Министерство науки и высшего образования Российской Федерации

федеральное государственное бюджетное образовательное учреждение высшего образования «Казанский национальный исследовательский технический университет им. А.Н.Туполева – КАИ»

Институт экономики, управления и социальных технологий

Методические указания
по выполнению практических работ по дисциплине Б1.Б.02
«Математическое моделирование инноваций»

Казань 2019

Практическое занятие №1
Тема: «Введение в математическое моделирование инноваций»

Задание: составить кроссворд на тему: «Введение в математическое моделирование инноваций».

Практическое занятие №2
Тема: «Постановка задачи линейного программирования и ее виды»

Задание: подготовить реферат на тему: «Постановка задачи линейного программирования и ее виды».

Практическое занятие №3
Тема: «Решение оптимизационных задач»

ЦЕЛЬ РАБОТЫ

Приобретение навыков решения одноиндексных задач линейного программирования (ЛП) в табличном редакторе Microsoft Excel.

ОСНОВНЫЕ ПОЛОЖЕНИЯ

Если в какой-либо системе (экономической, организационной, военной и т.д.) не хватает имеющихся в наличии ресурсов для эффективного выполнения каждой из намеченных работ, то возникают так называемые **распределительные задачи**. Цель решения распределительной задачи – отыскание оптимального распределения ресурсов по работам. Под оптимальностью распределения может пониматься, например, минимизация общих затрат, связаных с выполнением работ, или максимизация получаемого в результате общего дохода.

Для решения таких задач используются методы математического программирования. **Математическое программирование** – это раздел математики, занимающийся разработкой методов отыскания экстремальных значений функции, на аргументы которой наложены ограничения. Слово «программирование» заимствовано из зарубежной литературы, где оно используется в смысле «планирование».

Наиболее простыми и лучше всего изученными среди задач математического программирования являются задачи линейного программирования.

Характерные черты задач ЛП следующие:

1) показатель эффективности L представляет собой линейную функцию, заданную на элементах решения $x_1, x_2, ..., x_n$;

2) ограничительные условия, налагаемые на возможные решения, имеют вид линейных равенств или неравенств.

В общей форме записи модель задачи ЛП имеет вид:

целевая функция (ЦФ)

$$L = c_1 x_1 + c_2 x_2 + ... + c_n x_n \to max(min);$$

при ограничениях

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le (\ge, =) b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le (\ge, =) b_2, \\ .... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \le (\ge, =) b_m, \\ x_1, x_2, ..., x_k \ge 0 \left(k \le n\right). \end{cases}$$

Допустимое решение – это совокупность чисел $X = \left(x_1, x_2, ..., x_n\right)$, удовлетворяющих ограничениям задачи.

Оптимальное решение – это план $X^* = \left(x^*_1, x^*_2, ..., x^*_n\right)$, при котором ЦФ принимает свое максимальное (минимальное) значение.

ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

Для модели ЛП, соответствующей номеру Вашего варианта, найдите оптимальное решение в табличном редакторе Microsoft Excel и продемонстрируйте его преподавателю.

ИНСТРУКЦИЯ ПО ИСПОЛЬЗОВАНИЮ Microsoft Excel ДЛЯ РЕШЕНИЯ ОДНОИНДЕКСНЫХ ЗАДАЧ ЛП

Для того чтобы решить задачу ЛП в табличном редакторе Microsoft Excel, необходимо выполнить следующие действия.

1 Ввести условие задачи:

a) ***создать экранную форму для ввода условия задачи:***

- переменных,
- целевой функции (ЦФ),
- ограничений,
- граничных условий;

b) ***ввести исходные данные в экранную форму***:

- коэффициенты ЦФ,
- коэффициенты при переменных в ограничениях,
- правые части ограничений;

c) ***ввести зависимости из математической модели в экранную форму***:

- формулу для расчета ЦФ,
- формулы для расчета значений левых частей ограничений;

d) ***задать ЦФ*** (в окне **«Поиск решения»**):

- целевую ячейку,
- направление оптимизации ЦФ;

e) ***ввести ограничения и граничные условия*** (в окне **«Поиск решения»**):

- ячейки со значениями переменных,
- граничные условия для допустимых значений переменных,
- соотношения между правыми и левыми частями ограничений.

2 Решить задачу:

a) ***установить параметры решения задачи*** (в окне **«Поиск решения»***);*

b) ***запустить задачу на решение*** (в окне **«Поиск решения»***)*;

c) ***выбрать формат вывода решения*** (в окне **«Результаты поиска решения»**).

ПРИМЕР РЕШЕНИЯ ОДНОИНДЕКСНОЙ ЗАДАЧИ ЛП

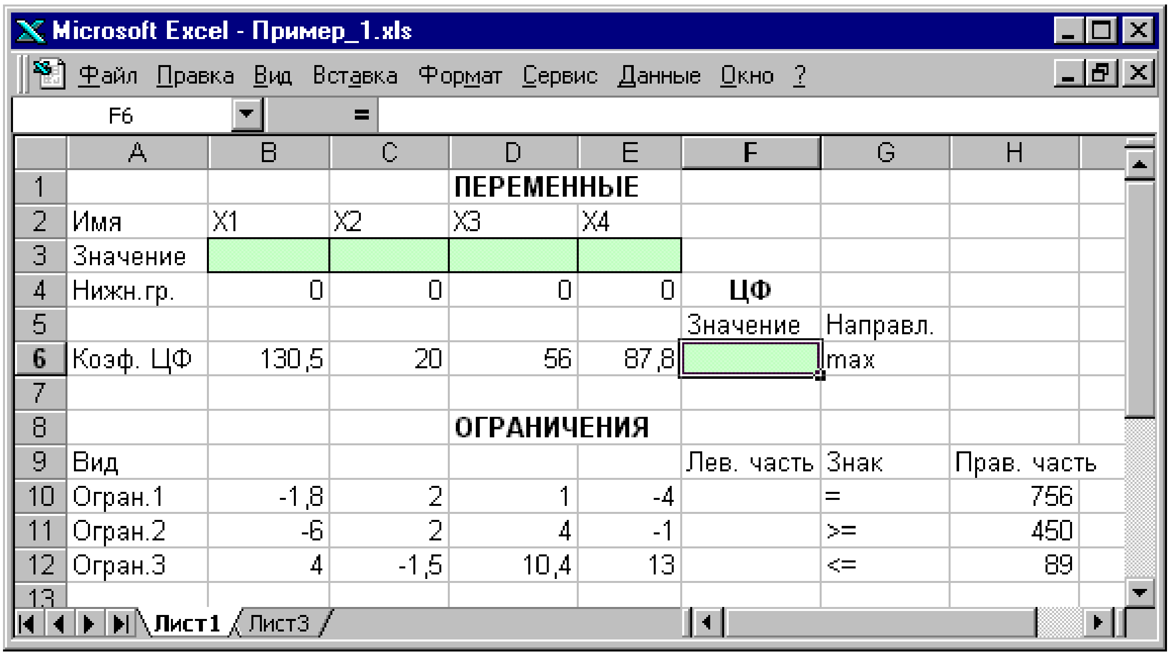
Рассмотрим пример нахождения решения для следующей одноиндексной задачи ЛП:

$$
\begin{aligned}
&L(X) = 130{,}5x_1 + 20x_2 + 56x_3 + 87{,}8x_4 \to max;\\
&\begin{cases}
-1{,}8x_1 + 2x_2 + x_3 - 4x_4 = 756,\\
-6x_1 + 2x_2 + 4x_3 - x_4 \ge 450,\\
4x_1 - 1{,}5x_2 + 10{,}4x_3 + 13x_4 \le 89,\\
x_j \ge 0;\ j = \overline{1{,}4}.
\end{cases}
\end{aligned}
\tag{3.1}
$$

ВВОД ИСХОДНЫХ ДАННЫХ

Создание экранной формы и ввод в нее условия задачи

Экранная форма для ввода условий задачи (3.1) вместе с введенными в нее исходными данными представлена на рис. 3.1.



	A	B	C	D	E	F	G	H
1				ПЕРЕМЕННЫЕ				
2	Имя	X1	X2	X3	X4			
3	Значение							
4	Нижн.гр.	0	0	0	0	ЦФ		
5						Значение	Направл.	
6	Коэф. ЦФ	130,5	20	56	87,8		max	
7								
8				ОГРАНИЧЕНИЯ				
9	Вид					Лев. часть	Знак	Прав. часть
10	Огран.1	-1,8	2	1	-4		=	756
11	Огран.2	-6	2	4	-1		>=	450
12	Огран.3	4	-1,5	10,4	13		<=	89
13								

Рис. 3.1. Экранная форма задачи (3.1) (курсор в ячейке F6)

В экранной форме на рис. 3.1 каждой переменной и каждому коэффициенту задачи поставлена в соответствие конкретная ячейка в Excel. Имя ячейки состоит из буквы, обозначающей столбец, и цифры, обозначающей строку, на пересечении которых находится объект задачи ЛП. Так, например, переменным задачи (3.1) соответствуют ячейки ***B*3** (x_1), ***C*3** (x_2), ***D*3** (x_3), ***E*3** (x_4), коэффициентам ЦФ соответствуют ячейки ***B*6** ($c_1 = 130{,}5$), ***C*6** ($c_2 = 20$), ***D*6** ($c_3 = 56$), ***E*6** ($c_4 = 87{,}8$), правым частям ограничений соответствуют ячейки ***H*10** ($b_1 = 756$), ***H*11** ($b_2 = 450$), ***H*12** ($b_3 = 89$) и т.д.

Ввод зависимостей из математической модели в экранную форму

Зависимость для ЦФ

В ячейку ***F*6**, в которой будет отображаться значение ЦФ, необходимо ввести **формулу,** по которой это значение будет рассчитано. Согласно (3.1) значение ЦФ определяется выражением

$$130{,}5x_1 + 20x_2 + 56x_3 + 87{,}8x_4\,. \qquad (3.2)$$

Используя обозначения соответствующих ячеек в Excel (см. рис. 3.1), формулу для расчета ЦФ (3.2) можно записать как **сумму произведений** каждой из ячеек, отведенных для значений переменных задачи (***B*3, *C*3, *D*3, *E*3**), на соответствующую ячейку, отведенную для коэффициентов ЦФ (***B*6, *C*6, *D*6, *E*6**), то есть

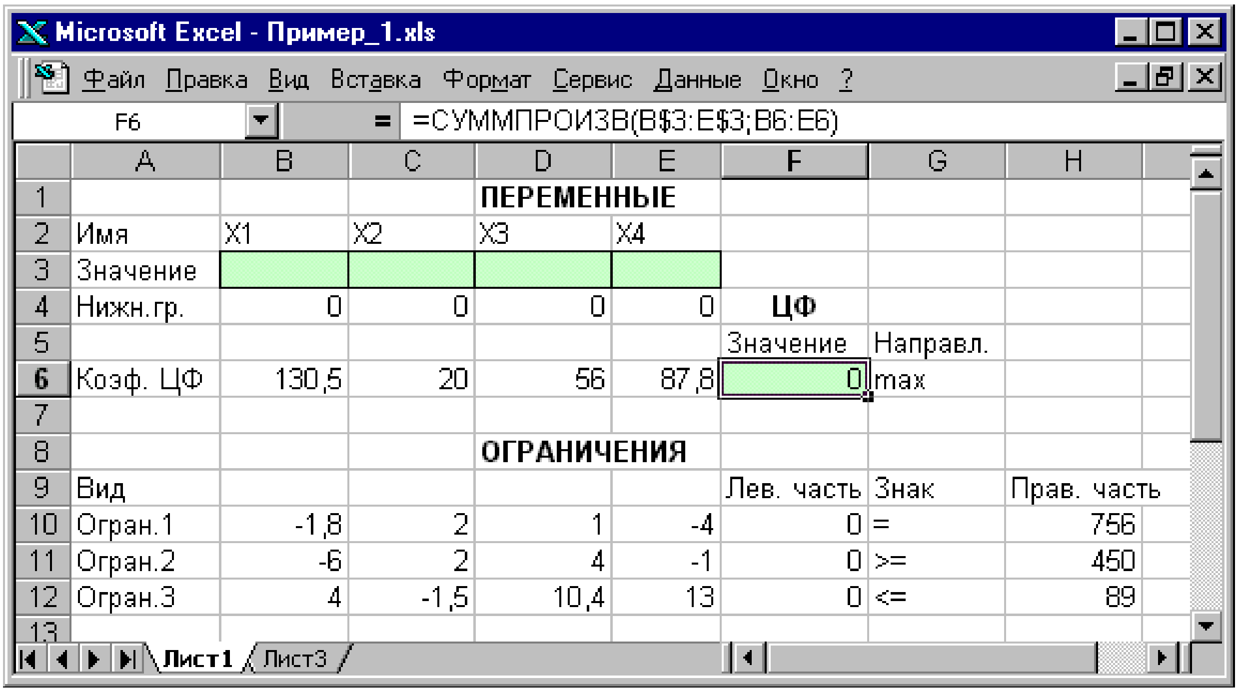
$$B6 \cdot B3 + C6 \cdot C3 + D6 \cdot D3 + E6 \cdot E3\,. \qquad (3.3)$$

Чтобы задать формулу (3.3) необходимо в ячейку ***F*6** ввести следующее выражение и нажать клавишу **«Enter»**

=СУММПРОИЗВ(B$3:E$3;B6:E6), (3.4)

где символ **$** перед номером строки 3 означает, что при копировании этой формулы в другие места листа Excel номер строки 3 не изменится;

символ **:** означает, что в формуле будут использованы **все** ячейки, расположенные между ячейками, указанными слева и справа от двоеточия (например, запись ***B6:E6*** указывает на ячейки ***B6, C6, D6* и *E6***). После этого в целевой ячейке появится 0 (нулевое значение) (рис. 3.2).



	A	B	C	D	E	F	G	H
1				ПЕРЕМЕННЫЕ				
2	Имя	X1	X2	X3	X4			
3	Значение							
4	Нижн.гр.	0	0	0	0	ЦФ		
5						Значение	Направл.	
6	Коэф. ЦФ	130,5	20	56	87,8	0	max	
7								
8				ОГРАНИЧЕНИЯ				
9	Вид					Лев. часть	Знак	Прав. часть
10	Огран.1	-1,8	2	1	-4	0	=	756
11	Огран.2	-6	2	4	-1	0	>=	450
12	Огран.3	4	-1,5	10,4	13	0	<=	89
13								

Рис. 3.2. Экранная форма задачи (3.1) после ввода всех необходимых формул (курсор в ячейке F6)

Примечание 1.1. Существует другой способ задания функций в Excel с помощью режима **«Вставка функций»,** который можно вызвать из меню **«Вставка»** или при нажатии кнопки « f_x « на стандартной панели инструментов. Так, например, формулу (3.4) можно задать следующим образом:

- курсор в поле ***F6*;**
- нажав кнопку « f_x**», вызовите окно «Мастер функций – шаг 1 из 2»;**
- выберите в окне **«Категория»** категорию **«Математические»;**
- в окне **«Функция»** выберите функцию **СУММПРОИЗВ*;***
- в появившемся окне **«СУММПРОИЗВ»** в строку **«Массив 1»** введите выражение **B$3:E$3**, а в строку **«Массив 2»** – выражение **B6:E6** (рис3.3);

- после ввода ячеек в строки **«Массив 1»** и **«Массив 2»** в окне **«СУММПРОИЗВ»** появятся числовые значения введенных массивов (см. рис. 3.3), а в экранной форме в ячейке ***F*6** появится текущее значение, вычисленное по введенной формуле, то есть 0 (так как в момент ввода формулы значения переменных задачи нулевые).

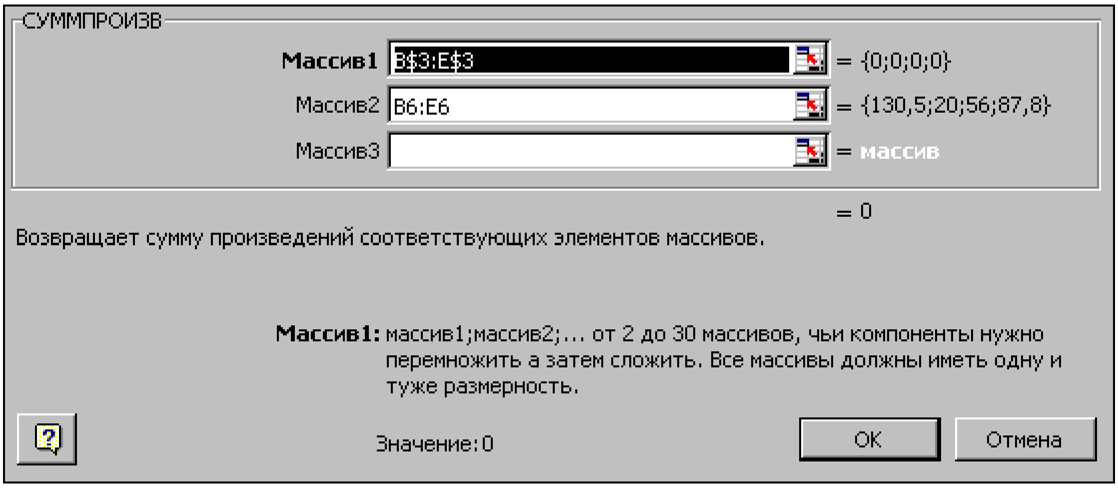


Рис. 3.3. Ввод формулы для расчета ЦФ в окно **«Мастер функций»**

Зависимости для левых частей ограничений

Левые части ограничений задачи (3.1) представляют собой *сумму произведений* каждой из ячеек, отведенных для значений переменных задачи (**B3, C3, D3, E3**), на соответствующую ячейку, отведенную для коэффициентов конкретного ограничения (**B10, C10, D10, E10 –** 1-е ограничение**; B11, C11, D11, E11** – 2-е ограничение и **B12, C12, D12, E12 –** 3-е ограничение). Формулы, соответствующие левым частям ограничений, представлены в табл. 3.1.

Таблица 3.1

Формулы, описывающие ограничения модели (3.1)

Левая часть ограничения	Формула Excel
$-1{,}8x_1 + 2x_2 + x_3 - 4x_4$ или $B10 \cdot B3 + C10 \cdot C3 + D10 \cdot D3 + E10 \cdot E3$	**=СУММПРОИЗВ(*B*$3:*E*$3;*B*10:*E*10)**

$-6x_1 + 2x_2 + 4x_3 - x_4$ или $B11 \cdot\ B3 +\ C11 \cdot\ C3 +\ D11 \cdot\ D3 + E11 \cdot E3$	**=СУММПРОИЗВ(*B*$3:**Е**$3;*B*11:*E* 11)**
$4x_1 - 1{,}5x_2 + 10{,}4x_3 + 13x_4$ или $B12 \cdot\ B3 +\ C12 \cdot\ C3 +\ D12 \cdot\ D3 + E12 \cdot E3$	**=СУММПРОИЗВ(*B*$3:*E*$3;*B*12:*E* 12)**

Как видно из табл. 1.1, формулы, задающие левые части ограничений задачи (3.1), отличаются друг от друга и от формулы (3.4) в целевой ячейке ***F*6** только номером строки во втором массиве. Этот номер определяется той строкой, в которой ограничение записано в экранной форме. Поэтому для задания зависимостей для левых частей ограничений достаточно скопировать формулу из целевой ячейки в ячейки левых частей ограничений. Для этого необходимо:

- поместить курсор в поле целевой ячейки ***F*6** и скопировать в буфер содержимое ячейки ***F*6** (клавишами **«Ctrl-Insert»**);
- помещать курсор поочередно в поля левой части каждого из ограничений, то есть в ***F*10, *F*11** и ***F*12,** и вставлять в эти поля содержимое буфера (клавишами **«Shift-Insert»**) (при этом номер ячеек во втором массиве формулы будет меняться на номер той строки, в которую была произведена вставка из буфера);
- на экране в полях **F10**, **F11** и **F12** появится 0 (нулевое значение) (см. рис. 3.2).

ПРОВЕРКА ПРАВИЛЬНОСТИ ВВЕДЕНИЯ ФОРМУЛ

Для проверки правильности введенных формул производите поочередно двойное нажатие левой клавиши мыши на ячейки с формулами. При этом на экране рамкой будут выделяться ячейки, используемые в формуле (рис. 3.4 и 3.5).

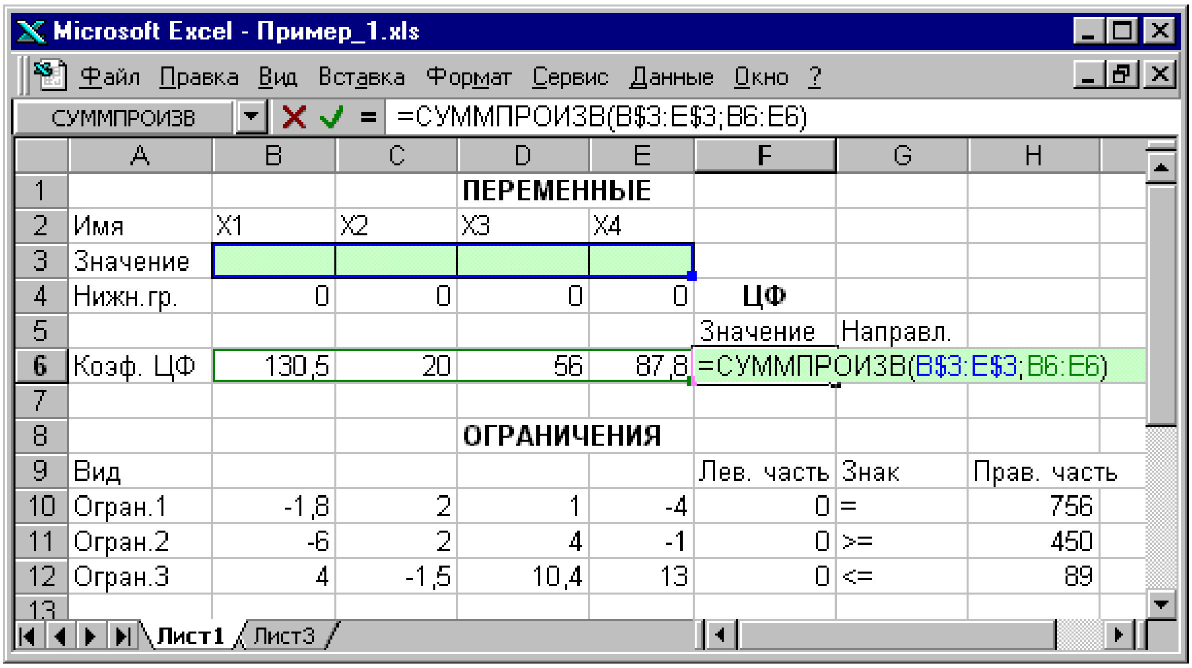


Рис. 3.4. Проверка правильности введения формулы в целевую ячейку **F6**

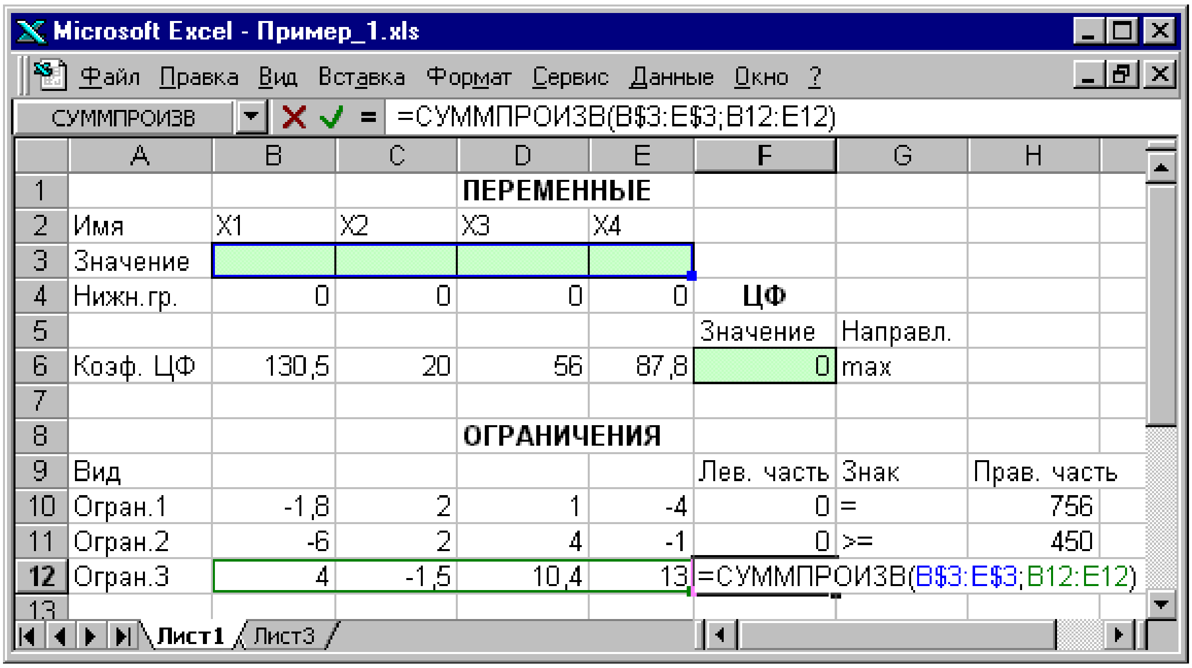


Рис. 3.5. Проверка правильности введения формулы в ячейку **F12** для левой части ограничения 3

Задание ЦФ

Дальнейшие действия производятся в окне **«Поиск решения»**, которое вызывается из меню **«Сервис»** (рис. 3.6):

- поставьте курсор в поле **«Установить целевую ячейку»**;

- введите адрес целевой ячейки **$*F*$6** или сделайте одно нажатие левой клавиши мыши на целевую ячейку в экранной форме - это будет равносильно вводу адреса с клавиатуры;
- введите направление оптимизации ЦФ, щелкнув один раз левой клавишей мыши по селекторной кнопке **«максимальному значению».**

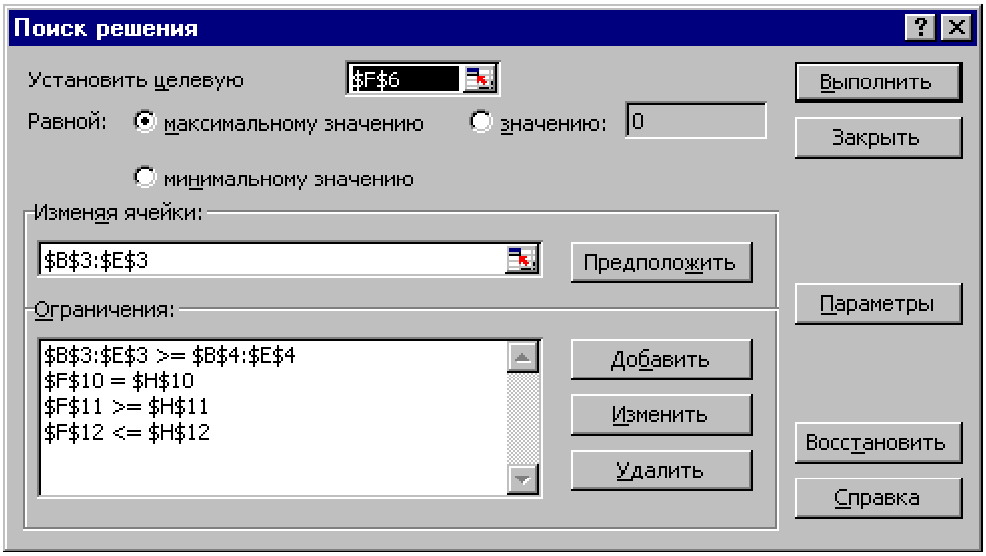


Рис. 3.6. Окно **«Поиск решения»** задачи (3.1)

Ввод ограничений и граничных условий

Задание ячеек переменных

В окно **«Поиск решения»** в поле **«Изменяя ячейки»** впишите адреса **$*B*$3:$*E*$3**. Необходимые адреса можно вносить в поле **«Изменяя ячейки»** и автоматически путем выделения мышью соответствующих ячеек переменных непосредственно в экранной форме.

Задание граничных условий для допустимых значений переменных

В нашем случае на значения переменных накладывается только граничное условие неотрицательности, то есть их нижняя граница должна быть равна нулю (см. рис. 3.1).

- Нажмите кнопку **«Добавить»**, после чего появится окно «**Добавление ограничения**» (рис.3.7).

- В поле **«Ссылка на ячейку»** введите адреса ячеек переменных **$B$3:$E$3**. Это можно сделать как с клавиатуры, так и путем выделения мышью всех ячеек переменных непосредственно в экранной форме.
- В поле знака откройте список предлагаемых знаков и выберите $\geq$.
- В поле **«Ограничение»** введите адреса ячеек нижней границы значений переменных, то есть **$*B*$4:$*E*$4**. Их также можно ввести путем выделения мышью непосредственно в экранной форме.

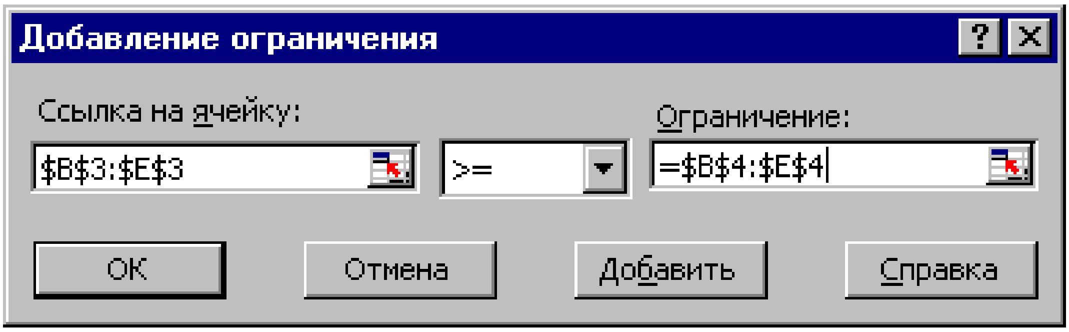


Рис. 3.7. Добавление условия неотрицательности переменных задачи (3.1)

Задание знаков ограничений $\leq$, $\geq$, =

- Нажмите кнопку **«Добавить»** в окне **«Добавление ограничения»**.
- В поле **«Ссылка на ячейку»** введите адрес ячейки левой части конкретного ограничения, например **$*F*$10**. Это можно сделать как с клавиатуры, так и путем выделения мышью нужной ячейки непосредственно в экранной форме.
- В соответствии с условием задачи (3.1) выбрать в поле знака необходимый знак, например =.
- В поле **«Ограничение»** введите адрес ячейки правой части рассматриваемого ограничения, например **$*H*$10**.
- Аналогично введите ограничения: **$*F*$11>=$*H*$11, $*F*$12<=$*H*$12.**

- Подтвердите ввод всех перечисленных выше условий нажатием кнопки **ОК.**

Окно **«Поиск решения»** после ввода всех необходимых данных задачи (3.1) представлено на рис. 3.6.

Если при вводе условия задачи возникает необходимость в изменении или удалении внесенных ограничений или граничных условий, то это делают, нажав кнопки **«Изменить»** или **«Удалить»** (см. рис. 3.6).

РЕШЕНИЕ ЗАДАЧИ

Установка параметров решения задачи

Задача запускается на решение в окне **«Поиск решения».** Но предварительно для установления конкретных параметров решения задач оптимизации определенного класса необходимо нажать кнопку **«Параметры»** и заполнить некоторые поля окна **«Параметры поиска решения»** (рис. 3.8).

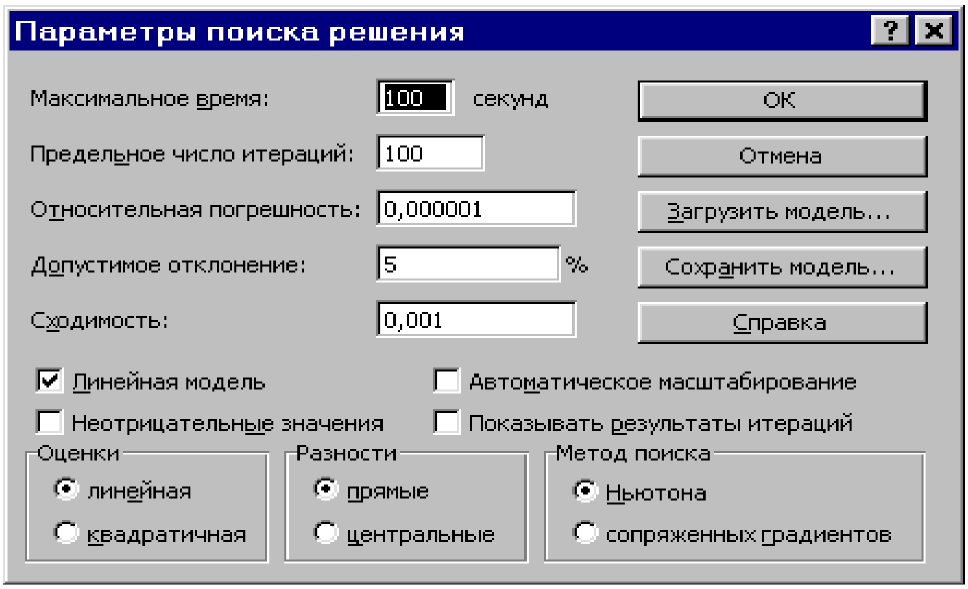


Рис. 3.8. Параметры поиска решения, подходящие для большинства задач ЛП

Параметр **«Максимальное время»** служит для назначения времени (в секундах), выделяемого на решение задачи. В поле можно ввести время, не превышающее 32 767 секунд (более 9 часов).

Параметр **«Предельное число итераций»** служит для управления временем решения задачи путем ограничения числа промежуточных вычислений. В поле можно ввести количество итераций, не превышающее 32 767.

Параметр **«Относительная погрешность»** служит для задания точности, с которой определяется соответствие ячейки целевому значению или приближение к указанным границам. Поле должно содержать число из интервала от 0 до 1. Чем *меньше* количество десятичных знаков во введенном числе, тем *ниже* точность. Высокая точность увеличит время, которое требуется для того, чтобы сошелся процесс оптимизации.

Параметр **«Допустимое отклонение»** служит для задания допуска на отклонение от оптимального решения в целочисленных задачах. При указании большего допуска поиск решения заканчивается быстрее.

Параметр **«Сходимость»** применяется только при решении нелинейных задач.

Установка флажка **«Линейная модель»** обеспечивает ускорение поиска решения линейной задачи за счет применение симплекс-метода.

Подтвердите установленные параметры нажатием кнопки **«ОК»**.

Запуск задачи на решение

Запуск задачи на решение производится из окна **«Поиск решения»** путем нажатия кнопки **«Выполнить».**

После запуска на решение задачи ЛП на экране появляется окно **«Результаты поиска решения»** с одним из сообщений, представленных на рис. 3.9, 3.10 и 3.11.

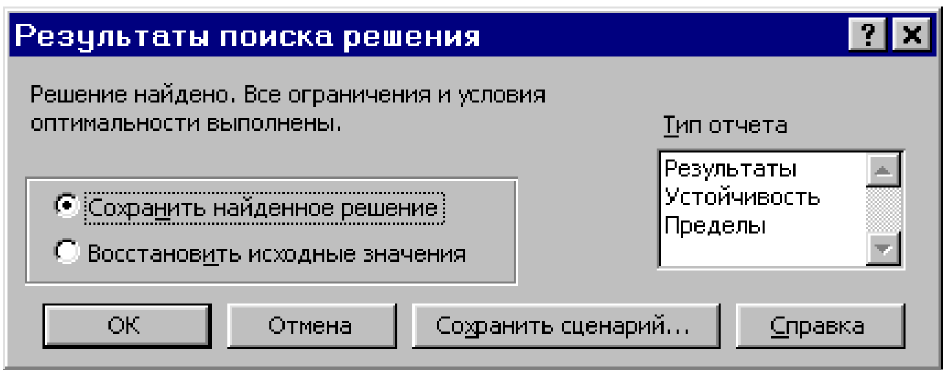


Рис. 3.9. Сообщение об успешном решении задачи

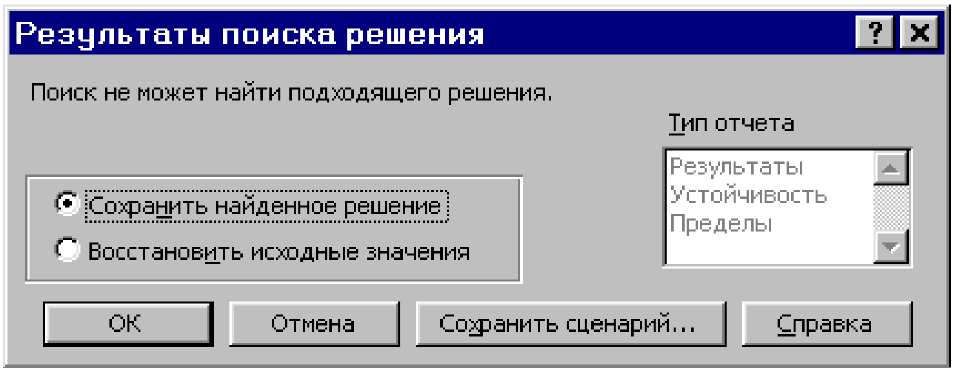


Рис. 3.10. Сообщение при несовместной системе ограничений задачи

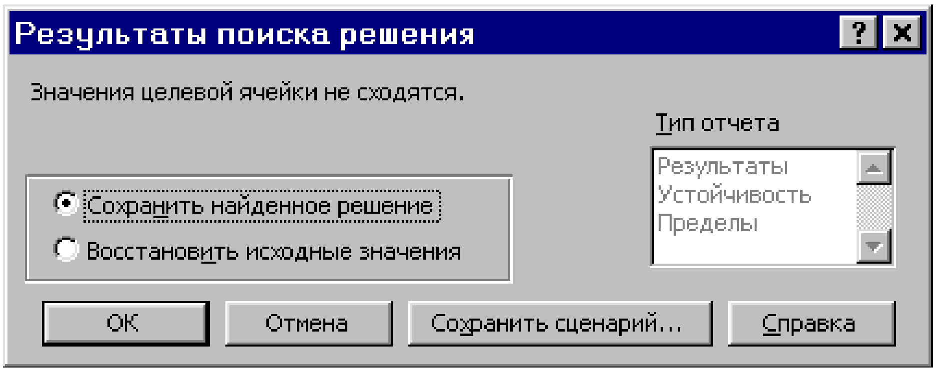


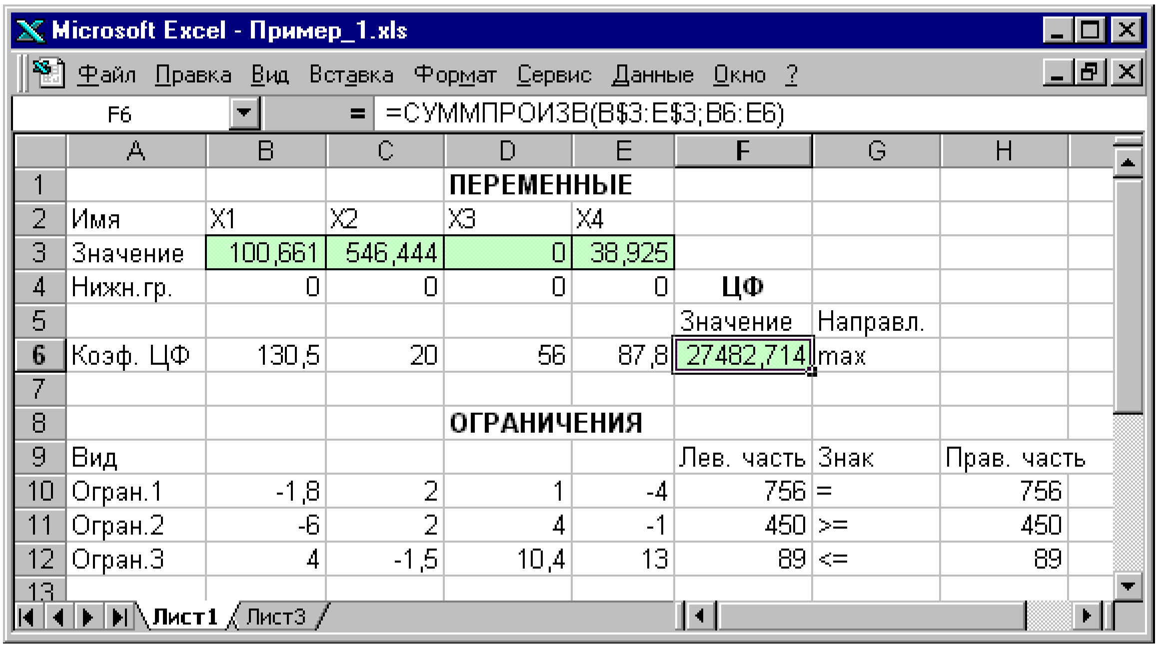
Рис. 3.11. Сообщение при неограниченности ЦФ в требуемом направлении

Иногда сообщения, представленные на рис. 3.10 и 3.11, свидетельствуют не о характере оптимального решения задачи, а о том, что при вводе условий задачи в Excel были допущены **ошибки**, не позволяющие Excel найти оптимальное решение, которое в действительности существует.

Если при решении задачи ЛП выдается сообщение о невозможности нахождения решения, то возможно, что причина заключается в ошибках ввода условия задачи в Excel. Поэтому, прежде чем делать вывод о принципиальной невозможности нахождения оптимального решения задачи, проверьте: не было ли допущено ошибок по ходу выполнения задания?

В том случае, когда при заполнении полей окна **«Поиск решения»** были допущены ошибки, не позволяющие Excel применить симплекс-метод для решения задачи или довести ее решение до конца, то после запуска задачи на решение на экран будет выдано соответствующее сообщение с указанием причины, по которой решение не найдено. Иногда слишком малое значение параметра **«Относительная погрешность»** не позволяет найти оптимальное решение. Для исправления этой ситуации увеличивайте погрешность поразрядно, например от 0,000001 до 0,00001 и т.д.

В окне **«Результаты поиска решения»** представлены названия трех типов отчетов: **«Результаты», «Устойчивость», «Пределы»**. Они необходимы при анализе полученного решения на чувствительность. Для получения же ответа (значений переменных, ЦФ и левых частей ограничений) прямо в экранной форме просто нажмите кнопку **«ОК».** После этого в экранной форме появляется оптимальное решение задачи (рис. 3.12).


Microsoft Excel - Пример_1.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно ?

F6 = =СУММПРОИЗВ(B$3:E$3;B6:E6)

	A	B	C	D	E	F	G	H
1				ПЕРЕМЕННЫЕ				
2	Имя	X1	X2	X3	X4			
3	Значение	100,661	546,444	0	38,925			
4	Нижн.гр.	0	0	0	0	ЦФ		
5						Значение	Направл.	
6	Коэф. ЦФ	130,5	20	56	87,8	27482,714	max	
7								
8				ОГРАНИЧЕНИЯ				
9	Вид					Лев. часть	Знак	Прав. часть
10	Огран.1	-1,8	2	1	-4	756	=	756
11	Огран.2	-6	2	4	-1	450	>=	450
12	Огран.3	4	-1,5	10,4	13	89	<=	89

Лист1 / Лист3

Рис. 3.12. Экранная форма задачи (3.1) после получения решения

ЦЕЛОЧИСЛЕННОЕ ПРОГРАММИРОВАНИЕ

Допустим, что к условию задачи (3.1) добавилось требование целочисленности значений всех переменных. В этом случае описанный выше процесс ввода условия задачи необходимо *дополнить* следующими шагами.

- В экранной форме укажите, на какие переменные накладывается требование целочисленности (этот шаг делается для наглядности восприятия условия задачи) (рис. 3.13).
- В окне «Поиск решения» (меню «Сервис»→«Поиск решения»), нажмите кнопку **«Добавить»** и в появившемся окне

«Добавление ограничений» введите ограничения следующим образом (рис. 3.14):

- в поле **«Ссылка на ячейку»** введите адреса ячеек переменных задачи, то есть **$*B*$3:$*E*$3**;

- в поле ввода знака ограничения установите **«целое»*;***

- подтвердите ввод ограничения нажатием кнопки «**OK».**

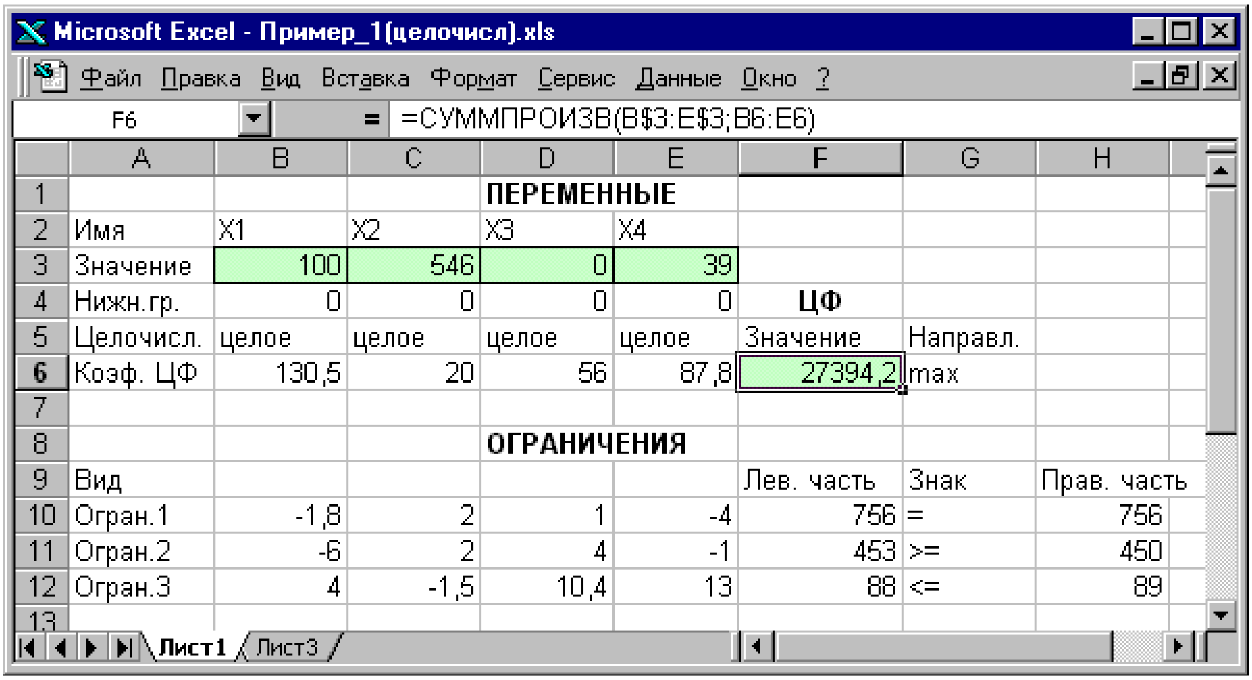


Рис. 3.13. Решение задачи (3.1) при условии целочисленности ее переменных

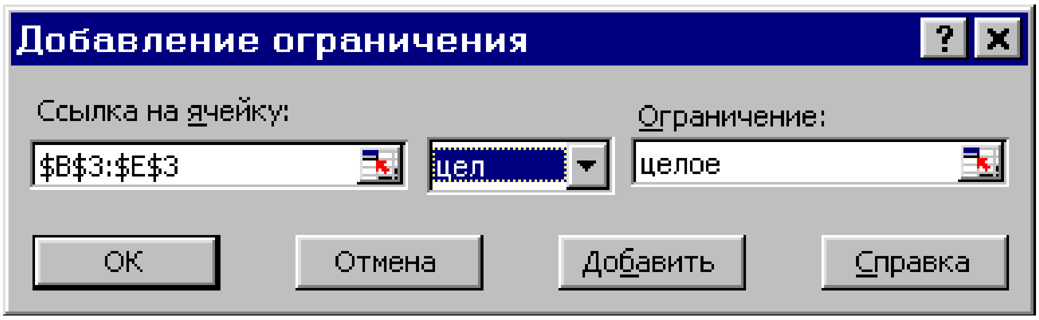


Рис. 3.14. Ввод условия целочисленности переменных задачи (3.1)

На рис. 3.13 представлено решение задачи (3.1), к ограничениям которой добавлено условие целочисленности значений ее переменных.

ЗАДАНИЯ

Используя MS Excel, найти решение для модели ЛП, соответствующей заданному варианту (табл.3.2).

Таблица 3.2

Варианты задач

1	$L(X) = 4x_1 + 6x_2 - 14x_3 + 49x_5 \rightarrow min;$ $\begin{cases} 21x_1 + 9x_2 - 2x_4 - 12x_5 \geq 58, \\ 110x_2 - 60x_3 + 80x_4 - 45x_5 = 290, \\ 5x_2 + 27x_3 - 14x_4 + x_5 \leq 72, \\ 87x_1 - 6{,}4x_2 + 130x_4 = 140, \\ x_j \geq 0 \left(j = \overline{1{,}5}\right). \end{cases}$
2	$L(X) = -38x_1 + 60x_2 + x_3 + 4x_4 + 8x_5 \rightarrow max;$ $\begin{cases} 18x_1 + 4x_2 + 2x_3 - 12x_5 \leq 86, \\ 2x_2 + 19x_3 - 7x_4 + 10x_5 = 130, \\ 0{,}4x_1 + 3x_2 - 4{,}2x_3 + 2x_4 - 5x_5 \leq 34, \\ 2{,}1x_1 + 13x_2 - 20x_3 + 6x_4 = 18, \\ x_j \geq 0 \left(j = \overline{1{,}5}\right). \end{cases}$
3	$L(X) = 10x_1 + 40x_3 + 13x_4 + 56x_5 \rightarrow min;$ $\begin{cases} 7x_1 + 16x_3 + 5x_4 + 25x_5 \leq 600, \\ 8x_1 + 1{,}7x_2 - 0{,}5x_4 + 4{,}7x_5 = 890, \\ 6x_1 + 4x_3 - 7x_4 + 6{,}3x_5 \leq 270, \\ 84x_1 + 62x_2 + 80x_3 + 14x_5 \geq 2300, \\ x_j \geq 0 \left(j = \overline{1{,}5}\right). \end{cases}$
4	$L(X) = 84x_1 + 5{,}7x_2 + 10x_4 - 3x_5 \rightarrow max;$ $\begin{cases} 4x_1 + 8{,}5x_2 + 16x_3 + 10x_5 \geq 50, \\ 10{,}4x_1 + 6x_3 + 2x_4 + 4x_5 \leq 120, \\ 19x_1 + 18x_2 - 20x_4 + 30x_5 = 600, \\ 200x_1 + 45x_2 - 8x_3 + 3{,}4x_4 \geq 210, \\ x_j \geq 0 \left(j = \overline{1{,}5}\right). \end{cases}$

5	$L(X)=0{,}84x_2-4x_3+3{,}8x_4+12x_5\to min;$ $\begin{cases}15x_1+9{,}6x_2+34x_4-8x_5\le 180,\\ 0{,}6x_1+11{,}1x_2-2{,}6x_3+1{,}5x_4-6{,}3x_5=68,\\ 14x_1+64x_3-38x_4+12x_5\le 81,\\ 190x_1-148x_2-7x_3+84x_5\ge 230,\\ x_j\ge 0\left(j=\overline{1,5}\right).\end{cases}$

Практическое занятие №4
Тема: «Введение в эконометрику»

Задание: составить кроссворд на тему: «Введение в эконометрику»

Практическое занятие №5.

Тема: «Модель линейной парной и множественной регрессии»

Вариант 1.

По данным 14 **наблюдений (см. табл.**5.4**) постройте модель парной линейной регрессии зависимости расходов на продовольственные товары на душу населения (тыс.руб.) от денежных доходов на душу населения (млн.руб.). Оцените качество построенной модели и рассчитайте прогноз расходов на продовольственные товары, если денежные доходы составляют** 130% **от их максимального значения.**

Таблица 5.1

№ п.п.	Расходы на продовольственные товары на душу населения, тыс. руб.	Денежные доходы на душу населения, млн. руб.
i	*y*	*x*
1	301,5	6,3
2	284,3	7,3
3	243,1	5,1
4	312,3	6,4
5	329,1	9,4
6	315,8	8,8
7	281,7	7,1
8	282,3	6
9	201,5	4,3
10	367,6	9,8
11	300	7,3
12	312,6	7,6
13	328,4	8,3
14	348,3	9,2

5

Вариант 2.

По данным 17 **наблюдений** (**табл**. 5.2) **постройте уравнение парной линейной регрессии зависимости доли расходов на товары длительного пользования** (%) **от дохода** (**млн**. **руб**.). **Оцените качество построенной модели и рассчитайте прогноз доли расходов на товары длительного пользования**, **если доход будет составлять** 85% **от его максимального уровня**.

Таблица 5.2

№ п.п.	**Доля расходов на товары длительного пользования**, %	**Доход**, **млн руб**.
i	*y*	*x*
1	11,1	1,5
2	10,9	1,2
3	13,7	1,7
4	13,6	2
5	15,3	2,4
6	14,4	1,9
7	14,9	2,6
8	17,6	3,1
9	14,9	2,3
10	15,2	2,1
11	13,8	1,9
12	12,1	1,5
13	15,1	2,4
14	13,8	2,1
15	13,3	2
16	15,8	2,7
17	13,8	1,7

Вариант 3.

По данным 18 **однородных предприятий** (**табл**. 5.3) **известны цена товара** (**тыс**. **руб.**) **и спрос** (**тыс**. **шт**.). **Постройте модель парной линейной регрессии и рассчитайте прогноз спроса**, **если цена товара будет на** 17% **меньше ее среднего уровня**.

Таблица 5.3

№ п.п.	**Спрос на товар**, **тыс.шт**.	**Цена товара**, **тыс.руб**.
i	*y*	*x*
1	3,7	220
2	7,2	115
3	8,6	105
4	4,3	180
5	2,8	265
6	2,1	304
7	3,1	245

6

i	*y*	*x*
8	2,9	250
9	2,5	267
10	3,8	202
11	4,4	176
12	3,1	248
13	3,3	229
14	3,1	237
15	4,4	176
16	2,9	276
17	7,5	114
18	2,4	300

1. Основные определения и формулы

Парная регрессия – регрессия (связь) между двумя переменными y и x, т.е. модель вида: $y = f(x) + \varepsilon$,

где y – зависимая переменная (результативный признак);

x – независимая объясняющая переменная (признак-фактор);

ε – возмущение или стохастическая переменная, включающая влияние неучтенных в модели факторов.

Практически в каждом отдельном случае величина y складывается из двух слагаемых: $y = \hat{y} + \varepsilon$,

где y – фактическое значение результативного признака;

$\hat{y}$ – теоретическое значение результативного признака, найденное исходя из уравнения регрессии. Знак «^» означает, что между переменными x и y нет строгой функциональной зависимости.

Различают *линейные* и *нелинейные* регрессии.

Линейная регрессия описывается уравнением прямой $\hat{y} = a + bx$.

Нелинейные регрессии делятся на два класса:

1) регрессии, *нелинейные по объясняющим переменным, но линейные по оцениваемым параметрам*, например:

• полиномы разных степеней $\hat{y} = a + b_1 x + b_2 x^2 + b_3 x^3$;

• равносторонняя гипербола $\hat{y} = a + \dfrac{b}{x}$;

2) регрессии, *нелинейные по оцениваемым параметрам*, например:

• □ степенная $\hat{y} = ax^b$;

• показательная $\hat{y} = ab^x$;

• □ экспоненциальная $\hat{y} = e^{a+bx}$.

Для построения парной линейной регрессии вычисляют вспомогательные величины (n – число наблюдений).

Выборочные средние: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ и $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Выборочная ковариация между x и y:

$$Cov(x,y) = \overline{yx} - \bar{y}\cdot\bar{x} \text{ или } Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

Ковариация – это числовая характеристика совместного распределения двух случайных величин.

Выборочная дисперсия для x:

$$Var(x) = \sigma_x^2 = \overline{x^2} - \bar{x}^2 \text{ или } Var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Выборочная дисперсия для y:

$$Var(y) = \sigma_y^2 = \overline{y^2} - \bar{y}^2 \text{ или } Var(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$

Выборочная дисперсия характеризует степень разброса значений случайной величины вокруг среднего значения (вариабельность, изменчивость).

Тесноту связи изучаемых явлений оценивает **выборочный коэффициент корреляции** между x и y:

$$r_{xy} = \frac{\overline{yx} - \bar{y}\cdot\bar{x}}{\sigma_x\sigma_y} = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}.$$

Коэффициент корреляции изменяется в пределах от -1 до +1. Чем ближе от по модулю к 1, тем ближе статистическая зависимость между x и y к линейной функциональной.

Если r_{xy}=0, то <u>линейная</u> связь между x и y отсутствует; $|r_{xy}|$<0,3 – связь слабая; $0,3 \le |r_{xy}| < 0,7$ – связь умеренная; $0,7 \le |r_{xy}| < 0,9$ – связь сильная; $0,9 \le |r_{xy}|$ $< 0,99$ – связь весьма сильная.

Положительное значение коэффициента свидетельствует о том, то связь между признаками прямая (с ростом x увеличивается значение y), отрицательное значение – связь обратная (с ростом x значение y уменьшается).

Построение линейной регрессии $\hat{y} = a + bx$ сводится к оценке ее параметров a и b. Классический подход к оцениванию параметров линейной регрессии основан на *методе наименьших квадратов* (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических $\hat{y}$ минимальна, т.е.

$$\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n} \varepsilon_i^2 \to \min.$$

Для линейной регрессии параметры a и b находятся из системы нормальных уравнений:

$$\begin{cases} na + b\sum x = \sum y, \\ a\sum x + b\sum x^2 = \sum yx. \end{cases}$$

Решая систему, находим **выборочный коэффициент линейной регрессии** y на x:

$$b = \frac{\overline{yx} - \overline{y} \cdot \overline{x}}{\overline{x^2} - \overline{x}^2} = \frac{Cov(x, y)}{Var(x)}$$

и параметр $a = \overline{y} - b\overline{x}$.

Коэффициент b при факторной переменной x показывает, насколько изменится в среднем величина y при изменении фактора x на единицу измерения.

Параметр $a = y$, когда $x = 0$. Если x не может быть равен 0, то a не имеет экономического смысла. Интерпретировать можно только знак при a: если $a > 0$, то относительное изменение результата происходит медленнее, чем изменение фактора, т.е. вариация результата меньше вариации фактора и наоборот.

Для оценки качества построенной модели регрессии можно использовать *коэффициент детерминации* либо *среднюю ошибку аппроксимации*.

Коэффициент детерминации

$$R^2 = r_{xy}^2 \text{ или } R^2 = b^2 \frac{Var(x)}{Var(y)} = \frac{Var(\hat{y})}{Var(y)}$$

показывает долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака y. Соответственно, величина $(1 - R^2)$ характеризует долю дисперсии показателя y, вызванную влиянием неучтенных в модели факторов и прочих причин.

Чем ближе R^2 к 1, тем качественнее регрессионная модель, т.е. построенная модель хорошо аппроксимирует исходные данные.

Средняя ошибка аппроксимации – это среднее относительное отклонение теоретических значений $\hat{y}$ от фактических y, т.е.

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

Построенное уравнение регрессии считается удовлетворительным, если значение $\overline{A}$ не превышает 10-12%.

Для линейной регрессии **средний коэффициент эластичности** находится по формуле:

$$\overline{Э} = f'(x) \frac{\overline{x}}{\overline{y}} = b \frac{\overline{x}}{\overline{y}}.$$

Средний коэффициент эластичности показывает на сколько процентов в среднем по совокупности изменится результат y от своей величины при изменении фактора x на 1% от своего значения.

Оценка значимости уравнения регрессии в целом дается с помощью F-критерия Фишера, который заключается в проверке гипотезы о статистической незначимости уравнения регрессии. Для этого выполняется сравнение *фактического* $F_{факт}$ и *критического* (табличного) $F_{табл}$ значений F-критерия Фишера.

$F_{факт}$ определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы, т.е.

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

$F_{табл}(\alpha; k_1; k_2)$ – максимально возможное значение критерия под влиянием случайных факторов при степенях свободы k_1=1, $k_2 = n$-2 и уровне значимости α. $F_{табл}$ находится из таблицы F-критерия Фишера (таблица 1 приложения).

Уровень значимости α – это вероятность отвергнуть правильную гипотезу при условии, что она верна.

Если $F_{табл} < F_{факт}$, то гипотеза об отсутствии связи изучаемого показателя с фактором отклоняется и делается вывод о существенности этой связи с уровнем значимости α (т.е. уравнение регрессии значимо).

Если $F_{табл} > F_{факт}$, то гипотеза принимается и признается статистическая незначимость и ненадежность уравнения регрессии.

Для линейной регрессии **значимость коэффициентов регрессии** оценивают с помощью t-критерия Стьюдента, согласно которому выдвигается гипотеза о случайной природе показателей, т.е. о незначимом их отличии от нуля. Далее рассчитываются фактические значения критерия $t_{факт}$ для каждого из оцениваемых коэффициентов регрессии, т.е.

$$t_b = \frac{b}{m_b}, \qquad t_a = \frac{a}{m_a},$$

где m_b и m_a – **стандартные ошибки** параметров линейной регрессии определяются по формулам:

$$m_b = \sqrt{\frac{\sum(y-\hat{y})^2}{(n-2)\sum(x-\bar{x})^2}} = \frac{S_{ост}}{\sigma_x\sqrt{n}},$$

$$m_a = \sqrt{\frac{\sum(y-\hat{y})^2}{(n-2)} \cdot \frac{\sum x^2}{n\sum(x-\bar{x})^2}} = \frac{S_{ост}}{n\sigma_x}\sqrt{\sum x^2}.$$

$t_{табл}(1-\alpha;k)$ – максимально возможное значение критерия Стьюдента под влиянием случайных факторов при данной степени свободы $k=n$-2 и уровне значимости α находится из таблицы критерия Стьюдента (таблица 2 приложения).

Если $t_{табл} < |t_{факт}|$, то гипотеза о несущественности коэффициента регрессии отклоняется с уровнем значимости α, т.е. коэффициент (a или b) не случайно отличается от нуля и сформировался под влиянием систематически действующего фактора x.

Если $t_{табл} > |t_{факт}|$, то гипотеза не отклоняется и признается случайная природа формирования параметра.

Значимость линейного коэффициента корреляции также проверяется с помощью t-критерия Стьюдента, т.е.

$$t_r = \frac{r_{xy}}{m_r} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}; \qquad m_r = \sqrt{\frac{1-r_{xy}^2}{n-2}}.$$

Гипотеза о несущественности коэффициента корреляции отклоняется с уровнем значимости α, если $t_{табл} < |t_r|$.

<u>*Замечание.*</u> Для линейной парной регрессии проверки гипотез о значимости коэффициента b и коэффициента корреляции r_{xy} равносильны проверке гипотезы о существенности уравнения регрессии в целом, т.е. $t_b = t_r = \sqrt{F}$.

Для расчета доверительного интервала определяют **предельную ошибку** для каждого показателя, т.е. $\Delta_a = t_{табл} m_a$, $\Delta_b = t_{табл} m_b$.

Доверительные интервалы для коэффициентов линейной регрессии:

$$a-\Delta_a < \gamma_a < a+\Delta_a, \; b-\Delta_b < \gamma_b < b+\Delta_b.$$

Если в границы доверительного интервала попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, т.к. он не может одновременно принимать и положительное, и отрицательное значения.

Прогнозное значение y_p определяется путем подстановки в уравнение регрессии $\hat{y} = a + bx$ соответствующего прогнозного значения x_p. Затем вычисляется **средняя стандартная ошибка прогноза**

$$m_{y_p} = \sigma_{ост}\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot Var(x)}}, \quad \text{где } \sigma_{ост} = \sqrt{\frac{\sum (y - \hat{y})^2}{(n-2)}};$$

и строится **доверительный интервал прогноза**

$$y_p - t_{табл} m_{y_p} < \gamma_{y_p} < y_p + t_{табл} m_{y_p}.$$

Интервал может быть достаточно широк за счет малого объема наблюдений.

Регрессии, нелинейные по включенным переменным, приводятся к линейному виду простой заменой переменных, а дальнейшая оценка параметров производится с помощью МНК.

Гиперболическая регрессия: $\hat{y} = a + b/x$.

Линеаризующее преобразование: $x' = 1/x; \quad y' = y$.

$$a = \frac{1}{n}\sum y - \frac{1}{n} b \sum \frac{1}{x}; \quad b = \frac{n\sum \frac{y}{x} - \sum \frac{1}{x}\sum y}{n\sum \frac{1}{x^2} - \left(\sum \frac{1}{x}\right)^2}.$$

Регрессии, нелинейные по оцениваемым параметрам, делятся на два типа: *внутренне нелинейные* $\hat{y} = a + bx^c$, $\hat{y} = a\left(1 - \frac{1}{1 - x^b}\right)$ и т.п. (к линейному виду не приводятся) и *внутренне линейные* (приводятся к линейному виду с помощью соответствующих преобразований), например:

Экспоненциальная регрессия: $\hat{y} = e^{a+bx}$.

Линеаризующее преобразование: $x' = x; \quad y' = \ln y$.

$$a=\frac{1}{n}\sum\ln y-\frac{1}{n}b\sum x; \qquad b=\frac{n\sum x\ln y-\sum x\sum\ln y}{n\sum x^2-\left(\sum x\right)^2}.$$

Степенная регрессия: $\hat{y}=ax^b, \quad (a>0).$

Линеаризующее преобразование: $x'=\ln x; \quad y'=\ln y.$

$$\ln a=\frac{1}{n}\sum\ln y-\frac{1}{n}b\sum x; \qquad b=\frac{n\sum(\ln x\cdot\ln y)-\sum\ln x\sum\ln y}{n\sum(\ln x)^2-\left(\sum\ln x\right)^2}.$$

Показательная регрессия: $\hat{y}=a\cdot b^x.$

Линеаризующее преобразование: $x'=x; \quad y'=\ln y.$

$$\ln a=\frac{1}{n}\sum\ln y-\frac{1}{n}\ln b\sum x; \qquad \ln b=\frac{n\sum x\ln y-\sum x\sum\ln y}{n\sum x^2-\left(\sum x\right)^2}.$$

Логарифмическая регрессия: $\hat{y}=a+b\ln x.$

Линеаризующее преобразование: $x'=\ln x; \quad y'=y.$

$$a=\frac{1}{n}\sum y-\frac{1}{n}b\sum\ln x; \qquad b=\frac{n\sum y\ln x-\sum\ln x\sum y}{n\sum(\ln x)^2-\left(\sum\ln x\right)^2}.$$

2. Решение типовых задач

__Пример 1.__ По 15 предприятиям отрасли (табл. 1.1) известны: x – объем произведенной продукции (тыс. ед.) и y – затраты на выпуск этой продукции (тыс. ден. ед.). Необходимо:

1) определить зависимость y от x;

2) построить корреляционные поля и график уравнения линейной регрессии y на x;

3) сделать вывод о качестве модели и рассчитать прогнозное значение y при прогнозном значении x, составляющем 107% от среднего уровня.

Таблица 5.3

i	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
x	2,7	1,5	8,2	4,5	3,3	5,8	3,0	7,1	1,2	10,4	4,9	5,2	11,5	9,4	6,5
y	110	70	310	120	75	170	100	180	30	440	190	150	390	310	230

__Решение:__

1) В Excel составим вспомогательную таблицу 5.4.

Таблица 5.4

	A	B	C	D	E	F
1	i	x	y	xy	x^2	y^2
2	1	2,7	110	297	7,29	12100
3	2	1,5	70	105	2,25	4900
4	3	8,2	310	2542	67,24	96100
5	4	4,5	120	540	20,25	14400
6	5	3,3	75	247,5	10,89	5625
7	6	5,8	170	986	33,64	28900
8	7	3,0	100	300	9,00	10000
9	8	7,1	180	1278	50,41	32400
10	9	1,2	30	36	1,44	900
11	10	10,4	440	4576	108,16	193600
12	11	4,9	190	931	24,01	36100
13	12	5,2	150	780	27,04	22500
14	13	11,5	390	4485	132,25	152100
15	14	9,4	310	2914	88,36	96100
16	15	6,5	230	1495	42,25	52900
17	сумма	85,2	2875	21512,5	624,48	758625
18	среднее	5,68	191,667	1434,167	41,632	50575
19	$n=$	15				

Вычислим количество измерений n. Для этого в ячейку **B19** поместим =**СЧЁТ(A2:A16)**.

С помощью функции $\sum$ (Автосумма) на панели инструментов ***Стандартная*** найдем сумму всех x (ячейка **B17**) и y (ячейка **C17**).

Вычислим выборочные средние: $\overline{x}$=5,68; $\overline{y}$=191,67. Таким образом, средний объем произведенной продукции по 15 предприятиям отрасли составляет 5,68 тыс. ед., а средние затраты на выпуск этой продукции – 191,67 тыс. ден. ед.

Заполним столбцы **D**, **E**, **F**. Например, в ячейку **D2** поместим =**B2*C2**, затем на клавиатуре нажмем ENTER. Щелкнем левой кнопкой мыши по ячейке **D2** и, ухватив за правый нижний угол этой ячейки (черный плюсик), потянем вниз до ячейки **D16**. Произойдет автоматическое заполнение диапазона **D3 – D16**.

Для вычисления в**ыборочной ковариации** между x и y используем формулу $Cov(x,y)=\overline{yx}-\overline{y}\cdot\overline{x}$, т.е. в ячейку **B21** поместим =**D18-B18*C18** и получим 345,5 (рис. 1.1).

Выборочную дисперсию для x найдем по формуле $Var(x)=\overline{x^2}-\overline{x}^2$, т.е. в ячейку **B22** поместим =**E18-B18*B18** и получим 9,37 (рис. 5.1). Аналогично определяем $Var(y)$=13838,89.

	A	B	C	D	E	F	G	H	I	J
1	i	x	y	xy	x^2	y^2	y^	y-y^	|(y-y^)/y|	(y-y^)^2
2	1	2,7	110	297	7,29	12100	81,78	28,22	0,257	796,34
3	2	1,5	70	105	2,25	4900	37,53	32,47	0,464	1054,24
4	3	8,2	310	2542	67,24	96100	284,59	25,41	0,082	645,64
5	4	4,5	120	540	20,25	14400	148,15	-28,15	0,235	792,69
6	5	3,3	75	247,5	10,89	5625	103,91	-28,91	0,385	835,51
7	6	5,8	170	986	33,64	28900	196,09	-26,09	0,153	680,77
8	7	3,0	100	300	9,00	10000	92,84	7,16	0,072	51,23
9	8	7,1	180	1278	50,41	32400	244,03	-64,03	0,356	4099,66
10	9	1,2	30	36	1,44	900	26,47	3,53	0,118	12,47
11	10	10,4	440	4576	108,16	193600	365,71	74,29	0,169	5518,31
12	11	4,9	190	931	24,01	36100	162,90	27,10	0,143	734,17
13	12	5,2	150	780	27,04	22500	173,97	-23,97	0,160	574,41
14	13	11,5	390	4485	132,25	152100	406,28	-16,28	0,042	264,93
15	14	9,4	310	2914	88,36	96100	328,8401	-18,84	0,061	354,95
16	15	6,5	230	1495	42,25	52900	221,90	8,10	0,035	65,55
17	сумма	85,2	2875	21512,5	624,48	758625	2875	0,00	2,730	16480,86
18	среднее	5,68	191,667	1434,167	41,632	50575	191,667	0,00	0,182	1098,724
19	n=	15								
20			проверка				Результат выполнения функции ЛИНЕЙН			
21	$Cov(x,y)$=	345,5	345,5		$A_{ср.}$ =	18,20		36,87	-17,78	
22	$Var(x)$=	9,3696	9,3696		$Э_{ср.}$ =	1,0928		3,003	19,379	
23	$Var(y)$=	13838,89	13838,9		$S_{ост.}$ =	35,606		**0,9206**	35,606	
24	r_{xy} =	0,959482	0,95948		$F_{табл.}$ =	4,67		150,74	13	
25	b=	36,87457			$F_{факт.}$ =	150,74		191102,48	16480,9	
26	a=	-17,7809			$t_{табл.}$ =	2,1604				
27	$Var(y$^$)$=	-36736,1								
28	R^2 =	**0,920606**		Уравнение парной линейной регрессии				y^=-17,78+36,87x		

Рис. 5.1. Решение примера 4.3 в Excel

Выборочный коэффициент корреляции r_{xy}=0,96 очень высокий, что указывает на прямую и весьма сильную связь между x и y, т.е. с ростом объема произведенной продукции (x) увеличиваются затраты на выпуск этой продукции (y).

Выборочный коэффициент линейной регрессии $b = \dfrac{Cov(x, y)}{Var(x)}$=36,87; параметр $a = \bar{y} - b\bar{x}$=-17,78. Значит, уравнение парной линейной регрессии имеет вид $\hat{y}$=-17,78+36,87$\cdot x$.

Коэффициент b показывает, что при увеличении объема произведенной продукции (x) на 1 тыс. ед. затраты на выпуск этой продукции (y) в среднем увеличатся на 36,87 тыс. ден. ед.

2) Подставляя в найденное уравнение регрессии фактические значения x, определим теоретические (расчетные) значения $\hat{y}$ в столбце **G**.

С помощью *Мастера диаграмм* строим корреляционные поля (выделяя столбцы со значениями x и y) и уравнение линейной регрессии (выделяя столбцы со значениями x и $\hat{y}$). Выбираем тип диаграммы – ***Точечная*** и, следуя рекомендациям *Мастера диаграмм,* вводим нужные параметры (название, подписи к осям, легенду и т.п.). В результате получим рис. 5.2.

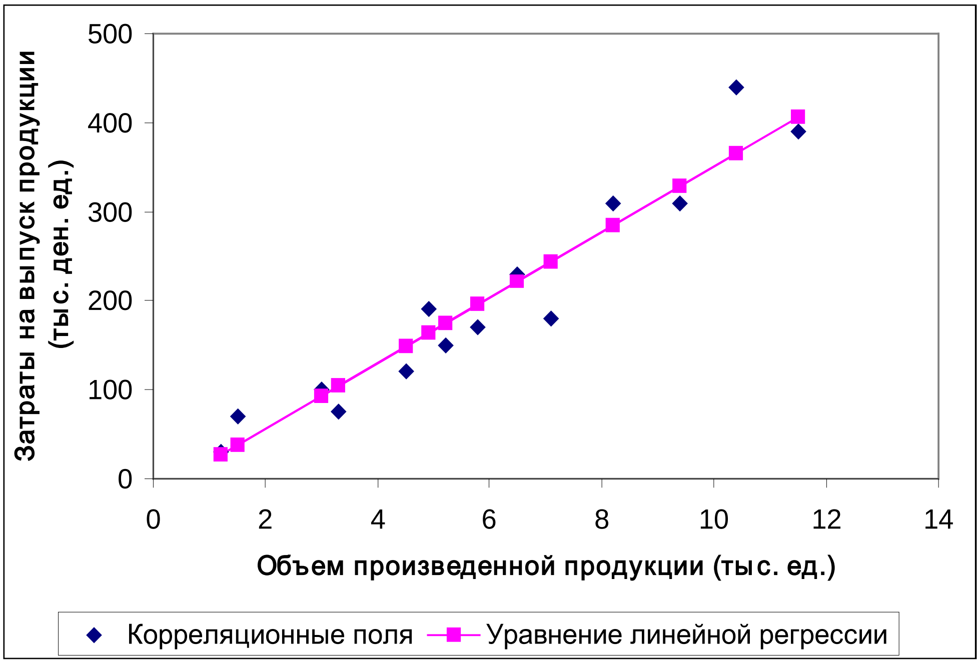


Рис. 5.2. График зависимости объема произведенной продукции от теоретических и фактических затрат на выпуск этой продукции

3) Для оценки качества построенной модели регрессии вычислим:

• **коэффициент детерминации** $R^2 = r_{xy}^2$=0,92, который показывает, что изменение затрат на выпуск продукции (y) на 92% объясняется изменением объема произведенной продукции (x), а 8% приходится на долю неучтенных в модели факторов, что указывает на качественность построенной регрессионной модели;

• **среднюю ошибку аппроксимации.** Для этого в столбце **H** вычислим

$\varepsilon = y - \hat{y},$ а в столбце **I** – выражение $\left|\dfrac{y-\hat{y}}{y}\right|$ (рис. 3.1). При умножении среднего значения (ячейка **I18**) на 100% получим $\overline{A} = 18{,}2\%$. Следовательно, в среднем теоретические значения $\hat{y}$ отклоняются от фактических y на 18,2%.

С помощью F-критерия Фишера оценим **значимость уравнения регрессии в целом**: $F_{факт} = \dfrac{R^2}{1-R^2}(n-2) = 150{,}74.$

На уровне значимости 0,05 $F_{табл}$=4,67 определяем по таблице F-критерия Фишера (таблица 1 приложения) либо с помощью встроенной статистической функции **FРАСПОБР** (рис. 5.3).

Так как $F_{табл} < F_{факт}$, то уравнение регрессии значимо при α=0,05.

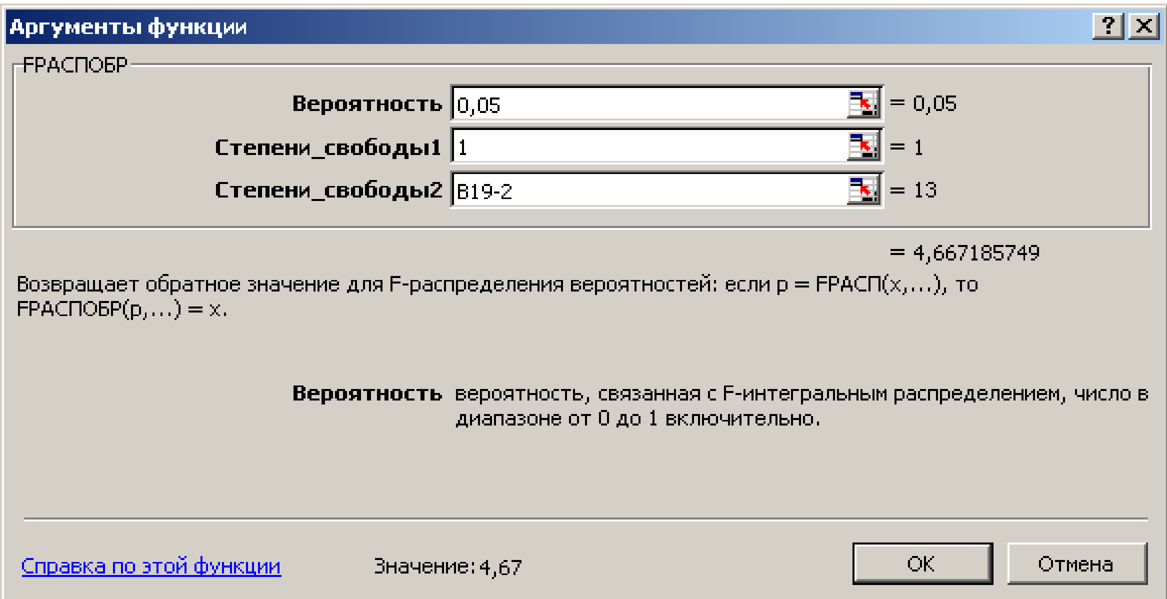


Рис. 5.3. Диалоговое окно функции **FРАСПОБР**

Средний коэффициент эластичности $\overline{Э}$=1,09 показывает, что с ростом объема произведенной продукции на 1% затраты на выпуск этой продукции в среднем по совокупности возрастут на 1,09%.

Рассчитаем **прогнозное значение** y_p путем подстановки в уравнение регрессии $\hat{y}$=-17,78+36,87$\cdot x$ прогнозного значения фактора $x_p = \overline{x} \cdot$1,07=6,08. Получим y_p =206,33. Следовательно, при выпуске продукции в количестве 6,08 тыс. ед. затраты на производство этой продукции составят 206,33 тыс. ден. ед.

Найдем $S_{ост} = \sigma_{ост}$=35,606 поместив в ячейку **F23 =КОРЕНЬ(J17/(B19-2))**.

Средняя стандартная ошибка прогноза:

$$m_{y_p} = 35{,}606 \cdot \sqrt{1 + \frac{1}{15} + \frac{(6{,}08 - 5{,}68)^2}{15 \cdot 9{,}37}} = 36{,}79.$$

На уровне значимости α=0,05 либо по таблице t-критерия Стьюдента (таблица 2 приложения) либо с помощью встроенной статистической функции **СТЬЮДРАСПОБР** определим $t_{табл}$=2,1604 и вычислим предельную ошибку прогноза, которая в 95% случаев не будет превышать $\Delta_{y_p} = t_{табл} m_{y_p}$ =79,48.

Доверительный интервал прогноза:

$$206{,}33\text{-}79{,}48 < \gamma_{y_p} < 206{,}33\text{+}79{,}48 \quad \text{или} \quad 126{,}85 < \gamma_{y_p} < 285{,}81.$$

Выполненный прогноз затрат на выпуск продукции оказался надежным (1-0,05=0,95), но неточным, так как диапазон верхней и нижней границ доверительного интервала составляет $D = \frac{285{,}81}{126{,}85}$=2,25 раза. Это произошло за счет малого объема наблюдений.

3. Решение задач с помощью электронных таблиц Excel

Решение эконометрических задач можно упростить, используя встроенные функции. Активизируем *Мастер функций* любым из способов:

• в главном меню выбрать ***Вставка / Функция***;

• на панели инструментов ***Стандартная*** щелкнуть по кнопке f_x ***Вставка функции***.

Для вычисления **выборочных средних** используем функцию **СРЗНАЧ(число1:числоN)** из категории ***Статистические***.

Выборочная ковариация между x и y находится с помощью функции **КОВАР(массив X;массив Y)** из категории ***Статистические***.

Выборочные дисперсии определяются статистической функцией **ДИСПР(число1:числоN)**.

Выборочный коэффициент корреляции между x и y вычисляется с помощью статистической функции **КОРРЕЛ(массив X;массив Y)**.

Параметры линейной регрессии $y = a + bx$ в Excel можно определить несколькими способами.

1 способ) С помощью встроенной функции **ЛИНЕЙН**. Порядок действий следующий:

1. Выделить область пустых ячеек 5х2 (5 строк, 2 столбца) для вывода результатов регрессионной статистики или область 1х2 – для получения только коэффициентов регрессии.

2. С помощью *Мастера функций* среди ***Статистических*** выбрать функцию **ЛИНЕЙН** и заполнить ее аргументы (рис. 5.4):

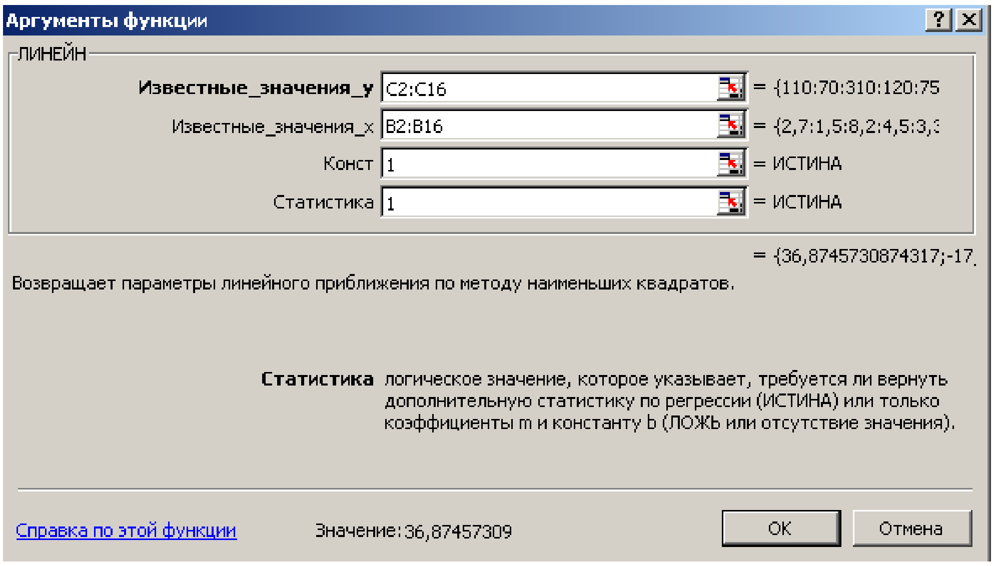


Рис. 5.4. Диалоговое окно ввода аргументов функции **ЛИНЕЙН**

Известные_значения_y – диапазон, содержащий данные результативного признака Y;

Известные_значения_x – диапазон, содержащий данные объясняющего признака X;

Конст – логическое значение (1 или 0), которое указывает на наличие или отсутствие свободного члена в уравнении; ставим 1;

Статистика – логическое значение (1 или 0), которое указывает, выводить дополнительную информацию по регрессионному анализу или нет; ставим 1.

3. В левой верхней ячейке выделенной области появится первое число таблицы. Для раскрытия всей таблицы нужно нажать на клавишу **<F2>**, а затем – на комбинацию клавиш **<CTRL>** + **<SHIFT>** + **<ENTER>**.

Дополнительная регрессионная статистика будет выведена в виде (табл. 2.3):

Таблица 5.4

Значение коэффициента b	Значение коэффициента a
Среднеквадратическое отклонение b	Среднеквадратическое отклонение a
Коэффициент детерминации R^2	Среднеквадратическое отклонение y
F -статистика $\left(F_{факт}\right)$	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов

В результате применения функции **ЛИНЕЙН** получим:

36,87	-17,78
3,003	19,379
0,9206	35,606
150,74	13
191102,48	16480,9

(2 способ) С помощью инструмента анализа данных ***Регрессия*** можно получить результаты регрессионной статистики, дисперсионного анализа, доверительные интервалы, остатки, графики подбора линий регрессии, графики остатков и нормальной вероятности. Порядок действий следующий:

1. Необходимо проверить доступ к *Пакету анализа*. Для этого в главном меню нужно выбрать ***Сервис / Настройки*** и напротив ***Пакета анализа*** установить флажок.

2. Выбрать в главном меню ***Сервис / Анализ данных / Регрессия*** и заполнить диалоговое окно (рис. 4.5):

Входной интервал Y – диапазон, содержащий данные результативного признака Y;

Входной интервал X – диапазон, содержащий данные объясняющего признака X;

Метки – флажок, который указывает, содержит ли первая строка названия столбцов или нет;

Константа-ноль – флажок, указывающий на наличие или отсутствие свободного члена в уравнении;

Выходной интервал – достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно задать произвольное имя нового листа, на который будут выведены результаты.

Для получения информации об остатках, графиков остатков, подбора и нормальной вероятности нужно установить соответствующие флажки в диалоговом окне. В результате получим итоги как на рис. 5.6.

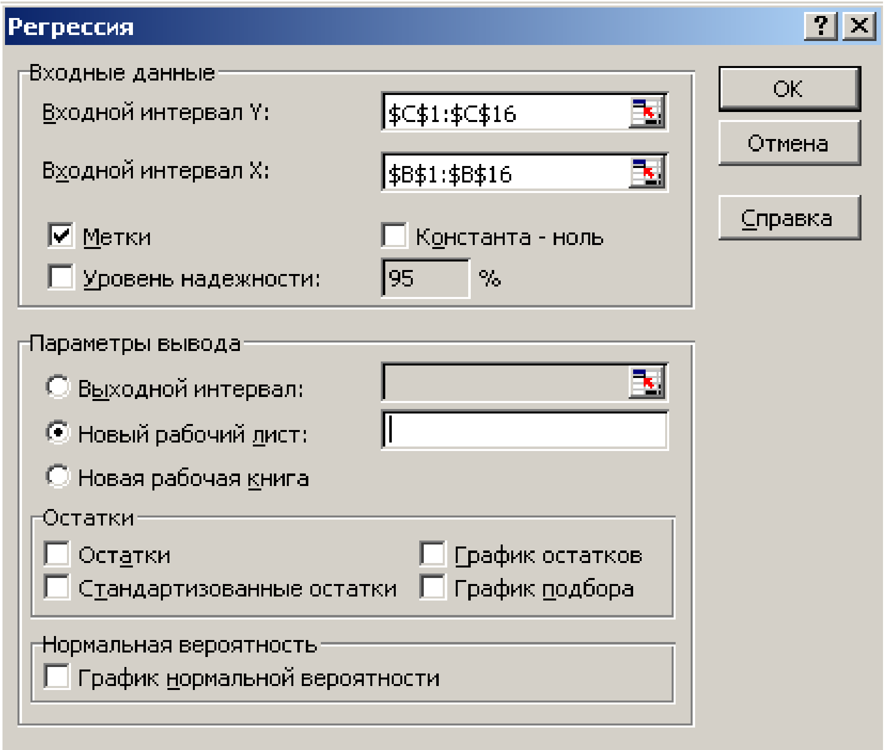


Рис.5.5. Диалоговое окно ввода параметров инструмента ***Регрессии***

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,9595
R-квадрат	0,9206
Нормированный R-квадрат	0,9145
Стандартная ошибка	35,6056
Наблюдения	15

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	191102,5	191102	150,74	1,59E-08
Остаток	13	16480,86	1267,76		
Итого	14	207583,3			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	-17,781	19,3787	-0,9175	0,3756	-59,6461	24,084	-59,6461	24,084
x	36,875	3,0034	12,2776	1,6E-08	30,3861	43,363	30,3861	43,363

Рис. 5.6. Результаты применения инструмента ***Регрессия***

В Excel **линия тренда** может быть добавлена в диаграмму с областями гистограммы или в график. Для этого:

1. Необходимо выделить область построения диаграммы и в главном меню выбрать ***Диаграмма / Добавить линию тренда***.

2. В появившемся диалоговом окне (рис. 5.7) выбрать вид линии тренда и задать соответствующие параметры.

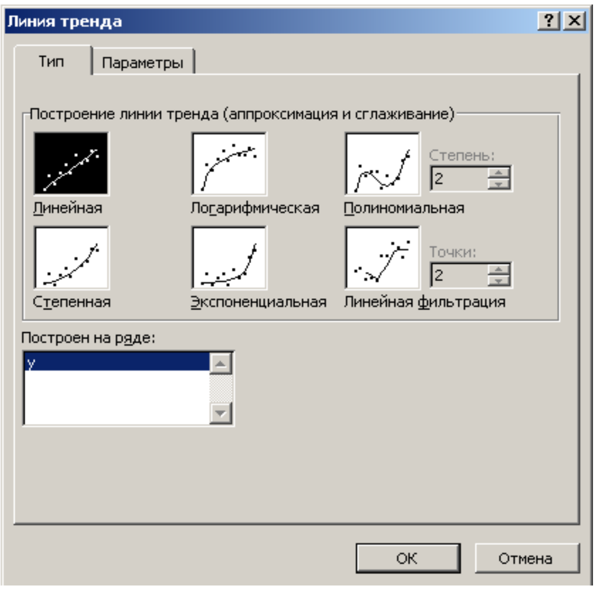


Рис. 5.7. Диалоговое окно типов линии тренда

Для полиноминального тренда необходимо задать степень аппроксимирующего полинома, для линейной фильтрации – количество точек усреднения.

Выбираем **Линейная** для построения уравнения линейной регрессии.

В качестве дополнительной информации можно *показать уравнение на диаграмме* и *поместить на диаграмму величину* R^2, установив соответствующие флажки на закладке ***Параметры*** (рис. 5.8).

В результате получим линейный тренд (рис. 5.9).

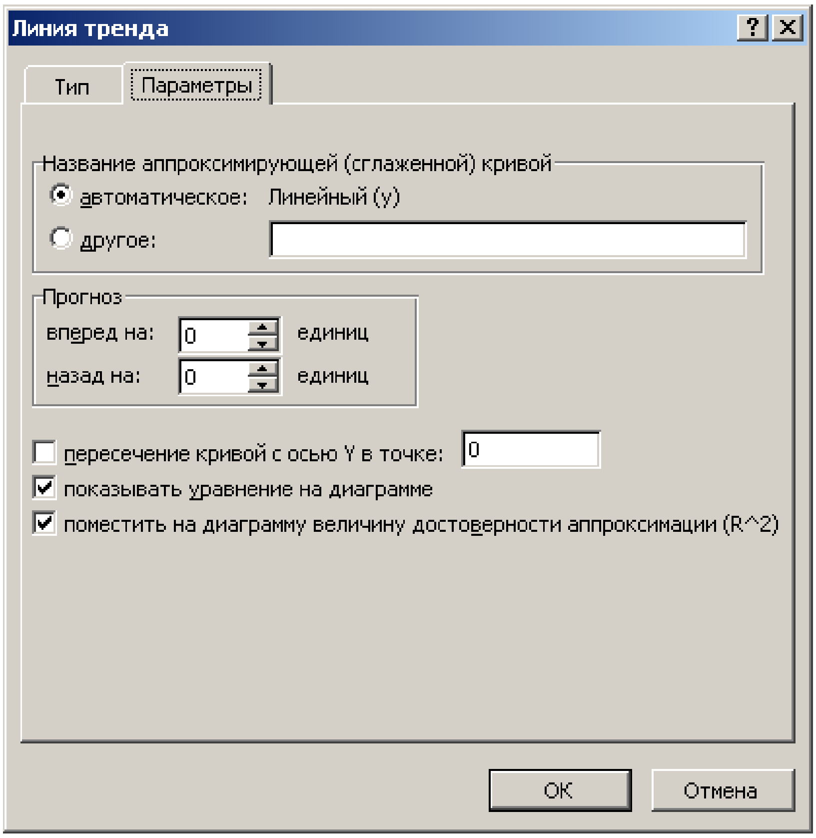


Рис. 5.8. Диалоговое окно параметров линии тренда

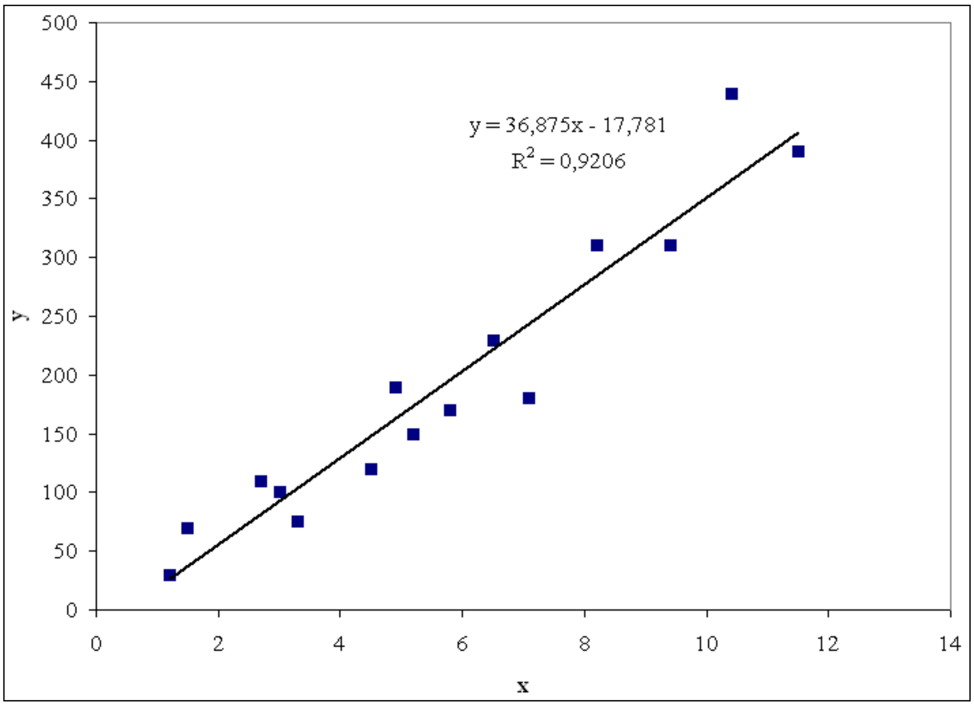


Рис. 5.9. Линейный тренд

Нелинейные модели регрессии иллюстрируются при вычислении параметров уравнения $\hat{y} = ab^x$ с применением выбранной в Excel статистической функции **ЛГРФПРИБЛ**. Порядок вычислений аналогичен применению функции ЛИНЕЙН.

Множественная регрессия и корреляция

Вариант 1

По данным 30 **наблюдений постройте модель множественной регрессии удовлетворительного качества** (**табл**. 5.5). **Рассчитайте прогноз результата, если прогнозные значения независимых факторов будут составлять** 112% **от их среднего уровня**.

Таблица 5.5

№	Валовой продукт, млн. руб.	Балансовая стоимость оборудования, млн. руб.	Объем промышленного производства, млн. руб.	Количество занятых, тыс.чел.
i	y	x_1	x_2	x_3
1	1365	3938	625	161
2	1398	3002	475	186
3	1969	6269	528	267
4	1000	2270	242	129
5	761	1879	197	100
6	1253	4709	407	149
7	1590	3976	702	144
8	1425	2946	521	153
9	1127	3151	369	128
10	1595	3424	447	178
11	1636	4110	645	207
12	2110	3910	717	171
13	1131	3045	299	99
14	2005	4575	464	166
15	768	2126	169	105
16	1682	4692	579	167
17	2146	5873	468	255
18	2865	6906	850	299
19	3133	8678	924	458
20	1706	3988	507	182
21	1456	3840	475	179
22	2616	6368	801	244
23	2657	6396	604	304
24	1538	3632	592	170
25	1249	2681	220	111
26	2960	6675	672	306
27	2255	5298	712	241
28	1423	3185	214	133
29	2488	7364	602	245
30	1426	3812	225	116

Вариант 2

По данным 30 **наблюдений постройте модель множественной регрессии удовлетворительного качества** (**табл**. 5.6). **Рассчитайте прогноз результата, если прогнозные значения независимых факторов будут составлять** 92% **от их среднего уровня**.

Таблица 5.6

№	Валовой продукт, млн. руб.	Балансовая стоимость оборудования, млн. руб.	Объем промышленного производства, млн. руб.	Количество занятых, тыс.чел.
i	y	x_1	x_2	x_3
1	4197	17125	1613	163
2	3562	14656	1273	177
3	5514	25701	1380	255
4	2710	10790	632	128
5	1975	10325	514	94
6	3209	13252	1073	142
7	4600	16397	2303	139
8	3478	12580	1395	146
9	2618	12854	907	122
10	3875	14790	1096	174
11	4148	16799	1745	201
12	4923	16228	1861	163
13	2690	11334	889	98
14	4756	18798	1595	160
15	1755	8672	477	103
16	4171	19553	1538	169
17	6059	23146	1260	247
18	6545	29168	2191	295
19	8782	40712	2475	452
20	4153	15059	1315	176
21	3920	17104	1210	174

i	y	x_1	x_2	x_3
22	6047	26266	2287	240
23	6811	30628	1598	292
24	3567	15688	1378	162
25	2581	11840	593	110
26	7677	27988	1621	289
27	6923	23448	2225	227
28	3348	12346	475	124
29	6432	31504	1873	237
30	2777	19119	559	123

Вариант 3

По данным 30 **наблюдений постройте модель множественной регрессии удовлетворительного качества** (**табл**. 5.7). **Рассчитайте прогноз результата, если прогнозные значения независимых факторов будут составлять** 106% **от их среднего уровня**.

Таблица 5.7

№	**Валовой продукт, млн. руб.**	**Балансовая стоимость оборудования, млн. руб.**	**Объем промышленного производства, млн. руб.**	**Количество занятых, тыс.чел.**
i	y	x_1	x_2	x_3
1	4987	45778	1932	162
2	4279	38461	1630	175
3	6721	70406	1853	253
4	3219	26535	818	129
5	2315	22163	642	93
6	4276	40832	1408	143
7	5021	41566	2649	138
8	4137	33209	1367	143
9	3355	36655	1149	122
10	4791	42714	1437	170
11	5115	50006	1987	202
12	6046	48984	2250	165
13	2989	30206	903	96
14	6104	49745	1856	158
15	2500	19844	594	105
16	5943	58668	1936	172
17	7355	61742	1475	253
18	9644	73772	3042	296
19	10452	95564	3337	442
20	6033	42046	1830	182
21	5018	45102	1517	175
22	8610	71929	3019	234
23	8262	75885	2118	288
24	4929	42592	1708	158
25	3161	18392	769	113
26	10624	76039	2285	279
27	9513	64083	3052	233
28	3860	32793	690	123
29	8329	74201	2412	233

i	y	x_1	x_2	x_3
30	4178	45298	726	118

1. Основные определения и формулы

Множественная регрессия – регрессия между переменными y и $x_1, x_2, ..., x_m$, т.е. модель вида: $y = f(x_1, x_2, ..., x_m) + \varepsilon$,

где y – зависимая переменная (результативный признак);

$x_1, x_2, ..., x_m$ – независимые объясняющие переменные;

ε – возмущение или стохастическая переменная, включающая влияние неучтенных в модели факторов;

m – число параметров при переменных x.

Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

Уравнение множественной линейной регрессии в случае m независимых переменных имеет вид $\hat{y} = a + b_1 x_1 + ... + b_m x_m$, а в случае двух независимых переменных – $\hat{y} = a + b_1 x_1 + b_2 x_2$ (двухфакторное уравнение).

Для оценки параметров уравнения множественной регрессии применяют *метод наименьших квадратов*. Строится система нормальных уравнений:

$$\begin{aligned}
&na + b_1 \sum x_1 + b_2 \sum x_2 + ... + b_m \sum x_m = \sum y, \\
&a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + ... + b_m \sum x_1 x_m = \sum y x_1, \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&a \sum x_m + b_1 \sum x_1 x_m + b_2 \sum x_2 x_m + ... + b_m \sum x_m^2 = \sum y x_m.
\end{aligned}$$

Решение этой системы позволяет получить оценки параметров регрессии с помощью метода определителей

$$a = \frac{\Delta a}{\Delta},\ b_1 = \frac{\Delta b_1}{\Delta}, \ldots,\ b_m = \frac{\Delta b_m}{\Delta},$$

где $\Delta = \begin{vmatrix} n & \sum x_1 & \sum x_2 & \dots & \sum x_m \\ \sum x_1 & \sum x_1^2 & \sum x_2 x_1 & \dots & \sum x_m x_1 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & \dots & \sum x_m x_2 \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_m & \sum x_1 x_m & \sum x_2 x_m & \dots & \sum x_m^2 \end{vmatrix}$ – определитель системы;

$\Delta a,\ \Delta b_1, \dots, \Delta b_m$ – частные определители, которые получаются путем замены соответствующего столбца матрицы определителя системы данными правой части системы.

Для двухфакторного уравнения **коэффициенты множественной линейной регрессии** можно вычислить по формулам:

$$b_1 = \frac{Cov(x_1, y)Var(x_2) - Cov(x_2, y)Cov(x_1, x_2)}{Var(x_1)Var(x_2) - \left(Cov(x_1, x_2)\right)^2},$$

$$b_2 = \frac{Cov(x_2, y)Var(x_1) - Cov(x_1, y)Cov(x_1, x_2)}{Var(x_1)Var(x_2) - \left(Cov(x_1, x_2)\right)^2},$$

$$a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2.$$

Частные уравнения регрессии характеризуют изолированное влияние фактора на результат, ибо другие факторы закреплены на неизменном уровне. Эффекты влияния других факторов присоединены в них к свободному члену уравнения множественной регрессии. Это позволяет на основе частных уравнений регрессии $(\hat{y}_{x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m})$ определять **частные коэффициенты эластичности**:

$$Э_i = b_i \frac{x_i}{\hat{y}_{x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m}}.$$

Средние коэффициентами эластичности показывают на сколько процентов в среднем изменится результат при изменении соответствующего фактора на 1%:

$$\bar{Э}_i = b_i \frac{\bar{x}_i}{\bar{y}}.$$

Их можно сравнивать друг с другом и соответственно ранжировать факторы по силе их воздействия на результат.

Тесноту совместного влияния факторов на результат оценивает **коэффициент (индекс) множественной корреляции**:

$$R_{yx_1...x_m} = \sqrt{1 - \frac{\sigma^2_{ост}}{\sigma^2_y}}.$$

Величина индекса множественной корреляции лежит в пределах от 0 до 1 и должна быть больше или равна максимальному парному индексу корреляции: $R_{yx_1...x_m} \geq r_{yx_i}(\max)$.

Чем ближе значение индекса множественной корреляции к 1, тем теснее связь результативного признака со всем набором исследуемых факторов.

Сравнивая индексы множественной и парной корреляции, можно сделать вывод о целесообразности (величина индекса множественной корреляции существенно отличается от индекса парной корреляции) включения в уравнение регрессии того или иного фактора.

При линейной зависимости совокупный **коэффициент множественной корреляции** определяется через матрицу парных коэффициентов корреляции:

$$R_{yx_1...x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}},$$

где $\Delta r = \begin{vmatrix} 1 & r_{yx_1} & ... & r_{yx_m} \\ r_{yx_1} & 1 & ... & r_{x_1x_m} \\ ... & ... & ... & ... \\ r_{yx_m} & r_{x_mx_1} & ... & 1 \end{vmatrix}$ – определитель матрицы парных коэффициентов корреляции;

$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & ... & r_{x_1x_m} \\ r_{x_2x_1} & 1 & ... & r_{x_2x_m} \\ ... & ... & ... & ... \\ r_{x_mx_1} & r_{x_mx_2} & ... & 1 \end{vmatrix}$ – определитель матрицы межфакторной корреляции.

Частные коэффициенты корреляции характеризуют тесноту линейной зависимости между результатом и соответствующим фактором при устранении влияния других факторов. Если вычисляется, например, $r_{yx_1|x_2}$ (частный коэффициент корреляции между y и x_1 при фиксированном влиянии x_2), это означает, что определяется количественная мера линейной зависимости между y и x_1, которая будет иметь место, если устранить влияние на эти признаки фактора x_2.

Частные коэффициенты корреляции, измеряющие влияние на y фактора x_i при неизменном уровне других факторов, можно определить как:

$$r_{yx_i|x_1...x_{i-1}x_{i+1}...x_m} = \sqrt{1 - \frac{1 - R^2_{yx_1...x_m}}{1 - R^2_{yx_1...x_{i-1}x_{i+1}...x_m}}}$$

или по рекуррентной формуле:

$$r_{yx_i|x_1...x_{i-1}x_{i+1}...x_m} = \frac{r_{yx_i|x_1...x_{m-1}} - r_{yx_m|x_1...x_{m-1}} r_{x_i x_m|x_1...x_{m-1}}}{\sqrt{\left(1 - r^2_{yx_m|x_1...x_{m-1}}\right)\left(1 - r^2_{x_i x_m|x_1...x_{m-1}}\right)}}.$$

Для двухфакторного уравнения:

$$r_{yx_1|x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_2}}}; \quad r_{yx_2|x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_1}}} \quad \text{или}$$

$$r_{yx_1|x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right)\left(1 - r^2_{x_1x_2}\right)}}; \quad r_{yx_2|x_1} = \frac{r_{yx_2} - r_{yx_1} r_{x_2x_1}}{\sqrt{\left(1 - r^2_{yx_1}\right)\left(1 - r^2_{x_1x_2}\right)}}.$$

Частные коэффициенты корреляции изменяются в пределах от -1 до +1.

Сравнение значений парного и частного коэффициентов корреляции показывает направление воздействия фиксируемого фактора. Если частный коэффициент корреляции $r_{yx_1|x_2}$ получится меньше, чем соответствующий парный коэффициент r_{yx_1}, значит взаимосвязь признаков y и x_1 в некоторой степени обусловлена воздействием на них фиксируемой переменной x_2. И наоборот, большее значение частного коэффициента по сравнению с парным свидетельствует о том, что фиксируемая переменная x_2 ослабляет своим воздействием связь y и x_1.

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. Например, $r_{yx_1|x_2}$ – коэффициент частной корреляции первого порядка.

Зная частные коэффициенты корреляции (последовательно первого, второго и более высокого порядка), можно определить **совокупный коэффициент множественной корреляции**:

$$R_{yx_1 \dots x_m} = \sqrt{1 - \left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{yx_2|x_1}\right) \cdot \dots \cdot \left(1 - r^2_{yx_m|x_1 \dots x_{m-1}}\right)}.$$

Качество построенной модели в целом оценивает **коэффициент (индекс) множественной детерминации**, который рассчитывается как квадрат индекса множественной корреляции: $R^2_{yx_1 \dots x_m}$. Индекс множественной детерминации фиксирует долю объясненной вариации результативного признака за счет рассматриваемых в регрессии факторов. Влияние других, не учтенных в модели факторов, оценивается как $1 - R^2$.

Если число параметров при x_i близко к объему наблюдений, то коэффициент множественной корреляции приблизится к единице даже при слабой связи факторов с результатом. Для того чтобы не допустить возможного преувеличения тесноты связи, используется **скорректированный индекс множественной корреляции**, который содержит поправку на число степеней свободы:

$$\hat{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-m-1}.$$

Чем больше величина m, тем сильнее различия $\hat{R}^2$ и R^2.

Значимость частных коэффициентов корреляции проверяется аналогично случаю парных коэффициентов корреляции. Единственным отличием является число степеней свободы, которое следует брать равным $k = n - m - 2$.

Значимость уравнения множественной регрессии в целом, так же как и в парной регрессии, оценивается с помощью F -*критерия Фишера*:

$$F_{факт} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}.$$

Мерой для оценки включения фактора в модель служит **частный F-критерий**. В общем виде для фактора x_i частный F-критерий определяется как

$$F_{x_i} = \frac{R^2_{yx_1 \dots x_m} - R^2_{yx_1 \dots x_{i-1} x_{i+1} \dots x_m}}{1 - R^2_{yx_1 \dots x_m}} \cdot \frac{n-m-1}{1}.$$

Для двухфакторного уравнения частные F-критерии имеют вид:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot (n-3), \quad F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot (n-3).$$

Если фактическое значение F_{x_i} превышает табличное, то дополнительное включение фактора x_i в модель статистически оправданно и коэффициент чистой регрессии b_i при факторе x_i статистически значим. Если же фактическое значение F_{x_i} меньше табличного, то фактор x_i нецелесообразно включать в модель, а коэффициент регрессии при данном факторе в этом случае статистически незначим.

Для оценки **значимости коэффициентов чистой регрессии** по t-критерию Стьюдента используется формула:

$$t_{b_i} = \frac{b_i}{m_{b_i}} = \sqrt{F_{x_i}},$$

где b_i – коэффициент чистой регрессии при факторе x_i;

m_{b_i} – **средняя квадратическая (стандартная) ошибка коэффициента регрессии** b_i, которая может быть определена по формуле:

$$m_{b_i} = \frac{\sigma_y \sqrt{1 - R^2_{yx_1 \dots x_m}}}{\sigma_{x_i} \sqrt{1 - R^2_{x_i x_1 \dots x_m}}} \cdot \frac{1}{\sqrt{n-m-1}}.$$

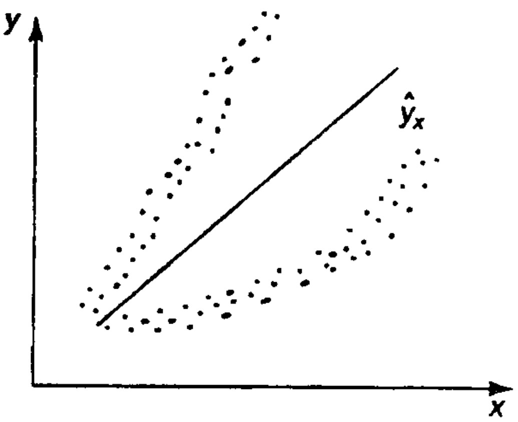
При дополнительном включении в регрессию нового фактора коэффициент детерминации должен возрастать, а остаточная дисперсия уменьшаться. Если это не так, то включаемый в анализ новый фактор не улучшает модель и практически является лишним фактором. Насыщение модели лишними факторами не только не снижает величину остаточной дисперсии и не увеличивает показатель детерминации, но и приводит к статистической незначимости параметров регрессии по t-критерию Стьюдента.

При построении уравнения множественной регрессии может возникнуть проблема **мультиколлинеарности** факторов. Считается, что две переменные явно коллинеарны, т.е. находятся между собой в линейной зависимости, если $r_{x_i x_j} \geq 0{,}7$. Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из регрессии. Предпочтение при этом отдается не фактору, более тесно связанному с результатом, а тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами.

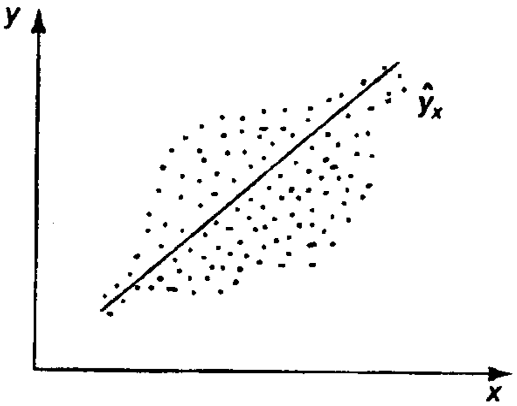
Для оценки мультиколлинеарности факторов может использоваться *определитель матрицы парных коэффициентов корреляции между факторами*. Чем ближе к 0 определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И наоборот, чем ближе к 1 определитель, тем меньше мультиколлинеарность факторов.

Для применения МНК требуется, чтобы дисперсия остатков была гомоскедастичной. Это означает, что для каждого значения фактора x_j остатки $\varepsilon_i = y_i - \bar{y}_i$ имеют одинаковую дисперсию. Если это условие применения МНК не соблюдается, то имеет место **гетероскедастичность**. При нарушении гомоскедастичности выполняются неравенства $\sigma^2_{\varepsilon_i} \neq \sigma^2_{\varepsilon_j} \neq \sigma^2, \; j \neq i.$

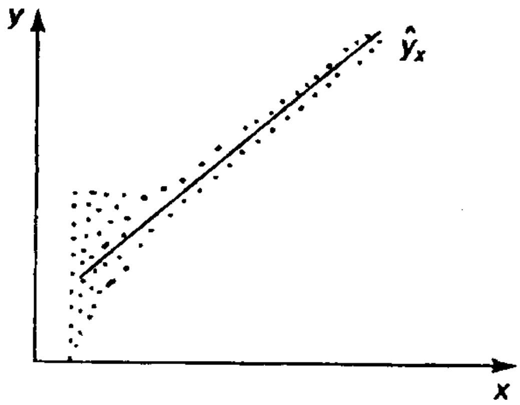
Наличие гетероскедастичности можно наглядно видеть из поля корреляции (рис. 5.10).



а)



б)



в)

Рис. 5.10. Примеры гетероскедастичности:

а) дисперсия остатков растет по мере увеличения x;

б) дисперсия остатков достигает максимальной величины при средних значениях переменной X и уменьшается при минимальных и максимальных значениях x;

в) максимальная дисперсия остатков при малых значениях x и дисперсия остатков однородна по мере увеличения значений x.

Для проверки выборки на гетероскедастичность можно использовать метод Гольдфельда-Квандта (при малом объеме выборки) или критерий Бартлетта (при большом объеме выборки).

Последовательность применения **теста Гольдфельда-Квандта**:

1) Упорядочить данные по убыванию той независимой переменной, относительно которой есть подозрение на гетероскедастичность.

2) Исключить из рассмотрения C центральных наблюдений. При этом $(n-C):2>p$, где p – число оцениваемых параметров. Из экспериментальных расчетов для случая однофакторного уравнения регрессии рекомендовано при n=30 принимать C=8, а при n=60 соответственно C=16.

3) Разделить совокупность из $(n-C)$ наблюдений на две группы (соответственно с малыми и большими значениями фактора x) и определить по

каждой из групп уравнение регрессии.

4) Вычислить остаточную сумму квадратов для первой (S_1) и второй (S_2) групп и найти их отношение $R = S_1 : S_2$, где $S_1 > S_2$. При выполнении нулевой гипотезы о гомоскедастичности отношение R будет удовлетворять F-критерию Фишера со степенями свободы $((n - C - 2p) : 2)$ для каждой остаточной суммы квадратов. Чем больше величина R превышает $F_{табл}$, тем более нарушена предпосылка о равенстве дисперсий остаточных величин.

Если необходимо включить в модель факторы, имеющие два или более качественных уровней (пол, профессия, образование, климатические условия, принадлежность к определенному региону и т.д.), то им должны быть присвоены *цифровые метки,* т.е. качественные переменные преобразованы в количественные. Такого вида сконструированные переменные называют ***фиктивными (искусственными) переменными***.

Коэффициент регрессии при фиктивной переменной интерпретируется как среднее изменение зависимой переменной при переходе от одной категории к другой при неизменных значениях остальных параметров. Значимость влияния фиктивной переменной проверяется с помощью t-критерия Стьюдента.

2. Решение типовых задач

Пример 2.1. По 15 предприятиям отрасли (табл. 5.8) изучается зависимость затрат на выпуск продукции y (тыс. ден. ед.) от объема произведенной продукции x_1 (тыс. ед.) и расходов на сырье x_2 (тыс. ден. ед). Необходимо:

1) Построить уравнение множественной линейной регрессии.

2) Вычислить и интерпретировать:

• средние коэффициенты эластичности;

• парные коэффициенты корреляции, оценить их значимость на уровне 0,05;

• частные коэффициенты корреляции;

• коэффициент множественной корреляции, множественный коэффициент детерминации, скорректированный коэффициент детерминации.

3) Оценить надежность построенного уравнения регрессии и целесообразность включения фактора x_1 после фактора x_2 и x_2 после x_1.

Таблица 5.8

i	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
x_1	2,7	1,5	8,2	4,5	3,3	5,8	3,0	7,1
x_2	55,7	52,0	120,1	80,3	29,8	110,5	90,4	118,2
y	110	70	310	120	75	170	100	180
i	*9*	*10*	*11*	*12*	*13*	*14*	*15*	
x_1	1,2	10,4	4,9	5,2	11,5	9,4	6,5	
x_2	24,7	298,9	58,2	120,3	224,7	271,2	102,0	
y	30	440	190	150	390	310	230	

<u>Решение:</u>

1) В Excel составим вспомогательную таблицу 5.9.

Таблица5.9

	A	B	C	D	E	F	G	H	I	J
1	i	x_1	x_2	y	yx_1	yx_2	x_1x_2	x_1^2	x_2^2	y^2
2	1	2,7	55,7	110	297,0	6127	150,39	7,29	3102,49	12100
3	2	1,5	52,0	70	105,0	3640	78,00	2,25	2704,00	4900
4	3	8,2	120,1	310	2542,0	37231	984,82	67,24	14424,01	96100
5	4	4,5	80,3	120	540,0	9636	361,35	20,25	6448,09	14400
6	5	3,3	29,8	75	247,5	2235	98,34	10,89	888,04	5625
7	6	5,8	110,5	170	986,0	18785	640,90	33,64	12210,25	28900
8	7	3,0	90,4	100	300,0	9040	271,20	9,00	8172,16	10000
9	8	7,1	118,2	180	1278,0	21276	839,22	50,41	13971,24	32400
10	9	1,2	24,7	30	36,0	741	29,64	1,44	610,09	900
11	10	10,4	298,9	440	4576,0	131516	3108,56	108,16	89341,21	193600
12	11	4,9	58,2	190	931,0	11058	285,18	24,01	3387,24	36100
13	12	5,2	120,3	150	780,0	18045	625,56	27,04	14472,09	22500
14	13	11,5	224,7	390	4485,0	87633	2584,05	132,25	50490,09	152100
15	14	9,4	271,2	310	2914,0	84072	2549,28	88,36	73549,44	96100
16	15	6,5	102,0	230	1495,0	23460	663,00	42,25	10404,00	52900
17	сумма	85,2	1757,0	2875,0	21512,5	464495,0	13269,5	624,5	304174,4	758625
18	среднее	5,68	117,133	191,667	1434,167	30966,333	884,633	41,632	20278,296	50575
19	n=	15								

Аналогично примеру 1 вычислим: $Cov(x_1, y)$=345,5; $Var(y)$=13838,89; $Cov(x_2, y)$=8515,78; $Cov(x_1, x_2)$=219,315; $Var(x_1)$=9,37; $Var(x_2)$=6558,08.

Затем найдем коэффициенты множественной линейной регрессии и оформим вывод результатов как на рис. 5.11.

Например, для вычисления значения коэффициента b_1 в ячейку **F20** поместим формулу **=(B20*B24-B21*B22)/(B23*B24-B22*B22)** и получим 29,83. В ячейке **F21** найдем b_2=0,3 как **=(B21*B23-B20*B22)/(B23*B24-B22*B22)**. Коэффициент a=-13,02 вычислим по формуле **=D18-B18*F20-C18*F21**.

Уравнение множественной линейной регрессии примет вид:

$\hat{y}$=-13,02+29,83$\cdot x_1$+0,3$\cdot x_2$.

Таким образом, при увеличении объема произведенной продукции (x_1) на 1 тыс. ед. затраты на выпуск этой продукции (y) в среднем увеличатся на 29,83 тыс. ден. ед., а при увеличении расходов на сырье (x_2) на 1 тыс. ден. ед. затраты увеличатся в среднем на 0,3 тыс. ден. ед.

	A	B	C	D	E	F	G	H	I	J	K
1	i	x_1	x_2	y	yx_1	yx_2	x_1x_2	x_1^2	x_2^2	y^2	
2	1	2,7	55,7	110	297,0	6127	150,39	7,29	3102,49	12100	
3	2	1,5	52,0	70	105,0	3640	78,00	2,25	2704,00	4900	
4	3	8,2	120,1	310	2542,0	37231	984,82	67,24	14424,01	96100	
5	4	4,5	80,3	120	540,0	9636	361,35	20,25	6448,09	14400	
6	5	3,3	29,8	75	247,5	2235	98,34	10,89	888,04	5625	
7	6	5,8	110,5	170	986,0	18785	640,90	33,64	12210,25	28900	
8	7	3,0	90,4	100	300,0	9040	271,20	9,00	8172,16	10000	
9	8	7,1	118,2	180	1278,0	21276	839,22	50,41	13971,24	32400	
10	9	1,2	24,7	30	36,0	741	29,64	1,44	610,09	900	
11	10	10,4	298,9	440	4576,0	131516	3108,56	108,16	89341,21	193600	
12	11	4,9	58,2	190	931,0	11058	285,18	24,01	3387,24	36100	
13	12	5,2	120,3	150	780,0	18045	625,56	27,04	14472,09	22500	
14	13	11,5	224,7	390	4485,0	87633	2584,05	132,25	50490,09	152100	
15	14	9,4	271,2	310	2914,0	84072	2549,28	88,36	73549,44	96100	
16	15	6,5	102,0	230	1495,0	23460	663,00	42,25	10404,00	52900	
17	сумма	85,2	1757,0	2875,0	21512,5	464495,0	13269,5	624,5	304174,4	758625	
18	среднее	5,68	117,133	191,667	1434,167	30966,333	884,633	41,632	20278,296	50575	
19	n=	15	проверка								
20	$Cov(x_1,y)$=	345,5	345,5		b_1=	29,83		$Э_{1\,ср.}$=	0,88		
21	$Cov(x_2,y)$=	8515,78	8515,78		b_2=	0,30		$Э_{2\,ср.}$=	0,18		
22	$Cov(x_1,x_2)$=	219,315	219,315		a=	-13,02					
23	$Var(x_1)$=	9,37	9,37		$r_{yx1\|x2}$=	0,3761		$F_{факт.}$=	79,62	$F_{табл.}$=	3,89
24	$Var(x_2)$=	6558,08	6558,08		$r_{yx2\|x1}$=	0,3426		F_{x1}=	22,41	$F_{x\ табл.}$=	4,75
25	$Var(y)$=	13838,89	13838,89		$r_{x1x2\|y}$=	0,2144		F_{x2}=	1,60		
26	r_{yx1}=	0,9595	0,9595		R_{x1x2y}=	0,9643		$\|t_{yx1}\|$=	12,2776	$t_{табл.}$=	2,1604
27	r_{yx2}=	0,8939	0,8939		R^2=	0,9299		$\|t_{yx2}\|$=	7,1896		
28	r_{x1x2}=	0,8847	0,8847		R^^{2}=	0,9182		$\|t_{x1x2}\|$=	6,8445		

Рис. 5.11. Решение примера 2.1 в Excel

2) Для вычисления **средних коэффициентов эластичности** воспользуемся формулой: $\overline{Э}_i = b_i \dfrac{\overline{x}_i}{\overline{y}}$. Вычисляем: $\overline{Э}_1$=0,88 и $\overline{Э}_1$=0,18. Т.е. увеличение только объема произведенной продукции (от своего среднего значения) или только расходов на сырье на 1% увеличивает в среднем затраты на выпуск продукции на 0,88% или 0,18% соответственно. Таким образом, фактор x_1 оказывает большее влияние на результат, чем фактор x_2.

Найдем значения **парных коэффициентов корреляции**:

$$r_{yx_1} = \frac{Cov(x_1, y)}{\sqrt{Var(x_1)Var(y)}} = 0{,}96; \quad r_{yx_2} = \frac{Cov(x_2, y)}{\sqrt{Var(x_2)Var(y)}} = 0{,}89;$$

$$r_{x_1x_2} = \frac{Cov(x_1, x_2)}{\sqrt{Var(x_1)Var(x_2)}} = 0{,}88.$$

Значения парных коэффициентов корреляции указывают на весьма тесную связь y с x_1 и на тесную связь с x_2. В то же время межфакторная связь $r_{x_1x_2}$ очень сильная ($r_{x_1x_2}$ =0,88>0,7), что говорит о том, что один из факторов является неинформативным, т.е. в модель необходимо включать или x_1, или x_2.

Значимость парных коэффициентов корреляции оценим с помощью t-критерия Стьюдента. $t_{табл}$=2,1604 определяем по таблице t-критерия Стьюдента (таблица 2 приложения) взяв α =0,05 и $k = n$ -2=13 либо с помощью встроенной статистической функции **СТЬЮДРАСПОБР** с такими же параметрами.

Фактическое значение t-критерия Стьюдента для каждого парного коэффициента определим по формулам: $t_{yx_1} = r_{yx_1}\sqrt{\frac{n-2}{1-r_{yx_1}^2}}$; $t_{yx_2} = r_{yx_2}\sqrt{\frac{n-2}{1-r_{yx_2}^2}}$;

$t_{x_1x_2} = r_{x_1x_2}\sqrt{\frac{n-2}{1-r_{x_1x_2}^2}}$. Получим t_{yx_1} =12,28; t_{yx_2} =7,19; $t_{x_1x_2}$ =6,84.

Так как фактические значения t-статистики превосходят табличные, то парные коэффициенты корреляции r_{yx_1}, r_{yx_2}, $r_{x_1x_2}$ не случайно отличаются от нуля, а статистически значимы.

Вычислим **частные коэффициенты корреляции** по формулам:

$$r_{yx_1|x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{\left(1-r_{yx_2}^2\right)\left(1-r_{x_1x_2}^2\right)}}; \quad r_{yx_2|x_1} = \frac{r_{yx_2} - r_{yx_1} r_{x_2x_1}}{\sqrt{\left(1-r_{yx_1}^2\right)\left(1-r_{x_1x_2}^2\right)}};$$

$$r_{x_1x_2|y} = \frac{r_{x_1x_2} - r_{x_1y} r_{x_2y}}{\sqrt{\left(1-r_{x_1y}^2\right)\left(1-r_{x_2y}^2\right)}}.$$

Получим $r_{yx_1|x_2}$ =0,38; $r_{yx_2|x_1}$ =0,34; $r_{x_1x_2|y}$ =0,21. Таким образом, фактор x_1 оказывает немного более сильное влияние на результат, чем x_2.

При сравнении значений коэффициентов парной и частной корреляции приходим к выводу, что из-за сильной межфакторной связи коэффициенты парной и частной корреляции отличаются довольно значительно.

Коэффициент множественной корреляции

$R_{yx_1x_2} = \sqrt{1-\left(1-r^2_{yx_1}\right)\cdot\left(1-r^2_{yx_2|x_1}\right)}$ =0,9643.

Следовательно, зависимость y от x_1 и x_2 характеризуется как очень тесная, в которой $R^2_{yx_1x_2}$ =93% вариации затрат на выпуск продукции определяются вариацией учтенных в модели факторов: объема произведенной продукции и расходов на сырье. Прочие факторы, не включенные в модель, составляют соответственно 7% от общей вариации y.

Скорректированный коэффициент множественной детерминации $\hat{R}^2 = 1-(1-R^2)\dfrac{n-1}{n-3}$ =0,9182 указывает на тесную связь между результатом и признаками.

3) Оценим **надежность уравнения регрессии в целом** с помощью F-критерия Фишера. Вычислим $F_{факт} = \dfrac{R^2}{1-R^2}\cdot\dfrac{(n-3)}{2}$ =79,62. $F_{табл}$=3,89 определяем по таблице F-критерия Фишера (таблица 1 приложения) взяв α =0,05, k_1=2, k_2 =15-2-1=12 либо с помощью встроенной статистической функции **FРАСПОБР** с такими же параметрами.

Так как фактическое значение больше табличного, то с вероятностью 95% делаем заключение о статистической значимости уравнения множественной линейной регрессии в целом.

Оценим целесообразность включения фактора x_1 после фактора x_2 и x_2 после x_1 с помощью частного F-критерия Фишера

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot (n-3)=22{,}41;\ F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot (n-3)=1{,}6.$$

$F_{табл}$=4,75 при α=0,05, k_1=1, k_2=12. Так как F_{x_1}=22,41>$F_{табл}$=4,75, а F_{x_2}=1,6<$F_{табл}$=4,75, то включение фактора x_1 в модель статистически оправдано и коэффициент чистой регрессии b_1 статистически значим, а дополнительное включение фактора x_2, после того, как уже введен фактор x_1, нецелесообразно. Низкое значение F_{x_2} (немногим больше 1) свидетельствует о статистической незначимости прироста $r^2_{yx_1}$ за счет включения в модель фактора x_2 после фактора x_1. Это означает, что парная регрессионная модель зависимости затрат на выпуск продукции от объема произведенной продукции является достаточно статистически значимой, надежной и что нет необходимости улучшать ее, включая дополнительный фактор x_2 (расходы на сырье).

3. Решение задач с помощью электронных таблиц Excel

Сводные данные основных характеристик для одного или нескольких массивов данных можно получить с помощью инструмента анализа данных **Описательная статистика.** Порядок действий следующий:

1. Необходимо проверить доступ к *Пакету анализа*. Для этого в главном меню нужно выбрать ***Сервис / Настройки*** и напротив ***Пакета анализа*** установить флажок.

2. Выбрать в главном меню ***Сервис / Анализ данных / Описательная статистика*** и заполнить диалоговое окно (рис. 5.12):

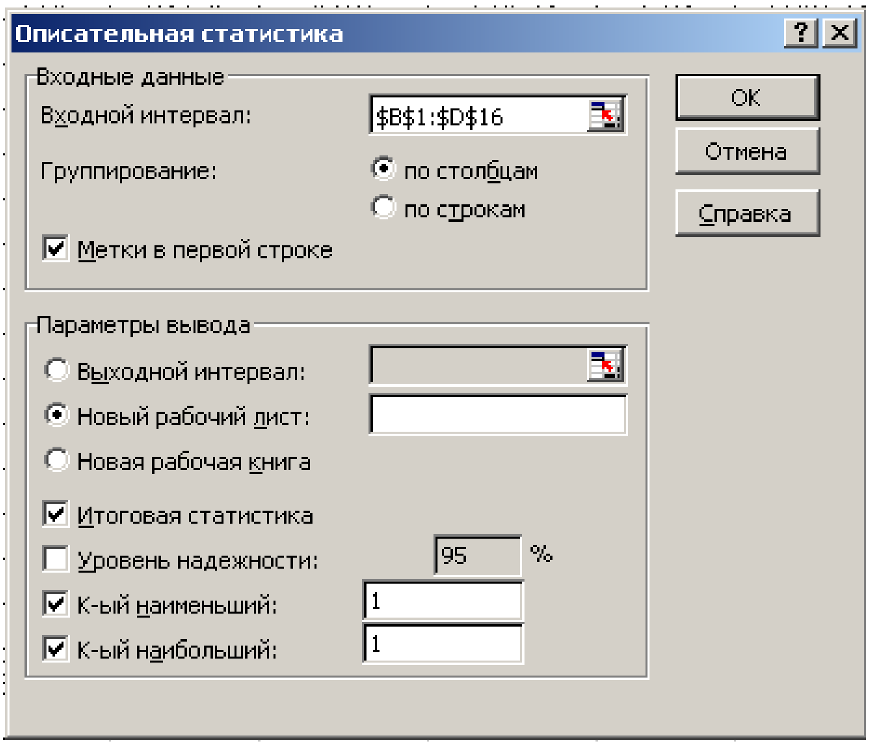


Рис. 5.12. Диалоговое окно ввода параметров инструмента
Описательная статистика

Входной интервал – диапазон, содержащий данные результативного и объясняющих признаков;

Группирование – указать, как расположены данные (в столбцах или строках);

Метки – флажок, который указывает, содержит ли первая строка названия столбцов или нет;

Выходной интервал – достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно задать произвольное имя нового листа, на который будут выведены результаты.

Для получения информации *Итоговой статистики, Уровня надежности,* k*-го наибольшего и наименьшего значений* нужно установить соответствующие флажки в диалоговом окне.

Получаем следующую статистику (рис. 5.13):

x_1		x_2		y	
Среднее	5,68	Среднее	117,133	Среднее	191,667
Стандартная ошибка	0,8181	Стандартная ошибка	21,643	Стандартная ошибка	31,440
Медиана	5,2	Медиана	102	Медиана	170
Мода	#Н/Д	Мода	#Н/Д	Мода	310
Стандартное отклонение	3,168	Стандартное отклонение	83,824	Стандартное отклонение	121,768
Дисперсия выборки	10,039	Дисперсия выборки	7026,512	Дисперсия выборки	14827,381
Эксцесс	-0,767	Эксцесс	0,608	Эксцесс	-0,282
Асимметричность	0,386	Асимметричность	1,213	Асимметричность	0,781
Интервал	10,3	Интервал	274,2	Интервал	410
Минимум	1,2	Минимум	24,7	Минимум	30
Максимум	11,5	Максимум	298,9	Максимум	440
Сумма	85,2	Сумма	1757	Сумма	2875
Счет	15	Счет	15	Счет	15
Наибольший(1)	11,5	Наибольший(1)	298,9	Наибольший(1)	440
Наименьший(1)	1,2	Наименьший(1)	24,7	Наименьший(1)	30

Рис. 5.13. Результат применения инструмента ***Описательная статистика***

Выборочные ковариации и **дисперсии** можно найти несколькими способами:

1 способ) Выборочные ковариации находятся с помощью статистической функции **КОВАР(диапазон1;диапазон2)**, а выборочные дисперсии – с помощью **ДИСПР(диапазон)**.

2 способ) Матрицу коэффициентов выборочных дисперсий и ковариаций можно получить с помощью инструмента анализа данных **Ковариация.** Для этого необходимо выбрать в главном меню ***Сервис / Анализ данных / Ковариация*** и заполнить диалоговое окно (рис. 5.14):

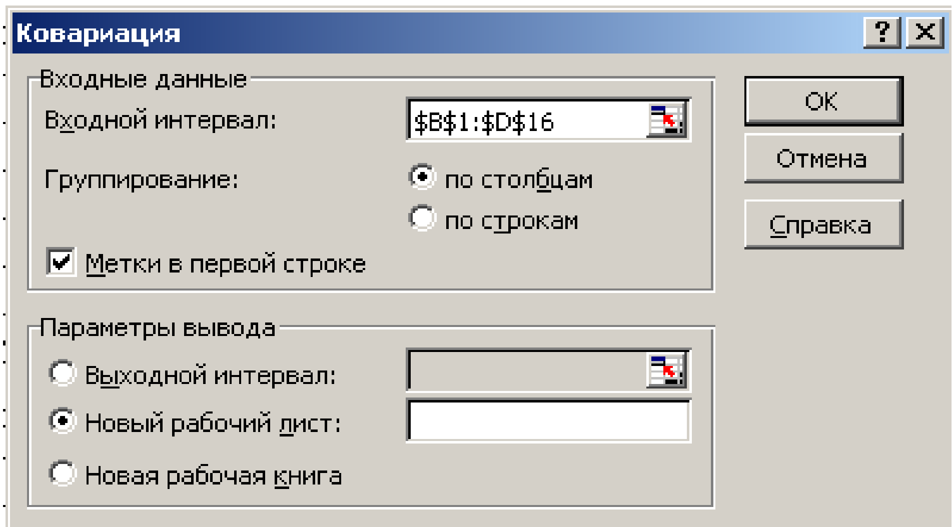


Рис. 5.14. Диалоговое окно ввода параметров инструмента ***Ковариация***

В результате получаем следующие данные (рис. 5.15):

	x_1	x_2	y
x_1	9,37		
x_2	219,315	6558,078	
y	345,5	8515,778	13838,89

Рис. 5.15. Результат применения инструмента ***Ковариация***

Из рис. 2.6 видно, что выборочная ковариация между x_1 и x_2 равна 219,315; между x_1 и y – 345,5; между x_2 и y – 8515,79. Выборочная дисперсия x_1 равна 9,37; x_2 – 6558,078; y – 13838,89.

Для проверки наличия коллинеарности или мультиколлинеарности и отбора факторов с помощью инструмента анализа данных **Корреляция** можно получить матрицу парных коэффициентов корреляции. Для этого необходимо выбрать в главном меню ***Сервис / Анализ данных / Корреляция*** и заполнить диалоговое окно (рис. 5.16):

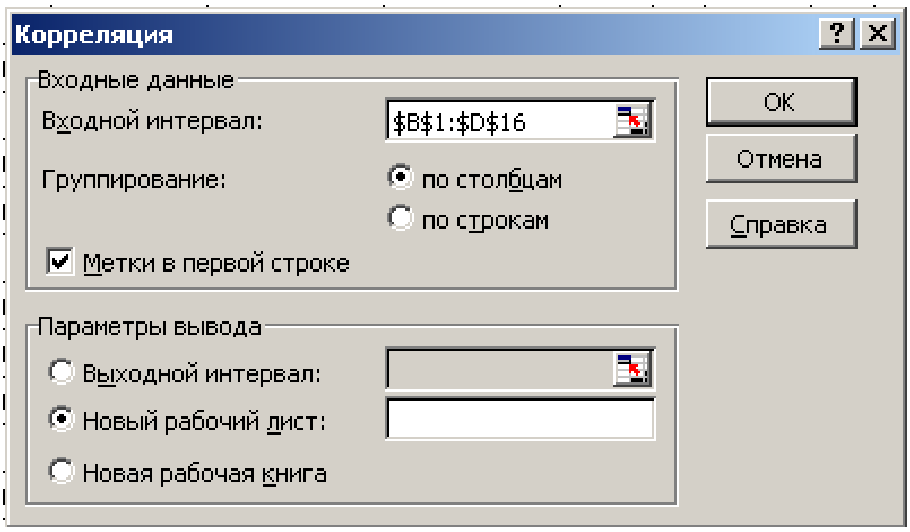


Рис. 5.16. Диалоговое окно ввода параметров инструмента ***Корреляция***

Получаем следующую матрицу коэффициентов парной корреляции (рис. 5.17), которые совпадают с вычисленными ранее.

	x_1	x_2	y
x_1	1		
x_2	0,884749	1	
y	0,959482	0,893893	1

Рис. 5.17. Матрица коэффициентов парной корреляции

Между факторами наблюдается мультиколлинеарность, так как значения всех коэффициентов корреляции больше 0,7.

Для вычисления параметров линейного уравнения множественной регрессии также как при вычислении параметров линейного уравнения парной регрессии используется инструмент анализа данных **Регрессия**, только во *входном интервале X* следует указать не один столбец, а все столбцы, содержащие значения факторных признаков (рис. 5.18).

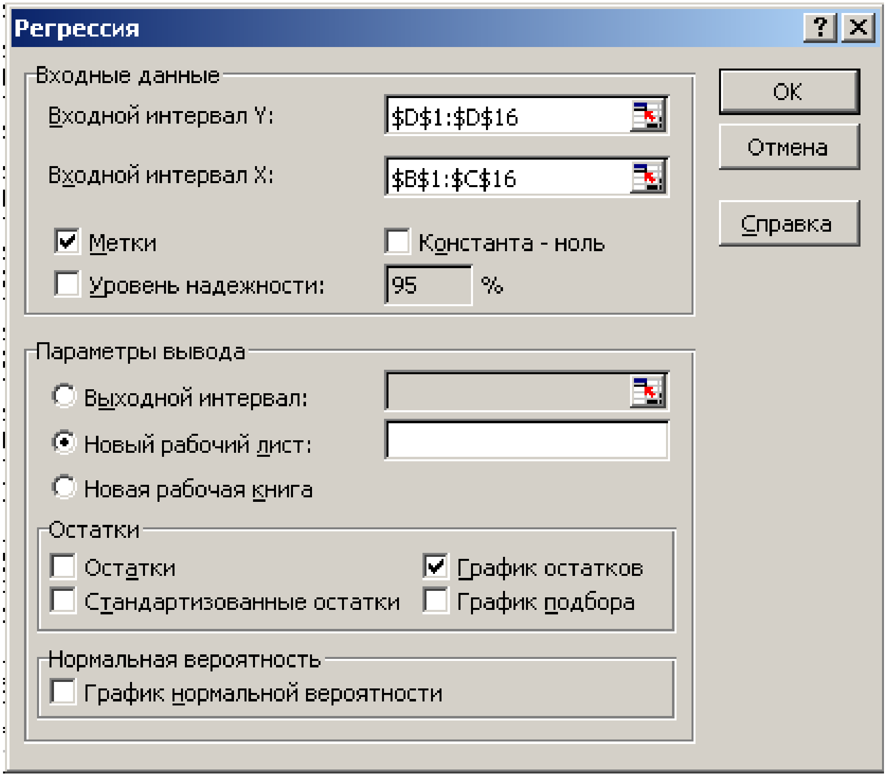


Рис. 5.18. Диалоговое окно ввода параметров инструмента ***Регрессия***

Результаты анализа представлены на рис. 5.19.

ВЫВОД ИТОГОВ

Регрессионная ст ат ист ика	
Множественный R	0,9643
R-квадрат	0,9299
Нормированный R-квадрат	0,9182
Стандартная ошибка	34,8167
Наблюдения	15

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значи-мост ь F*
Регрессия	2	193036,931	96518,47	79,62255	1,18E-07
Остаток	12	14546,402	1212,2		
Итого	14	207583,333			

	Коэффи-циент ы	*Ст андарт - ная ошибка*	*t-ст а- т ист ика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	-13,021	19,3203	-0,674	0,5131	-55,117	29,074	-55,117	29,074
x1	29,832	6,3013	4,734	0,0005	16,102	43,561	16,102	43,561
x2	0,301	0,2382	1,263	0,2305	-0,218	0,820	-0,218	0,820

Рис. 5.19 Результат применения инструмента ***Регрессия***

Следовательно, множественный коэффициент корреляции равен 0,96; множественный коэффициент детерминации 0,93; скорректированный коэффициент детерминации 0,918. На уровне значимости 0,05 фактическое значение F-критерия Фишера 79,62. Уравнение множественной линейной регрессии имеет вид: $\hat{y}$=-13,02+29,83$\cdot x_1$+0,3$\cdot x_2$.

Из рис. 2.10, на котором получены графики остатков x_1 и x_2, видим, что расположение остатков не имеет направленности. Следовательно, они независимы от значений x_i, значит построенная модель адекватна.

<u>Пример 2.</u> По данным табл. 2.1 проверить наличие гетероскедастичности остатков.

<u>Решение:</u>

<u>1 способ)</u> графический. В примере 1 было получено уравнение парной линейной регрессии $\hat{y}$=-17,78+36,87$\cdot x$; вычислены теоретические значения $(\hat{y})$ и остатки $\varepsilon = y - \hat{y}$. Эти результаты можно получить намного проще, используя регрессионный анализ. Для этого в диалоговом окне ввода параметров инструмента ***Регрессия*** нужно поставить флажок напротив «Остатки» и в Excel дополнительно будет выведены сведения, как на рис. 5.20.

ВЫВОД ОСТАТКА

Наблюдение	*Предсказанное y*	*Ост ат ки*
1	81,780	28,2196
2	37,531	32,4690
3	284,591	25,4094
4	148,155	-28,1547
5	103,905	-28,9052
6	196,092	-26,0916
7	92,843	7,1572
8	244,029	-64,0286
9	26,469	3,5314
10	365,715	74,2853
11	162,904	27,0955
12	173,967	-23,9669
13	406,277	-16,2767
14	328,840	-18,8401
15	221,904	8,0962

Рис. 5.20. Результат применения инструмента ***Регрессия*** (вывод остатка)

Предсказанное y – это теоретические значения $(\hat{y})$.

Таким образом, во вспомогательных таблицах можно не рассчитывать столбцы $\hat{y}$ и ε, а использовать результаты регрессионного анализа.

По данным рис. 5.20 построим график остатков (рис. 5.21). Остаточные величины ε не обнаруживают тенденцию по мере увеличения x и y. Следовательно, есть равенство дисперсий остаточных величин, т.е. не наблюдается гетероскедастичности остатков.

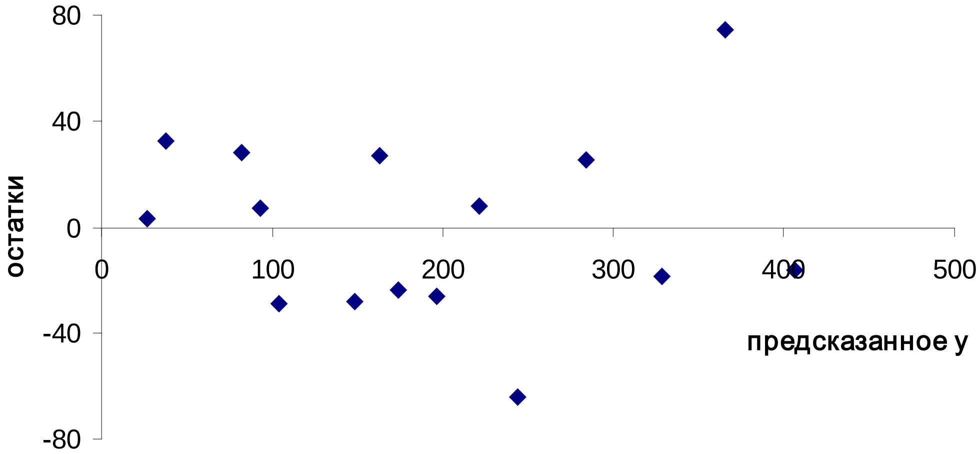


Рис. 5.21. График остатков

2 способ) теоретический. Для применения теста Гольдфельда-Квандта упорядочим данные по фактору x. Исключим из рассмотрения C=5 центральных наблюдений. Разделим оставшуюся совокупность из 15-5=10 наблюдений на две группы (по 5) и определим по каждой из групп уравнение регрессии. Это легко сделать, используя результаты регрессионного анализа (в главном меню выбрать ***Сервис / Анализ данных / Регрессия*** и установить флажок напротив «Остатки»). Результаты расчетов сведем в табл. 5.9, вычислив ε^2 и суммы по каждой группе.

Таблица 5.9

Уравнения регрессии	x	y	$\hat{y}$	ε	ε^2
Первая группа с первыми 5 наблюдениями: $\hat{y}$ =20,31+24,23$\cdot x$	1,2	30	49,38	-19,38	375,640
	1,5	70	56,65	13,35	178,236
	2,7	110	85,72	24,28	589,438
	3,0	100	92,99	7,01	49,144
	3,3	75	100,26	-25,26	637,953
Сумма (S_2)					1830,412
Вторая группа	7,1	180	215,37	-35,37	1250,803

с последними 5 наблюдениями: $\hat{y}$=-138,46+49,83$\cdot x$	8,2	310	270,19	39,81	1585,234
	9,4	310	329,99	-19,99	399,472
	10,4	440	379,82	60,18	3621,439
	11,5	390	434,64	-44,64	1992,721
Сумма (S_1)					8849,670

Величина $R = S_1 : S_2$=8849,670:1830,412=4,83 $(S_1 > S_2)$. $F_{табл}$=9,28 при 5%-ном уровне значимости и числе степеней свободы $k_1 = k_2$=(15-5-2·2):2=3.

Так как $R < F_{табл}$, то гетероскедастичность остатков отсутствует.

Практическое занятие №6

Тема: «Временные ряды»

Вариант 1.

Имеются поквартальные данные (**табл**. 6.1) **по розничному товарообороту в** 1995 - 1999 **гг** (**в** % **к предыдущему периоду**). **Постройте мультипликативную модель временного ряда**. **Рассчитайте прогноз розничного товарооборота на** 1 **квартал** 2000 **года**.

Таблица 6.1

№ квартала	*t*	y_t
1 **кв**. 1995**г**.	1	113,1
2 **кв**. 1995**г**.	2	95,9
3 **кв**. 1995**г**.	3	98
4 **кв**. 1995**г**.	4	101,8

№ квартала	t	y_t
1 кв. 1996г.	5	107,8
2 кв. 1996г.	6	96,3
3 кв. 1996г.	7	95,7
4 кв. 1996г.	8	99,8
1 кв. 1997г.	9	104
2 кв. 1997г.	10	95,8
3 кв. 1997г.	11	95,5
4 кв. 1997г.	12	99,3
1 кв. 1998г.	13	104
2 кв. 1998г.	14	96,2
3 кв. 1998г.	15	95,1
4 кв. 1998г.	16	97,5
1 кв. 1999г.	17	101
2 кв. 1999г.	18	93,5
3 кв. 1999г.	19	92
4 кв. 1999г.	20	94,6

Вариант 2.

Имеются поквартальные данные (табл. 6.2) **о разрешениях на строительство нового частного жилья, выданных в США в** 1990 - 1994 **гг в** % к **уровню** 1987 **года. Постройте аддитивную модель временного ряда. Рассчитайте прогноз на** 1 **квартал** 1995 **года**.

Таблица 6.2

№ квартала	t	y_t
1 кв.1990г.	1	68,3
2 кв.1990г.	2	61,9
3 кв.1990г.	3	65,1
4 кв.1990г.	4	74,1
1 кв. 1991г.	5	67,5
2 кв. 1991г.	6	66,8
3 кв. 1991г.	7	65,5
4 кв. 1991г.	8	73,6
1 кв. 1992г.	9	73,7
2 кв. 1992г.	10	61,0
3 кв. 1992г.	11	71,4
4 кв. 1992г.	12	82,3
1 кв. 1993г.	13	79,0
2 кв. 1993г.	14	72,7
3 кв. 1993г.	15	73,9
4 кв. 1993г.	16	85,3
1 кв. 1994г.	17	83,4
2 кв. 1994г.	18	76,2
3 кв. 1994г.	19	81,6
4 кв. 1994г.	20	89,3

Вариант 3.

Имеются данные об объеме продаж компании (млн руб) на основе поквартальных данных за 1993 - 1998 **гг** (**табл.** 6.3). **Построить аддитивную модель временного ряда**. **Рассчитать прогноз объема продаж на** 1 **квартал** 1999 **года**.

Таблица 6.3

№ квартала	*t*	y_t
1 **кв.** 1993г.	1	60,1
2 **кв.** 1993г.	2	95,7
3 **кв.** 1993г.	3	116
4 **кв.** 1993г.	4	101,8
1 **кв.** 1994г.	5	97
2 **кв.** 1994г.	6	116,7
3 **кв.** 1994г.	7	150,9
4 **кв.** 1994г.	8	123,2
1 **кв.** 1995г.	9	94,2
2 **кв.** 1995г.	10	124,1
3 **кв.** 1995г.	11	157,2
4 **кв.** 1995г.	12	115,8
1 **кв.** 1996г.	13	101,6
2 **кв.** 1996г.	14	149,2
3 **кв.** 1996г.	15	162,3
4 **кв.** 1996г.	16	142,3
1 **кв.** 1997г.	17	114,6
2 **кв.** 1997г.	18	152,4
3 **кв.** 1997г.	19	153
4 **кв.** 1997г.	20	142,2
1 **кв.** 1998г.	21	106,1
2 **кв.** 1998г.	22	144,7
3 **кв.** 1998г.	23	156,2
4 **кв.** 1998г.	24	157

1. Основные определения и формулы

При построении эконометрической модели используются два типа данных:

1) данные, характеризующие совокупность различных объектов в определенный момент времени *(пространственные модели)*;

2) данные, характеризующие один объект за ряд последовательных моментов времени *(модели временных рядов)*.

Временной ряд (ряд динамики) – это совокупность значений какоголибо показателя за несколько последовательных моментов или периодов времени. Каждый уровень временного ряда формируется из *трендовой* (T), *циклической* (S) и *случайной* (E) компонент.

Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется **аддитивной моделью** временного ряда; как произведение – **мультипликативной моделью** временного ряда.

Аддитивная модель: $Y = T + S + E$;

мультипликативная модель: $Y = T \cdot S \cdot E$.

Основная задача эконометрического исследования отдельного временного ряда – выявление и придание количественного выражения каждой из перечисленных выше компонент с тем, чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

Построение аддитивной и мультипликативной моделей сводится к расчету значений T, S и E для каждого уровня ряда.

Процесс построения модели следующий:

Шаг 1. Выравнивание исходного ряда методом скользящей средней.

Шаг 2. Расчет значений сезонной компоненты S.

Шаг 3. Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных в аддитивной $(T + E)$ или в мультипликативной $(T \cdot E)$ модели.

Шаг 4. Аналитическое выравнивание уровней $(T + E)$ или $(T \cdot E)$ и расчет значений T с использованием полученного уравнения тренда.

Шаг 5. Расчет полученных по модели значений $(T + S)$ или $(T \cdot S)$.

Шаг 6. Расчет абсолютных и/или относительных ошибок.

Автокорреляция уровней ряда – корреляционная зависимость между последовательными уровнями временного ряда.

Коэффициент автокорреляции уровней ряда первого порядка:

$$r_1 = \frac{\sum_{t=2}^{n} (y_t - \bar{y}_1) \cdot (y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n} (y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^{n} (y_{t-1} - \bar{y}_2)^2}},$$

где $\bar{y}_1 = \frac{1}{n-1} \sum_{t=2}^{n} y_t, \quad \bar{y}_2 = \frac{1}{n-1} \sum_{t=2}^{n} y_{t-1}.$

Аналогично определяются коэффициенты автокорреляции второго и более высоких порядков. Например, коэффициент автокорреляции второго порядка:

$$r_2 = \frac{\sum_{t=3}^{n} (y_t - \bar{y}_3) \cdot (y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n} (y_t - \bar{y}_3)^2 \cdot \sum_{t=3}^{n} (y_{t-2} - \bar{y}_4)^2}},$$

где $\bar{y}_3 = \frac{1}{n-2} \sum_{t=3}^{n} y_t, \quad \bar{y}_4 = \frac{1}{n-2} \sum_{t=3}^{n} y_{t-2}.$

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют **лагом.** С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается.

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют **автокорреляционной функцией** временного ряда, а график зависимости ее значений от величины лага (порядка коэффициента автокорреляции) – **коррелограммой**.

Построение аналитической функции для моделирования тенденции (тренда) временного ряда называют *аналитическим выравниванием временного ряда*.

Поскольку зависимость от времени может принимать разные формы, для ее формализации можно использовать различные виды функций. Для построения трендов чаще всего применяются линейные, гиперболические, экспоненциальные, степенные функции.

Параметры трендов определяются обычным МНК, в качестве независимой переменной выступает время $t = 1, 2, \ldots, n,$ а в качестве зависимой переменной – фактические уровни временного ряда $\hat{y}_t$. Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации. Критерием отбора наилучшей формы тренда является наибольшее значение скорректированного коэффициента детерминации.

При построении моделей регрессии по временным рядам для устранения тенденции можно использовать следующие методы:

1) *метод отклонений от тренда* предполагает вычисление трендовых значений для каждого временного ряда модели, например $\hat{y}_t$ и $\hat{x}_t$, и расчет отклонений от трендов: $y_t - \hat{y}_t$ и $x_t - \hat{x}_t$, которые используют для дальнейшего анализа.

2) *метод последовательных разностей* заключается в следующем:

• если ряд содержит линейный тренд, то исходные данные заменяются первыми разностями: $\Delta_t = y_t - y_{t-1}$;

• если параболический тренд – вторыми разностями: $\Delta_t^2 = \Delta_t - \Delta_{t-1}$;

• если экспоненциальный или степенной тренд, то этот метод применяется к логарифмам исходных данных.

Простейший подход к моделированию сезонных колебаний – это расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Автокорреляция в остатках – корреляционная зависимость между значениями остатков ε_t за текущий и предыдущий моменты времени.

Для определения автокорреляции остатков используют **критерий Дарбина-Уотсона**:

$$d = \frac{\sum_{t=2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n} \varepsilon_t^2}, \qquad 0 \le d \le 4.$$

Коэффициент автокорреляции остатков первого порядка:

$$r_1^{\varepsilon} = \frac{\sum_{t=2}^{n} \varepsilon_t \cdot \varepsilon_{t-1}}{\sum_{t=2}^{n} \varepsilon_t^2}, \qquad -1 \le r_1^{\varepsilon} \le 1.$$

Между критерием Дарбина-Уотсона и коэффициентом автокорреляции остатков первого порядка существует соотношение: $d = 2 \cdot \left(1 - r_1^{\varepsilon}\right)$.

Алгоритм выявления автокорреляции остатков на основе *критерия Дарбина-Уотсона* следующий. Выдвигается гипотеза H_0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H_1 и H_1^* состоят, соответственно, в наличии положительной или отрицательной автокорреляции в остатках. Далее по статистическим таблицам (таблица 3 приложения) определяются критические значения критерия Дарбина-Уотсона d_L и d_U для заданного числа наблюдений n, числа независимых переменных модели m и уровня значимости α. По этим значениям числовой промежуток $[0; 4]$ разбивают на пять отрезков. Принятие или отклонение каждой из гипотез с вероятностью $1-\alpha$ осуществляется следующим образом:

- $0 < d < d_L$ – есть положительная автокорреляция остатков, H_0 отклоняется, с вероятностью $1-\alpha$ принимается H_1;
- $d_L < d < d_U$ – зона неопределенности;
- $d_U < d < 4 - d_U$ – нет оснований отклонять H_0, т.е. автокорреляция остатков отсутствует;
- $4 - d_U < d < 4 - d_L$ – зона неопределенности;
- $4 - d_L < d < 4$ – есть отрицательная автокорреляция остатков, H_0 отклоняется, с вероятностью $1-\alpha$ принимается H_1^*.

Если фактическое значение критерия Дарбина-Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу H_0.

Эконометрические модели, содержащие не только текущие, но и лаговые значения факторных переменных, называются **моделями с распределенным лагом**, например, $y_t = a + b_0 x_t + b_1 x_{t-1} + ... + b_p x_{t-p} + \varepsilon_t$.

Оценка параметров моделей с распределенными лагами проводится по методу Койка или методу Алмон.

Модели, содержащие в качестве факторов лаговые значения зависимой переменной, называются **моделями авторегрессии**, например, $y_t = a + b_0 x_t + c_1 y_{t-1} + \varepsilon_t$.

2. Решение типовых задач

<u>Пример 3.</u> По данным табл. 2.4 требуется:

1) Рассчитать критерий Дарбина-Уотсона.

2) Оценить полученный результат при 5%-ном уровне значимости.

3) Указать, пригодно ли уравнение для прогноза.

<u>Решение:</u>

1) В Excel составим вспомогательную таблицу как на рис. 6.1.

	A	B	C	D	E	F	G	H	I	J
1	i	x_1	x_2	y	y^	ε_i=y-y^	ε_{i-1}	$\varepsilon_i-\varepsilon_{i-1}$	$(\varepsilon_i-\varepsilon_{i-1})^2$	ε_i^2
2	1	2,7	55,7	110	84,284	-194,284	–	–	–	37746,173
3	2	1,5	52,0	70	47,372	-117,372	-194,284	76,911	5915,364	13776,266
4	3	8,2	120,1	310	267,735	-577,735	-117,372	-460,363	211934,129	333778,168
5	4	4,5	80,3	120	145,383	-265,383	-577,735	312,353	97564,221	70427,957
6	5	3,3	29,8	75	94,390	-169,390	-265,383	95,993	9214,599	28692,959
7	6	5,8	110,5	170	193,251	-363,251	-169,390	-193,861	37581,958	131951,018
8	7	3,0	90,4	100	103,674	-203,674	-363,251	159,577	25464,733	41483,057
9	8	7,1	118,2	180	234,349	-414,349	-203,674	-210,675	44383,889	171684,879
10	9	1,2	24,7	30	30,209	-60,209	-414,349	354,140	125415,158	3625,089
11	10	10,4	298,9	440	387,163	-827,163	-60,209	-766,954	588218,989	684198,750
12	11	4,9	58,2	190	150,666	-340,666	-827,163	486,497	236679,525	116053,237
13	12	5,2	120,3	150	178,300	-328,300	-340,666	12,366	152,910	107781,024
14	13	11,5	224,7	390	397,653	-787,653	-328,300	-459,352	211004,588	620396,559
15	14	9,4	271,2	310	348,997	-658,997	-787,653	128,656	16552,293	434276,845
16	15	6,5	102,0	230	211,575	-441,575	-658,997	217,421	47272,096	194988,812
17	сумма	85,2	1757,0	2875,0	2875,0	-5750,0	-5308,4	-247,3	1657354,5	2990860,8

Рис. 6.1. Вспомогательная таблица к примеру 3

Теоретические значения $\hat{y}$ можно определить несколькими способами:

1 способ) Используя полученное в примере 2 уравнение множественной линейной регрессии $\hat{y}$=-13,02+29,83$\cdot x_1$+0,3$\cdot x_2$ можно вычислить для каждого набора x_1 и x_2 свое значение $\hat{y}$.

2 способ) Применить инструмент анализа данных ***Регрессия*** (***Сервис / Анализ данных***), поставив флажок для вывода остатков. В столбце с названием «Предсказанное y» будут находиться теоретические значения y.

Заполним столбец **G**, который получается смещением столбца **F** на одно значение вниз. Столбец **H** найдем как разность между столбцами **F** и **G**. Заполним столбцы **I** и **J**, возведя в квадрат значения столбцов **F** и **H** соответственно.

Критерий Дарбина-Уотсона рассчитываем по формуле:

$$d = \frac{\sum(\varepsilon_i - \varepsilon_{i-1})^2}{\sum \varepsilon_i^2} = \frac{1657354{,}5}{2990860{,}8} = 0{,}554$$ или в Excel **=I17/J17**.

2) Фактическое (найденное выше) значение сравним с табличными значениями при 5%-ном уровне значимости. При n=15 и m=2 нижнее значение d_L=0,95, а верхнее d_U=1,54 (таблица 3 приложения). Так как

$0 < d < d_L$, то это означает наличие в остатках положительной автокорреляции.

3) Уравнение регрессии не может быть использовано для прогноза, так как в нем не устранена автокорреляция в остатках, которая может иметь разные причины. Автокорреляция в остатках может означать, что в уравнение не включен какой-либо существенный фактор; возможно, что форма связи неточна или в рядах динамики имеется общая тенденция.

3. Решение задач помощью электронных таблиц Excel

Пример 4. Динамика объема платных услуг населению региона по кварталам 2004-2007 гг. характеризуется данными, представленными в табл. 6.4

Таблица 6.4

Год	Квартал	t	Объем платных услуг населению (млн. руб.), y_t
2004	I	1	375
	II	2	371
	III	3	869
	IV	4	1015
2005	I	5	357
	II	6	471
	III	7	992
	IV	8	1020
2006	I	9	390
	II	10	355
	III	11	992
	IV	12	905
2007	I	13	461
	II	14	454
	III	15	920
	IV	16	927

Необходимо построить аддитивную модель временного ряда, оценить качество построенной модели и сделать прогноз об объеме платных услуг населению на I и II кварталы 2008 г.

Решение:

Занесем данные t и y_t в Excel и построим поле корреляции (рис.6.2):

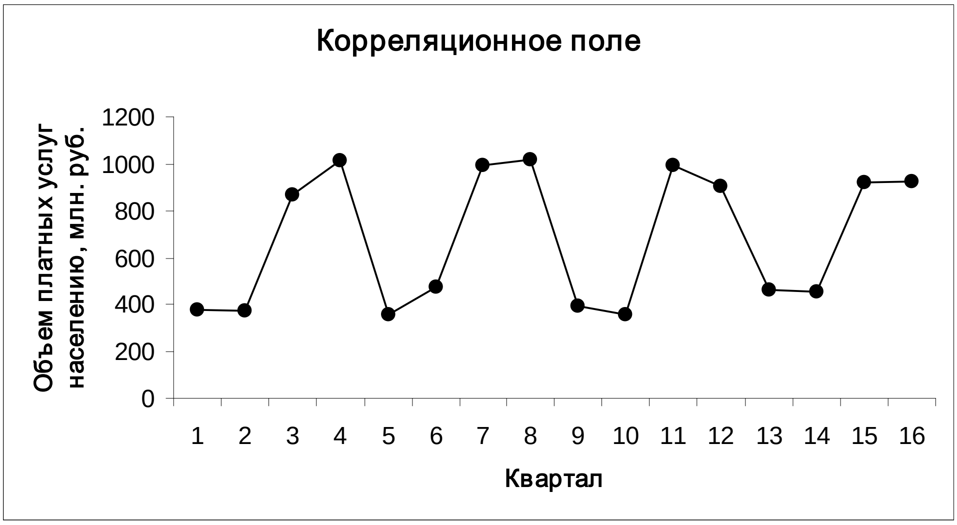


Рис. 6.2. Корреляционное поле

Рассчитаем коэффициенты автокорреляции. Для этого составляем первую вспомогательную таблицу 6.5.

Таблица 6.5

A	B	C	D	E	F	G	H
t	y_t	y_{t-1}	$y_t-\bar{y}_1$	$y_{t-1}-\bar{y}_2$	$(y_t-\bar{y}_1)\cdot$ $\cdot(y_{t-1}-\bar{y}_2)$	$(y_t-\bar{y}_1)^2$	$(y_{t-1}-\bar{y}_2)^2$
1	375	–	–	–	–	–	–
2	371	375	-308,63	-288,13	88925,15	95249,39	83020,82
3	869	371	189,38	-292,13	-55322,75	35862,89	85341,88
4	1015	869	335,38	205,87	69042,53	112476,39	42381,08
5	357	1015	-322,63	351,87	-113520,98	104086,89	123810,15
6	471	357	-208,63	-306,13	63867,07	43524,39	93717,62
7	992	471	312,38	-192,13	-60017,65	97578,14	36915,22
8	1020	992	340,38	328,87	111937,99	115855,14	108153,28
9	390	1020	-289,63	356,87	-103357,51	83882,64	127353,82
10	355	390	-324,63	-273,13	88665,91	105381,39	74601,82
11	992	355	312,38	-308,13	-96253,15	97578,14	94946,15
12	905	992	225,38	328,87	74118,33	50793,89	108153,28
13	461	905	-218,63	241,87	-52878,10	47796,89	58499,48
14	454	461	-225,63	-202,13	45606,33	50906,64	40857,88
15	920	454	240,38	-209,13	-50270,43	57780,14	43736,75
16	927	920	247,38	256,87	63542,39	61194,39	65980,48
Сумма	10874	9947	304,6	0,0	74085,13	1159947,36	1187469,73
Среднее	679,63	663,13	–	–	–	–	–

Столбец **C** получается сдвигом данных столбца **B** на одно значение вниз. Столбец **D** вычисляется как разность между столбцом **B** и его средним значением (679,63).

Замечание. Для вычисления средних используйте статистическую функцию **СРЗНАЧ(диапазон),** поскольку среднее значение y_{t-1} получается путем деления суммы не на 16, а на 15.

Например, в ячейке **I5** вычислим коэффициент автокорреляции первого порядка: =**F18/КОРЕНЬ(G18*H18)** или

$$r_1 = \frac{74085{,}13}{\sqrt{1159947{,}36 \cdot 118769{,}73}} = 0{,}063125.$$

Аналогично составляем вспомогательную таблицу для расчета коэффициента автокорреляции второго порядка. Для этого скопируем табл. 4.2 и вставим ее ниже (например, начиная с ячейки **A21**), а затем немного исправим значения. В результате получим табл. 6.6.

Таблица 6.6

t	y_t	y_{t-2}	$y_t - \bar{y}_3$	$y_{t-2} - \bar{y}_4$	$(y_t - \bar{y}_3)\cdot(y_{t-2} - \bar{y}_4)$	$(y_t - \bar{y}_3)^2$	$(y_{t-2} - \bar{y}_4)^2$
1	375	–	–	–	–	–	–
2	371	–	–	–	–	–	–
3	869	375	189,38	-288,13	-54565,25	35862,89	83020,82
4	1015	371	335,38	-292,13	-97974,22	112476,39	85341,88
5	357	869	-322,63	205,87	-66417,73	104086,89	42381,08
6	471	1015	-208,63	351,87	-73408,18	43524,39	123810,15
7	992	357	312,38	-306,13	-95628,40	97578,14	93717,62
8	1020	471	340,38	-192,13	-65397,38	115855,14	36915,22
9	390	992	-289,63	328,87	-95248,01	83882,64	108153,28
10	355	1020	-324,63	356,87	-115847,84	105381,39	127353,82
11	992	390	312,38	-273,13	-85320,03	97578,14	74601,82
12	905	355	225,38	-308,13	-69445,55	50793,89	94946,15
13	461	992	-218,63	328,87	-71898,48	47796,89	108153,28
14	454	905	-225,63	241,87	-54571,17	50906,64	58499,48
15	920	461	240,38	-202,13	-48587,80	57780,14	40857,88
16	927	454	247,38	-209,13	-51734,36	61194,39	43736,75
Сумма	10874	9027	613,25	-256,87	-1046044,39	1064697,97	1121489,25
Среднее	679,63	644,79	–	–	–	–	–

Следовательно,

$$r_2 = \frac{-1046044{,}39}{\sqrt{1064697{,}97 \cdot 1121489{,}25}} = -0{,}957281.$$

Аналогично находим коэффициенты автокорреляции более высоких порядков, и все полученные значения заносим в сводную табл. 3.7, на основании которой построим коррелограмму (рис. 6.3).

Таблица 6.7

1. Лаг	r_t
1	0,063125
2	-0,957281
3	-0,037521
4	0,964635
5	0,046429
6	-0,973828
7	-0,073964
8	0,961077
9	0,146166
10	-0,972632
11	-0,114039
12	0,983594
13	0,241605
14	-0,999972

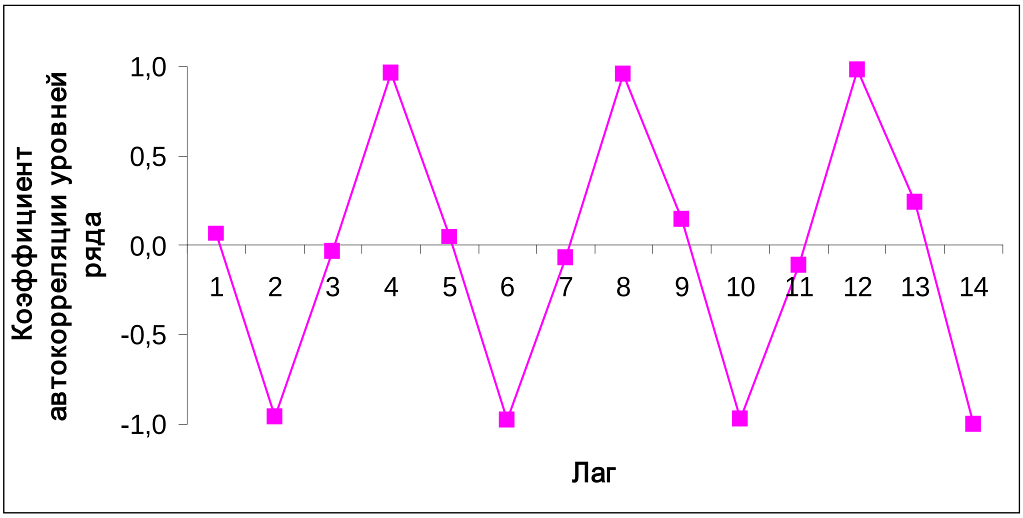


Рис.6.3. Коррелограмма

Анализ коррелограммы (рис. 6.3) и графика (рис. 6.2) исходных уровней временного ряда позволяет сделать вывод о наличии в изучаемом временном ряде сезонных колебаний периодичностью в четыре квартала.

Таким образом, данный временной ряд содержит сезонные колебания периодичностью 4, т.к. объем платных услуг населению в первый-второй кварталы ниже, чем в третий-четвертый.

Рассчитаем компоненты *аддитивной модели* временного ряда.

Шаг 1. Проведем выравнивание исходных уровней ряда методом скользящей средней. Для этого заполним табл. 6.8:

- Просуммируем уровни ряда последовательно за каждые четыре квартала со сдвигом на один момент времени и определим условные годовые объемы платных услуг населению (столбец **С**). Для этого в ячейку **С3** поместим =**СУММ(В2:В5)** и протянем за правый нижний уголок ячейки до ячейки **С15**. В результате произойдет автоматическое заполнение диапазона **С3 – С15.**

- Разделив полученные суммы на 4, найдем скользящие средние (столбец **D**). Например, в ячейку **D3** поместим =**С3/4**. Полученные таким образом выровненные значения уже не содержат сезонной компоненты.

- Приведем эти значения в соответствие с фактическими моментами времени, для чего найдем средние значения из двух последовательных скользящих средних – центрированные скользящие средние (столбец **Е**). Например, в ячейку **Е4** поместим =**СРЗНАЧ(D3:D4)**.

Таблица 6.8

	A	B	C	D	E	F
1	t	y_t	Итого за четыре квартала	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
2	1	375	–	–	–	–
3	2	371	2630	657,50	–	–
4	3	869	2612	653,00	655,250	213,750
5	4	1015	2712	678,00	665,500	349,500
6	5	357	2835	708,75	693,375	-336,375
7	6	471	2840	710,00	709,375	-238,375
8	7	992	2873	718,25	714,125	277,875
9	8	1020	2757	689,25	703,750	316,250
10	9	390	2757	689,25	689,250	-299,250
11	10	355	2642	660,50	674,875	-319,875
12	11	992	2713	678,25	669,375	322,625
13	12	905	2812	703,00	690,625	214,375

14	13	461	2740	685,00	694,000	-233,000
15	14	454	2762	690,50	687,750	-233,750
16	15	920	–	–	–	–
17	16	927	–	–	–	–

Шаг 2. Найдем оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними (столбец **F** табл. 6.5). Так, в ячейку **F4** поместим =**B4-E4**.

Составим таблицу 6.6, распределив значения столбца **F** таблицы 6.5 по кварталам и годам. С помощью статистической функции **СРЗНАЧ(диапазон)** найдем средние за каждый квартал (по всем годам) оценки сезонной компоненты $\overline{S}_i$.

В ячейке **G25** рассчитаем сумму средних с помощью встроенной статистической функции =**СУММ(C25:F25)**.

Вычислим корректирующий коэффициент: $k = 11{,}25/4 = 2{,}8125$. Рассчитаем скорректированные значения сезонной компоненты $S_i = \overline{S}_i - k$ и занесем полученные данные в табл. 6.9.

В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем. кварталам должна быть равна нулю (ячейка **G26**).

Таблица 6.9

	A	B	C	D	E	F	G
19	**Показатели**	Год	№ квартала, i				Сумма
20			I	II	III	IV	
21		2004	–	–	213,750	349,500	
22		2005	-336,375	-238,375	277,875	316,250	
23		2006	-299,250	-319,875	322,625	214,375	
24		2007	-233,000	-233,750	–	–	
25	$\overline{S}_i$		-289,542	-264,000	271,417	293,375	11,25
26	S_i		-292,354	-266,813	268,604	290,563	0,0

Составим табл. 6.10, в которой в столбец **C** поместим вычисленные S_i. Причем через каждые четыре квартала эти значения будут повторяться.

Таблица 6.10

	A	B	C	D	E	F	G	H	I
28	t	y_t	S_i	$y_t - S_i$	T	$T+S$	E	E^2	$(y_t - \overline{y})^2$
29	1	375	-292,35	667,35	672,69	380,33	-5,33	28,42	92796,39
30	2	371	-266,81	637,81	673,61	406,80	-35,80	1281,53	95249,39
31	3	869	268,60	600,40	674,54	943,14	-74,14	5496,84	35862,89
32	4	1015	290,56	724,44	675,46	966,02	48,98	2398,60	112476,39
33	5	357	-292,35	649,35	676,39	384,03	-27,03	730,80	104086,89
34	6	471	-266,81	737,81	677,31	410,50	60,50	3660,19	43524,39
35	7	992	268,60	723,40	678,24	946,84	45,16	2039,18	97578,14
36	8	1020	290,56	729,44	679,16	969,73	50,27	2527,42	115855,14
37	9	390	-292,35	682,35	680,09	387,74	2,26	5,13	83882,64
38	10	355	-266,81	621,81	681,02	414,20	-59,20	3504,94	105381,39
39	11	992	268,60	723,40	681,94	950,54	41,46	1718,54	97578,14
40	12	905	290,56	614,44	682,87	973,43	-68,43	4682,46	50793,89
41	13	461	-292,35	753,35	683,79	391,44	69,56	4838,96	47796,89
42	14	454	-266,81	720,81	684,72	417,90	36,10	1302,89	50906,64
43	15	920	268,60	651,40	685,64	954,25	-34,25	1172,83	57780,14
44	16	927	290,56	636,44	686,57	977,13	-50,13	2513,07	61194,39
45	Сумма	10874	–	–	–	–	–	37901,81	1252743,75
46	Среднее	679,625	–	–	–	–	–	–	–

Шаг 3. Исключим влияние сезонной компоненты, вычитая ее значение из каждого уровня исходного временного ряда. В столбце **D** табл. 6.10 получим величины $T + E = Y - S,$ которые рассчитываются за каждый момент времени и содержат только тенденцию и случайную компоненту.

Шаг 4. Определим компоненту T данной модели аналитическим выравниванием ряда $(T+E)$ с помощью линейного тренда. Порядок действий:

• Выделим диапазон значений **D29:D44**, а затем в главном меню выберем ***Вставка / Диаграмма*** и, следуя рекомендациям *Мастера Диаграмм*, построим *График с маркерами, помечающими точки данных*.

• На полученной диаграмме выделим *Область построения диаграммы* и в главном меню выберем ***Диаграмма / Добавить линию тренда***. В диалогом окне на вкладке *Тип* выберем **Линейная**, а на вкладке *Параметры* поставим флажок «**показать уравнение на диаграмме**». Получим рис. 6.4:

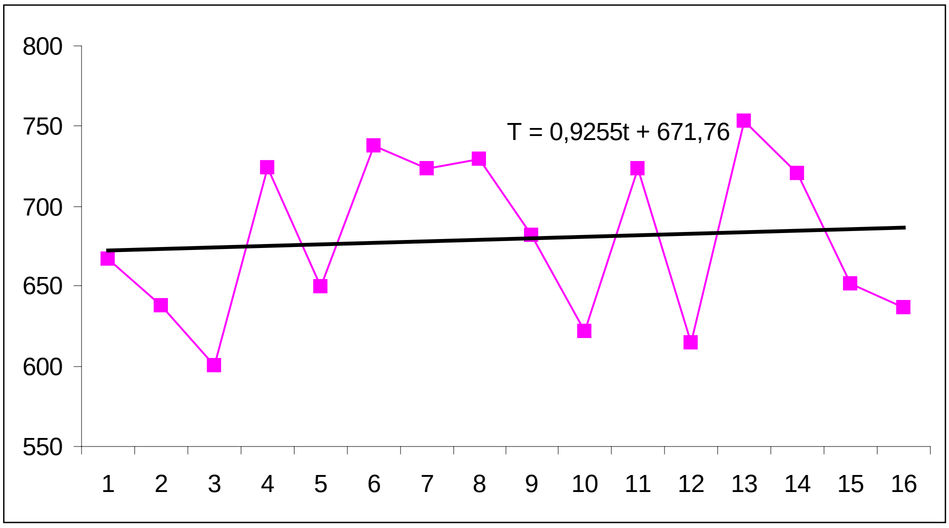


Рис. 6.4. Линейный тренд аддитивной модели

В результате аналитического выравнивания линейный тренд имеет вид: T =671,76+0,9255$\cdot t$. Подставляя в это уравнение значения t =1,2, …, 16, найдем уровни T для каждого момента времени (столбец **E** табл. 6.10).

Шаг 5. Найдем значения уровней ряда. Для этого прибавим к уровням T значения сезонной компоненты для соответствующих кварталов (столбец **F** табл. 6.10). Например, **F29 =C29+E29**.

На одном графике (рис. 6.5) построим фактические значения уровней временного ряда (y_t) и теоретические $(T+S)$, полученные по аддитивной модели.

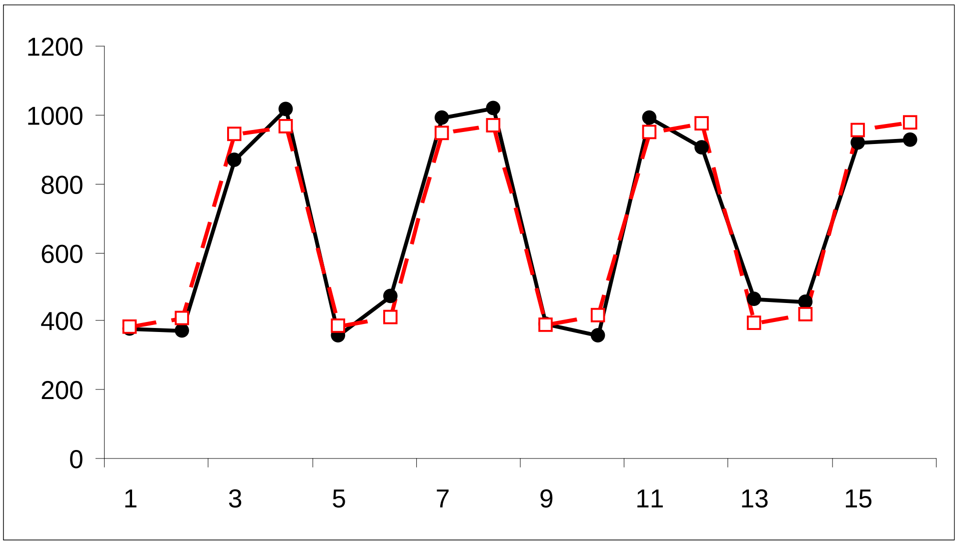


Рис. 6.5. Фактические и теоретические значения уровней временного ряда, полученные по аддитивной модели

Вычислим абсолютные ошибки (столбец **G** табл. 6.10): $E = y_t - (T + S)$. Например, в ячейку **G29** поместим =**B29-F29**. В столбце **H** найдем E^2.

Для оценки качества построенной модели вычислим:

$$R^2 = 1 - \frac{E^2}{(y_t - \bar{y})^2} = 0{,}97.$$

Следовательно, аддитивная модель объясняет 97% общей вариации уровней временного ряда объема платных услуг населению по кварталам за 4 года.

Шаг 6. Прогнозирование по аддитивной модели. Прогнозное значение уровня временного ряда в аддитивной модели $F_t = T + S$. Для определения T воспользуемся уравнением тренда T =671,76+0,9255$\cdot t$. Получим:

T_{17}=671,76+0,9255·17=687,494; T_{18}=671,76+0,9255·18=688,419.

Значения сезонных компонент за соответствующие кварталы: S_1=-292,35 и S_2=-266,81. Значит, $F_{17} = T_{17} + S_1$ =395,14 и $F_{18} = T_{18} + S_2$ = 421,61.

Т.е. в первые два квартала 2008 г. следовало ожидать предоставления объема платных услуг населению на 395,14 и 421,61 млн. руб. соответственно.

Пример 5 По данным табл. 6.1 построить мультипликативную модель временного ряда, оценить качество построенной модели и сделать прогноз об объеме платных услуг населению на I и II кварталы 2008 г.

Решение:

Шаг 1. Полностью совпадает с методикой построения аддитивной модели. Можно скопировать табл. 6.8 без столбца **F** на новый лист Excel (табл. 6.11).

Таблица 6.11

	A	B	C	D	E	F
1	t	y_t	Итого за	Скользящая	Центрированная	Оценка

			четыре квартала	средняя за четыре квартала	скользящая средняя	сезонной компоненты
2	1	375	–	–	–	–
3	2	371	2630	657,50	–	–
4	3	869	2612	653,00	655,250	1,326
5	4	1015	2712	678,00	665,500	1,525
6	5	357	2835	708,75	693,375	0,515
7	6	471	2840	710,00	709,375	0,664
8	7	992	2873	718,25	714,125	1,389
9	8	1020	2757	689,25	703,750	1,449
10	9	390	2757	689,25	689,250	0,566
11	10	355	2642	660,50	674,875	0,526
12	11	992	2713	678,25	669,375	1,482
13	12	905	2812	703,00	690,625	1,310
14	13	461	2740	685,00	694,000	0,664
15	14	454	2762	690,50	687,750	0,660
16	15	920	–	–	–	–
17	16	927	–	–	–	–

Шаг 2. Найдем оценки сезонной компоненты как частное от деления фактических уровней ряда на центрированные скользящие средние (столбец **F** табл. 4.8). Например, в ячейку **F4** поместим =**B4/E4**.

Составим табл. 6.12, распределив значения столбца **F** таблицы 6.11 по кварталам и годам. С помощью статистической функции **СРЗНАЧ(диапазон)** найдем средние за каждый квартал (по всем годам) оценки сезонной компоненты $\overline{S}_i$. В ячейке **G25** рассчитаем сумму средних с помощью статистической функции =**СУММ(C25:F25)**.

Таблица 6.12

	A	B	C	D	E	F	G
19	Показа-тели	Год	№ квартала, i				Сумма
20			I	II	III	IV	
21		2004	–	–	1,326	1,525	
22		2005	0,515	0,664	1,389	1,449	
23		2006	0,566	0,526	1,482	1,310	
24		2007	0,664	0,660	–	–	
25	$\overline{S}_i$		0,582	0,617	1,399	1,428	4,026
26	S_i		0,578	0,613	1,390	1,419	4,0

Вычислим корректирующий коэффициент: $k=4/4{,}026=0{,}9936$.

Рассчитаем скорректированные значения сезонной компоненты $S_i = \overline{S}_i \cdot k$ и занесем полученные данные в табл. 6.12.

Так же как и в аддитивной модели считается, что сезонные воздействия за период взаимопогашаются. В мультипликативной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов в цикле. В нашем случае число периодов одного цикла равно 4, как и сумма S_i в ячейке **G26**.

Составим табл. 6.13, в которой в столбец **C** поместим вычисленные S_i. Причем через каждые четыре квартала эти значения будут повторяться.

Таблица 6.13

	A	B	C	D	E	F	G	H	I
28	t	y_t	S_i	y_t / S_i	T	$T \cdot S$	E	$(y_t - T \cdot S)^2$	$(y_t - \overline{y})^2$
29	1	375	0,578	648,87	654,91	378,49	0,9908	12,210	92796,39
30	2	371	0,613	605,46	658,19	403,31	0,9199	1043,975	95249,39
31	3	869	1,390	625,12	661,47	919,54	0,9450	2554,536	35862,89
32	4	1015	1,419	715,21	664,76	943,40	1,0759	5126,314	112476,39
33	5	357	0,578	617,72	668,04	386,08	0,9247	845,629	104086,89
34	6	471	0,613	768,66	671,32	411,35	1,1450	3557,749	43524,39
35	7	992	1,390	713,60	674,60	937,79	1,0578	2938,908	97578,14
36	8	1020	1,419	718,73	677,88	962,03	1,0603	3360,680	115855,14
37	9	390	0,578	674,82	681,16	393,67	0,9907	13,433	83882,64
38	10	355	0,613	579,35	684,44	419,40	0,8465	4146,800	105381,39
39	11	992	1,390	713,60	687,72	956,03	1,0376	1293,539	97578,14
40	12	905	1,419	637,70	691,01	980,66	0,9229	5723,761	50793,89
41	13	461	0,578	797,67	694,29	401,25	1,1489	3569,985	47796,89
42	14	454	0,613	740,92	697,57	427,44	1,0621	705,530	50906,64
43	15	920	1,390	661,80	700,85	974,28	0,9443	2946,327	57780,14
44	16	927	1,419	653,20	704,13	999,28	0,9277	5224,753	61194,39
45	Сумма	10874						43064,13	1252743,75
46	Среднее	679,625	–	–	–	–	–	–	–

Шаг 3. Разделим каждый уровень исходного ряда на соответствующие значения сезонной компоненты. В результате в столбце **D** табл. 6.13 получим величины $T \cdot E = Y / S$, которые содержат только тенденцию и случайную компоненту.

Шаг 4. Определим компоненту T в мультипликативной модели выравниванием ряда $(T \cdot E)$ с помощью линейного тренда. Порядок действий:

• Выделим диапазон значений **D29:D44**, а затем в главном меню выберем ***Вставка / Диаграмма*** и, следуя рекомендациям *Мастера Диаграмм*, построим *График с маркерами, помечающими точки данных*.

• На полученной диаграмме выделим *Область построения диаграммы* и в главном меню выберем ***Диаграмма / Добавить*** линию тренда. В диалогом окне на вкладке *Тип* выберем **Линейная**, а на вкладке *Параметры* поставим флажок «**показать уравнение на диаграмме**». Получим рис. 6.6:

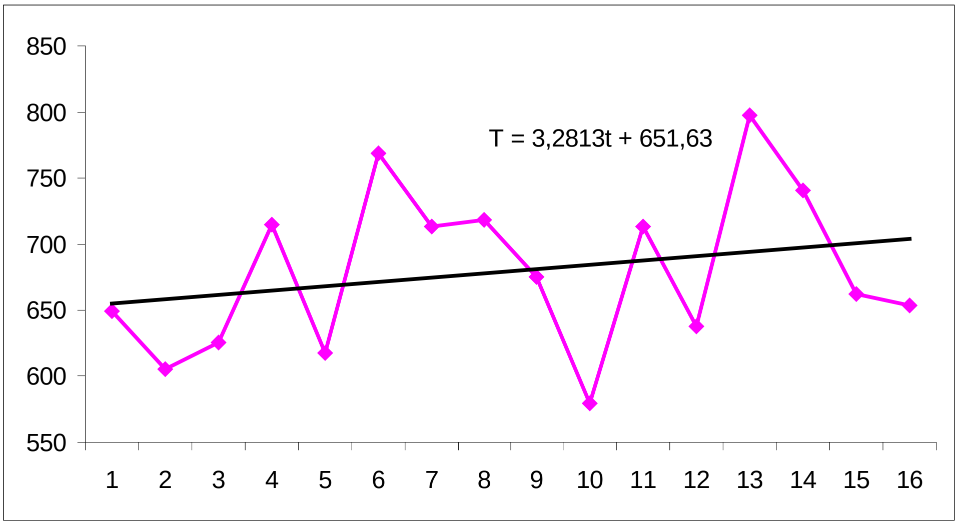


Рис. 6.6. Линейный тренд мультипликативной модели

В результате аналитического выравнивания линейный тренд имеет вид: T =651,63+3,2813$\cdot t$. Подставляя в это уравнение значения t =1,2, …, 16, найдем уровни T для каждого момента времени (столбец **E** табл. 6.13).

Шаг 5. Найдем уровни ряда, умножив значения T на соответствующие значения сезонной компоненты (столбец **F** табл.6.13). Например, в ячейку **F29** поместим =**C29*E29**.

На одном графике (рис. 6.10) построим фактические значения уровней временного ряда (y_t) и теоретические $(T \cdot S)$, полученные по мультипликативной модели.

Вычислим абсолютные ошибки (столбец **G** табл. 6.13): $E = y_t / (T \cdot S)$. Например, в ячейку **G29** поместим =**B29/F29**.

В столбце **H** найдем $(y_t - T \cdot S)^2$.

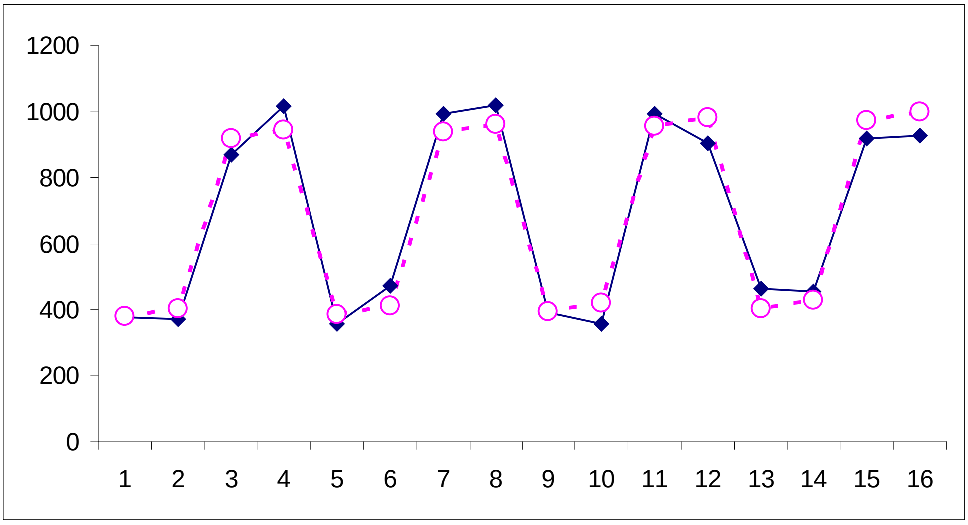


Рис. 6.7. Фактические и теоретические значения уровней временного ряда, полученные по мультипликативной модели

Для оценки качества построенной мультипликативной модели вычислим:

$$R^2 = 1 - \frac{(y_t - T \cdot S)^2}{(y_t - \bar{y})^2} = 0{,}966.$$

Следовательно, мультипликативная модель объясняет 96,6% общей вариации уровней временного ряда объема платных услуг населению по кварталам за 4 года.

Сравнивая показатели детерминации аддитивной и мультипликативной моделей, делаем вывод, что они примерно одинаково аппроксимируют исходные данные.

Шаг 6. Прогнозирование по мультипликативной модели. Прогнозное значение F_t уровня временного ряда в мультипликативной модели есть произведение трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда T =651,63+3,2813$\cdot t$. Получим:

$$T_{17} = 651{,}63 + 3{,}2813 \cdot 17 = 707{,}412;$$

$$T_{18} = 651{,}63 + 3{,}2813 \cdot 18 = 710{,}693.$$

Значения сезонных компонент за соответствующие кварталы равны: S_1 =0,578 и S_2=0,613. Таким образом,

$$F_{17} = T_{17} \cdot S_1 = 408{,}84;$$

$$F_{18} = T_{18} \cdot S_2 = 435{,}48.$$

Т.е. в первые два квартала 2008 г. следовало ожидать предоставления объема платных услуг населению на 408,84 и 435,48 млн. руб. соответственно.

Таким образом, аддитивная и мультипликативная модели дают примерно одинаковый результат по прогнозу.

ПРИЛОЖЕНИЕ

СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

1. Таблица значений F -критерия Фишера при уровне значимости $\alpha = 0{,}05$

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

Продолжение

$k_2 \backslash k_1$	1	2	3	4	5	6	8	12	24	∞
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1

2. Таблица значений t -критерия Стьюдента (двухсторонний)

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

3. Значения статистик Дарбина-Уотсона при 5%-ном уровне значимости

n	$m=1$		$m=2$		$m=3$		$m=4$		$m=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1,78				
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,85	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,99
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83