеризует чувствительность величины y к изменению X_j. b_0 – свободный член, определяющий значение y в случае, когда все независимые переменные равны 0.

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

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

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

Вычислим частные производные.

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

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

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

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

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

Такая система называется системой линейных нормальных уравнений.

Обычно она имеет единственное решение. Здесь y – результирующая переменная, x_1, x_2, … x_m – независимые переменные (факторы), n – число наблюдений, m – число факторов. Для решения системы может быть применён метод Крамера (метод определителей).

$$b_0 = \frac{\Delta b_0}{\Delta}, b_1 = \frac{\Delta b_1}{\Delta}, ..., \frac{\Delta b_m}{\Delta}, \text{ где } \Delta - \text{определитель системы,}$$

$$\Delta b_0, \Delta b_1, ..., \Delta b_m - \text{частные определители.}$$

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

Коэффициенты уравнения регрессии также могут быть найдены по формуле:

$$B = (X^T \cdot X)^{-1} \cdot (X^T Y), \quad \text{где} \quad B = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{m1} \\ 1 & x_{12} & \cdots & x_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & \cdots & x_{mn} \end{bmatrix} \; - \; \text{матрица,}$$

первый столбец которой состоит из единиц, остальные столбцы – это наблюдаемые значения факторов x_1, x_2, … x_m.

Матрица X^T – транспонированная матрица X.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \text{наблюдаемые значения переменной Y.}$$

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

Для m = 1 система имеет вид:

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

Для m = 2 система имеет вид:

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

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

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

Таблица 2.2.1 – Расчётная таблица

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

Продолжение табл. 2.2.1

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

Из неё составляется система линейных уравнений:

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

Вычисление коэффициентов уравнения

Система линейных уравнений может быть решена различными способами.

Методом Крамера по формулам:

$$b_0 = \frac{\Delta_{b0}}{\Delta} \quad b_1 = \frac{\Delta_{b1}}{\Delta} \quad b_2 = \frac{\Delta_{b2}}{\Delta}, \text{ где } \Delta, \Delta_0, \Delta_1, \Delta_2 - \text{определители}.$$

Вычисление коэффициентов парной линейной корреляции по формулам:

$$r_{yx_1} = \frac{\overline{x_1 y} - \overline{x}_1 \cdot \overline{y}}{\sigma_{x_1} \cdot \sigma_y}; \quad r_{yx_2} = \frac{\overline{x_2 y} - \overline{x}_2 \cdot \overline{y}}{\sigma_{x_2} \cdot \sigma_y}; \quad r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \overline{x}_1 \cdot \overline{x}_2}{\sigma_{x_1} \cdot \sigma_{x_2}}.$$

Вычисление средних квадратических отклонений:

$$\sigma_{x_1} = \sqrt{\overline{x_1^2} - (\overline{x_1})^2}; \quad \sigma_{x_2} = \sqrt{\overline{x_2^2} - (\overline{x_2})^2}; \quad \sigma_y = \sqrt{\overline{y}^2 - (\overline{y})^2}.$$

Вычисление коэффициента множественной корреляции и коэффициента детерминации выполняется по формулам:

$$r_{yx} = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2}};$$

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

Или по формулам $r_{x_1 x_2}(y) = \dfrac{r_{x_1 x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{1 - r^2_{yx_1}} \cdot \sqrt{1 - r^2_{yx_2}}}$ $\quad R^2 = (r_{x_1 x_2}(y))^2$.

Определение значимости коэффициентов регрессии с помощью t-критерия Стьюдента.

Выдвинем гипотезу H_0: полученные коэффициенты уравнения регрессии незначимы.

Для каждого коэффициента определим расчётные значения t-статистик Стьюдента.

$$t_{b_i} = \frac{|b_i|}{\sigma_{b_i}}, i = 1,2, \text{ где } \sigma_{bi} - \text{стандартные ошибки уравнения регрессии.}$$

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

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

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

Критическое значение критерия Стьюдента может быть вычислено с помощью статистической функции =СТЬЮДРАСПОБР (α, df) (рис. 2.2.3).

Вычисляем расчётные значения t-статистик Стьюдента:

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

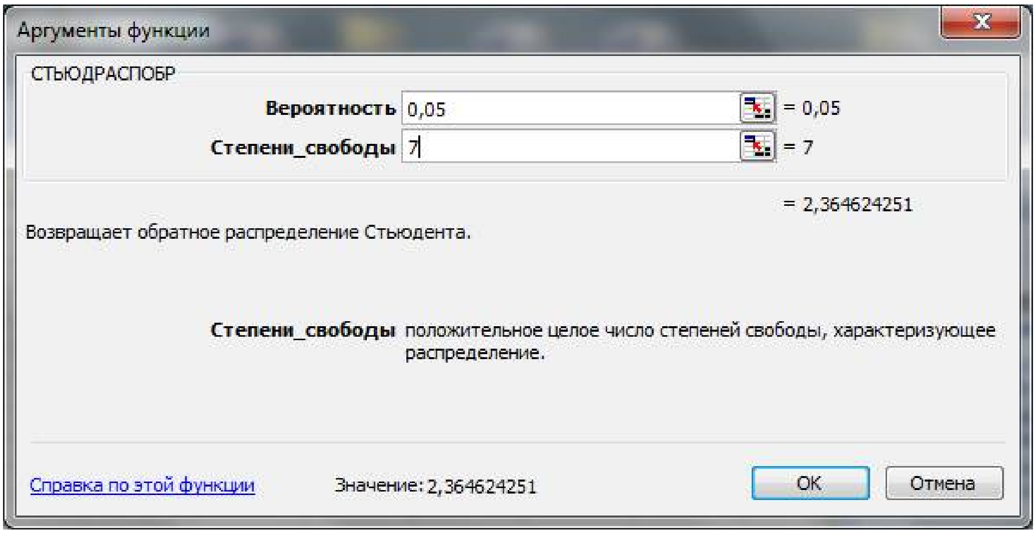


Рисунок 2.2.3 – Функция СТЬЮДРАСПОБР

Если:

- ✓ $t_{в1} > t_{крит}$, коэффициент b_1 статистически значим, гипотеза H_0 отвергается.
- ✓ $t_{в1} < t_{крит}$, коэффициент b_1 статистически незначим, гипотеза H_0 принимается.
- ✓ $t_{в2} > t_{крит}$, коэффициент b_2 статистически значим, гипотеза H_0 отвергается.
- ✓ $t_{в2} < t_{крит}$, коэффициент b_2 статистически незначим, гипотеза H_0 принимается.

С помощью критерия Фишера оценить статистическую надёжность построенного уравнения (при уровне значимости $\alpha = 0{,}05$).

Выдвинем гипотезу H_0: полученное уравнение регрессии незначимо.

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

$$F_{факт} = \frac{R^2}{1-R^2} \cdot \frac{N-m-1}{m},$$ где n – число наблюдений; m – число факторов.

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

$F_{крит}$ можно определить с использованием статистической функции =FРАСПОБР(α; $к_1$; $к_2$), где $к_1 = m$ (число факторов); $к_2 = n - m - 1$ (n – число наблюдений).

Например:

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

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

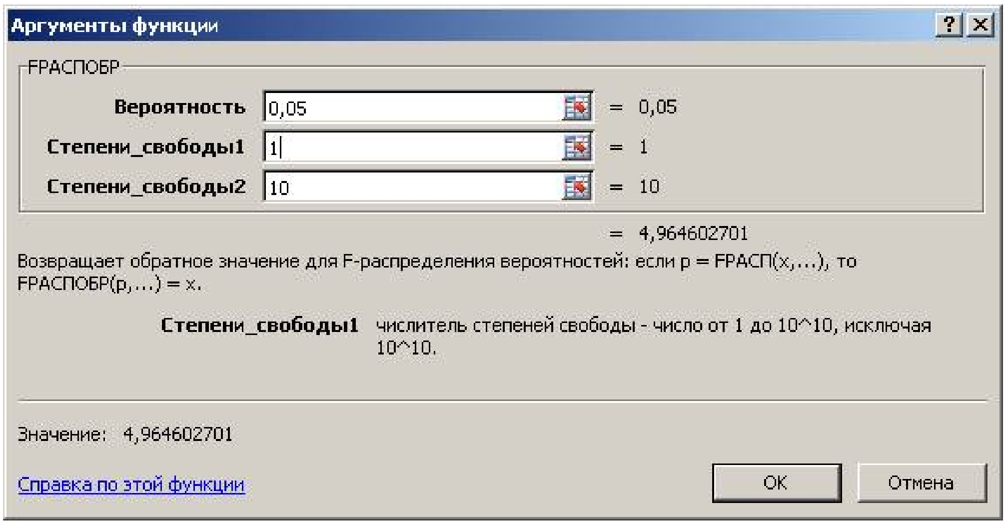


Рисунок 2.2.4 – Функция FРАСПОБР

2.2.3 Задания для выполнения

Задание 1. Линейная регрессия

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

Для этого выполнить следующие пункты.

- Построить поле корреляции.
- Выбрать и обосновать вид уравнения регрессии.
- Рассчитать коэффициенты линейной парной регрессии.
- Проанализировать коэффициенты a_0 и a_1.
- Рассчитать коэффициент парной линейной корреляции и сделать выводы о тесноте связи между переменными построенного уравнения.
- Оценить качество построенного уравнения регрессии. Проверку значимости коэффициентов уравнения и линейного коэффициента корреляции выполнить с помощью t-критерия Стьюдента. При уровне значимости $\alpha = 0{,}05$.
- Рассчитать коэффициент детерминации.
- Оценить качество построенного уравнения регрессии с помощью F-критерия Фишера.
- Рассчитать среднюю ошибку аппроксимации.
- Выполнить точечный и интервальный прогнозы для Y при

$$X_{прогнозном} = X_{среднее} \cdot 1{,}1.$$

- Расчёты выполнить в Excel с использованием инструмента Регрессия из пакета Анализ данных, а также с помощью функции ЛИНЕЙН.

Вариант 1

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	97	161
2	73	131
3	79	135
4	99	147
5	86	139
6	91	151
7	85	135
8	77	132
9	89	161
10	95	159
11	72	120
12	115	160

Вариант 2

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	69	124
2	83	133
3	92	146
4	97	153
5	88	138
6	93	159
7	74	145
8	79	152
9	105	168
10	99	154
11	85	127
12	94	155

Вариант 3

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	77	123
2	85	152
3	79	140
4	93	142
5	89	157
6	81	181
7	79	133
8	97	163
9	73	134
10	95	155
11	84	132
12	108	165

Вариант 4

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	83	137
2	88	142
3	75	128
4	89	140
5	85	133
6	79	153
7	81	142
8	97	154
9	79	132
10	90	150
11	84	132
12	112	166

Вариант 5

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	79	134
2	91	154
3	77	128
4	87	138
5	84	133
6	76	144
7	84	160
8	94	149
9	79	125
10	98	163
11	81	120
12	115	162

Вариант 6

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	92	147
2	78	133
3	79	128
4	88	152
5	87	138
6	75	122
7	81	145
8	96	141
9	80	127
10	102	151
11	83	129
12	94	147

Вариант 7

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	75	133
2	78	125
3	81	129
4	93	153
5	86	140
6	77	135
7	83	141
8	94	152
9	88	133
10	99	156
11	80	124
12	112	156

Вариант 8

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	74	122
2	81	134
3	90	136
4	79	125
5	89	120
6	87	127
7	77	125
8	93	148
9	70	122
10	93	157
11	87	144
12	121	165

Вариант 9

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	78	133
2	94	139
3	85	141
4	73	127
5	91	154
6	88	142
7	73	122
8	82	135
9	99	142
10	113	168
11	69	124
12	83	130

Вариант 10

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	81	124
2	77	131
3	85	146
4	79	139
5	93	143
6	100	159
7	72	135
8	90	152
9	71	127
10	89	154
11	82	127
12	111	162

Вариант 11

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	87	162
2	63	133
3	69	135
4	89	147
5	76	139
6	81	151
7	75	135
8	67	132
9	79	161
10	85	159
11	62	123
12	95	162

Вариант 12

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	79	126
2	93	135
3	102	146
4	107	153
5	98	138
6	103	159
7	84	145
8	89	152
9	115	168
10	109	154
11	95	127
12	104	155

Вариант 13

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	67	125
2	75	157
3	89	141
4	93	146
5	89	159
6	81	181
7	79	133
8	98	163
9	73	134
10	95	155
11	84	138
12	109	167

Вариант 14

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	85	139
2	88	144
3	77	129
4	89	146
5	87	133
6	79	153
7	83	142
8	97	154
9	79	135
10	95	150
11	84	135
12	112	168

Вариант 15

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	78	132
2	90	154
3	78	128
4	87	138
5	86	135
6	76	144
7	87	164
8	94	149
9	79	127
10	97	163
11	86	122
12	116	164

Вариант 16

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	91	145
2	76	133
3	79	129
4	86	152
5	87	137
6	75	122
7	82	146
8	96	141
9	83	127
10	101	151
11	83	129
12	96	146

Вариант 17

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	76	135
2	79	125
3	83	129
4	92	155
5	87	140
6	78	1347
7	83	141
8	94	153
9	87	133
10	99	157
11	85	124
12	114	158

Вариант 18

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	73	125
2	81	134
3	90	138
4	78	125
5	89	120
6	86	128
7	77	125
8	94	148
9	70	123
10	95	156
11	87	147
12	120	166

Вариант 19

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	75	134
2	94	138
3	84	141
4	73	126
5	91	154
6	86	143
7	74	122
8	82	137
9	98	142
10	112	167
11	68	125
12	85	132

Вариант 20

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	83	126
2	75	132
3	85	147
4	78	138
5	93	145
6	102	159
7	73	135
8	93	153
9	75	127
10	87	152
11	83	126
12	114	165

Вариант 21

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	95	163
2	74	131
3	77	137
4	98	147
5	86	138
6	93	151
7	85	136
8	75	132
9	89	165
10	99	159
11	73	122
12	98	164

Вариант 22

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	67	126
2	85	135
3	93	146
4	98	155
5	88	138
6	94	159
7	75	147
8	79	153
9	104	168
10	99	155
11	86	127
12	95	156

Вариант 23

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	78	124
2	86	151
3	79	140
4	95	142
5	89	157
6	81	181
7	79	135
8	98	163
9	73	136
10	96	155
11	84	133
12	107	164

Вариант 24

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	85	138
2	88	142
3	75	128
4	89	142
5	87	135
6	79	155
7	83	142
8	97	154
9	79	132
10	91	155
11	84	133
12	114	167

Вариант 25

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	75	135
2	92	155
3	77	128
4	88	139
5	84	133
6	76	146
7	85	160
8	94	149
9	79	126
10	97	163
11	81	123
12	114	165

Вариант 26

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	95	145
2	79	132
3	81	128
4	88	152
5	87	138
6	75	122
7	81	145
8	97	142
9	81	127
10	102	152
11	83	129
12	96	148

Вариант 27

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	74	135
2	78	125
3	81	129
4	95	152
5	85	143
6	77	135
7	84	142
8	95	152
9	88	135
10	99	156
11	82	128
12	114	159

Вариант 28

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	73	124
2	80	136
3	93	138
4	78	124
5	88	120
6	87	127
7	77	124
8	93	149
9	74	121
10	95	158
11	87	143
12	123	165

Вариант 29

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	76	134
2	95	139
3	85	142
4	73	126
5	92	156
6	88	142
7	74	123
8	82	135
9	98	145
10	112	168
11	68	122
12	82	131

Вариант 30

Номер региона	Средняя добыча нефти по скважине, т, x	Среднедневная заработная плата, руб., y
1	83	126
2	78	133
3	86	146
4	79	139
5	95	143
6	100	158
7	73	135
8	90	153
9	71	128
10	88	154
11	82	126
12	114	165

Пример выполнения задания.

Имеются данные о зависимости материалоёмкости продукции (у, кг) от размеров предприятия (выпуск продукции х, у. е.). Выполнить пункты задания, указанные выше, при этом вычислить точечный и интервальный прогнозы для Y при $X_{прогнозном}$ = 50 у. е.

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

1. Построим поле корреляции. При построении использовать диаграмму «точечный график».

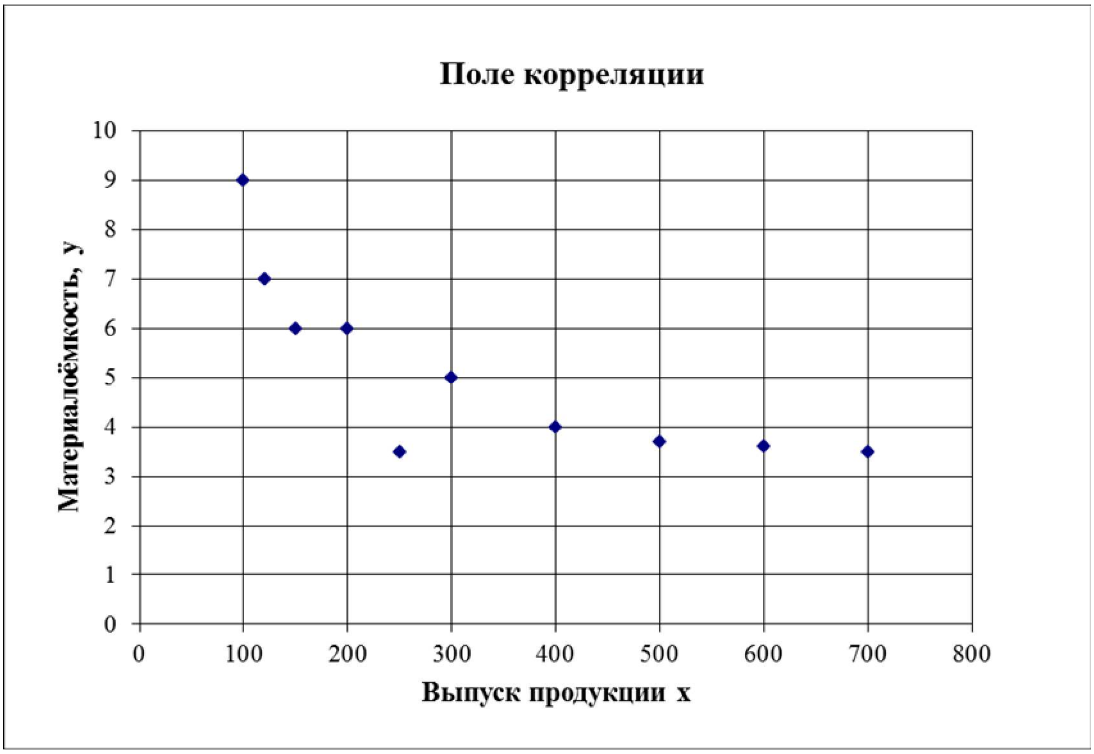


Рисунок 2.2.5 – Поле корреляции

2. Выбрать и обосновать спецификацию уравнения регрессии.

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

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

Рассчитать коэффициенты a_0 и a_1 выбранного уравнения. Для этого составить систему нормальных уравнений и решить её методом определителей. Расчёты проведём в электронных таблицах Excel.

Таблица 2.2.2 – Расчётная таблица

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

Продолжение табл. 2.2.2

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

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

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

Найдём коэффициенты методом определителей.

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

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

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

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

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

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

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

Для расчётов используем формулы:

Среднее значение x

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

Среднее значение y

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

Среднее значение произведений xy

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

Среднее значение квадратов x

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

Среднее значение квадратов y

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

Среднее квадратическое отклонение по x

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

Среднее квадратическое отклонение по y

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

Коэффициент корреляции.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$t_{крит}\ (0{,}05;\ 8) = 2{,}3.$$

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

Для оценки качества построенного уравнения рассчитывается коэффициент детерминации.

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

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

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

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

Сравним его с критическим значением.

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

$$F_{крит}(0,05;\ 1:\ 8) = 5,32.$$

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

Рассчитаем среднюю ошибку аппроксимации:

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

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

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

Выполним точечный и интервальный прогнозы для материалоёмкости продукции при х = 50 у. е.

Выполним точечный прогноз

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

Найдём дисперсию ошибки:

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

Выполним интервальный прогноз.

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

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

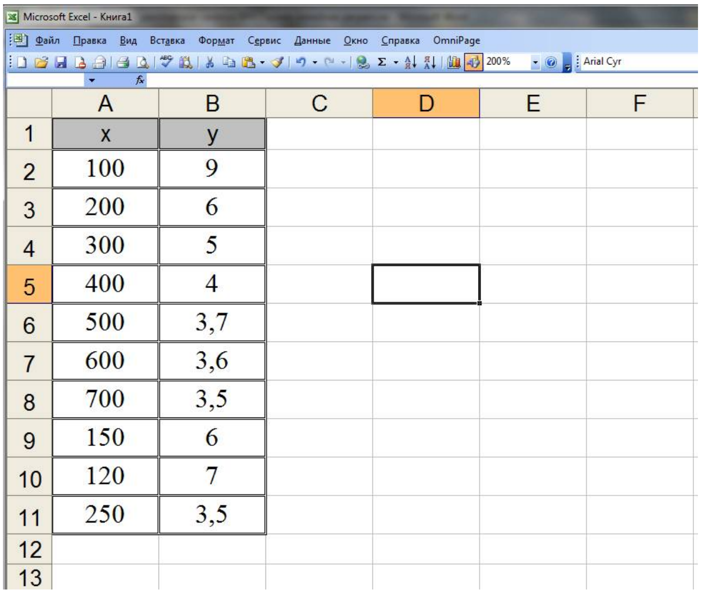
Следовательно, при выпуске продукции, равном 50 у.е. материалоёмкость продукции будет находиться в пределах:

$$3,9799 \le у \le 10,1946.$$

3. Использование инструмента Регрессия из пакета «Анализ данных» электронных таблиц Excel

Рассмотрим последовательность действий для использования инструмента Регрессия.

1. Введём исходные данные на лист Excel (рис. 3.1).



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

Рисунок 3.1 – Исходные данные

2. Войти в меню Данные (версия Excel-2007 или 2010), в версии 2003 в меню Сервис. Выбрать Анализ данных. Откроется окно программы Анализ данных с названиями инструментов (рис. 3.2).

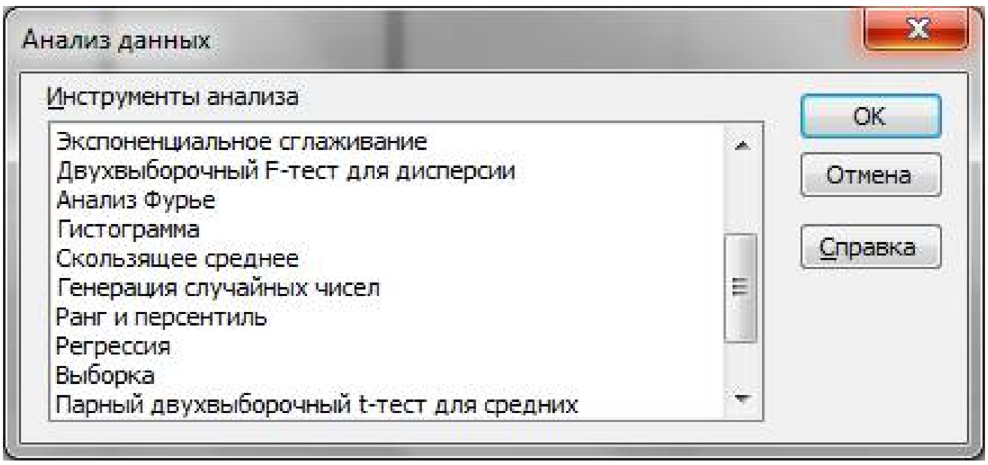


Рисунок 3.2 – Анализ данных

3. Выбрать в этом окне пункт Регрессия.
4. Откроется окно инструмента Регрессия (рис. 3.3).

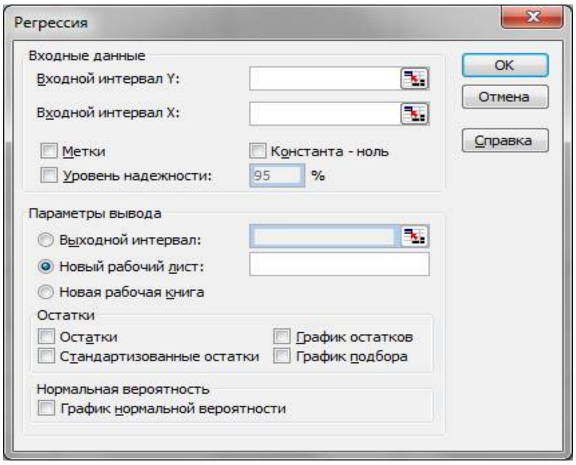


Рисунок 3.3 – Регрессия

5. Ввести исходные данные в интервал Y блок ячеек B1 : B11 и интервал X блок ячеек A1 : A11, в переключатель метки поставить флажок, в уровень надёжности поставить флажок (уровень надёжности будет 95 %), в выходной интервал поставить точку и ввести номер ячейки, начиная с которой будут выведены результаты на том же листе (рис. 3.4.).

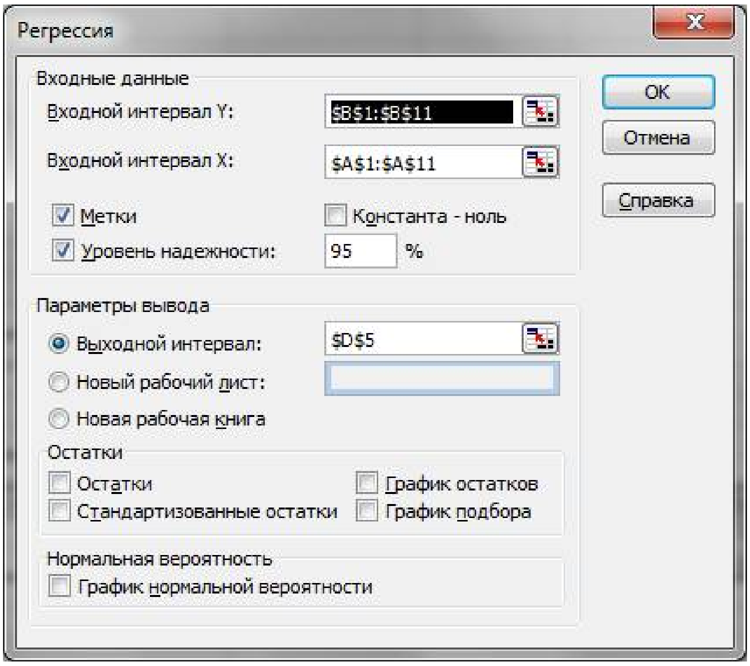


Рисунок 3.4 – Регрессия

6. После щелчка по ОК появятся результаты.

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

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

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

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

Таблица 3.1 – Результаты

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

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

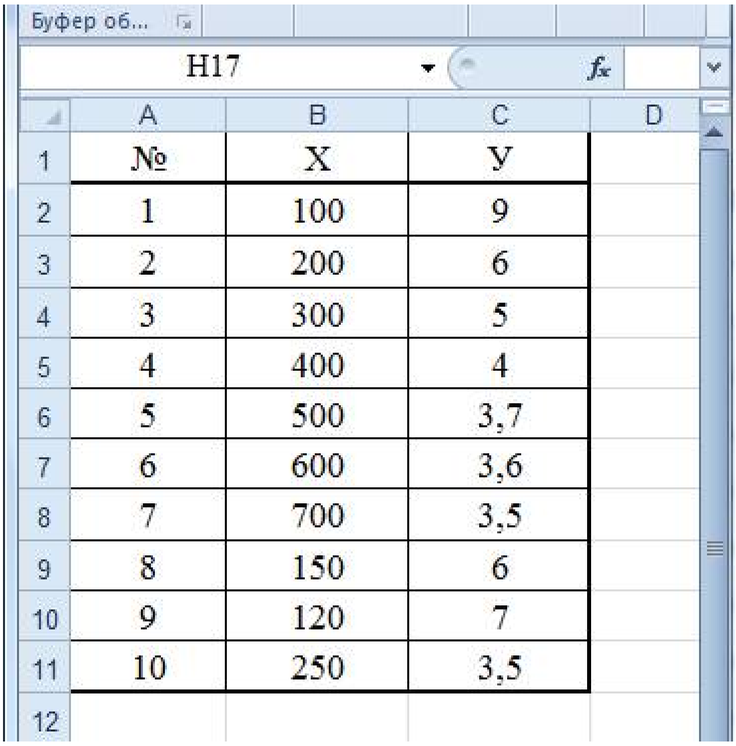
Результаты совпадают с результатами, вычисленными по формулам.

3.1 Применение статистической функции ЛИНЕЙН для составления и анализа уравнения парной линейной регрессии

Последовательность действий

1. Введите исходные данные на лист Excel (рис. 3.5).

2. Выделите область пустых ячеек 5x2 (5 строк и 2 столбца) для вывода результатов регрессионной статистики.



№	X	У
1	100	9
2	200	6
3	300	5
4	400	4
5	500	3,7
6	600	3,6
7	700	3,5
8	150	6
9	120	7
10	250	3,5

Рисунок 3.5 – Исходные данные

3. В главном меню выберите Вставка/Функция.

4. В окне Категория выберите Статистические, в окне Выберите функцию – **ЛИНЕЙН**. Щёлкните по кнопке ОК (рис. 3.6). Откроется окно функции **ЛИНЕЙН** (рис. 3.7).

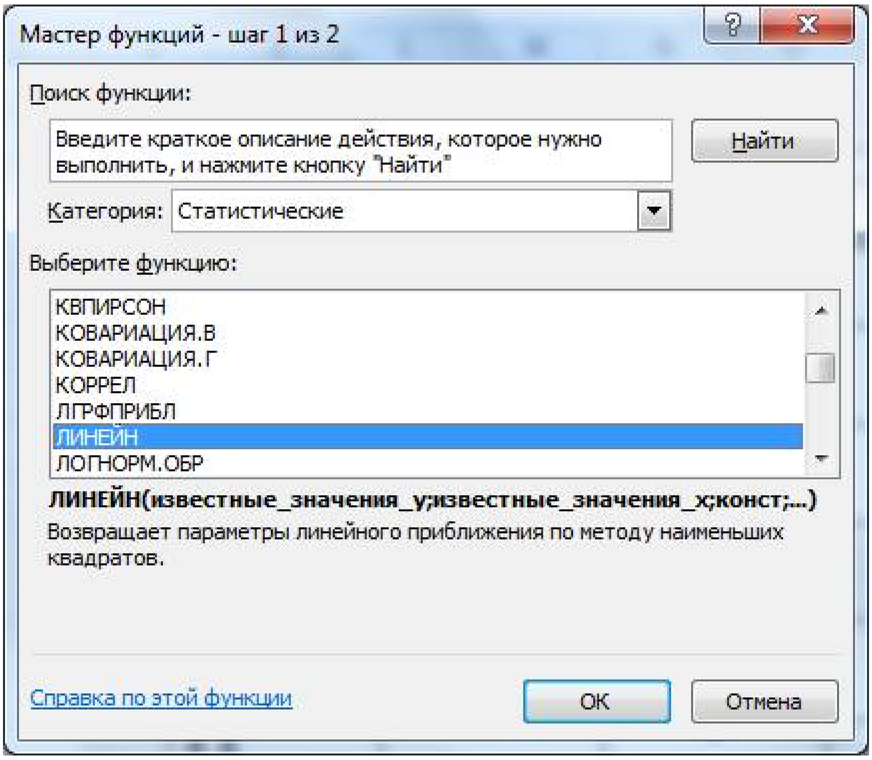


Рисунок 3.6 – Функция ЛИНЕЙН

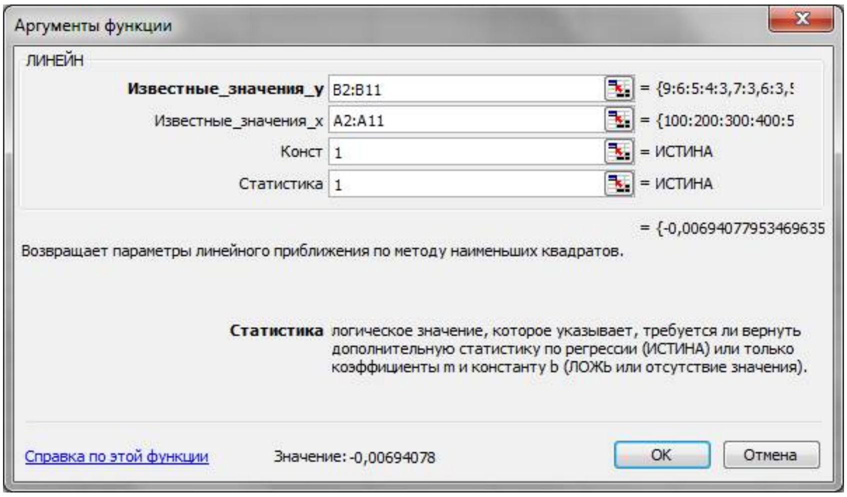


Рисунок 3.7 – Аргументы функции

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

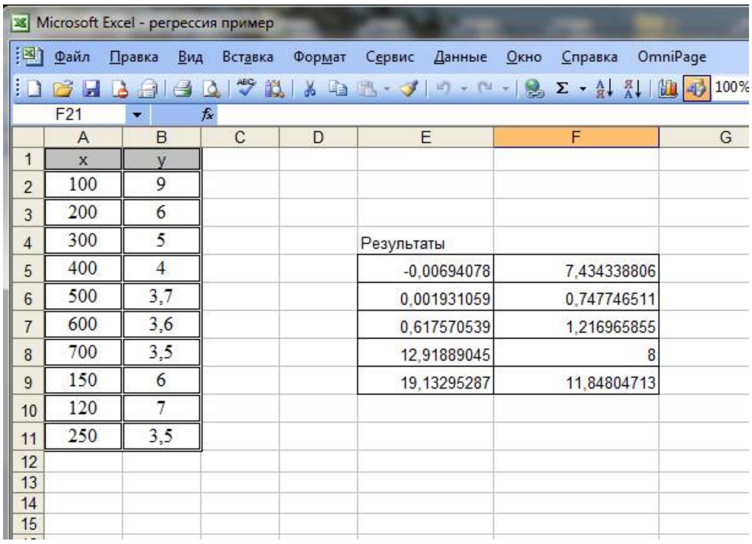
– известные значения функции Y – диапазон, содержащий данные результативного признака;

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

- Константа – логическое значение, которое указывает на наличие или отсутствие свободного члена в уравнении. Если константа = 1, то свободный член рассчитывается обычным образом. Если константа = 0, то свободный член равен 0.
- Статистика – логическое значение, которое указывает, как выводить дополнительную информацию по регрессионному анализу. Если статистика = 1, то дополнительная информация выводится, если статистика = 0, то выводятся только коэффициенты уравнения.
- Щёлкните по кнопке **ОК**.
- В левой верхней ячейке выделенной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу F2, а затем – на комбинацию клавиш CTRL + SHIFT + ENTER. Появятся результаты (рис. 3.8).
- Дополнительная регрессионная статистика будет выводиться в порядке, указанном в следующей схеме:

Таблица 3.2 – Результат функции ЛИНЕЙН

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



	A	B	C	D	E	F	G
1	x	y					
2	100	9					
3	200	6					
4	300	5			Результаты		
5	400	4			-0,00694078	7,434338806	
6	500	3,7			0,001931059	0,747746511	
7	600	3,6			0,617570539	1,216965855	
8	700	3,5			12,91889045	8	
9	150	6			19,13295287	11,84804713	
10	120	7					
11	250	3,5					
12							
13							
14							
15							

Рисунок 3.8 – Результат

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

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

- t_{a1}, t-статистика Стьюдента:

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

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

Оба коэффициента значимы, так t-статистики Стьюдента для них больше чем $t_{крит}$. Результаты совпадают с результатами, вычисленными по формулам.

3.2 Применение инструмента дисперсионного однофакторного анализа пакета «Анализ данных» электронных таблиц Excel

Теоретическая часть. Рассмотрим применение дисперсионного однофакторного анализа и пакета «Анализ данных» для решения производственных и экономических задач о влиянии новых технологий на какую-либо производственную или экономическую величину. Работа может быть использована при преподавании дисциплин «Программные комплексы общего назначения», «Пакеты прикладных программ, информационные технологии в экономике и менеджменте». Рассмотрим сначала математический аппарат дисперсионного однофакторного анализа, затем применим пакет «Анализ данных».

В качестве величины, на которую влияет фактор, может быть качество продукции, урожайность, качество подготовки специалистов, качество знаний по дисциплинам в результате применения балльно-рейтинговой системы, качество очистки воды и др. Дисперсионный анализ применяют, чтобы установить, оказывает ли существенное влияние некоторый качественный фактор F, который имеет p уровней F_1, F_2, …, F_p на изучаемую величину X. Основная идея дисперсионного анализа состоит в сравнении факторной дисперсии, порождаемой воздействием фактора, и остаточной дисперсии, обусловленной случайными причинами. Если различие между этими дисперсиями значимо, то фактор оказывает существенное влияние на X [1].

Опишем математический аппарат на примере.

Для проверки влияния новой технологии на качество однотипной продукции проведена выборочная проверка процентов брака за пять месяцев на трёх производственных объектах. Результаты проверки представлены в таблице (матрице наблюдений) (табл. 3.2.1). Методом дисперсионного анализа проверить нулевую гипотезу о существенном влиянии технологии на качество продукции.

Таблица 3.2.1 – Исходные данные

Номер испытания	Уровни фактора		
	F_1	F_2	F_3
1	3	1	2
2	5	4	4
3	4	5	3
4	3	2	10
5	2	5	3
Групповая средняя	3,4	3,4	4,4

Примем нулевую гипотезу H_0: технология существенно влияет на качество продукции. Конкурирующей гипотезой является гипотеза H_1: технология не влияет существенно на качество продукции. В условии задачи число уровней фактора равно числу производственных участков, на которых проведена выборочная проверка продукции: Р = 3, а число наблюдений на каждом уровне q = 5, т. к. проверка проводилась за пять месяцев.

Вычислим значения групповых средних $\overline{x}_{гр,j}$ для каждого из уровней фактора F_j (j = 1, 2, 3) и запишем их в дополнительной строке таблицы.

$$\overline{x}_{гр,1} = \frac{\sum_{i=1}^{q} x_{i1}}{q} = \frac{3+5+4+3+2}{5} = 3,4; \quad \overline{x}_{гр,2} = \frac{\sum_{i=1}^{q} x_{i2}}{q} = \frac{1+4+5+2+5}{5} = 3,4;$$

$$\overline{x}_{гр,3} = \frac{\sum_{i=1}^{q} x_{i3}}{q} = \frac{2+4+3+10+3}{5} = 4,4.$$

Вычислим общую среднюю $\overline{x}$:

$$\overline{x} = \frac{\sum_{j=1}^{p}\sum_{i=1}^{q} x_{ij}}{pq} = \frac{\sum_{j=1}^{p} \overline{x}_{гр,j}}{p} = \frac{\sum_{j=1}^{3} \overline{x}_{гр,j}}{3} = \frac{3,4+3,4+4,4}{3} = \frac{11,2}{3} = 3,73.$$

Вычислим общую сумму квадратов отклонений вариант от общей средней.

$$S_{общ} = \sum_{i=1}^{p}\sum_{j=1}^{q}(x_{ij} - \overline{x})^2 = \sum_{i=1}^{p}\sum_{j=1}^{q} x_{ij}^2 - pq\overline{x}^2.$$

Для расчёта $S_{общ}$ составим таблицу квадратов вариант (табл. 3.2.2).

Таблица 3.3.2 – Квадраты

Номер испытания	Уровни фактора		
	F_1	F_2	F_3
1	9	1	4
2	25	16	16
3	16	25	9
4	9	4	100
5	4	25	9
Σ	63	71	138

Получим $S_{общ} = 63 + 71 + 138 - 3 \cdot 5 \cdot 3,73^2 = 62,93$.

Вычислим факторную сумму квадратов отклонений групповых средних от общей средней:

$$S_{факт} = q \cdot \sum_{j=1}^{p}(x_{гр,j} - \overline{x})^2 = q(\sum_{j-1}^{p} x_{гр,j} - p \cdot \overline{x}^2) = 5 \cdot (3,4^2 + 3,4^2 + 4,4^2 - 3 \cdot 3,73^2) = 3,33.$$

Вычислим остаточную сумму квадратов отклонений.

$$S_{осг} = S_{общ} - S_{факт} = 62,93 - 3,33 = 59,6.$$

Определим факторную и остаточную дисперсии.

$$s^2_{факт} = \frac{S_{факт}}{p-1} = \frac{3,33}{3-1} = 1,66;$$

$$s^2_{ост} = \frac{S_{ост}}{p(q-1)} = \frac{59,6}{3(5-1)} = 4,96.$$

Сравним факторную и остаточную дисперсии по критерию Фишера–Снедекора. Для этого найдём наблюдаемое значение критерия $F_{набл.}$

$$F_{набл} = \frac{s^2_{факт}}{s^2_{ост}} = \frac{1,66}{4,96} = 0,33.$$

По таблице критических точек распределения Фишера–Снедекора находим значение $F_{кр}(\alpha, к_1, к_2)$, соответствующее заданному уровню значимости $\alpha = 0,05$ и значениям степеней свободы факторной дисперсии $к_1 = p - 1 = 3 - 1 = 2$ и $к_2 = p(q - 1) = 3(5 - 1) = 12$;

$$F_{кр}(0,05; 2; 12) = 3,88.$$

Так как $F_{набл} = 0,33 < F_{кр} = 3,88$, то нулевую гипотезу о существенном влиянии технологий на процент брака не отвергаем. Применительно к рассматриваемой задаче принятие нулевой гипотезы означает существенное влияние технологии на качество продукции на заданном уровне значимости.

Используем теперь пакет «Анализ данных», встроенный в электронные таблицы Excel. Подготовим исходные данные на листе Excel (рис. 3.2.1).

	A	B	C	D	E
1					
2		Номер испытания	Уровни фактора		
3			F_1	F_2	F_3
4		1	3	1	2
5		2	5	4	4
6		3	4	5	3
7		4	3	2	10
8		5	2	5	3
9		Групповая средняя	3,4	3,4	4,4
10					
11					
12					

Рисунок 3.2.1 – Исходные данные

В окне пакета «Анализ данных» выберем категорию **однофакторный дисперсионный анализ** и щёлкнем по кнопке ОК (рис. 3.2.2).

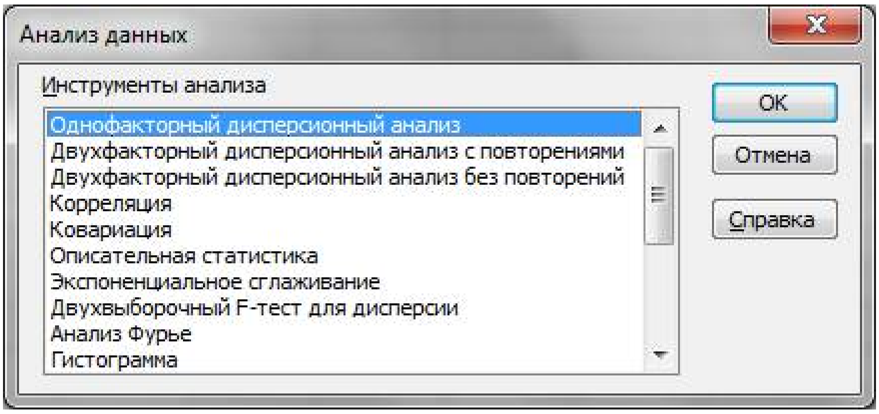


Рисунок 3.2.2 – Анализ данных

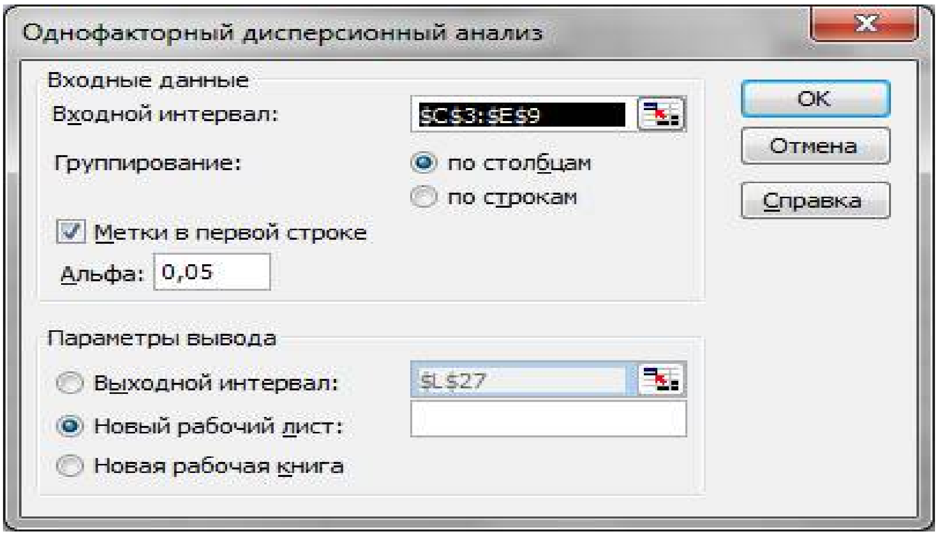


Рисунок 3.2.3 – Параметры анализа

Откроется окно **однофакторный дисперсионный анализ** (рис. 3.2.4). Введём исходные данные из интервала С3 : Е9, поставим переключатель **группирование по столбцам**. Установим флажок на вкладке **Метки в первой строке**. Зададим уровень значимости Альфа = 0.05. В параметрах вывода укажем в выходном интервале ячейку, начиная с которой будут выведены результаты расчёта.

Таблица 3.2.2 – Результаты

Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	*Счёт*	*Сумма*	*Среднее*	*Дисперсия*		
F1	5	17	3,4	1,3		
F2	5	17	3,4	3,3		
F3	5	22	4,4	10,3		
Дисперсионный анализ						
Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Между группами	3,333333333	2	1,666667	0,3355705	0,721428254	3,885293835
Внутри групп	59,6	12	4,966667			
Итого	62,93333333	14				

Пояснение к таблице результатов. В таблице Дисперсионный анализ во втором столбце SS в строке первой вычислена факторная сумма $S_{факт} = 3{,}333$, во второй строке вычислена остаточная сумма $S_{осг} = 59{,}6$, в третьей строке общая сумма квадратов отклонений вариант от общей средней $S_{общ} = 62{,}933$. В столбце df показано число степеней свободы. В столбце MS в первой строке выведена факторная дисперсия $s^2_{факт} = 1{,}66$, во второй строке остаточная дисперсия $s^2_{ост} = 4{,}96$. В столбце F вычислено наблюдаемое значение критерия $F_{набл} = 0{,}335$. В столбце *Р-Значение* вычислено значение $p = 0{,}721$, по которому можно судить о принятии гипотезы H_0. Если значение p > уровня значимости Альфа = 0,05, то гипотеза H_0 принимается, в противном случае не принимается. В столбце *F критическое* вычислено значение $F_{кр}(0{,}05;\ 2;\ 12) = 3{,}885$, соответствующее заданному уровню значимости $\alpha = 0{,}05$ и значениям степеней свободы факторной дисперсии $к_1 = p - 1 = 3 - 1 = 2$ и $к_2 = p(q - 1) = 3(5 - 1) = 12$. Это критическая точка распределения Фишера–Снедекора. По полученным результатам можно судить, что нулевую гипотезу о существенном влиянии технологий на процент брака не отвергаем. Применительно к рассматриваемой задаче принятие нулевой гипотезы означает существенное влияние технологии на качество продукции на заданном уровне значимости.

Задание по теме «Однофакторный дисперсионный анализ»

Для проверки влияния новой технологии на качество однотипной продукции проведена выборочная проверка процентов брака за пять месяцев на трёх производственных объектах F_1, F_2, F_3 Результаты проверки представлены в таблице (матрице наблюдений) (табл. 3.2.3). Методом дисперсионного анализа проверить нулевую гипотезу о существенном влиянии технологии на качество продукции. Вычисления выполнить: 1) в таблицах Excel по заданным формулам; 2) Используя пакет «Анализ данных», категорию «Однофакторный дисперсионный анализ».

Таблица 3.2.3 – Варианты задания

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
1	1	4	1	2		2	1	5	2	1
	2	2	2	5			2	4	3	3
	3	4	4	1			3	1	2	1
	4	2	3	3			4	2	3	3
	5	3	1	2			5	1	1	3
Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
3	1	3	4	1		4	1	1	2	3
	2	2	3	4			2	4	5	1
	3	3	4	5			3	5	1	2
	4	4	5	1			4	3	4	5
	5	5	1	2			5	1	3	2

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
5	1	2	3	4		6	1	3	4	5
	2	5	1	2			2	1	2	3
	3	3	4	5			3	4	5	1
	4	1	2	3			4	2	3	4
	5	4	5	1			5	5	1	2

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
7	1	4	5	6		8	1	5	1	2
	2	1	2	3			2	3	4	5
	3	4	5	1			3	1	2	3
	4	2	3	4			4	4	5	1
	5	5	1	2			5	2	3	4

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
9	1	1	2	3		10	1	2	4	5
	2	4	5	2			2	1	2	3
	3	1	2	3			3	5	3	1
	4	4	5	1			4	2	3	4
	5	2	3	4			5	3	2	1

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
11	1	2	2	3		12	1	5	4	3
	2	4	4	5			2	2	3	3
	3	5	5	1			3	4	1	2
	4	3	3	5			4	5	2	4
	5	4	2	1			5	1	5	1

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
13	1	4	5	1		14	1	2	1	3
	2	2	4	3			2	4	2	2
	3	1	2	3			3	3	2	3
	4	4	5	1			4	1	4	4
	5	2	3	4			5	5	5	1

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
15	1	5	3	4		16	1	4	3	6
	2	3	4	5			2	5	4	7
	3	3	5	2			3	2	2	3
	4	4	2	6			4	3	1	8
	5	1	1	2			5	1	5	1

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
17	1	5	4	6		18	1	3	5	4
	2	4	3	3			2	4	4	5
	3	3	2	2			3	2	3	3
	4	6	1	1			4	1	2	2
	5	1	3	5			5	3	1	2

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
19	1	5	6	2		20	1	3	4	4
	2	4	4	3			2	4	3	3
	3	4	3	4			3	4	5	5
	4	3	2	1			4	2	2	3
	5	5	2	1			5	1	2	2

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
21	1	4	3	2		22	1	4	3	4
	2	5	3	4			2	4	4	3
	3	4	2	3			3	5	5	3
	4	3	5	3			4	3	2	2
	5	2	3	1			5	2	1	2

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
23	1	2	3	4		24	1	3	6	3
	2	6	7	8			2	2	4	1
	3	4	5	3			3	5	4	2
	4	6	2	4			4	1	3	2
	5	1	5	1			5	6	2	7

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
25	1	3	6	7		26	1	4	5	7
	2	2	3	4			2	2	7	2
	3	6	4	1			3	5	3	4
	4	5	3	2			4	8	5	2
	5	1	2	3			5	2	3	4

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
27	1	6	3	2		28	1	8	7	5
	2	5	4	3			2	4	6	5
	3	2	5	4			3	2	3	4
	4	4	4	2			4	2	3	5
	5	3	5	6			5	1	4	6

Вариант	Номер испытания	F_1	F_2	F_3		Вариант	Номер испытания	F_1	F_2	F_3
29	1	3	4	5		30	1	2	5	4
	2	2	3	6			2	6	7	3
	3	4	5	7			3	2	4	6
	4	3	5	2			4	7	6	3
	5	1	2	3			5	2	3	1

4. Математический пакет Mathcad

Mathcad – математический пакет, т. е. пакет программ для математических вычислений. Кроме Mathcad существуют и другие математические пакеты, например, MathLab, Mathematika и др., однако Mathcad – самый распространённый. Пакет Mathcad используется при изучении дисциплин «Моделирование систем», «Системы искусственного интеллекта», «Системы реального времени» и «Информационные технологии».

Mathcad – это мощная и в то же время простая универсальная среда для решения задач в различных отраслях науки и техники, финансов и экономики, физики и астрономии, математики и статистики.

Mathcad остаётся единственной системой, в которой описание решения математических задач задаётся с помощью привычных математических формул и знаков.

Mathcad позволяет выполнять как численные, так и аналитические (символьные) вычисления, имеет чрезвычайно удобный математико-ориентированный интерфейс и прекрасные средства научной графики.

Система Mathcad существует в нескольких основных вариантах:

• Mathcad Standard – идеальная система для повседневных технических вычислений. Предназначена для массовой аудитории и широкого использования в учебном процессе;

• Mathcad Professional – промышленный стандарт прикладного использования математики в технических приложениях. Ориентирована на математиков и научных работников, проводящих сложные и трудоёмкие расчёты.

• Mathcad Professional Academic – пакет программ для профессионального использования математического аппарата с электронными учебниками и ресурсами.

Mathcad – универсальная программа, представляющая автоматизированную систему, которая позволяет динамически обрабатывать данные в числовом и аналитическом виде. Она сочетает в себе возможности проведения расчётов и подготовки форматирования научно-технических документов, которые обычно содержат формулы, результаты расчётов в виде таблиц и графиков, текстовые комментарии.

В Mathcad этим документам соответствуют два вида объектов:

- формулы;

- текстовые блоки.

Документ данной программы называется **рабочим листом**. Он может содержать перечисленные выше виды объектов. В ходе расчётов формулы обрабатываются последовательно слева направо и сверху вниз.

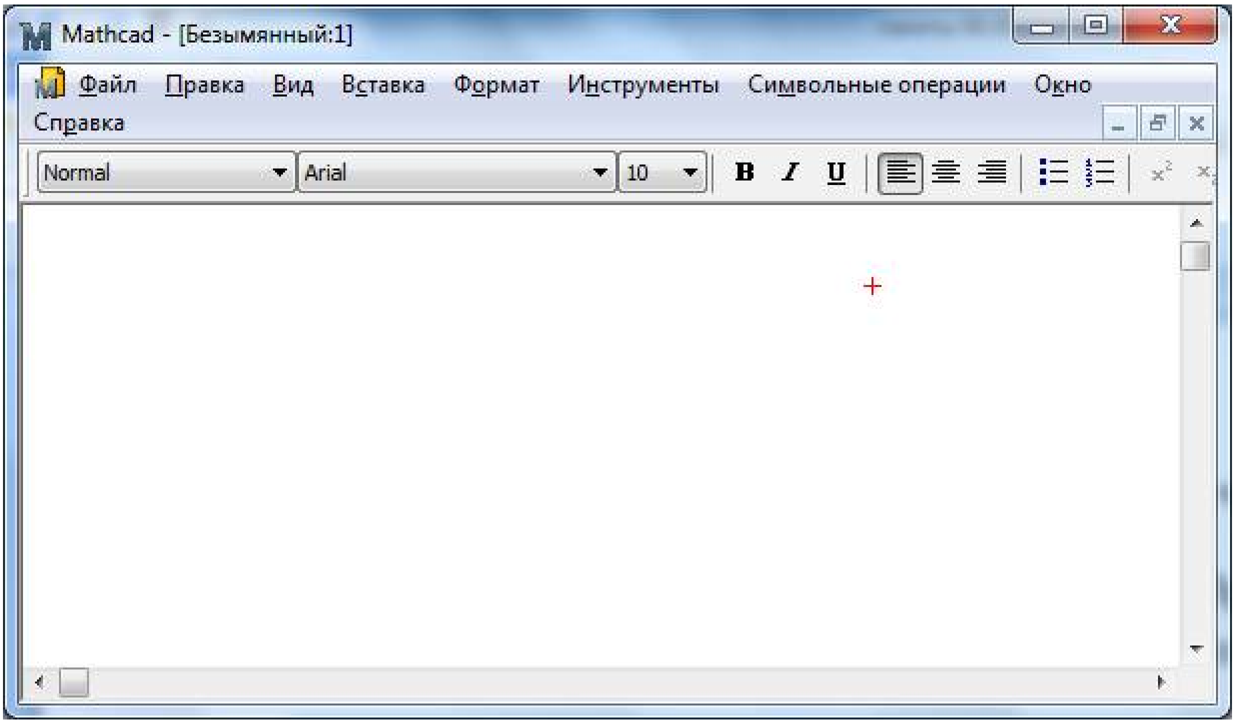


Рисунок 4.1 – Рабочий лист

Ввод информации производится в месте расположения курсора, который имеет три вида:

- крестообразный – определяет место создания следующего объекта;
- угловой – появляется при вводе формул;
- вертикальная черта – при вводе текста.

Для ввода формул используются следующие панели:

1) панель Арифметика – содержит арифметические действия и функции (цифры, скобки, тригонометрические функции и т. д.);
2) панель Управление вычислениями;
3) панель Графики;
4) панель Математика;
5) панель Матрица;
6) панель Булево – логические операции типа равенства и т. п.;
7) панель Программирование;
8) панель Греческий алфавит;
9) панель Символы.

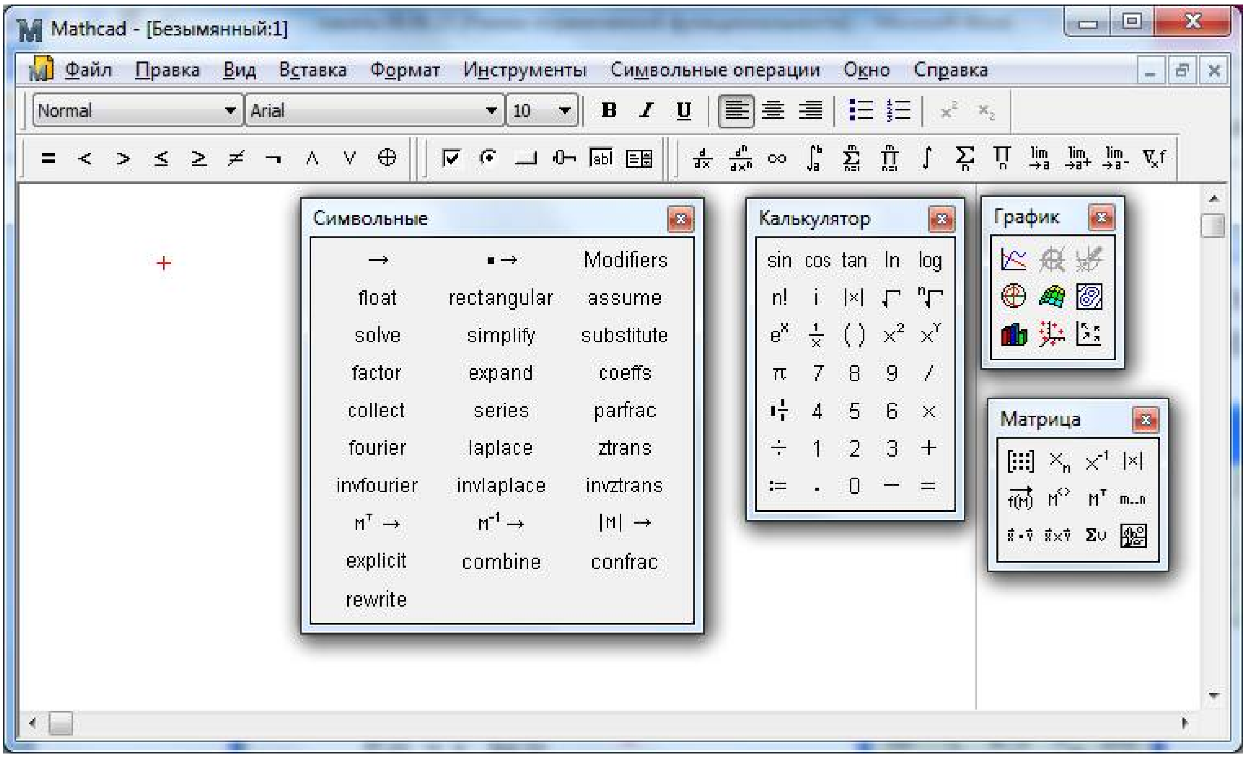


Рисунок 4.2 – Панели

Рассмотрим задачи, которые часто встречаются в нефтегазовом деле.

4.1 Табулирование функций, построение графиков функций в Mathcad

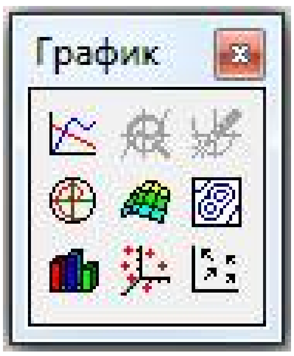


Рисунок 4.1.1 – Панель графиков

Одним из многих достоинств Mathcad является лёгкость построения графиков. Панель графиков вызывается нажатием кнопки с изображением графиков на математической панели (рис. 4.1.1).

На панели графиков расположены девять кнопок с изображением различных типов графиков (название графиков каждой кнопки высвечивается при подводе к ней курсора и ожидании в течение 3–5 секунд): X-Y Plot – графики в декартовых координатах, Polar Plot – графики в полярных координатах, 3D Bar Chart – столбиковые диаграммы, Surface Plot – трёхмерный график, Cunter Plot – карта линий уровня, Vector Field Plot – векторное поле, 3D Scatter Plot – трёхмерный точечный график. Сначала нас будет интересовать левая верхняя кнопка X-Y графиков в декартовой системе координат (по-английски X-Y Plot).

Пример 1. Составить таблицу значений функции:

$$y = x \ln x,\ x \in [4, 9],\ h = 0{,}5.$$

Решение:

На листе Mathcad набираем x:=4,4.5..9. Чтобы получить знак **:=** нажимаем на клавиши SHIFT и **;**. Чтобы получить «**..**» нажимаем на клавишу **;**. Далее набрать f(x):=x · ln(x) и для вывода значений функции набираем f(x) = и щелкаем по чистой области листа, в результате получим таблицу значений функции y (рис. 4.1.2).

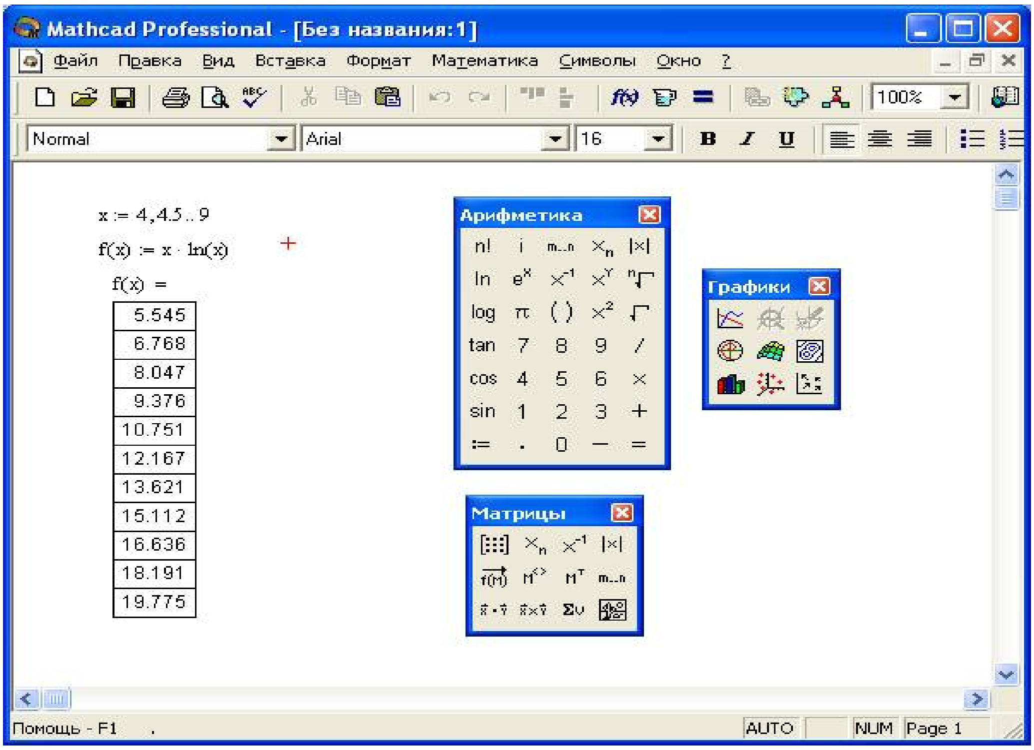


Рисунок 4.1.2 – Пример

Пример 2. Построить график функции $y = x^2$ на промежутке [–5;5] с шагом h = 0,1.

Решение. На листе вводим x:=–5,–4,9..5 и f(x):=x^2, далее – войти в меню Вставка | График | X-Y Зависимость либо использовать панель инструментов График. На листе появится область построения графика. На ней задать аргумент x и пределы изменения x, слева ввести текст «f(x)». Затем щёлкнуть вне области графика. Появится график функции (рис. 4.1.2).

Для этого нужно:

1) набрать, как и в прошлый раз, условие задачи;

2) провести ранжировку x, набрав пределы его изменения x, например x:= 0,0.01,5. Здесь мы задали шаг изменения x гораздо меньше, чем в предыдущем разделе. Чем меньше шаг изменения аргумента, тем более гладким получается график;

3) вызвав панель графиков, нажать на кнопку с изображением декартовых графиков. Появятся два вложенных друг в друга квадрата, внутри которых есть несколько точек;

4) сначала нужно подвести курсор к средней точке оси абсцисс и набрать там аргумент x;

5) затем следует подвести курсор к средней точке около оси y и набрать там наименование функции в виде y (x);

6) щёлкнем несколько раз мышью вне графика. На экране появится график параболы. Поместим курсор внутрь графика и щёлкнем левой клавишей мыши. Появится окно, показанное на рисунке 4.1.3. Оно состоит из трёх страниц. На рисунке 4.1.3 представлена первая страница.

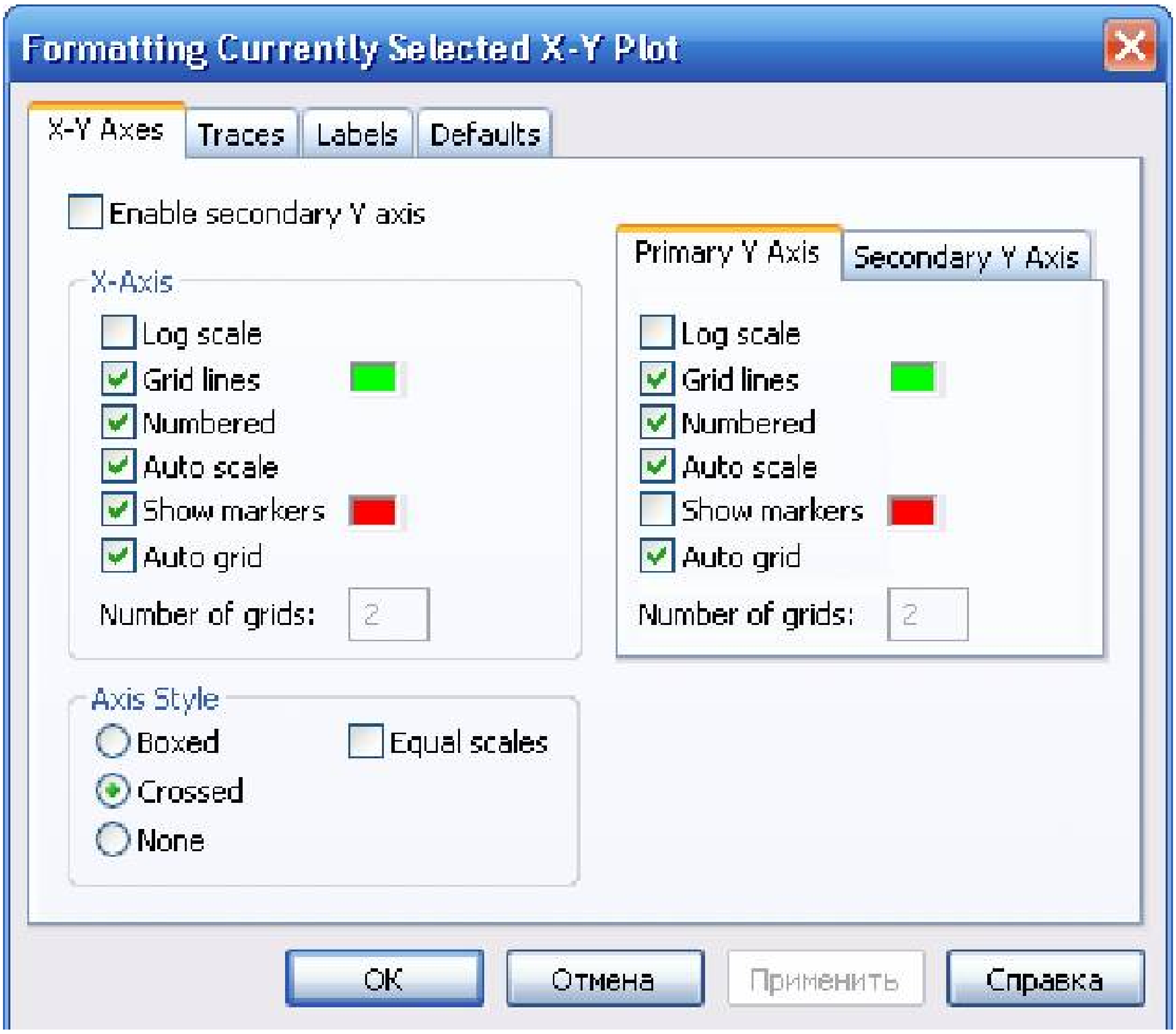


Рисунок 4.1.3 – Первая страница окна формирования графиков

В левом нижнем углу страницы имеются точки Boxed (коробочка), Crossed (оси), None (нет). Нажав на первую точку, введём в график оси координат. На первой странице имеются строки X-Axes (ось X) и Primary Y-Axes (первая ось Y), а под ними ряд надписей, левая часть которых относится к оси X, а правая – к оси Y: Log Scale (логарифмическая шкала) вводит логарифмический масштаб для соответствующей оси; Grid lines (сетка) – её нажатие вводит сетку на график; Numbered (оцифровка) – оцифровка сетки; Auto scale (автоматическая оцифровка); Show markers (показ маркеров); Auto grid (автоматическое разбиение сетки). Наличие надписей Enable secondary Y-Axes (возможность второй оси Y) и Secondary YAxes (вторая ось Y) даёт возможность формировать графики различного масштаба для различных функций.

На рисунке 4.1.4 представлена вторая страница того же окна.

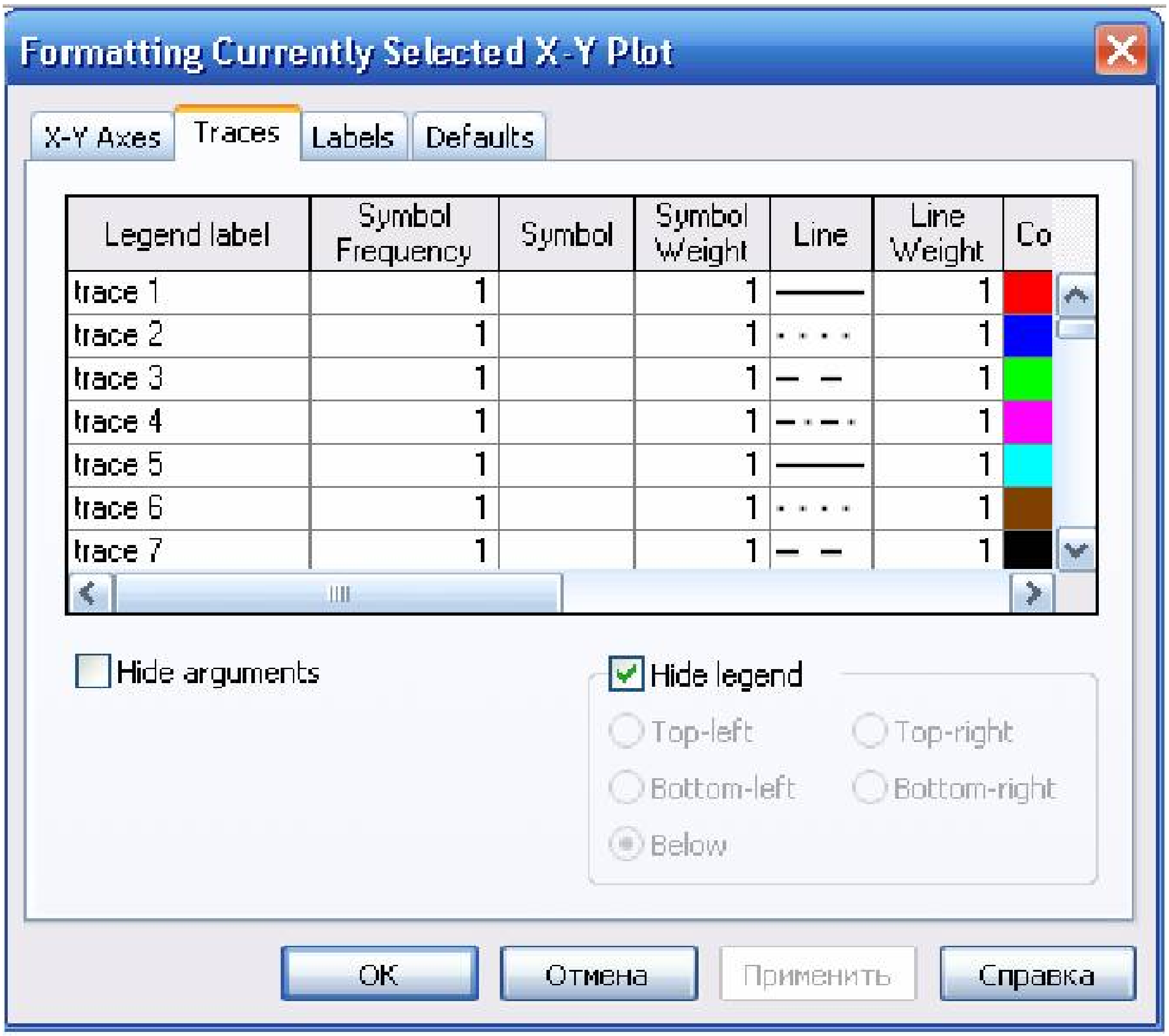


Рисунок 4.1.4. – Вторая страница окна формирования графика

Из её левого столбца (trace (след) 1, trace 2 и т. д.) следует, что на одном графике можно наносить до 16 различных функций.

Вводя соответствующие значения в остальные столбцы, можно изменять вид (сплошная линия, пунктир, точки), цвет, толщину и т. д. каждой функции.

На третьей странице окна задаётся заголовок (Title), место его расположения Above (сверху), Belove (снизу), наименования осей (Axis Labels).

Выбрав те или иные требования к графику, нажмём ОК и получим желаемый график. Заполнив графы двух страниц, как показано на рисунках 4.1.3, 4.14, получим график, показанный на рисунке 4.1.5.

$x := -5, -4.9 .. 5$

$f(x) := x^2$

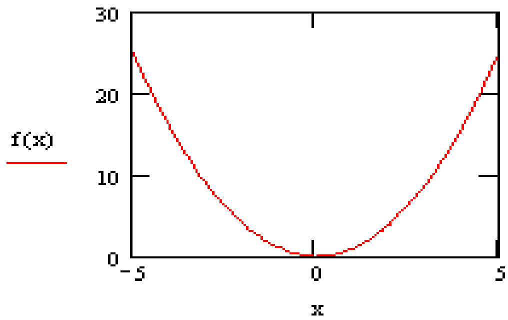


Рисунок 4.1.5 – График

4.2 Задания. Табулирование функций, построение графиков функций

№ варианта	*Заданная функция*	*Интервал*
1.	$y = x^2 + \sin(1/x)$	0,5; 3.
2.	$y = \sin(2x)/(3 - \sin(x))$	0; π.
3.	$y = x^3 + x^2 + 3 - 3x$	–1; 3.
4.	$y = e^{-\frac{x}{3}} \sin(2{,}2x)$	0; 2.
5.	$y = \sqrt{x^2 + 2} - 0{,}1x$	0; 3.
6.	$y = \sin(-1/x)\cos(x)$	1; 4.
7.	$y = x \sin(4x)$	0,3; π/2.
8.	$y = e^{\sqrt{x}} \sin(2x)$	0; π.
9.	$y = 2 + 2\cos(x)$	0; 2π.
10.	$y = 9/(4 - 5\cos(x))$	0; 2π.
11.	$y = 3/(1 - \cos(x))$	0; 2π.
12.	$y = 3 - \sin(3x)$	0; 2π.
13.	$y = 5\sin(3x)$	0; 2π.
14.	$y = x/(x^2 + 1)$	–10; 10.
15.	$y = x/(x-1)^2$	–10; 10.
16.	$y = x^3/2(x + 1)^2$	–10; 10.
17.	$y = (x^2 + 16)/x$	–10; 10.
18.	$y = 2 - \sin(x)$	0; 2π.
19.	$y = (1 - x^3)/x^2$	–10; 10.

№ варианта	*Заданная функция*	*Интервал*
20.	$y = 2\cos(x)$	0; 2π.
21.	$y = 5\sin^3(x/3)$	0; 2π.
22.	$y = ((x + 1)/(x – 1))^2$	–10; 10.
23.	$y = e^{\frac{1}{(x+2)}}$	–10; 10.
24.	$y = \sin^4(x) + \cos^4(x)$	–10; 10.
25.	$y = (x^3 + 16)/x$	–10; 10.
26.	$y = 4(1 + \cos(x))$	0; 2π.
27.	$y = 1/(2\sqrt{3}\cos(x))$	0; 2π.
28.	$y = 3 + 2\cos(2x)$	0; 2π.
29.	$y = e^x / x$	–10; 10.
30.	$y = 10(1 – \cos(x))$	0; 2π.

Библиографический список

1. Туманова, О. Н. Использование компьютерных технологий для решения экономических задач : учеб. пособие / О. Н. Туманова. – Ухта : УГТУ, 2010. – 94 с.

2. Туманова, О. Н. Программные комплексы общего назначения : метод. указания / О. Н. Туманова. – Ухта : УГТУ, 2014. – 50 с.

3. Информатика. Базовый курс : учеб. пособие для студентов высш. техн. заведений / С. В. Симонович [и др.] ; под ред. С. В. Симоновича. – 2-е изд. – СПб. : Питер, 2007. – 640 с.

4. Кудинов, Ю. А. Основы современной информатики / Ю. А. Кудинов. – СПб. : Лань, 2011. – 256 с.

5. Практикум по основам современной информатики / Ю. А. Кудинов [и др.]. – СПб. : Лань, 2011. – 352 с.

6. Безручко, В. Т. Информатика [Электронный ресурс] : курс лекций / В. Т. Безручко. – М. : ИД ФОРУМ : НИЦ ИНФРА-М, 2014. – 432 с. – Режим доступа : http://www.znanium.com

7. Экономическое моделирование в Microsoft Excel / М. Джеффери [и др.]. – М. : Вильямс, 2012. – 562 с.

8. Электронные словари и энциклопедии на Академике [Электронный ресурс]. – Режим доступа : http://dic.academic.ru

9. Орлова, И. В. Экономико-математические методы и модели: компьютерное моделирование [Электронный ресурс] : учеб. пособие / И. В. Орлова, В. А. Половников. – М. : Вузовский учебник : ИНФРА-М, 2013. – 389 с. – Режим доступа : www.alleng.ru/

Учебное издание

Ольга Николаевна Туманова
Валентина Ивановна Серкова

Прикладные программные продукты

Учебное пособие

Корректор О. В. Мойсеня
Технический редактор Л. П. Коровкина

План 2015 г., позиция 087. Подписано в печать 29.01.2016 г.
Компьютерный набор. Гарнитура Times New Roman.
Формат 60x84 1/16. Бумага офсетная. Печать трафаретная.
Усл. печ л. 4,6. Уч.- изд. л. 4,2. Тираж 120 экз. Заказ № 302.

Ухтинский государственный технический университет.
169300, Республика Коми, г. Ухта, ул. Первомайская, 13.
Типография УГТУ.
169300, Республика Коми, г. Ухта, ул. Октябрьская, 13.