САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ АЭРОКОСМИЧЕСКОГО
ПРИБОРОСТРОЕНИЯ

ЭКОНОМЕТРИКА

В данном учебном пособии рассмотрен ряд вопросов, раскрывающих основное содержание дисциплины «Эконометрика», формулируются цели и задачи этого направления. Приводятся темы практических работ, а также вопросы для контроля знаний, рекомендуемая литература. Основное внимание уделяется построению эконометрических моделей на основе пространственных данных и временных рядов. Приводятся краткие методические положения, включающие основные понятия, определения, формулы. Рассмотрены примеры решения типовых задач, представлены процедуры, математический аппарат и программные средства моделирования задач эконометрического анализа.

СОДЕРЖАНИЕ

Предисловие

Целью преподавания дисциплины является получение знаний в области построения эконометрических моделей и определения возможностей использования моделей для описания, анализа и прогнозирования реальных экономических процессов как на микро-, так и на макроуровне. Основными задачами изучения дисциплины являются:

- методология принятия решений о спецификации и идентификации моделей;
- ознакомление с методами и приемами интерпретации результатов эконометрического моделирования;
- изучение принципов выбора метода оценки параметров моделей;
- выработка устойчивых практических навыков разработки прогнозных оценок.

Каждый студент должен в соответствии с индивидуальным вариантом выполнить три практических задачи (аналогично приведенным примерам) и ответить на вопросы в конце каждой главы.

Введение

Эконометрика – быстроразвивающаяся отрасль науки, цель которой состоит в том, чтобы придать количественные меры экономическим отношениям. Слово «эконометрика» представляет собой комбинацию двух слов: «экономика» и «метрика» (от греч. «метрон»). Таким образом, сам термин подчеркивает специфику, содержание эконометрики как науки: количественное выражение тех связей и соотношений, которые раскрыты и обоснованы экономической теорией.

Зарождение эконометрики является следствием междисциплинарного подхода к изучению экономики. Эта наука возникла в результате взаимодействия и объединения в особый «сплав» трех компонент: экономической теории, статистических и математических методов. Впоследствии к ним присоединилось развитие вычислительной техники как условие развития эконометрики.

Существуют различные варианты определения эконометрики:

1) Расширенные, при которых к эконометрике относят все, что связано с измерениями в экономике;

2) Узко инструментально ориентированные, при которых понимают определенный набор математико-статистических средств, позволяющих верифицировать модельные соотношения между анализируемыми экономическими показателями.

Эконометрика – это самостоятельная научная дисциплина, объединяющая совокупность теоретических результатов, приемов, методов и моделей, предназначенных для того, чтобы на базе экономической теории, экономической статистики и экономических измерений, математико-статистического инструментария придавать конкретное количественное выражение общим (качественным) закономерностям, обусловленным экономической теорией.

Становление и развитие эконометрического метода происходили на основе так называемой высшей статистики – на методах парной и множественной регрессии, парной, частной и множественной корреляции, выделения тренда и других компонент временного ряда, на статистическом оценивании. Основной базой для эконометрических исследований служат данные официальной статистики, либо данные бухгалтерского учета.

Эконометрическое моделирование реальных социально-экономических процессов и систем обычно преследует два типа конечных прикладных целей (или одну из них): 1) прогноз экономических и социально-экономических показателей, характеризующих состояние и развитие анализируемой системы; 2) имитацию различных возможных сценариев социально-экономического развития анализируемой системы (многовариантные сценарные расчеты, ситуационное моделирование).

1 Парная регрессия и корреляция
1.1 Методические указания

В экономике широко используются методы статистики. Ставя цель дать количественное описание взаимосвязей между экономическими переменными, эконометрика, прежде всего, связана с методами регрессии и корреляции.

В зависимости от количества факторов, включенных в уравнение регрессии, принято различать простую (парную) и множественную регрессии.

Парная регрессия – уравнение связи двух переменных y и x:

$$y = f(x),$$

где y – зависимая переменная (результативный признак),

x – независимая, объясняющая переменная (признак-фактор).

Различают *линейные* и *нелинейные* регрессии.

Линейная регрессия: $y = a + b \cdot x + \varepsilon$

Нелинейные регрессии делятся на два класса:

- регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам,
- регрессии, нелинейные по оцениваемым параметрам.

Регрессии, *нелинейные по объясняющим переменным*:

- Полиномы разных степеней $y = a + b_1 x + b_2 x^2 + b_3 x^3 + \varepsilon$
- Равносторонняя гипербола $y = a + \frac{b}{x} + \varepsilon$

Регрессии, *нелинейные по оцениваемым параметрам*:

- Степенная $y = a \cdot x^b + \varepsilon$
- Показательная $y = a \cdot b^x + \varepsilon$
- Экспоненциальная $y = e^{a+bx} + \varepsilon$

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессий, линейных по параметрам, используют *метод* *наименьших квадратов* (МНК). МНК позволяет получить такие оценки параметров a и b, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических $\hat{y}_x$ минимальна:

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

Знак «^» означает, что между переменными x и y нет строгой функциональной зависимости, поэтому практически в каждом отдельном случае величина y складывается из двух слагаемых:

$$y = \hat{y}_x + \varepsilon$$

где y – фактическое значение результативного признака; $\hat{y}_x$ – теоретическое значение результативного признака, найденное исходя из уравнения регрессии; ε – случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии.

Случайная величина ε называется также возмущением. Она включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения. Ее присутствие в модели порождено тремя источниками: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных.

От правильно выбранной спецификации модели зависит величина случайных ошибок: они тем меньше, чем в большей мере теоретические значения результативного признака $\hat{y}_x$, подходят к фактическим данным y.

К ошибкам спецификации относятся неправильный выбор той или иной математической функции для $\hat{y}_x$ и недоучет в уравнении регрессии какого-либо существенного фактора, т. е. использование парной регрессии вместо множественной.

Для линейных и нелинейных уравнений, приводимых к линейным, решается следующая система относительно a и b:

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

Можно воспользоваться готовыми формулами, которые вытекают из этой системы:

$$a = \bar{y} - b\bar{x},$$

$$b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}.$$

Тесноту связи изучаемых явлений оценивает линейный коэффициент *парной корреляции* r_{xy} и *индекс корреляции* ρ_{xy}. Для линейной регрессии $(-1 \le r_{xy} \le 1)$, причем, если коэффициент регрессии $b > 0$, то $0 \le r_{xy} \le 1$ и, наоборот, при $b < 0$, $-1 \le r_{xy} \le 0$.

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y}$$

Оценку качества построенной модели дает коэффициент детерминации, а также средняя ошибка аппроксимации.

Коэффициент детерминации (квадрат линейного коэффициента корреляции ${r_{xy}}^2$) характеризует долю дисперсии результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака:

$${r_{xy}}^2 = \frac{\sigma^2_{y\,объясн}}{\sigma^2_{y\,общ}} = \frac{\sum(\hat{y}_x - \bar{y})^2}{\sum(y - \bar{y})^2}$$

Средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических.

Фактические значения результативного признака отличаются от теоретических, рассчитанных по уравнению регрессии, т.е. y и $\hat{y}_x$. Чем меньше это отличие, тем ближе теоретические значения подходят к эмпирическим данным, лучше качество модели. Величина отклонений фактических и расчетных значений результативного признака $(y - \hat{y}_x)$ по каждому наблюдению представляет собой ошибку аппроксимации. Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации как среднюю арифметическую простую:

$$\overline{A} = \frac{1}{n}\sum\left|\frac{y - \hat{y}}{y}\right| \cdot 100, \%.$$

Допустимый предел значений $\overline{A}$ - не более 8-10% (это свидетельствует о хорошем подборе модели к исходным данным).

Средний коэффициент эластичности $\overline{Э}$ показывает, на сколько процентов в среднем по совокупности изменится результат y от своей средней величины $\bar{y}$ при изменении фактора x на 1% от своего среднего значения:

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

Оценка значимости уравнения регрессии в целом дается с помощью *F-критерия Фишера*. При этом выдвигается нулевая гипотеза, что коэффициент регрессии равен нулю, т.е. $b=0$, и, следовательно, фактор x не оказывает влияния на результат y.

Непосредственному расчету F-критерия предшествует анализ дисперсии. Центральное место в нем занимает разложение общей суммы квадратов отклонений переменной y от среднего значения $\bar{y}$ на две части: «объясненную» и «необъясненную»:

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

где $\sum(y - \bar{y})^2$ – общая сумма квадратов отклонений;

$\sum\left(\hat{y}_x - \bar{y}\right)^2$ - сумма квадратов отклонений, обусловленная регрессией (*объясненная* или *факторная*);

$\sum\left(y - \hat{y}_x\right)^2$ - остаточная сумма квадратов отклонений (*необъясненная*).

Если сумма квадратов отклонений, обусловленная регрессией, будет больше остаточной суммы квадратов, то уравнение регрессии статистически значимо и фактор *x* оказывает существенное влияния на результат *y*. Это равносильно тому, что коэффициент $r_{xy}^{\ 2}$ будет приближаться к единице.

При расчете объясненной суммы квадратов $\sum\left(\hat{y}_x - \bar{y}\right)^2$ используются теоретические (расчетные) результативного признака $\hat{y}_x$, найденные по линии регрессии: $\hat{y}_x = a + b \cdot x$.

Сумма квадратов отклонений, обусловленных линейной регрессией, составляет:

$$\sum\left(\hat{y}_x - \bar{y}\right)^2 = b^2 \cdot \sum\left(x - \bar{x}\right)^2.$$

Поскольку при заданном объеме наблюдений по *x* и *y* факторная сумма квадратов при линейной регрессии зависит только от одной константы коэффициента регрессии *b*, то данная сумма квадратов имеет одну степень свободы. Число степеней свободы – это число свободы независимого варьирования признака; оно связано с числом единиц совокупности *n* и с числом определяемых по ней констант. Существует равенство между числом степеней свободы общей, факторной и остаточной суммы квадратов. Число степеней свободы остаточной суммы квадратов при линейной регрессии составляет *n-2*. Число степеней для общей суммы квадратов составляет *n-1*, так как для $\sum\left(y - \bar{y}\right)^2$ требуется *n-1* независимых отклонений (из *n* единиц после расчета среднего уровня свободно варьируются лишь *n-1*, число отклонений).

Следовательно, имеем два равенства:

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

$$n-1=1+(n-2).$$

Разделив каждую сумму квадратов на соответствующее ей число степеней свободы, получим средний квадрат отклонений, или, что то же самое, дисперсию на одну степень свободы D.

$$D_{общ}=\frac{\sum(y-\bar{y})^2}{n-1};$$

$$D_{факт}=\frac{\sum(\hat{y}_x-\bar{y})^2}{1};$$

$$D_{ост}=\frac{\sum(y-\hat{y}_x)^2}{n-2}.$$

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину F-критерия для проверки нулевой гипотезы $(H_0: D_{факт}=D_{ост})$:

$$F=\frac{D_{факт}}{D_{ост}}.$$

Если нулевая гипотеза справедлива, то факторная и остаточная дисперсии не отличаются друг от друга. Фактическое значение F-критерия Фишера сравнивается с табличным значением $F_{табл}(\alpha;k_1;k_2)$ при уровне значимости α и степенях свободы $k_1=m$ и $k_2=n-m-1$.

Табличное значение F-критерия – это максимальная величина отношения дисперсий, которая может иметь место при случайном их расхождении для данного уровня вероятности наличия нулевой гипотезы. Вычисленное значение F-критерия признается достоверным (отличным от единицы), если оно больше табличного. В этом случае нулевая гипотеза об отсутствии связи признаков отклоняется и делается вывод о существенности этой связи: $F_{факт}>F_{табл}$. H_0отклоняется.

Если же величина окажется меньше табличной $F_{факт}<F_{табл}$, то вероятность нулевой гипотезы выше заданного уровня (например, 0.05) и она не может быть отклонена без серьезного риска сделать неправильный вывод

о наличии связи. В этом случае уравнение регрессии считается статистически незначимым, H_0 не отклоняется.

$$F=\frac{\sum(\hat{y}-\bar{y})^2/m}{\sum(y-\hat{y})^2/(n-m-1)}=\frac{r_{xy}^2}{1-r_{xy}^2}(n-2),$$

где n – число единиц совокупности;

m – число параметров при переменных x.

1.2 Решение типовых задач

Задача 1.1 По территориям региона приводятся данные (таблица 1.1.1).

Таблица 1.1.1.

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, у.е.	Среднедневная заработная плата, у.е.
1	17,82	13,75
2	18,59	15,29
3	15,29	8,69
4	15,51	9,02
5	18,15	14,41
6	20,13	18,37
7	15,07	8,14
8	15,62	9,35
9	17,49	12,98
10	17,16	12,43

Требуется:

— построить линейное уравнение парной регрессии y от x;

— рассчитать линейный коэффициент парной корреляции и среднюю ошибку аппроксимации;

— оценить модель через ошибку аппроксимации $\overline{A}$ и F-критерий.

Решение. Для расчета параметров уравнения линейной регрессии строим расчетную таблицу (таблица 1.1.2).

Таблица 1.1.2

	y	x	yx	x^2	y^2	$\hat{y}_x$	$y - \hat{y}_x$	A_i
1	17,82	13,75	245,03	189,06	317,55	17,84	-0,02	0,084175
2	18,59	15,29	284,24	233,78	345,59	18,61	-0,02	0,080689
3	15,29	8,69	132,87	75,52	233,78	15,31	-0,02	0,098103
4	15,51	9,02	139,90	81,36	240,56	15,47	0,04	0,257898
5	18,15	14,41	261,54	207,65	329,42	18,17	-0,01	0,082645
6	20,13	18,37	369,79	337,46	405,22	20,15	-0,02	0,074516
7	15,07	8,14	122,67	66,26	227,10	15,03	0,04	0,265428
8	15,62	9,35	146,05	87,42	243,98	15,64	-0,02	0,096031
9	17,49	12,98	227,02	168,48	305,90	17,45	0,04	0,228702
10	17,16	12,43	213,30	154,50	294,47	17,18	-0,02	0,087413
Итого	170,83	122,43	2142,40	1601,50	2943,58	170,82	-	1,36
Ср.знач.	17,08	12,24	214,24	160,15	294,36	17,08	-	0,14

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - (\bar{x})^2} = \frac{214{,}24 - 17{,}08 \cdot 12{,}24}{160{,}15 - 12{,}24^2} = 0{,}50,$$

$$a = \bar{y} - b \cdot \bar{x} = 17{,}08 - 0{,}50 \cdot 12{,}24 = 10{,}96.$$

Получено уравнение регрессии: $\hat{y} = 10{,}96 + 0{,}50 \cdot x.$

С увеличением среднедушевого прожиточного минимума на одну у.е. среднедневная заработная плата возрастет в среднем на 0,50 у.е.

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

Тесноту линейной связи оценит коэффициент корреляции:

$$r_{xy} = b\frac{\sqrt{\overline{x^2} - (\bar{x})^2}}{\sqrt{\overline{y^2} - (\bar{y})^2}} = 0{,}999;$$

$$r_{xy}^2 = 0{,}99.$$

Это означает, взаимосвязь между параметрами прямая и тесная, и что 99% вариации заработной платы (*y*) объясняется вариацией фактора *x* – среднедушевого прожиточного минимума.

Качество модели определяет средняя ошибка аппроксимации:

$$\bar{A} = \frac{1}{n}\sum A_i = \frac{1{,}36}{10} = 0{,}14\%$$

Качество построенной модели хорошее, так как $\bar{A}$ не превышает 8-10%.

Рассчитаем *F*-критерий:

$$F_{факт.} = \frac{0{,}99}{1 - 0{,}99} \cdot 8 = 19941{,}81,$$

Полученное значение указывает на необходимость применять гипотезу *H*₁ о неслучайной природе выявленной зависимости, т. к. $F_{факт} \rangle F_{табл} = 5{,}32$ (табличные значения *F*-критерия приведены в приложении).

Задача 1.2 По группе предприятий, производящих однородную продукцию известно, как зависит себестоимость единицы продукции *y* от факторов, приведенных в таблице 1.2.1.

Таблица 1.2.1

Признак-фактор	Уравнение парной регрессии	Среднее значение фактора
Объем производства, млн. руб., x_1	$\hat{y}_{x_1} = 0{,}65 + 58{,}85 \cdot 1/x_1$	$\bar{x}_1 = 2{,}55$
Трудоемкость единицы продукции, чел.-час., x_2	$\hat{y}_{x_2} = 9{,}50 + 9{,}85x_2$	$\bar{x}_2 = 1{,}58$
Оптовая цена за 1т энергоносителя, млн. руб, x_3	$\hat{y}_{x_3} = 15{,}75 \cdot x_3^{1,55}$	$\bar{x}_3 = 1{,}53$
Доля прибыли, изымаемой государством, %, x_4	$\hat{y}_{x_4} = 14{,}07 \cdot 1{,}01^{x_4}$	$\bar{x}_4 = 28{,}35$

Требуется:

— определить с помощью коэффициентов эластичности силу влияния каждого фактора на результат;

— ранжировать факторы по силе влияния.

Решение:

Для уравнения равносторонней гиперболы $\hat{y}_{x_1} = 0{,}65 + 58{,}85 \cdot \dfrac{1}{x_1}$:

$$\bar{Э}_{yx_1} = f'(x_1)\frac{\bar{x}_1}{\bar{y}} = -\frac{b}{\bar{x}_1^2} \cdot \frac{\bar{x}_1}{a + b/\bar{x}_1} = -\frac{b}{a \cdot \bar{x}_1 + b} = -\frac{58{,}85}{0.65 \cdot 2{,}55 + 58{,}85} = -0{,}973\%.$$

Для уравнения прямой $\hat{y}_{x_2} = 9{,}50 + 9{,}85x_2$:

$$\bar{Э}_{yx_2} = f'(x_2)\frac{\bar{x}_2}{\bar{y}} = \frac{b \cdot \bar{x}_2}{a + b \cdot \bar{x}_2} = \frac{9{,}85 \cdot 1{,}58}{9{,}50 + 9{,}85 \cdot 1{,}58} = 0{,}62\%.$$

Для уравнения степенной зависимости $\hat{y}_{x_3} = 15{,}75 \cdot x_3^{1,55}$:

$$\bar{Э}_{yx_3} = f'(x_3)\frac{\bar{x}_3}{\bar{y}} = a \cdot b \cdot \bar{x}_3^{b-1} \cdot \frac{\bar{x}_3}{a \cdot \bar{x}_3^b} = b = 1{,}55\%.$$

Для уравнения показательной зависимости $\hat{y}_{x_4} = 14{,}07 \cdot 1{,}01^{x_4}$:

$$\bar{Э}_{yx_4} = f'(x_4)\frac{\bar{x}_4}{\bar{y}} = a \cdot b^{\bar{x}_4} \cdot \ln b \cdot \frac{\bar{x}_4}{a \cdot b^{\bar{x}_4}} = \ln b \cdot \bar{x}_4 = \ln 1{,}01 \cdot 26{,}35 = 0{,}26\%.$$

Сравнивая значения $\bar{Э}_{yx_i}$, ранжируем x_i по силе их влияния на себестоимость единицы продукции:

а) $\bar{Э}_{yx_3} = 1{,}55\%$,

б) $\overline{Э}_{yx_1} = -0{,}973\%$,

в) $\overline{Э}_{yx_2} = 0{,}62\%$,

г) $\overline{Э}_{yx_4} = 0{,}26\%$.

Для формирования уровня себестоимости продукции группы предприятий первоочередное значение имеют цены на энергоносители; в гораздо меньшей степени влияют трудоемкость продукции и отчисляемая часть прибыли. Фактором снижения себестоимости выступает размер производства: с ростом его на 1% себестоимость единицы продукции снижается на 0,973%.

1.3 Решение с помощью ППП Excel

Задача 1.3 По территориям региона приводятся данные (таблица 1.3.1).

Таблица 1.3.1.

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, у.е.	Среднедневная заработная плата, у.е.
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

1) Встроенная статистическая функция **ЛИНЕЙН** определяет параметры линейной регрессии $\hat{y} = a + b \cdot x$. Порядок вычисления следующий:

1. Введите исходные данные или откройте существующий файл, содержащий анализируемые данные;
2. Выделите область пустых ячеек 5х2 (5 строк, 2 столбца) для вывода результатов регрессионной статистики или область 1х2 получения только оценок коэффициентов регрессии;
3. Активизируйте **Мастер функций** любым из способов:

   а) в главном меню выберите **Вставка/Функция**;

   б) на панели инструментов **Стандартная** щелкните по кнопке **Вставка Функции**;
4. В окне категория (рис. 1.1) выберите **Статистические**, в окне функция – **ЛИНЕЙН**. Щелкните по кнопке ОК;

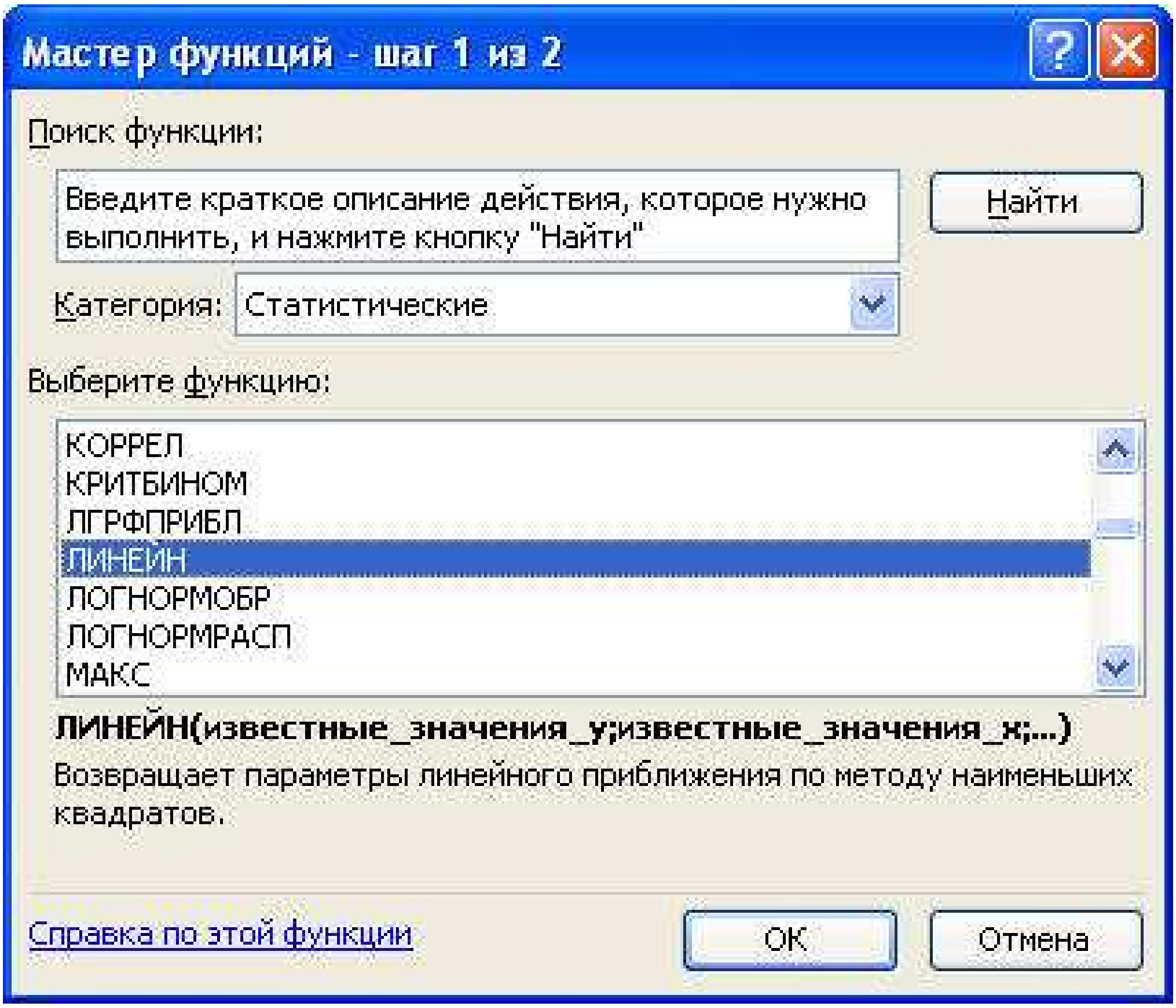


Рис. 1.1 Диалоговое окно «Мастер функции»

5. заполните аргументы функции (рис.1.2):

Известные_значения_y – диапазон, содержащий данные результативного признака;

Известные_значения_x – диапазон, содержащий данные факторов независимого признака;

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

Статистика – логическое значение, которое указывает, выводить дополнительную информацию по регрессионному анализу или нет. Если *Статистика*=1, то дополнительная информация выводится, если

Статистика=0, то выводятся только оценки параметров уравнения. Щелкните по кнопке **ОК**;

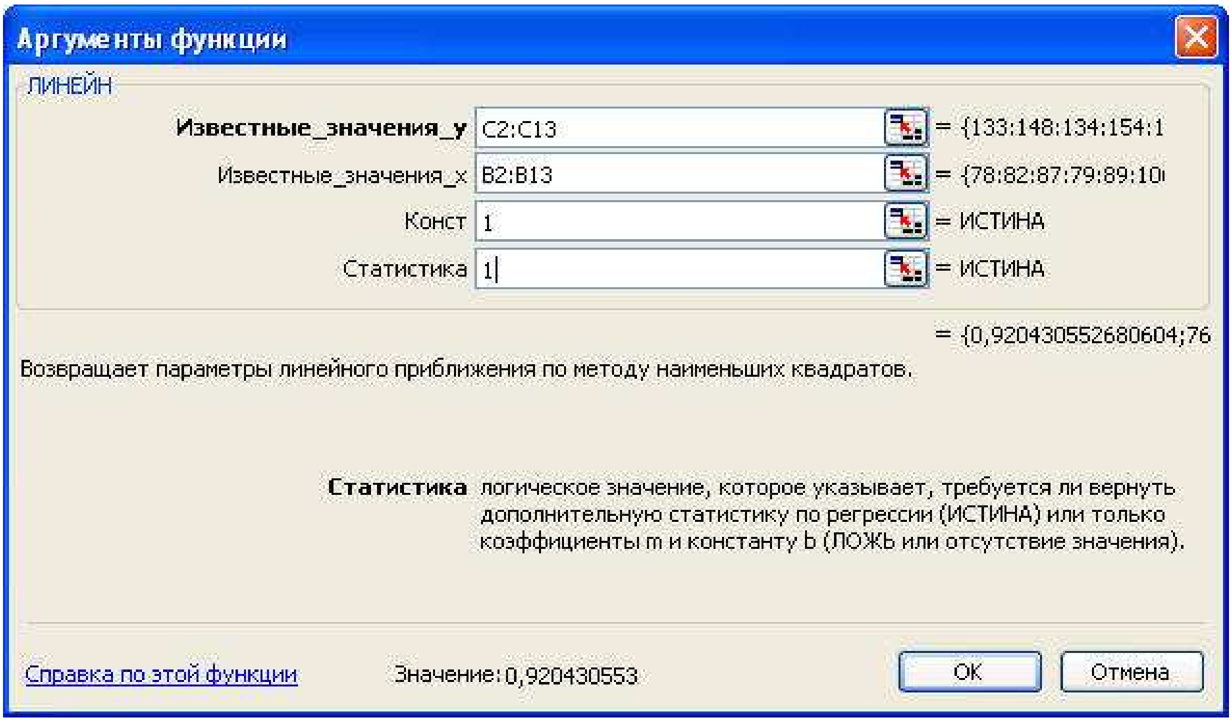


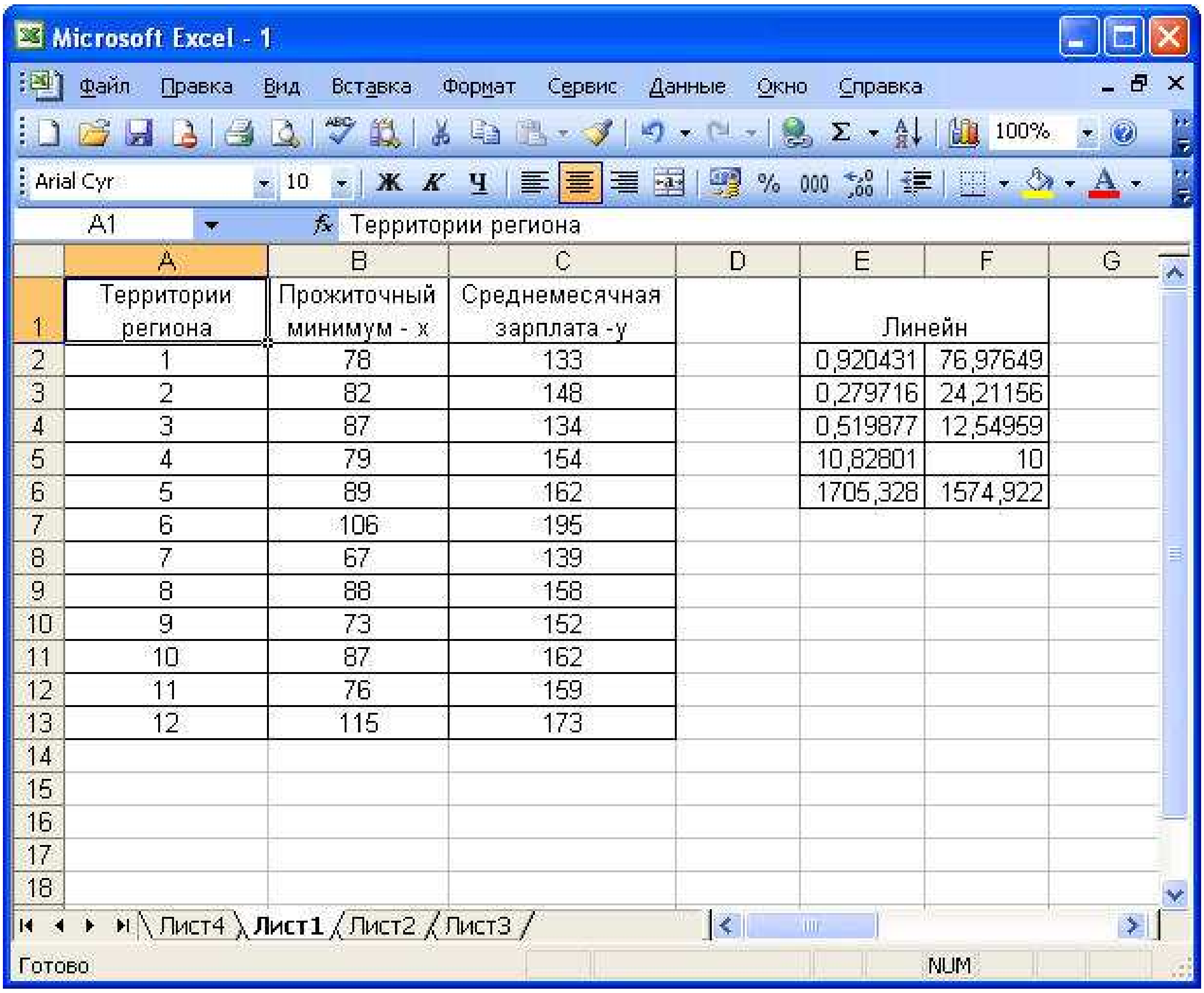
Рис. 1.2 Диалоговое окно ввода аргументов функции **ЛИНЕЙН**

6. В левой верхней ячейке выделенной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу <F2>, а затем – на комбинацию клавиш <CTRL>+<SHIFT>+<ENTER>.

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

Значение коэффициента *b*	Значение коэффициента *a*
Среднеквадратическое отклонение *b*	Среднеквадратическое отклонение *a*
Коэффициент детерминации r_{xy}^2	Среднеквадратическое отклонение *y*
F-статистика	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов

Для данных из задачи 1.4 результат вычисления функции ЛИНЕЙН представлен на рис.1.3.



Территории региона	Прожиточный минимум - x	Среднемесячная зарплата -y		Линейн	
1	78	133		0,920431	76,97649
2	82	148		0,279716	24,21156
3	87	134		0,519877	12,54959
4	79	154		10,82801	10
5	89	162		1705,328	1574,922
6	106	195			
7	67	139			
8	88	158			
9	73	152			
10	87	162			
11	76	159			
12	115	173			

Рис. 1.3 Результат вычисления функции **ЛИНЕЙН**

2) С помощью инструмента анализа данных **Регрессия,** помимо результатов регрессионной статистики, можно получить остатки и графики подбора линии регрессии и остатков. Порядок действий следующий:

1. Проверьте доступ к пакету анализа. В главном меню последовательно выберите **Сервис/Настройки.** Установите флажок **Пакет анализа** (рис.1.4);

2. В главном меню выберите **Сервис/Анализ данных/Регрессия**. Щелкните по кнопке **ОК**;

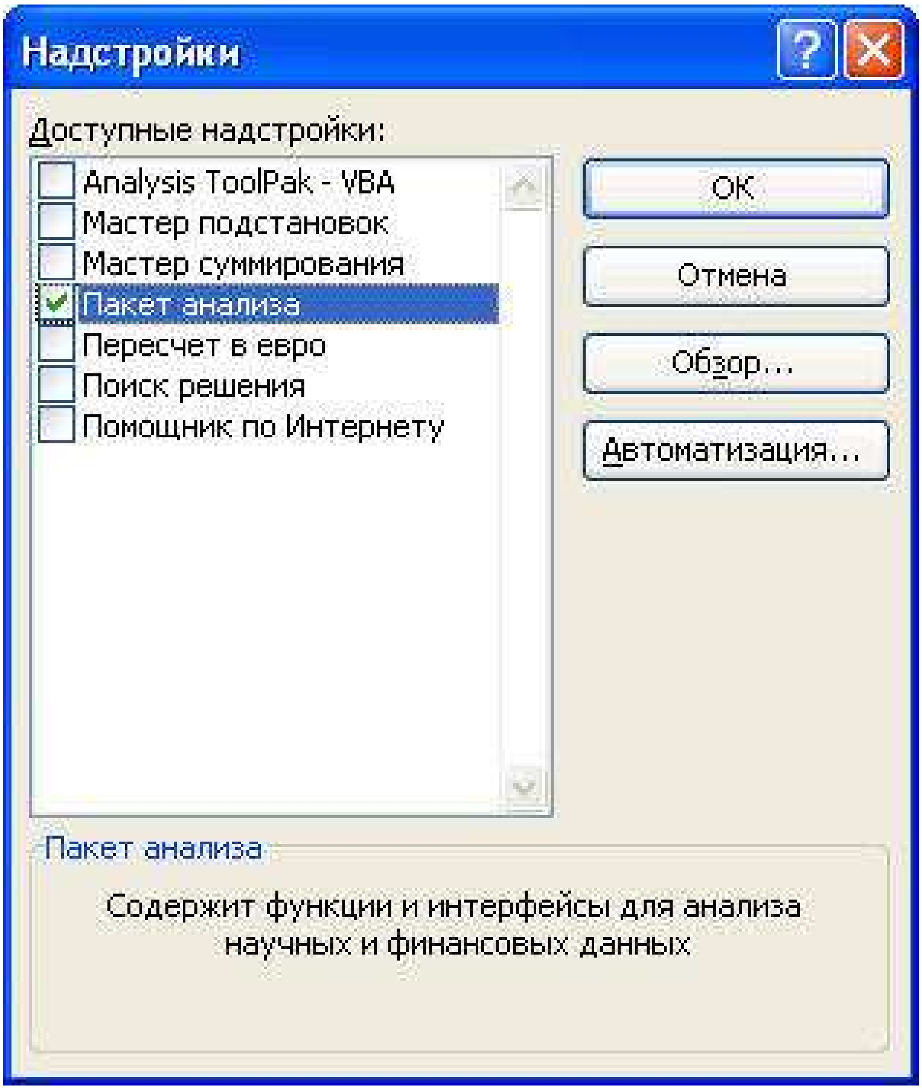


Рис.1.4 Подключение надстройки **Пакет анализа**

3. Заполните диалоговое окно ввода данных и параметров вывода (рис. 1.5):

Входной интервал Y – диапазон, содержащий данные результативного признака;

Входной интервал X – диапазон, содержащий данные факторов независимого признака;

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.

Если необходимо получить информацию и графики остатков, установите соответствующие флажки в диалоговом окне. Щелкните по кнопке **ОК.**

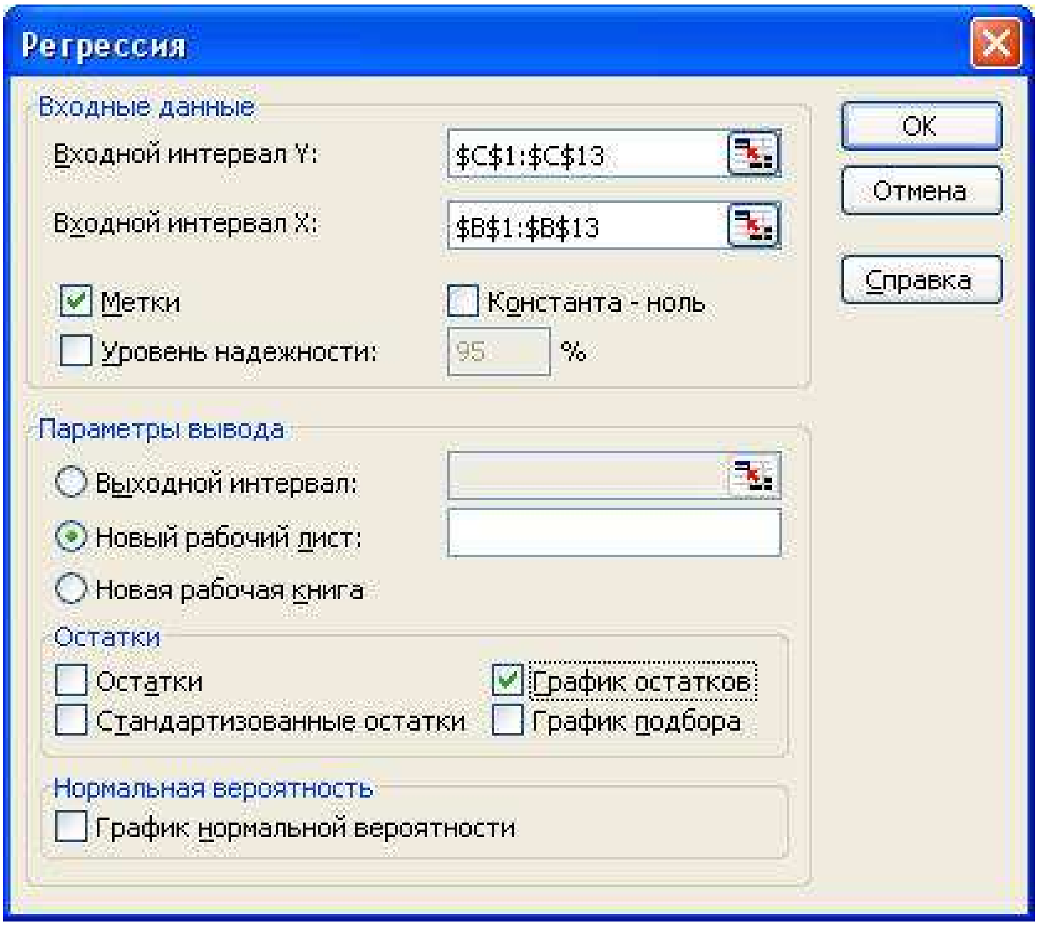
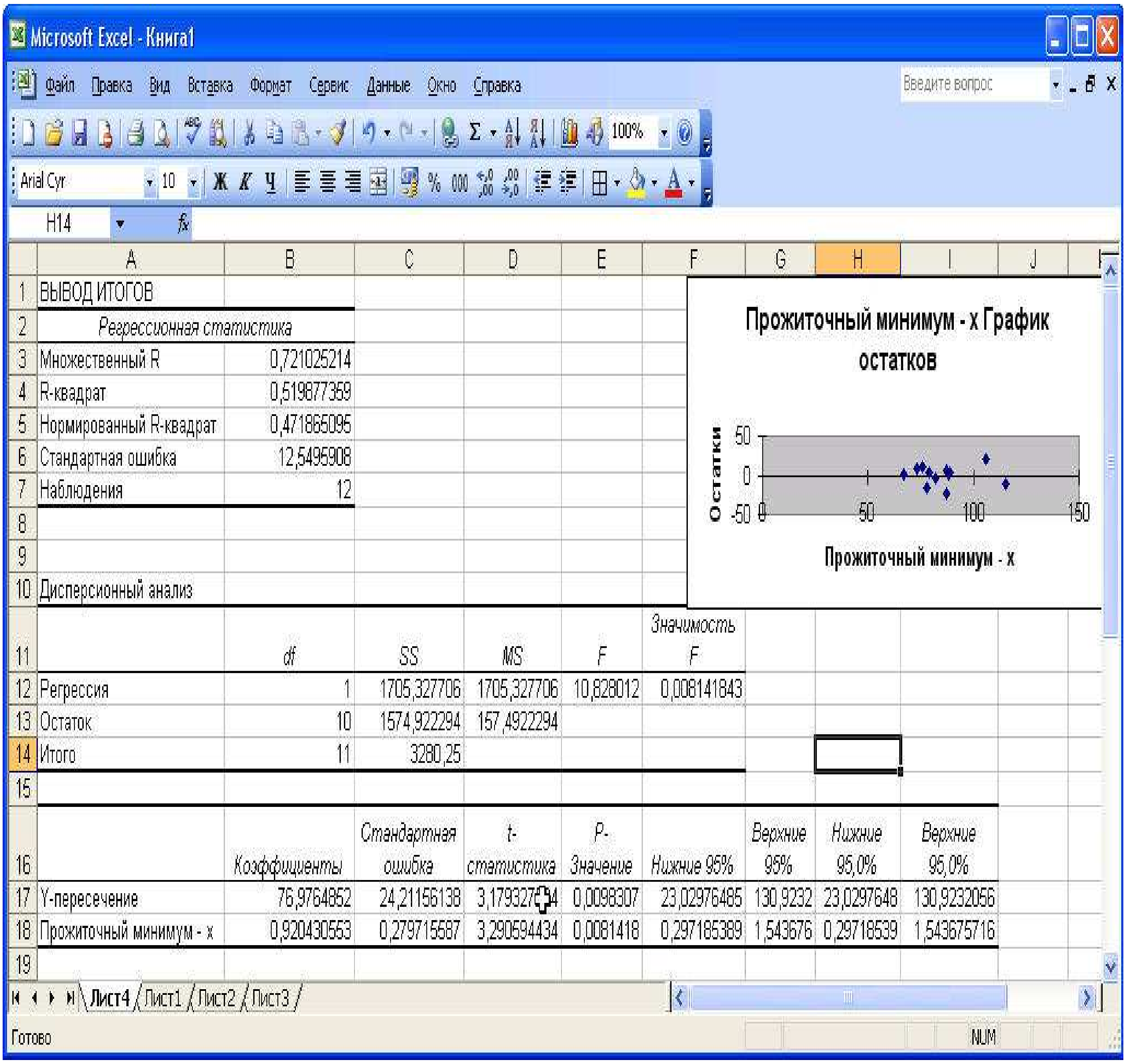


Рис. 1.5 Диалоговое окно ввода параметров инструмента **Регрессия**

Результаты регрессионного анализа для данных из задачи 1.3 представлены на рис. 1.6.



ВЫВОД ИТОГОВ					
Регрессионная статистика					
Множественный R	0,721025214				
R-квадрат	0,519877359				
Нормированный R-квадрат	0,471865095				
Стандартная ошибка	12,5495908				
Наблюдения	12				
Дисперсионный анализ					
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	1705,327706	1705,327706	10,828012	0,008141843
Остаток	10	1574,922294	157,4922294		
Итого	11	3280,25			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	76,9764852	24,21156138	3,179327[illegible]4	0,0098307	23,02976485	130,9232	23,0297648	130,9232056
Прожиточный минимум - x	0,920430553	0,279715587	3,290594434	0,0081418	0,297185389	1,543676	0,29718539	1,543675716

Рис. 1.6 Результат применения инструмента **Регрессия.**

1.4 Контрольные вопросы

1) Что такое коэффициент регрессии? Каковы способы его оценивания?
2) В чем смысл коэффициента детерминации?
3) Для чего применяется критерий Фишера? В чем его суть?
4) Для чего применяется анализ дисперсии в критерии Фишера?
5) Для чего необходим расчет дисперсии на одну степень свободы?
6) Перечислите виды нелинейных моделей.
7) Как определяются коэффициенты эластичности по различным видам регрессионных моделей?
8) Как определяется средняя ошибка аппроксимации? В чем ее смысл?
9) Как определяется коэффициент корреляции? В чем его смысл?
10) Какие виды регрессий существуют? В каких случаях они применяются?

1.5 Пример варианта промежуточного тестирования

1. Вариация уровня заработной платы на 56% объясняет вариацию расходов на продукты питания в общих расходах. Это означает, что:

а) линейный коэффициент парной регрессии равен 56%;
б) коэффициент детерминации равен 44%;
в) линейный коэффициент парной регрессии равен 44%;
г) коэффициент детерминации равен 56%;
д) средняя ошибка аппроксимации равна 56%.

2. После получения уравнения парной линейной регрессии выяснилось, что в среднем расчетные значения отклоняются от фактических на 1,5%. Это означает, что:

а) коэффициент детерминации равен 1,5%;
б) средняя ошибка аппроксимации равна 0,015;
в) линейный коэффициент парной регрессии равен 0,015;
г) значение $F_{расч}$ равно 1,5
д) с вероятностью 1,5% уравнение регрессии статистически значимо.

3. $F_{расч}$=4,5 $F_{табл}$=4,3 Какой вывод можно сделать из анализа предложенных значений:

а) уравнение регрессии статистически незначимо;
б) необходим переход в расчетах от линейной регрессии к нелинейной;
в) уравнение регрессии статистически значимо;
г) в расчетах необходимо использовать множественную, а не парную регрессию.
д) получено неверное значение $F_{расч}$.

4. Каково соотношение между числом степеней свободы общей, факторной и остаточной сумм квадратов:

а) их сумма больше единицы;
б) сумма числа степеней свободы факторной и остаточной равна числу степеней свободы общей суммы квадратов;
в) их сумма равна нулю;
г) сумма числа степеней свободы факторной и остаточной больше числа степеней свободы общей суммы квадратов;
д) сумма числа степеней свободы факторной и остаточной меньше числа степеней свободы общей суммы квадратов.

5. Если при анализе дисперсии сумма квадратов отклонений, обусловленная регрессией, будет больше остаточной суммы квадратов, то уравнение регрессии:

а) статистически незначимо;
б) не рекомендуется использовать в дальнейших расчетах;
в) неверно вследствие неточного расчета коэффициентов;
г) статистически значимо;
д) отклоняется с доверительной вероятностью 95%.

6. Величина F-критерия для проверки гипотезы $H_0 : D_{факт} = D_{ост}$ это:

а) отношение факторной дисперсии к остаточной;
б) произведение факторной и остаточной дисперсий;
в) отношение общей дисперсии к остаточной;
г) произведение общей и остаточной дисперсий;
д) отношение общей дисперсии к произведению факторной и остаточной.

7. Что следует из линейного уравнения y=500+300x, если y - это издержки, а x - объем продукции?

а) с уменьшением объема продукции на 1 ед. издержки производства возрастают в среднем на 500 ед;
б) с увеличением объема продукции на 1 ед. издержки производства возрастают на 800 ед;
в) с увеличением объема продукции в среднем на 1 ед. издержки производства возрастают в среднем на 300 ед;
г) с уменьшением объема продукции в среднем на 300 ед. издержки производства возрастают на 1 ед;
д) с увеличением объема продукции в среднем на 500 ед. издержки производства возрастают в среднем на 300 ед.

2 Множественная регрессия и корреляция
2.1 Методические указания

Множественная регрессия – уравнение связи с несколькими независимыми переменными:

$$y = f(x_1, x_2, ..., x_p),$$

где y - зависимая переменная (результативный признак);

$x_1, x_2, ..., x_p$ - независимые переменные (факторы).

Для построения множественной регрессии используются линейная, степенная, экспоненциальная и гиперболическая функции. Можно использовать и другие функции, приводимые к линейному виду.

Для линейных и нелинейных уравнений, приводимых к линейным, строится следующая система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии:

$$\begin{aligned}
&\sum y = na + b_1\sum x_1 + b_2\sum x_2 + ... + b_p\sum x_p,\\
&\sum yx_1 = a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 + ... + b_p\sum x_1x_p,\\
&..\\
&\sum yx_p = a\sum x_p + b_1\sum x_1x_p + b_2\sum x_2x_p + ... + b_p\sum x_p^2.
\end{aligned}$$

Для ее решения может быть применен метод определителей:

$$a = \frac{\Delta a}{\Delta},\quad b_1 = \frac{\Delta b_1}{\Delta},\ \ldots\ldots,\ b_p = \frac{\Delta b_p}{\Delta},$$

где $\Delta = \begin{vmatrix} n & \sum x_1 & \sum x_2 & ... & \sum x_p \\ \sum x_1 & \sum x_1^2 & \sum x_2x_1 & ... & \sum x_px_1 \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 & ... & \sum x_px_2 \\ ... & ... & ... & ... & ... \\ \sum x_p & \sum x_1x_p & \sum x_2x_p & ... & \sum x_p^2 \end{vmatrix}$ - определитель системы.

Другой вид уравнения множественной регрессии – *уравнение регрессии в стандартизованном масштабе*:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} +,,,+ \beta_p t_{x_p}$$

где $t_y = \dfrac{y - \bar{y}}{\sigma_y}, t_{x_i} = \dfrac{x_i - \bar{x}_i}{\sigma_{x_i}}$ - стандартизованные переменные,

β_i - стандартизованные коэффициенты регрессии.

Связь коэффициентов множественной регрессии b_i со стандартизованными коэффициентами β_i описывается соотношением

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}},$$

Параметр *a* определяется как $a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - ... - b_p\bar{x}_p$.

Для расчета *частных коэффициентов эластичности* применяется следующая формула

$$Э_{y_{x1}} = b_i \frac{x_i}{\hat{y}_{x_i \cdot x_1, x_2, ..., x_{i-1}, ..., x_{i+1}, ..., x_p}}.$$

При построении уравнения множественной регрессии может возникнуть проблема *мультиколлинеарности* факторов, их тесной линейной связи. Чем сильнее мультиколлинеарность факторов, тем менее надежна оценка распределения суммы объясненной вариации по отдельным факторам с помощью метода наименьших квадратов.

Считается, что две переменные *коллинеарны*, т.е. находятся между собой в линейной зависимости, если коэффициент корреляции больше или равен 0,7.

Для оценки мультиколлинеарности факторов используется определитель матрицы парных коэффициентов корреляции между факторами.

Если между факторами существует полная линейная зависимость и все коэффициенты корреляции равны 1, то определитель такой матрицы равен 0:

$$Det|R| = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0$$

Если факторы не коррелированны между собой, то матрица коэффициентов корреляции имеет определитель, равный 1.

Оценки параметров регрессии должны отвечать определенным критериям. Они должны быть несмещенными, состоятельными и эффективными.

Несмещенность оценки означает, что математическое ожидание остатков равно нулю. Если оценки обладают свойством несмещенности, то их можно сравнивать по разным исследованиям.

Оценки считаются *эффективными*, если они характеризуются наименьшей дисперсией. В практических исследованиях это означает возможность перехода от точечного оценивания к интервальному.

Состоятельность оценок характеризует увеличение их точности с увеличением объема выборки. Большой практический интерес представляют те результаты регрессии, для которых доверительный интервал ожидаемого значения параметра регрессии b_i имеет предел значений вероятности, равный единице. Иными словами, вероятность получения оценки на заданном расстоянии от истинного значения параметра близка к единице.

Указанные критерии оценок (несмещенность, состоятельность и эффективность) обязательно учитываются при разных способах оценивания. Метод наименьших квадратов строит оценки регрессии на основе минимизации суммы квадратов остатков. Поэтому очень важно исследовать поведение остаточных величин регрессии ε_i. Условия, необходимые для получения несмещенных, состоятельных и эффективных оценок, представляют собой предпосылки МНК, соблюдение которых желательно для получения достоверных результатов регрессии.

Исследования остатков ε_i предполагают проверку наличия следующих пяти предпосылок МНК:

1) случайный характер остатков;
2) нулевая средняя величина остатков, не зависящая от x_i;
3) гомоскедастичность – дисперсия каждого отклонения ε_i, одинакова для всех значений x;
4) отсутствие автокорреляции остатков – значения остатков ε_i распределены независимо друг от друга;
5) остатки подчиняются нормальному распределению.

Если распределение случайных остатков ε_i не соответствует некоторым предпосылкам МНК, то следует корректировать модель.

Для применения МНК требуется, чтобы дисперсия остатков была *гомоскедастичной*. Это означает, что для каждого значения фактора x_j остатки ε_i имеют одинаковую дисперсию. Если это условие не соблюдается, то имеет место *гетероскедастичность*.

Уравнения множественной регрессии могут включать в качестве независимых переменных качественные признаки (например, профессия, пол, образование, климатические условия, отдельные регионы и т.д.). Чтобы ввести такие переменные в регрессионную модель их необходимо упорядочить и присвоить им те или иные значения, т.е. качественные переменные преобразовать в количественные. Такие переменные в эконометрике принято называть *фиктивными переменными*. Например: 1 – мужской пол, 0 – женский.

Коэффициент регрессии при фиктивной переменной интерпретируется как среднее изменение зависимой переменной при переходе от одной категории (женский пол) к другой (мужской пол) при неизменных значениях остальных параметров. На основе *t*-критерия Стьюдента делается вывод о значимости влияния фиктивной переменной, существенности расхождения между категориями.

2.2 Решение типовых задач

Задача 2.1 Зависимость спроса на компьютеры x_1 от цены на них x_2 и от цены на ноутбуки x_3 представлена уравнением:

$$\lg x_1 = 0{,}1274 - 0{,}2245 \cdot \lg x_2 + 2{,}8557 \cdot \lg x_3.$$

Требуется:

— представить данное уравнение в естественной форме (не в логарифмах);

— оценить значимость параметров данного уравнения, если известно, что *t*-критерий для параметра b_2 при $x2$ составил 0,8, а для параметра b_3 при x_3составил 1,1.

Решение:

Представленное степенное уравнение множественной регрессии приводим к естественной форме путем потенцирования обеих частей уравнения:

$$x_1 = 10^{0{,}1274} \cdot x_2^{-0{,}2245} \cdot x_3^{2{,}8557};$$

$$x_1 = 1{,}3409 \cdot \frac{1}{x_2^{0{,}2245}} \cdot x_3^{2{,}8557}.$$

Значения коэффициентов регрессии b_1 и b_2 в степенной функции равны коэффициентам эластичности результата x_1 от x_2 и x_3.

$$\overline{Э}_{x_1 x_2} = -0{,}2245\%; \qquad \overline{Э}_{x_1 x_3} = 2{,}8557\%.$$

Спрос на компьютеры x_1 сильнее связан с ценой на ноутбуки – он увеличивается в среднем на 2,86% при росте цен на 1%. С ценой на компьютеры спрос на них связан обратной зависимостью – с ростом цен на 1% потребление снижается в среднем на 0,22%.

Табличные значения *t*-критерия обычно лежит в интервале от 2 до 3 (табличные значения приведены в приложении). Поэтому в данном примере *t*-критерий меньше табличного значения, что свидетельствует о случайной природе взаимосвязи, о статистической ненадежности всего уравнения. Применять полученное уравнение для прогноза не рекомендуется.

Задача 2.2 Имеются следующие данные о ценах и дивидендах по обыкновенным акциям, также о доходности кампании.

Таблица 2.2.1

№	Цена акции, у.е.	Доходность капитала, %	Уровень дивидендов, %
1	25	15,2	2,6
2	20	13,9	2,1
3	15	15,8	1,5
4	34	12,8	3,1
5	20	6,9	2,5
6	33	14,6	3,1
7	28	15,4	2,9
8	30	17,3	2,8
9	23	13,7	2,4
10	24	12,7	2,4
11	25	15,3	2,6
12	26	15,2	2,8
13	26	12,0	2,7
14	20	15,3	1,9
15	20	13,7	1,9
16	13	13,3	1,6
17	21	15,1	2,4
18	31	15,0	3,0
19	26	11,2	3,1
20	11	12,1	2,0

Задание: построить линейное уравнение множественной регрессии и пояснить экономический смысл его параметров.

Решение.

Необходимо построить расчетную таблицу:

Таблица 2.2.2

№	y	x_1	x_2	$x_2 * x_2$	$x_1 * x_1$	$y * x_1$	$y * x_2$	$x_1 * x_2$
1	25	15,2	2,6	6,76	231,04	380,0	65,0	39,52
2	20	13,9	2,1	4,41	193,21	278,0	42,0	29,19
3	15	15,8	1,5	2,25	249,64	237,0	22,5	23,70
4	34	12,8	3,1	9,61	163,84	435,2	105,4	39,68
5	20	6,9	2,5	6,25	47,61	138,0	50,0	17,25
6	33	14,6	3,1	9,61	213,16	481,8	102,3	45,26
7	28	15,4	2,9	8,41	237,16	431,2	81,2	44,66
8	30	17,3	2,8	7,84	299,29	519,0	84,0	48,44
9	23	13,7	2,4	5,76	187,69	315,1	55,2	32,88
10	24	12,7	2,4	5,76	161,29	304,8	57,6	30,48
11	25	15,3	2,6	6,76	234,09	382,5	65,0	39,78
12	26	15,2	2,8	7,84	231,04	395,2	72,8	42,56
13	26	12,0	2,7	7,29	144,0	312,0	70,2	32,40
14	20	15,3	1,9	3,61	234,09	306,0	38,0	29,07
15	20	13,7	1,9	3,61	187,69	274,0	38,0	26,03
16	13	13,3	1,6	2,56	176,89	172,9	20,8	21,28
17	21	15,1	2,4	5,76	228,01	317,1	50,4	36,24
18	31	15,0	3,0	9,0	225,0	465,0	93,0	45,0
19	26	11,2	3,1	9,61	125,44	291,2	80,6	34,72
20	11	12,1	2,0	4,0	146,41	133,1	22,0	24,20
Итого	471	276,5	49,4	126,7	3916,59	6569,1	1216	682,34
Ср. значение	23,55	-	-	-	-	325,455	60,8	34,117

По данным табл. 2.2.2 строится система нормальных уравнений с тремя неизвестными:

$$\begin{cases} \sum y = n \cdot a + b_1 \cdot \sum x_1 + b_2 \cdot \sum x_2 \\ \sum y \cdot x_1 = a \cdot \sum x_1 + b_1 \cdot \sum x_1^2 + b_2 \cdot \sum x_1 \cdot x_2 \\ \sum y \cdot x_2 = a \cdot \sum x_2 + b_1 \cdot \sum x_1 \cdot x_2 + b_2 \cdot \sum x_2^2 \end{cases}$$

$$\begin{cases} 471 = 20 \cdot a + 276.5 \cdot b_1 + 49.4 \cdot b_2 \\ 6569.1 = 276.5 \cdot a + 3916.59 \cdot b_1 + 682.34 \cdot b_2 \\ 1216 = 49.4 \cdot a + 682.34 \cdot b_1 + 126.7 \cdot b_2 \end{cases}$$

Из этой системы находятся коэффициенты a, b_1, b_2:

$a = -13.925$

$b_1 = 0.686$

$b_2 = 11.331$

Таким образом, уравнение множественной регрессии имеет вид:

$\hat{y} = -13.925 + 0686 \cdot x_1 + 11.331 \cdot x_2$.

Экономический смысл коэффициентов b_1 и b_2 в том, что это показатели силы связи, характеризующие изменение цены акции при изменении какого-либо факторного признака на единицу своего измерения при фиксированном влиянии другого фактора.

2.3 Решение с помощью ППП Excel

Задача 2.3 По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%).

Таблица 2.3.1

Номер предприятия	y	x_1	x_2
1	7,0	3,9	10,0
2	7,0	3,9	14,0
3	7,0	3,7	15,0
4	7,0	4,0	16,0
5	7,0	3,8	17,0
6	7,0	4,8	19,0
7	8,0	5,4	19,0
8	8,0	4,4	20,0
9	8,0	5,3	20,0
10	10,0	6,8	20,0
11	9,0	6,0	21,0
12	11,0	6,4	22,0
13	9,0	6,8	22,0
14	11,0	7,2	25,0
15	12,0	8,0	28,0
16	12,0	8,2	29,0
17	12,0	8,1	30,0
18	12,0	8,5	31,0
19	14,0	9,6	32,0
20	14,0	9,0	36,0

Сводную таблицу основных статистических характеристик для одного или нескольких массивов данных можно получить с помощью инструмента анализа данных **Описательная статистика.** Для этого выполните следующие шаги:

1) введите исходные данные или откройте существующий файл, содержащий анализируемые данные;

2) в главном меню выберите последовательно пункты **Сервис / Анализ данных / Описательная статистика**, после чего щелкните по кнопке **ОК**;

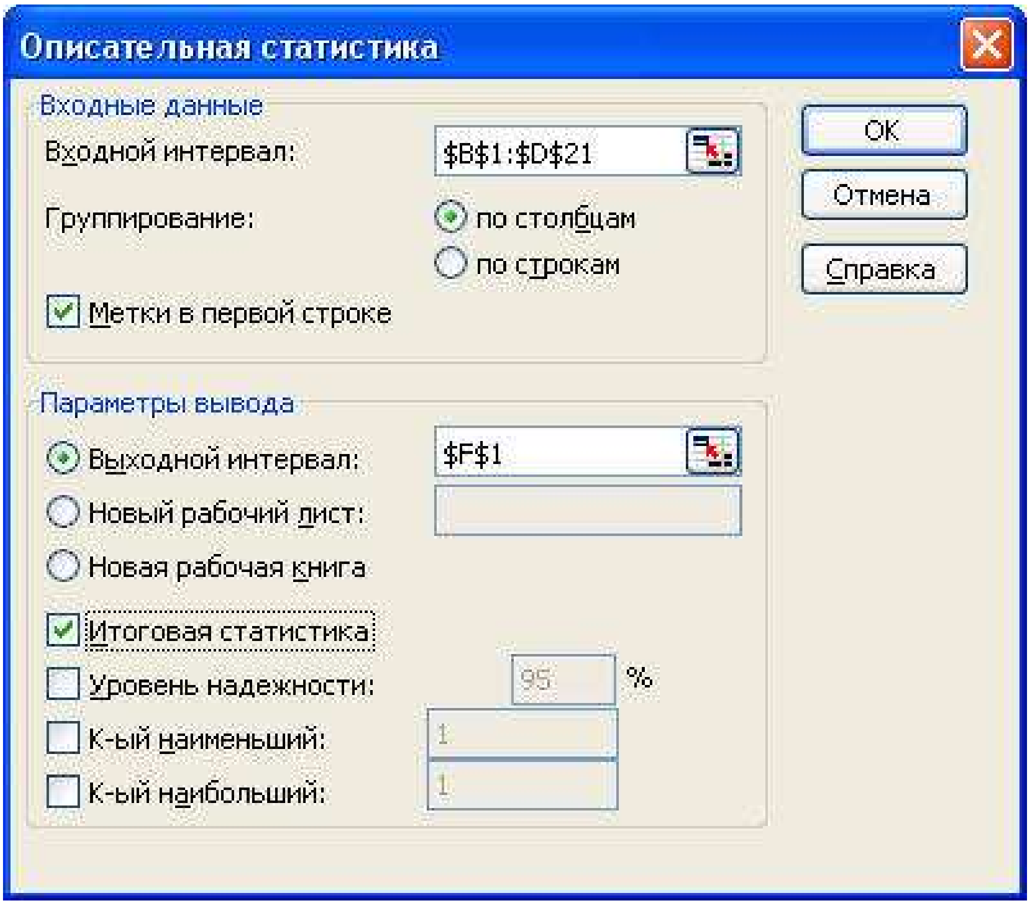


Рис. 2.1 Диалоговое окно ввода параметров инструмента **Описательная статистика**

3) заполните диалоговое окно ввода данных и параметров вывода (рис. 2.1):

Входной интервал – диапазон, содержащий анализируемые данные, это может быть одна или несколько строк (столбцов);

Группирование – по столбцам или по строкам – необходимо указывать дополнительно;

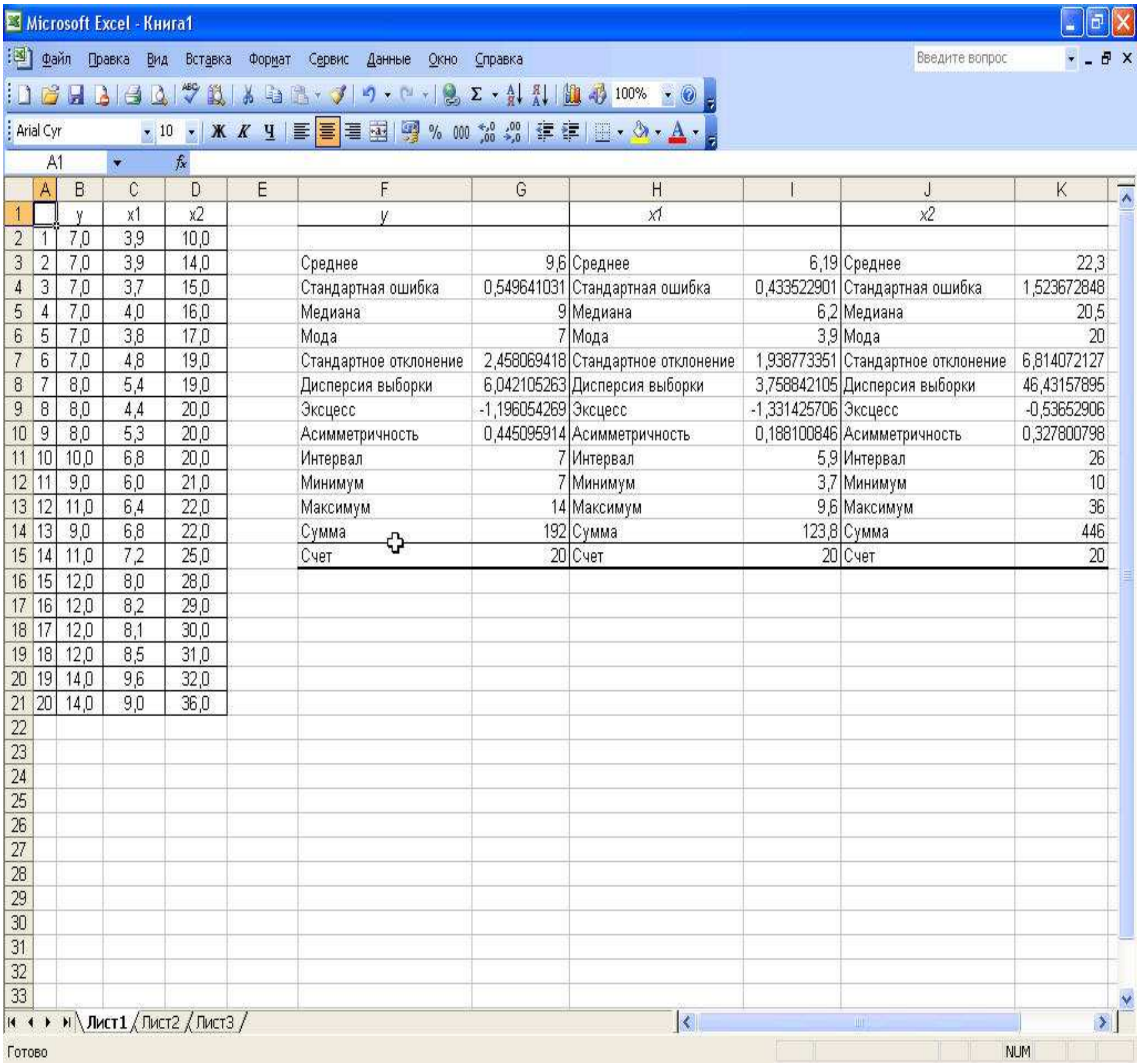
Метки – флажок, который указывает, содержит ли первая строка названия столбцов или нет;

Выходной интервал – достаточно указывать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно задать произвольное имя нового листа.

Если необходимо получить дополнительную информацию *Итоговой статистики, Уровня надежности, к-го наибольшего и наименьшего значений*, установите соответствующие флажки в диалоговом окне. Щелкните по кнопке **ОК**.

Результат вычисления соответствующих показателей для каждого признака представлены на рис. 2.2.



A	y	x1	x2
1	7,0	3,9	10,0
2	7,0	3,9	14,0
3	7,0	3,7	15,0
4	7,0	4,0	16,0
5	7,0	3,8	17,0
6	7,0	4,8	19,0
7	8,0	5,4	19,0
8	8,0	4,4	20,0
9	8,0	5,3	20,0
10	10,0	6,8	20,0
11	9,0	6,0	21,0
12	11,0	6,4	22,0
13	9,0	6,8	22,0
14	11,0	7,2	25,0
15	12,0	8,0	28,0
16	12,0	8,2	29,0
17	12,0	8,1	30,0
18	12,0	8,5	31,0
19	14,0	9,6	32,0
20	14,0	9,0	36,0

y		*x1*		*x2*	
Среднее	9,6	Среднее	6,19	Среднее	22,3
Стандартная ошибка	0,549641031	Стандартная ошибка	0,433522901	Стандартная ошибка	1,523672848
Медиана	9	Медиана	6,2	Медиана	20,5
Мода	7	Мода	3,9	Мода	20
Стандартное отклонение	2,458069418	Стандартное отклонение	1,938773351	Стандартное отклонение	6,814072127
Дисперсия выборки	6,042105263	Дисперсия выборки	3,758842105	Дисперсия выборки	46,43157895
Эксцесс	-1,196054269	Эксцесс	-1,331425706	Эксцесс	-0,53652906
Асимметричность	0,445095914	Асимметричность	0,188100846	Асимметричность	0,327800798
Интервал	7	Интервал	5,9	Интервал	26
Минимум	7	Минимум	3,7	Минимум	10
Максимум	14	Максимум	9,6	Максимум	36
Сумма	192	Сумма	123,8	Сумма	446
Счет	20	Счет	20	Счет	20

Рис. 2.2 Результат применения инструмента **Описательная статистика**

Матрица парных коэффициентов корреляции переменных рассчитывается с использованием инструмента анализа данных Корреляция. Для этого:

1) в главном меню последовательно выберите пункты **Сервис / Анализ данных / Корреляция**. Щелкните по кнопке **ОК**;

2) заполните диалоговое окно ввода данных и параметров вывода (рис. 2.3);

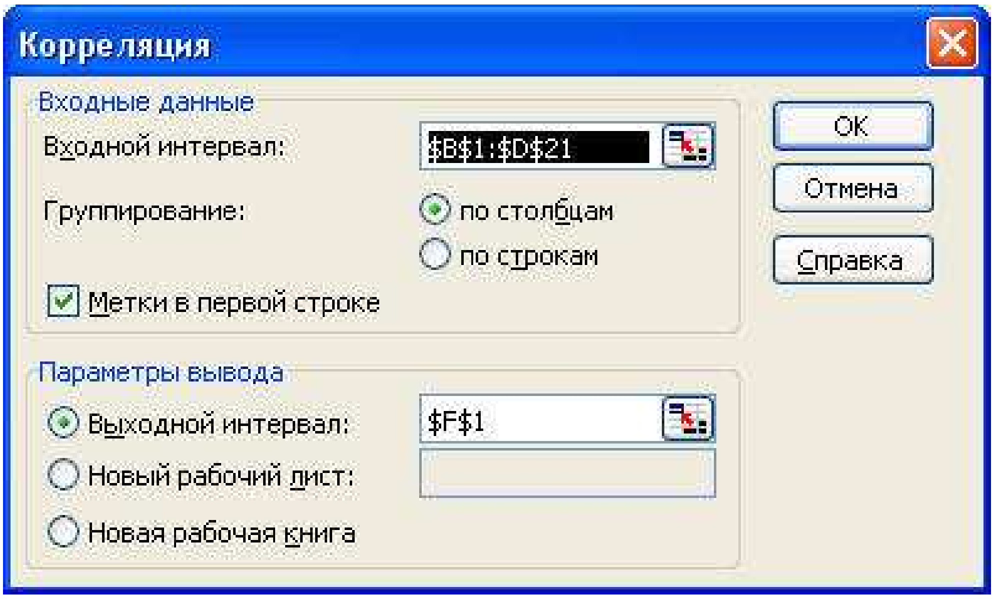
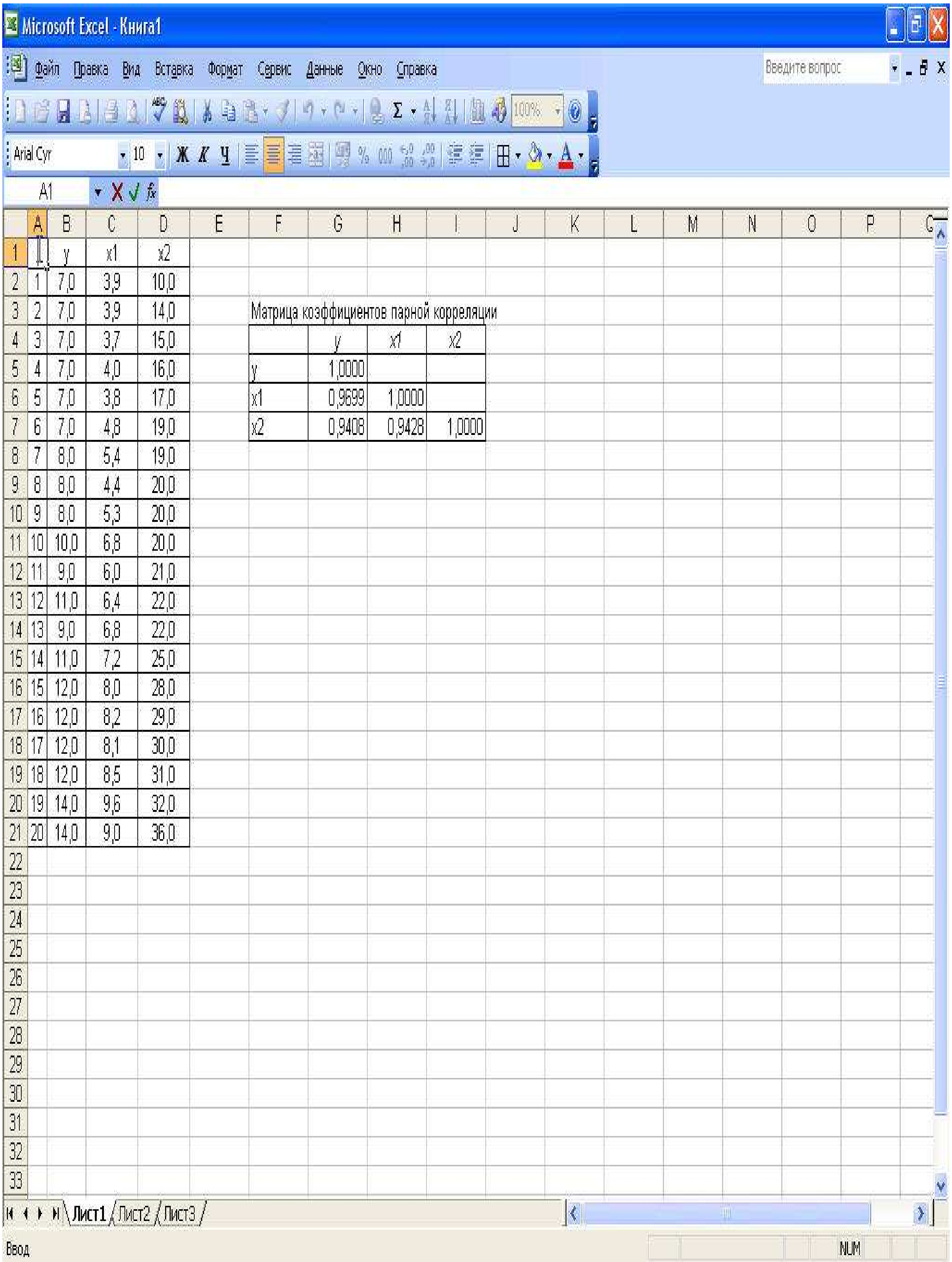


Рис. 2.3 Диалоговое окно ввода параметров инструмента **Корреляция**

3) результаты вычислений – матрица коэффициентов парной корреляции – представлены на рис. 2.4.



A	y	x1	x2
1	7,0	3,9	10,0
2	7,0	3,9	14,0
3	7,0	3,7	15,0
4	7,0	4,0	16,0
5	7,0	3,8	17,0
6	7,0	4,8	19,0
7	8,0	5,4	19,0
8	8,0	4,4	20,0
9	8,0	5,3	20,0
10	10,0	6,8	20,0
11	9,0	6,0	21,0
12	11,0	6,4	22,0
13	9,0	6,8	22,0
14	11,0	7,2	25,0
15	12,0	8,0	28,0
16	12,0	8,2	29,0
17	12,0	8,1	30,0
18	12,0	8,5	31,0
19	14,0	9,6	32,0
20	14,0	9,0	36,0

Матрица коэффициентов парной корреляции

	y	x1	x2
y	1,0000		
x1	0,9699	1,0000	
x2	0,9408	0,9428	1,0000

Рис. 2.4 Матрица коэффициентов парной корреляции

Для вычисления параметров линейного уравнения множественной регрессии используется инструмент анализа данных **Регрессия.** Она аналогична расчету параметров парной линейной регрессии, описанной выше, только в отличие от парной регрессии в диалоговом окне при заполнении параметра *входной интервал X* следует указывать не один столбец, а все столбцы, содержащие значения факторных признаков (рис. 2.5).

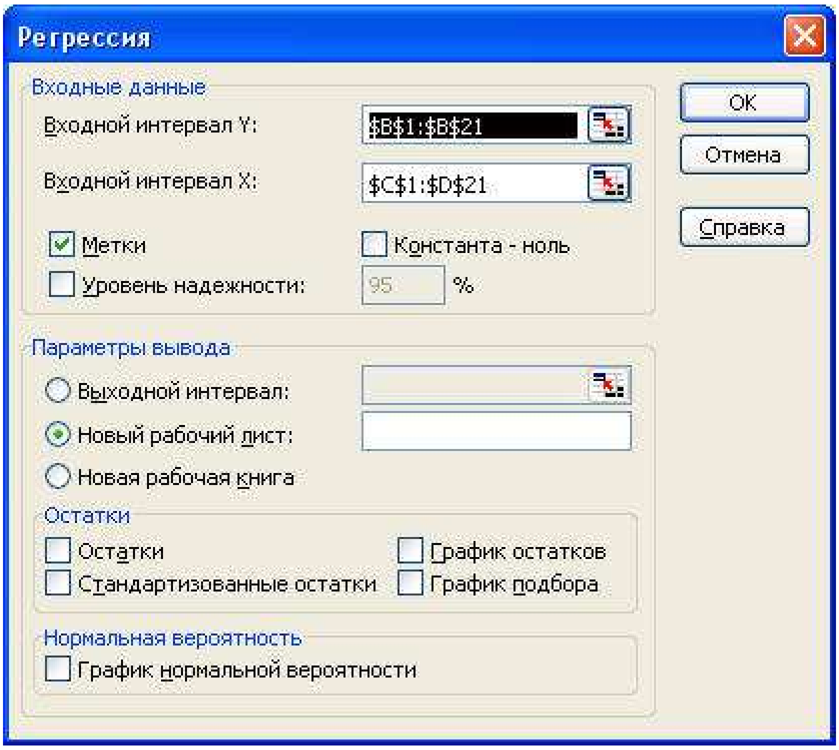
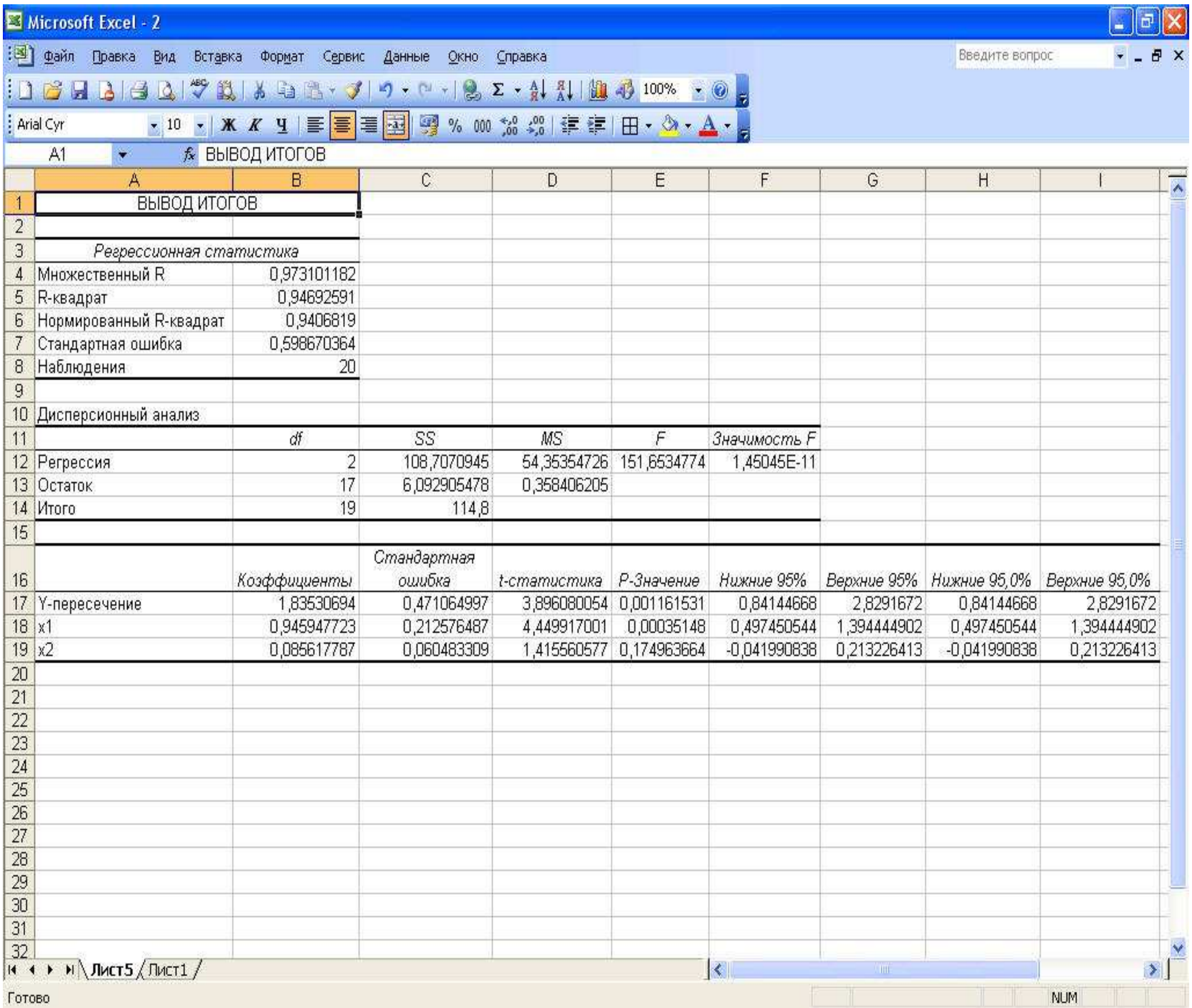


Рис. 2.5 Диалоговое окно ввода параметров инструмента **Регрессия**

Результаты анализа представлены на рис. 2.6.



ВЫВОД ИТОГОВ								
Регрессионная статистика								
Множественный R	0,973101182							
R-квадрат	0,94692591							
Нормированный R-квадрат	0,9406819							
Стандартная ошибка	0,598670364							
Наблюдения	20							
Дисперсионный анализ								
	df	*SS*	*MS*	*F*	*Значимость F*			
Регрессия	2	108,7070945	54,35354726	151,6534774	1,45045E-11			
Остаток	17	6,092905478	0,358406205					
Итого	19	114,8						
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	1,83530694	0,471064997	3,896080054	0,001161531	0,84144668	2,8291672	0,84144668	2,8291672
x1	0,945947723	0,212576487	4,449917001	0,00035148	0,497450544	1,394444902	0,497450544	1,394444902
x2	0,085617787	0,060483309	1,415560577	0,174963664	-0,041990838	0,213226413	-0,041990838	0,213226413

Рис. 2.6 Результат применения инструмента **Регрессия**

2.4 Контрольные вопросы

1) Что представляет собой множественная регрессия?

2) Какие функции применяются для построения множественной регрессии?

3) Чему равен определитель матрицы, если между всеми факторами существует полная линейная связь?

4) Какие факторы включаются в модель множественной регрессии?

5) Как определяется множественный коэффициент корреляции?

6) Что такое фиктивные переменные? Как они входят в уравнение множественной регрессии?

7) Каким критериям должны отвечать оценки параметров регрессии?

8) Для чего применяется анализ остатков при наличии регрессионной модели?

9) Что представляет собой и для чего применяется метод наименьших квадратов?

10) Каким критериям должны отвечать оценки параметров регрессии?

2.5 Пример варианта промежуточного тестирования

1. Зависимость расходов на продукты питания по совокупности семей характеризуется уравнением $y_x = 0,5 + 0,35x_1 + 0,73x_2$, где y - расходы семьи за месяц на продукты питания (тыс.руб), x_1 - месячный доход на одного члена семьи (тыс.руб.), x_2 - размер семьи (чел.). Справедливо ли утверждение?

а) с ростом дохода на одного члена семьи на одну тысячу руб. расходы на питание возрастут в среднем на 350 руб. при том же размере семьи;

б) с ростом дохода на одного члена семьи на одну тысячу руб. расходы на питание возрастут в среднем на 730 руб. при том же размере семьи;

в) с ростом дохода на одного члена семьи на одну тысячу руб. расходы на питание уменьшатся в среднем на 350 руб. при том же размере семьи;

г) с ростом дохода на одного члена семьи на одну тысячу руб. расходы на питание уменьшатся в среднем на 730 руб. при том же размере семьи;

д) с ростом дохода на одного члена семьи на одну тысячу руб. расходы на питание изменятся в 0,5 раз при том же размере семьи.

2. В функции потребления $C_t = a + b_0R_t + b_1R_{t-1}$ коэффициент b_0 - краткосрочная предельная склонность к потреблению характеризует:

а) рост потребления;

б) эффект единичного возрастания дохода R_t при неизменном уровне предыдущего дохода R_{t-1};

в) эффект единичного возрастания дохода R_t при изменении уровня предыдущего дохода R_{t-1};

г) снижение потребления;

д) стабилизации потребления.

3. Производственная функция имеет вид $P = 2,0F_1^{0,3} \cdot F_2^{0,2} \cdot F_3^{0,5}$, где P - выпуск продукции, F_1 - стоимость основных производственных фондов, F_2 - отработка человеко-дней, F_3 - затраты на производство. Можно ли утверждать, что коэффициент эластичности выпуска продукции составляет:

а) 0,2%;

б) 0,5%;

в) 0,7%;

г) 0,3%;

д) 1%.

4. При исследовании спроса на масло получено уравнение $y = 0,056x_1^{-0,858}x_2^{1,126}$, где y - количество масла на душу населения (кг.), x_1 - цена (руб.), x_2 - доход на душу населения (тыс.руб.). Справедливо ли утверждение, что:

а) с падением цен на 1% спрос повысится в среднем на 0,126%;

б) с уменьшением дохода на 1% спрос понизится в среднем на 0,858%;

в) с ростом цены на 1% при том же доходе спрос снизится в среднем на 0,858%, а рост дохода на 1% при неизменных ценах вызовет увеличение спроса в среднем на 1,126%;

г) зависимость цены и спроса не определена;

д) зависимость дохода на душу населения и спроса не определена.

5. Зависимость себестоимости от цены на расходные материалы имеет вид $\hat{y}_{x_3} = 15{,}75 \cdot x_3^{1,55}$. На сколько процентов в среднем по совокупности изменится себестоимость от своей средней величины при изменении цены на 1% от своего среднего значения?

а) на 15,75%;
б) нельзя дать ответ, т.к. не задано среднее значение себестоимости;
в) на 1,55%;
г) на 17,3%;
д) на $15{,}75^{1,55}$.

6. Множественная регрессия – это уравнение связи с несколькими независимыми переменными. Какое из уравнений является уравнением множественной регрессии:

а) $y = 0.9 + 65x + 32x^2 + 100x^3$;

б) $x_1 = 1{,}3409 \cdot \frac{1}{x_2^{0,2245}} \cdot x_3^{2,8557}$;

в) $y = e^{50+35x}$;

г) $y = 0.4 + \frac{77}{x}$;

д) $x_1 = 12.5 + 0.25x_2$.

7. Зависимость спроса на компьютеры x_1 от цены на них x_2 и от цены на ноутбуки x_3 представлена уравнением: $\lg x_1 = 0{,}12 - 0{,}22 \cdot \lg x_2 + 2{,}85 \cdot \lg x_3$. **С каким параметром спрос на компьютеры связан сильнее:**

а) с ценой на компьютеры;
б) влияние одинаково;
в) со спросом на ноутбуки;
г) не достаточно информации для ответа;
д) с ценой на ноутбуки.

3 Временные ряды в экономических исследованиях
3.1 Методические указания

Модели, построенные по данным, характеризующим один объект за ряд последовательных моментов (периодов) времени, называются *моделями временных рядов*.

Временной ряд – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов. Каждый уровень временного ряда формируется из трендовой (T), циклической (S) и случайной (E) компонент. Модели, в которых временной ряд представлен как сумма перечисленных компонент – *аддитивные модели*, как произведение – *мультипликативные модели*. 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 и более временных рядов.

Аддитивная модель имеет вид:

$$Y=T+S+E$$

Мультипликативная модель:

$$Y=T\cdot S\cdot E$$

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

Построение модели включает следующие шаги:

1. Выравнивание исходного ряда методом скользящей средней;
2. Расчет значений сезонной компоненты S;
3. Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных в аддитивной (T+E) или в мультипликативной (T·E) модели.
4. Аналитическое выравнивание уровней (T+E) или (T·E) и расчет значений T с использованием полученного уравнения тренда;

5. Расчет полученных по модели значений (T+S) или (T·S);
6. Расчет абсолютных или относительных ошибок.

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.

3.2 Решение типовых задач

Задача 3.1 Имеются данные об общем количестве правонарушений на таможне одного из субъектов РФ.

Таблица 3.1.1

Год	Квартал	t	Количество возбужденных дел, y_t
2007	I	1	375
	II	2	371
	III	3	869
	IV	4	1015
2008	I	5	357
	II	6	471
	III	7	992
	IV	8	1020
2009	I	9	390
	II	10	355
	III	11	992
	IV	12	905
2010	I	13	461
	II	14	454
	III	15	920
	IV	16	927

Данный временной ряд содержит сезонные колебания периодичностью 4, т.к. количество правонарушений в первый-второй кварталы ниже, чем в третий-четвертый. Необходимо рассчитать компоненты аддитивной модели временного ряда.

Шаг 1. Проводится выравнивание исходных уровней ряда методом скользящей средней. Для этого:

1.1. Суммируются уровни ряда последовательно за каждые четыре квартала со сдвигом на один момент времени.

1.2. Разделив полученные суммы на 4, находятся скользящие средние. Полученные таким образом выровненные значения уже не содержат сезонной компоненты.

1.3. Необходимо привести эти значения в соответствие с фактическими моментами времени, для чего находятся средние значения из двух последовательных скользящих средних – центрированные скользящие средние.

Таблица 3.1.2

№ квартала, *t*	Количество правонарушений y_t	Итого за четыре квартала	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	**2**	**3**	**4**	**5**	**6**
1	375	–	–	–	–
2	371	2630	657,5	–	–
3	869	2612	653	655,25	213,75
4	1015	2712	678	665,5	349,5
5	357	2835	708,75	693,75	-336,75
6	471	2840	710	709,375	-238,375
7	992	2873	718,25	714,125	277,875
8	1020	2757	689,25	703,75	316,25
9	390	2757	689,25	689,25	-299,25
10	355	2642	660,5	674,875	-319,875
11	992	2713	678,25	669,375	322,625
12	905	2812	703	690,625	214,375
13	461	2740	685	694	-233
14	454	2762	690,5	687,75	-233,75
15	920	–	–	–	–
16	927	–	–	–	–

Шаг 2. Находятся оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними. Эти оценки используются для расчета значений сезонной компоненты *S*. Для этого находятся средние за каждый квартал (по всем годам) оценки сезонной компоненты S_i. 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.

Таблица 3.1.3

Показатели	Год	№ квартала, i			
		I	II	III	IV
	2007	–	–	213,75	349,5
	2008	-336,75	-238,375	277,875	316,25
	2009	-299,25	-319,875	322,625	214,375
	2010	-233	-233,75	–	–
Всего за i-й квартал		-869	-792	814,25	880,125
Средняя оценка сезонной компоненты для i-го квартала, $\overline{S}_i$		-289,667	-264	271,417	293,375
Скорректированная сезонная компонента, S_i		-292,448	-266,781	268,636	290,593

Для данной модели имеем:

-298,667-264+271,417+293,375=11,125.

Корректирующий коэффициент: k=11,125/4=2,781.

Расчет скорректированных значений сезонной компоненты ($S_i = \overline{S}_i - k$).

Проверка равенства нулю суммы значений сезонной компоненты:

-292,448-266,781+268,636+290,593=0.

Шаг 3. Исключается влияние сезонной компоненты, путем вычитания ее значения из каждого уровня исходного временного ряда. Получаются величины *T+E=Y-S*. Эти значения рассчитываются за каждый момент времени и содержат только тенденцию и случайную компоненту.

Таблица 3.1.4

t	y_t	S_i	y_t-S_i	T	$T+S$	E= y_t-(T+S)	E^2
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	375	-292,448	667,448	672,700	380,252	-5,252	27,584
2	371	-266,781	637,781	673,624	406,843	-35,843	1284,721
3	869	268,636	600,364	674,547	943,183	-74,183	5503,117
4	1015	290,593	724,407	675,470	966,063	48,937	2394,830
5	357	-292,448	649,448	676,394	383,946	-26,946	726,087
6	471	-266,781	737,781	677,317	410,536	60,464	3655,895
7	992	268,636	723,364	678,240	946,876	45,124	2036,175
8	1020	290,593	729,407	679,163	969,756	50,244	2524,460
9	390	-292,448	682,448	680,087	387,639	2,361	5,574
10	355	-266,781	621,781	681,010	414,229	-59,229	3508,074
11	992	268,636	723,364	681,933	950,569	41,431	1716,528
12	905	290,593	614,407	682,857	973,450	-68,450	4685,403
13	461	-292,448	753,448	683,780	391,332	69,668	4853,630
14	454	-266,781	720,781	684,703	417,922	36,078	1301,622
15	920	268,636	651,364	685,627	954,263	-34,263	1173,953
16	927	290,593	636,407	686,550	977,143	-50,143	2514,320

Шаг 4. Определение компоненты *T* данной модели. Для этого проводится аналитическое выравнивание ряда (*T+E*) с помощью линейного тренда. Результаты аналитического выравнивания следующие:

*T=671,777+0,9233*t.*

Подставляя в это уравнение значения *t=1,2,…,16,* находятся уровни *T* для каждого момента времени.

Шаг 5. Находятся значения уровней ряда, полученные по аддитивной модели. Для этого к уровням *T* прибавляются значения сезонной компоненты для соответствующих кварталов.

Для оценки качества построенной модели применяется сумма квадратов полученных абсолютных ошибок.

$$R^2 = 1 - \frac{E^2}{(y_t - \bar{y})^2} = 1 - \frac{37911,973}{1252743,75} = 0,970$$

Следовательно, можно сказать, что аддитивная модель объясняет 97% общей вариации уровней временного ряда количества правонарушений по кварталам за 4 года.

Шаг 6. Прогнозирование по аддитивной модели. Необходимо дать прогноз об общем объеме правонарушений на I и II кварталы 2011 года. Прогнозное значение F_t уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда

$$T = 671,777 + 0,9233 \cdot t$$

Получим

$$T_{17} = 671,777 + 0,9233 \cdot 17 = 687,473$$

$$T_{18} = 671,777 + 0,9233 \cdot 18 = 688,396$$

Значения сезонных компонент за соответствующие кварталы равны: $S_1=-292,448$ и $S_2=-266,781$. Таким образом,

$$F_{17} = T_{17} + S_1 = 687,473 - 292,448 \approx 395$$

$$F_{18} = T_{18} + S_2 = 688,396 - 266,781 \approx 422$$

Т.е., в первые два квартала 2011 г. следует ожидать порядка 395 и 422 правонарушений соответственно.

Задача 3.2 На основе помесячных данных о числе браков (тыс.) в регионе за последние три года была построена аддитивная модель временного ряда. Скорректированные значения сезонной компоненты за соответствующие месяцы приводятся в таблице 3.2.1.

Таблица 3.2.1

Месяц	Скорректированные значения сезонной компоненты	Месяц	Скорректированные значения сезонной компоненты
Январь	-1,0	Июль	3,0
Февраль	2,0	Август	1,0
Март	-0,5	Сентябрь	2,5
Апрель	0,3	Октябрь	1,0
Май	-2,0	Ноябрь	-3,0
Июнь	-1,1	Декабрь	?

Уравнение тренда выглядит следующим образом:

$$\hat{y}_t = 2{,}5 + 0{,}03 \cdot t\,.$$

При расчете параметров тренда использовались фактические моменты времени (t от 1 до 36 мес.).

Требуется:

- определить значение сезонной компоненты за декабрь;
- на основе постоянной модели дать прогноз, заключенных в течение 1 квартала следующего года.

Решение:

Сумма значений сезонной компоненты внутри одного цикла должна быть равна 0 (в соответствии с методикой построения аддитивной модели временного ряда). Следовательно, значение сезонной компоненты за декабрь составит:

S_{12} = 0 – (-1 + 2 – 0,5 + 0,3 – 2 – 1,1 + 3 + 1 +2 ,5 + 1 – 3) = - 2,2 .

Прогнозное значение уровня временного ряда F_t в аддитивной модели есть сумма трендового значения T_t и соответствующего значения сезонной компоненты S_t.

Число браков, заключенных в 1 квартале следующего года, есть сумма числа браков, заключенных в январе F_{37}, феврале F_{38} и марте F_{39}.

Для расчета трендовых значений воспользуемся уравнением тренда, указанным в условии задачи:

$$\hat{y}_t = 2{,}5 + 0{,}03 \cdot t$$
$$T_{37} = 2{,}5 + 0{,}03 \cdot 37 = 3{,}61,$$
$$T_{38} = 2{,}5 + 0{,}03 \cdot 38 = 3{,}64,$$
$$T_{39} = 2{,}5 + 0{,}03 \cdot 39 = 3{,}67.$$

Соответствующие значения сезонных компонент составят:

$S_1=-1$	Январь
$S_2=2$	Февраль
$S_3=-0{,}5$	Март

Таким образом,

$$F_{37} = T_{37} + S_1 = 3{,}61 - 1{,}0 = 2{,}61;$$
$$F_{38} = T_{38} + S_2 = 3{,}64 + 2{,}0 = 5{,}64;$$
$$F_{39} = T_{39} + S_3 = 3{,}67 - 0{,}5 = 3{,}17;$$

Количество браков, заключенных в 1 квартале следующего года, составит: 2,61 +5,64 + 3,17 = 11,42 тыс.

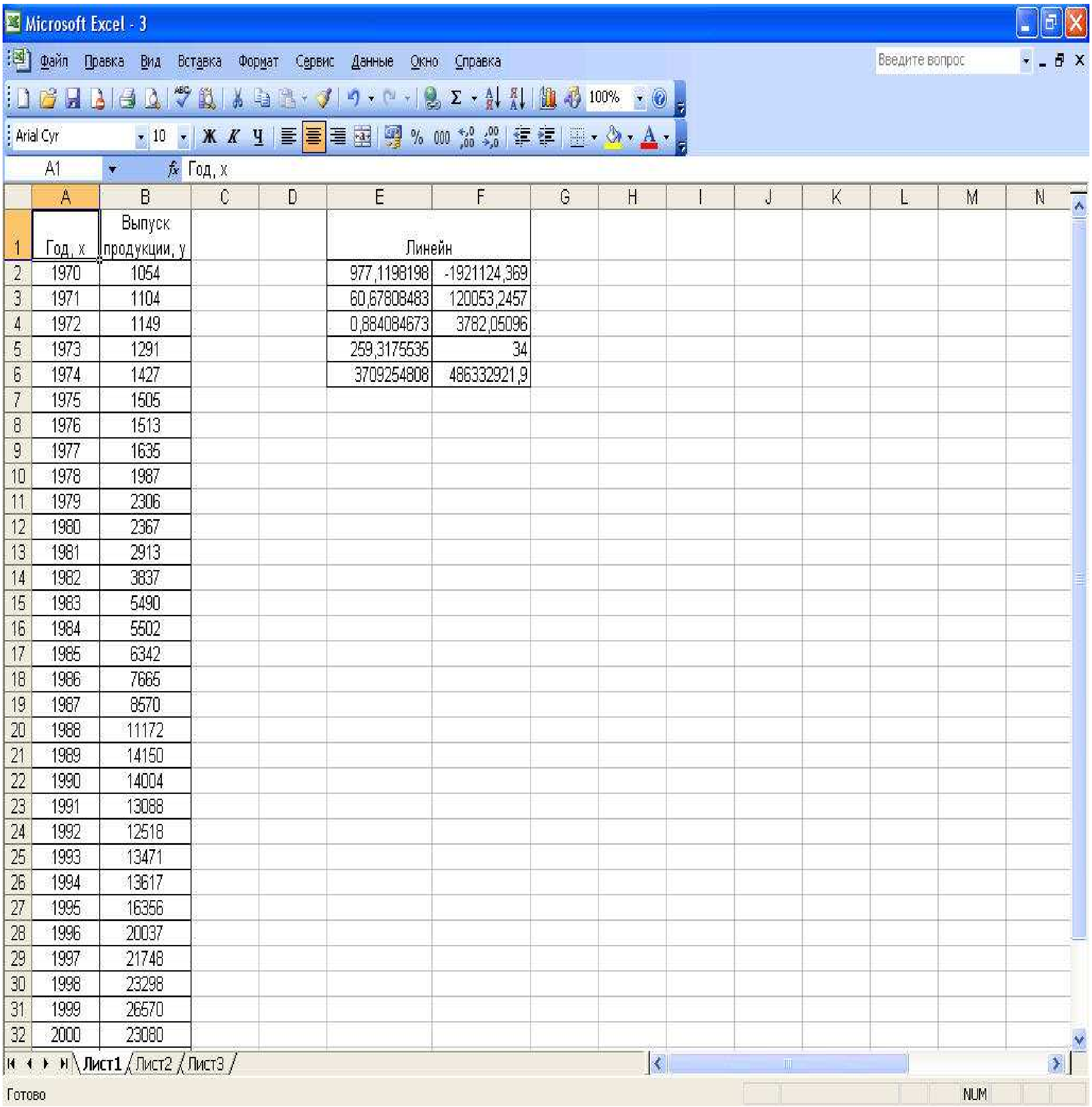
3.3 Решение с помощью ППП Excel

Задача 3.3 Динамика выпуска продукции Швеции характеризуется данными (млн. долл.), представленными в табл. 3.3.1.

Таблица 3.3.1

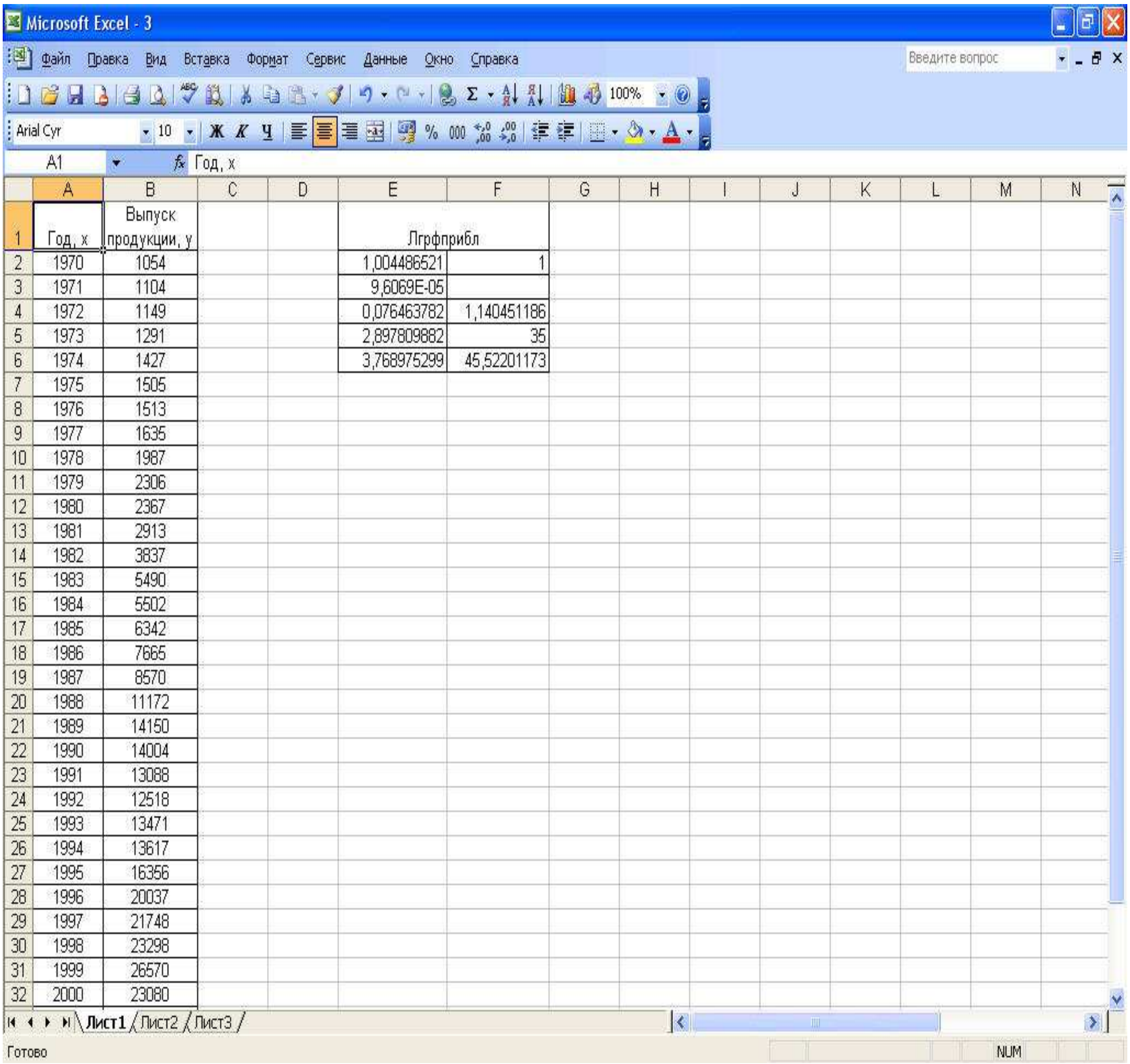
Год, x	Выпуск продукции, y
1970	1054
1971	1104
1972	1149
1973	1291
1974	1427
1975	1505
1976	1513
1977	1635
1978	1987
1979	2306
1980	2367
1981	2913
1982	3837
1983	5490
1984	5502
1985	6342
1986	7665
1987	8570
1988	11172
1989	14150
1990	14004
1991	13088
1992	12518
1993	13471
1994	13617
1995	16356
1996	20037
1997	21748
1998	23298
1999	26570
2000	23080
2001	23981
2002	23446
2003	29658
2004	39573
2005	38435

1. Для определения параметров линейного тренда по методу наименьших квадратов используется статистическая функция **ЛИНЕЙН**, для определения экспоненциального тренда – **ЛГРФПРИБЛ**. В качестве зависимой переменной в данном примере выступает время (t = 1, 2, …, n). Приведем результаты вычисления функции **ЛИНЕЙН** и **ЛГРФПРИБЛ** (рис. 3.1 и 3.2).



	A	B	E	F
1	Год, x	Выпуск продукции, y	Линейн	
2	1970	1054	977,1198198	-1921124,369
3	1971	1104	60,67808483	120053,2457
4	1972	1149	0,884084673	3782,05096
5	1973	1291	259,3175535	34
6	1974	1427	3709254808	486332921,9
7	1975	1505		
8	1976	1513		
9	1977	1635		
10	1978	1987		
11	1979	2306		
12	1980	2367		
13	1981	2913		
14	1982	3837		
15	1983	5490		
16	1984	5502		
17	1985	6342		
18	1986	7665		
19	1987	8570		
20	1988	11172		
21	1989	14150		
22	1990	14004		
23	1991	13088		
24	1992	12518		
25	1993	13471		
26	1994	13617		
27	1995	16356		
28	1996	20037		
29	1997	21748		
30	1998	23298		
31	1999	26570		
32	2000	23080		

Рис. 3.1 Результат вычисления функции **ЛИНЕЙН**



Год, x	Выпуск продукции, y	Лгрфприбл	
1970	1054	1,004486521	1
1971	1104	9,6069E-05	
1972	1149	0,076463782	1,140451186
1973	1291	2,897809882	35
1974	1427	3,768975299	45,52201173
1975	1505		
1976	1513		
1977	1635		
1978	1987		
1979	2306		
1980	2367		
1981	2913		
1982	3837		
1983	5490		
1984	5502		
1985	6342		
1986	7665		
1987	8570		
1988	11172		
1989	14150		
1990	14004		
1991	13088		
1992	12518		
1993	13471		
1994	13617		
1995	16356		
1996	20037		
1997	21748		
1998	23298		
1999	26570		
2000	23080		

Рис. 3.2 Результат вычисления функции **ЛГРФПРИБЛ**

Запишем уравнение линейного и экспоненциального тренда, используя данные рис. 3.1 и 3.2:

$$\hat{y}_t = -1921124{,}37 + 977{,}12 \cdot t,$$

$$\hat{y}_t = -1{,}0045^t.$$

2. Построение графиков осуществляется с помощью **Мастера диаграмм**.

Порядок построения следующий:

1) введите исходные данные или откройте существующий файл, содержащий анализируемые данные;

2) активизируйте **Мастер диаграмм** любым из следующих способов:

а) в главном меню выберите **Вставка / Диаграмма**;

б) на панели инструментов **Стандартная** щелкните по кнопке **Мастер диаграмм**;

3) в окне **Тип** выберите **График** (рис. 3.3); вид графика выберите в поле рядом со списком типов. Щелкните по кнопке **Далее**;

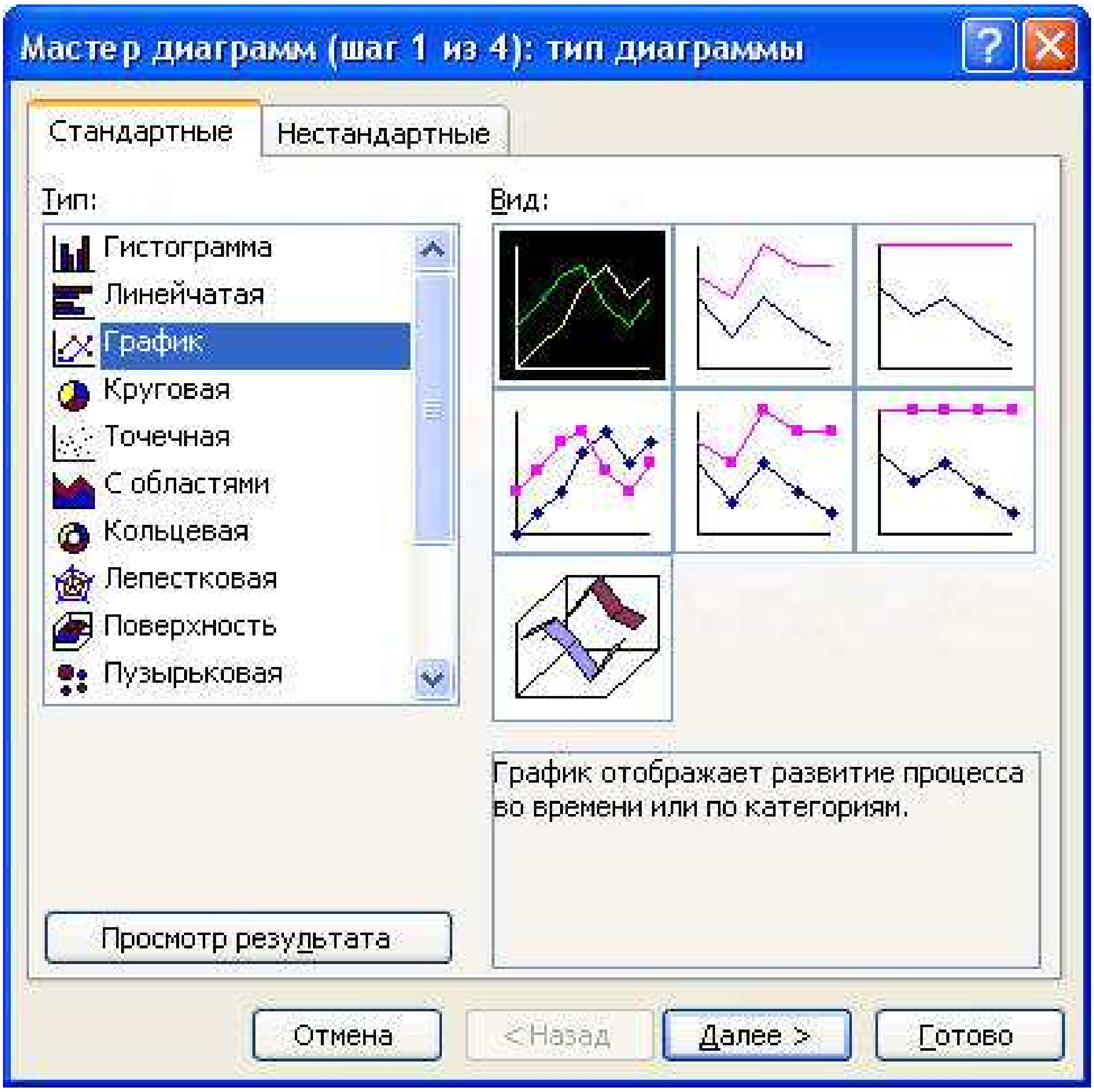


Рис. 3.3 Диалоговое окно **Мастера диаграмм**: тип диаграммы

4) заполните диапазон данных, как показано на рис. 3.4. Установите флажок размещения данных в столбцах (строках). Щелкните по кнопке **Далее**;

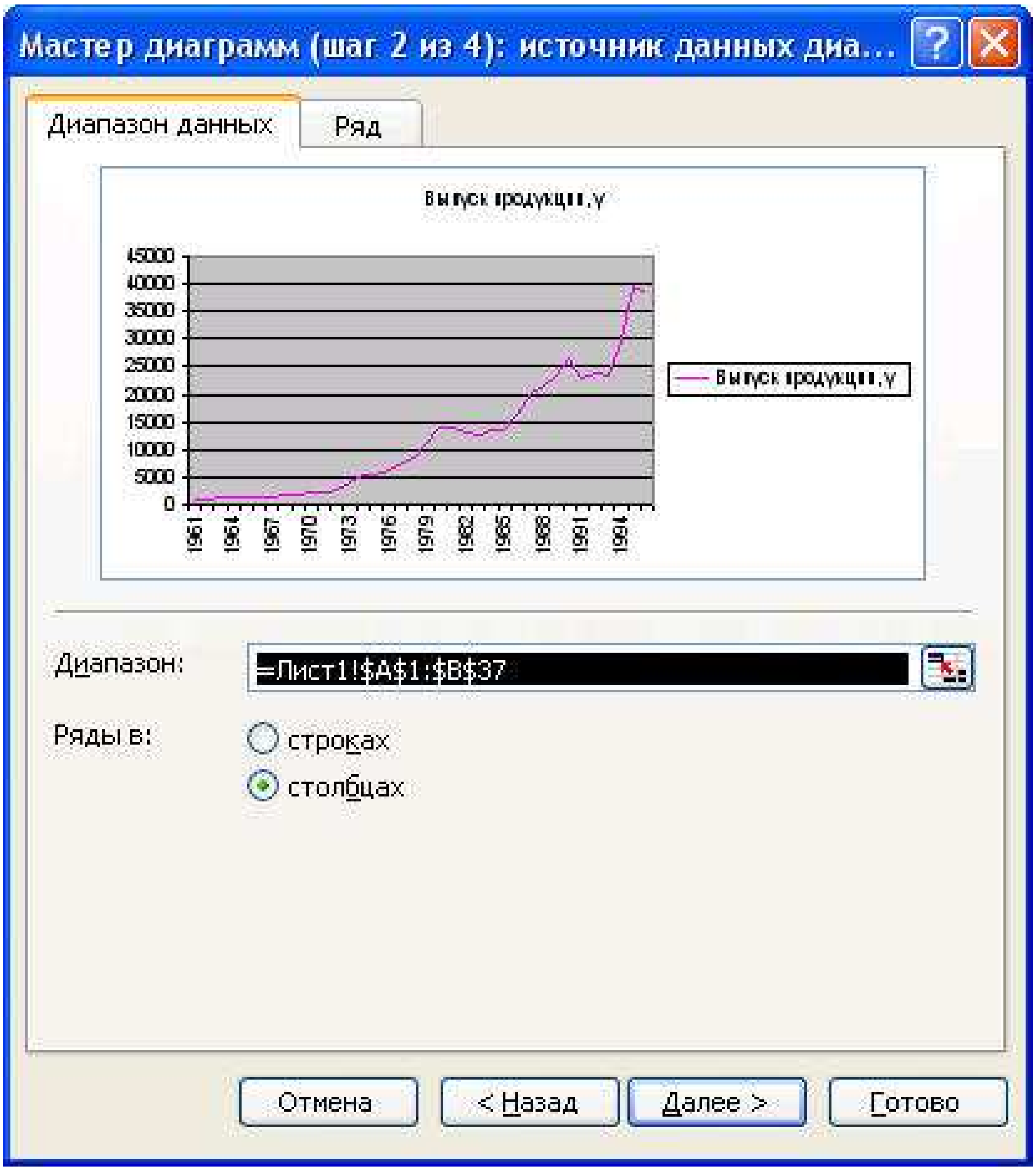


Рис. 3.4 Диалоговое окно **Мастера диаграмм**: источник данных

5) заполните параметры диаграммы на разных закладках (рис. 3.5): название диаграммы и осей, значение осей, линии сетки, параметры легенды, таблица и подписи данных. Щелкните по кнопке **Далее**;

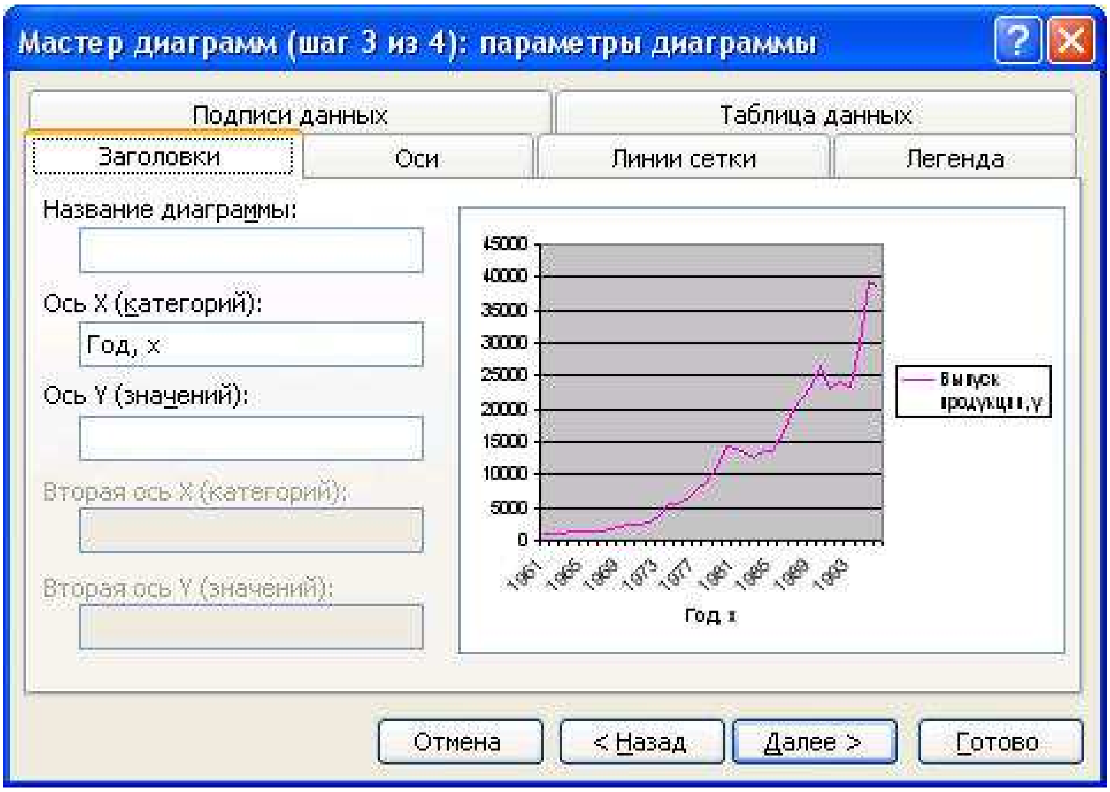


Рис. 3.5 Диалоговое окно **Мастера диаграмм**: параметры диаграммы

6) укажите место размещения диаграммы на отдельном или имеющемся листе (рис. 3.6). Щелкните по кнопке **Далее**. Готовая диаграмма, отражающая динамику уровня изучаемого ряда, представлена на рис. 3.7.

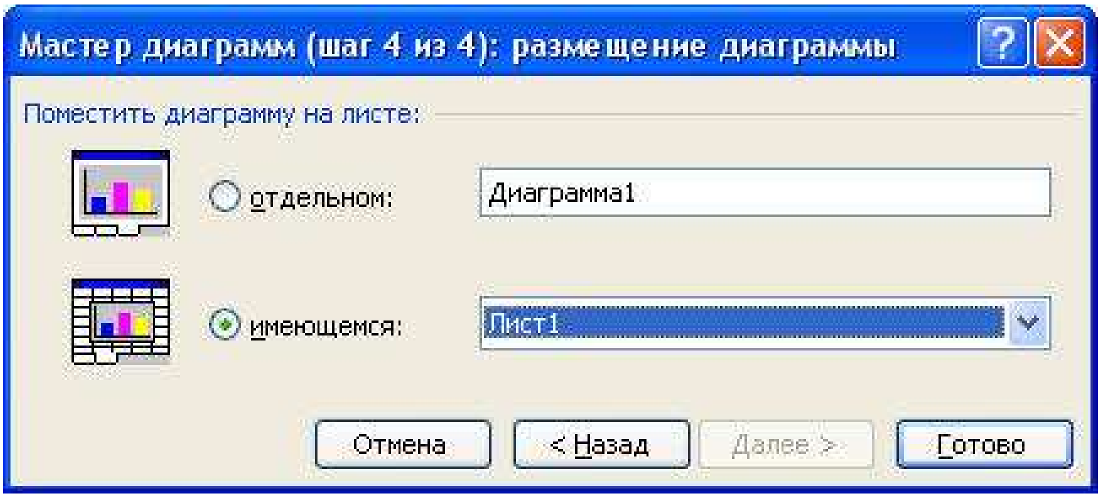


Рис. 3.6 Диалоговое окно **Мастера диаграмм**: размещение диаграммы

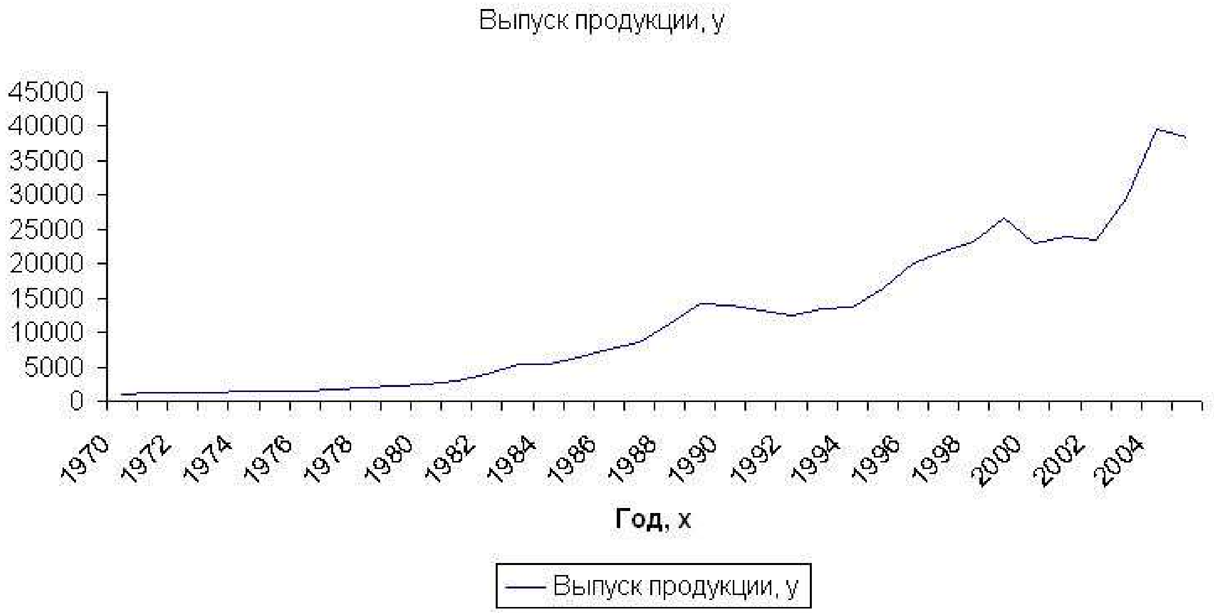


Рис. 3.7 Динамика выпуска продукции

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

1) выделите область построения диаграммы; в главном меню выберите Диаграмма / Добавить линию тренда;

2) в появившемся диалоговом окне (рис. 3.8) выберите вид линии тренда и задайте соответствующие параметры. Для полиномиального тренда необходимо задать степень аппроксимирующего полинома, для скользящего среднего – количество точек усреднения.

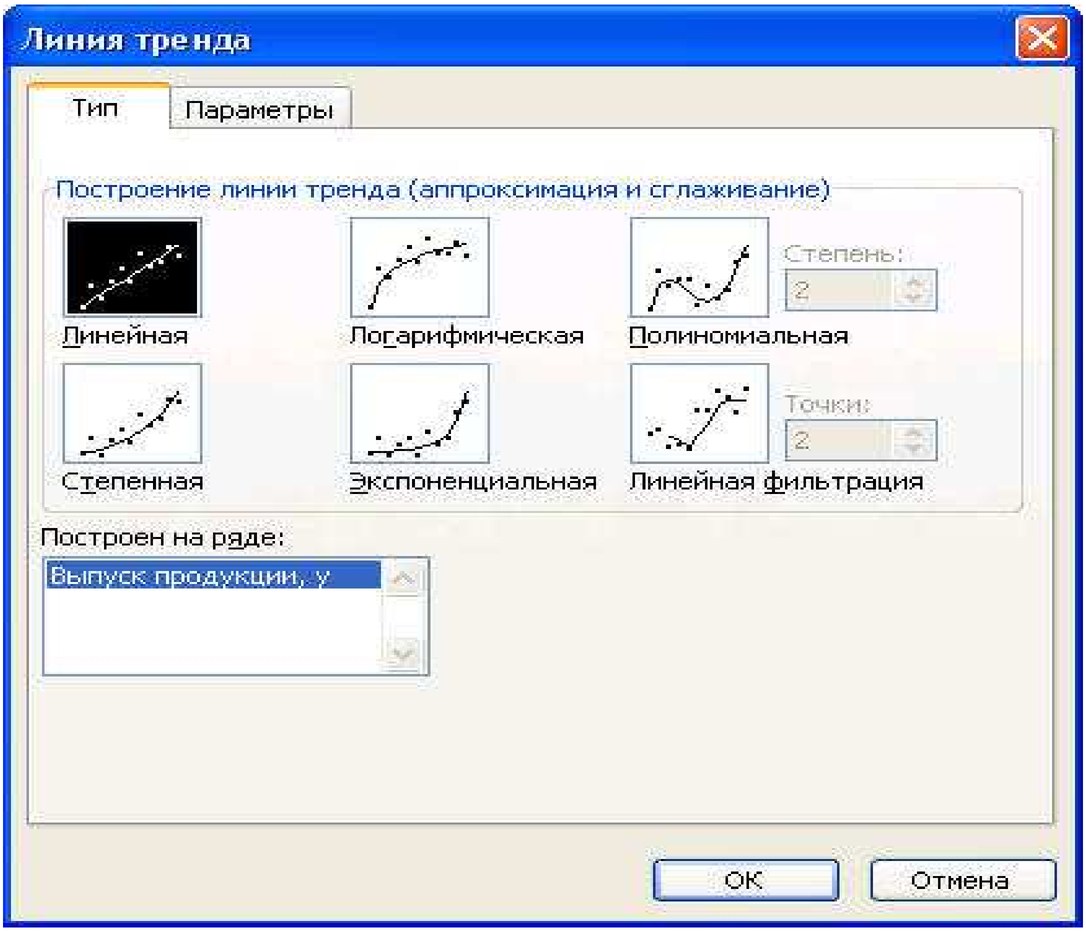


Рис. 3.8 Диалоговое окно типов линий тренда

В качестве дополнительной информации на диаграмме можно отобразить уравнение регрессии и значение среднеквадратического отклонения, установив соответствующие флажки на закладке **Параметры** (рис. 8.9). Щелкните по кнопке **ОК**.

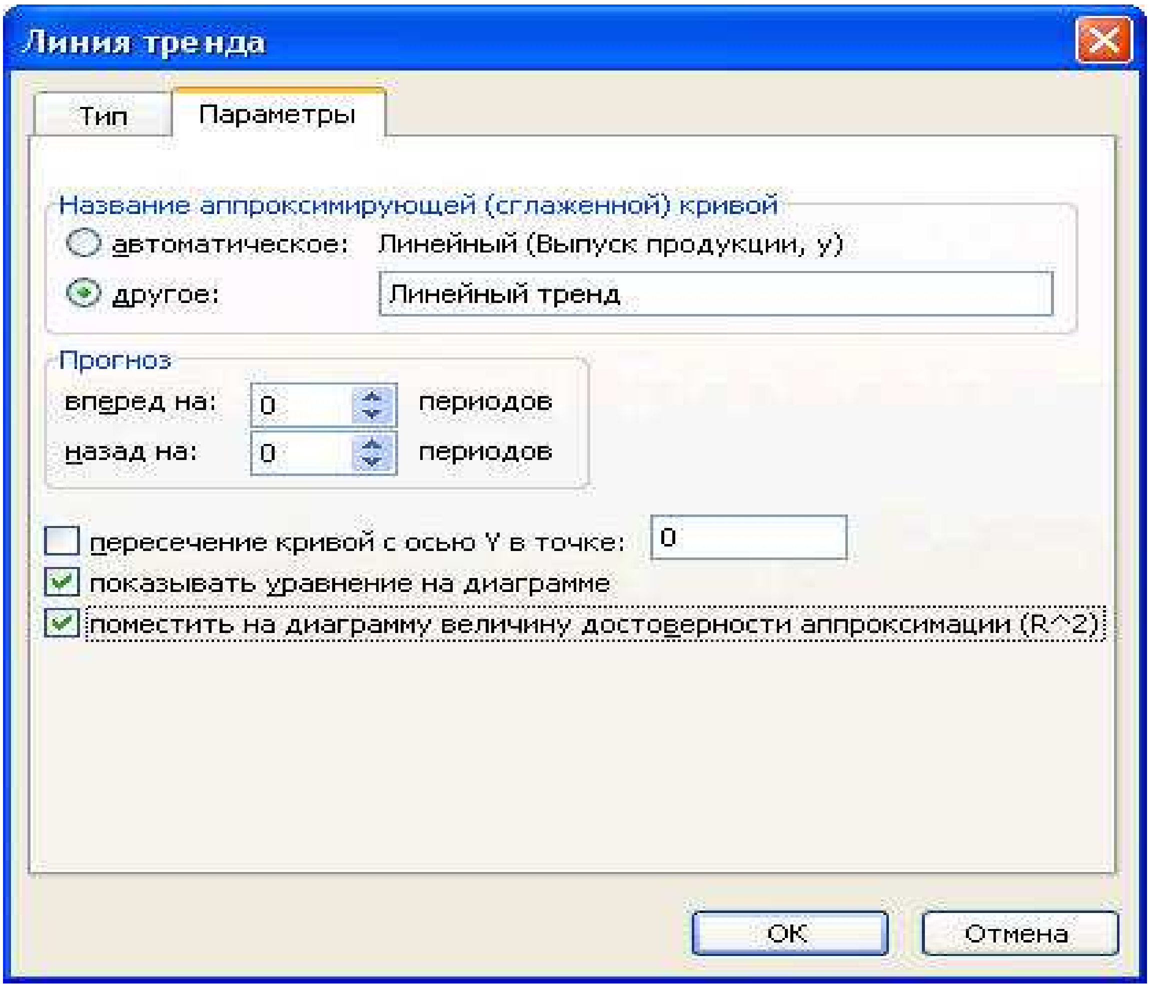


Рис. 3.9 Диалоговое окно параметров линии тренда

На рис. 3.10-3.14 представлены различные виды трендов, описывающие исходные данные задачи.

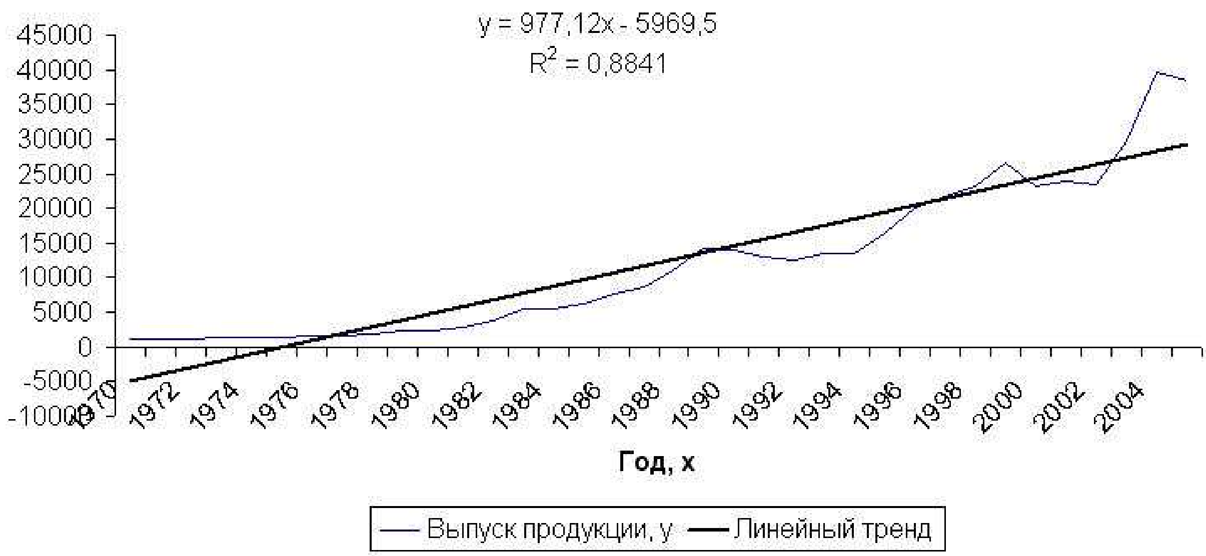


Рис.3.10 Линейный тренд

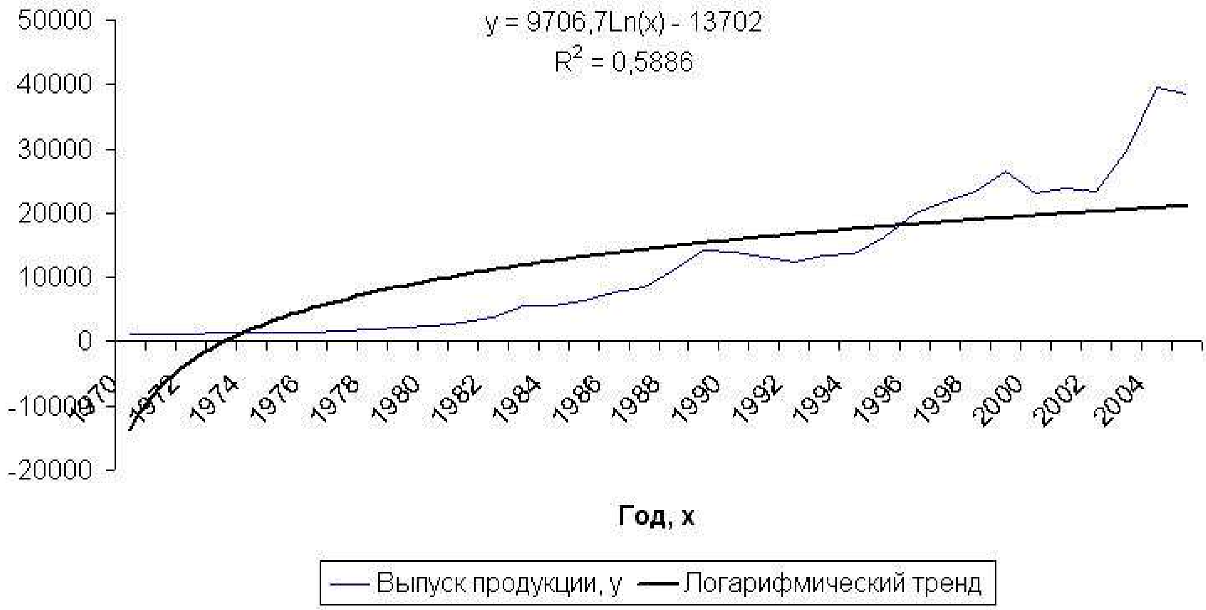


Рис.3.11 Логарифмический тренд

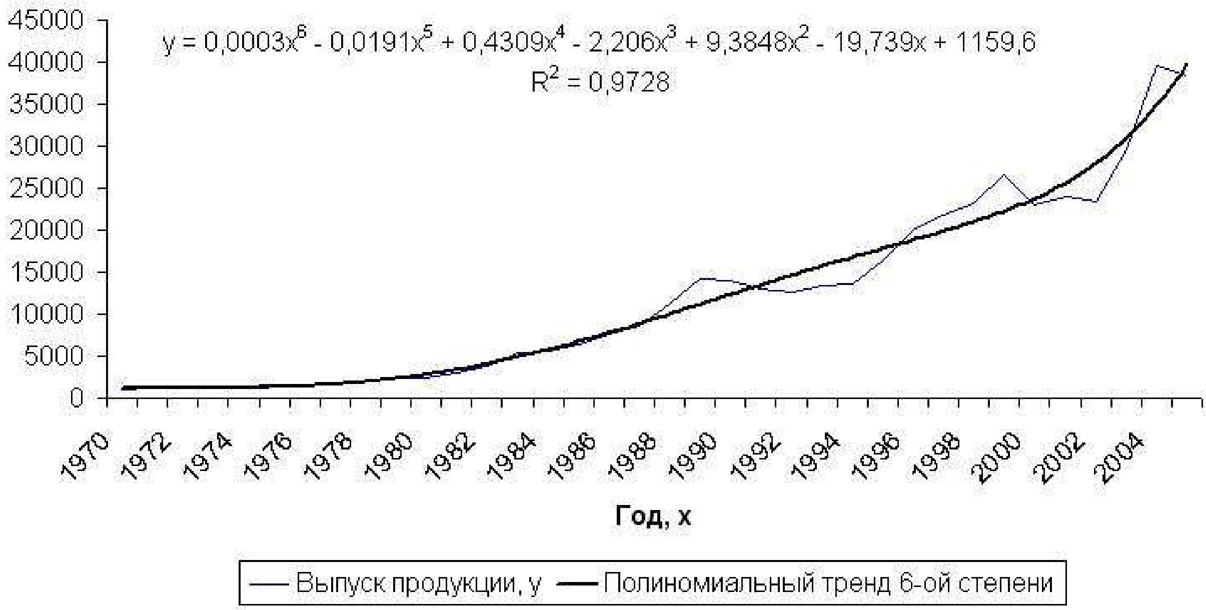


Рис.3.12 Полиномиальный тренд

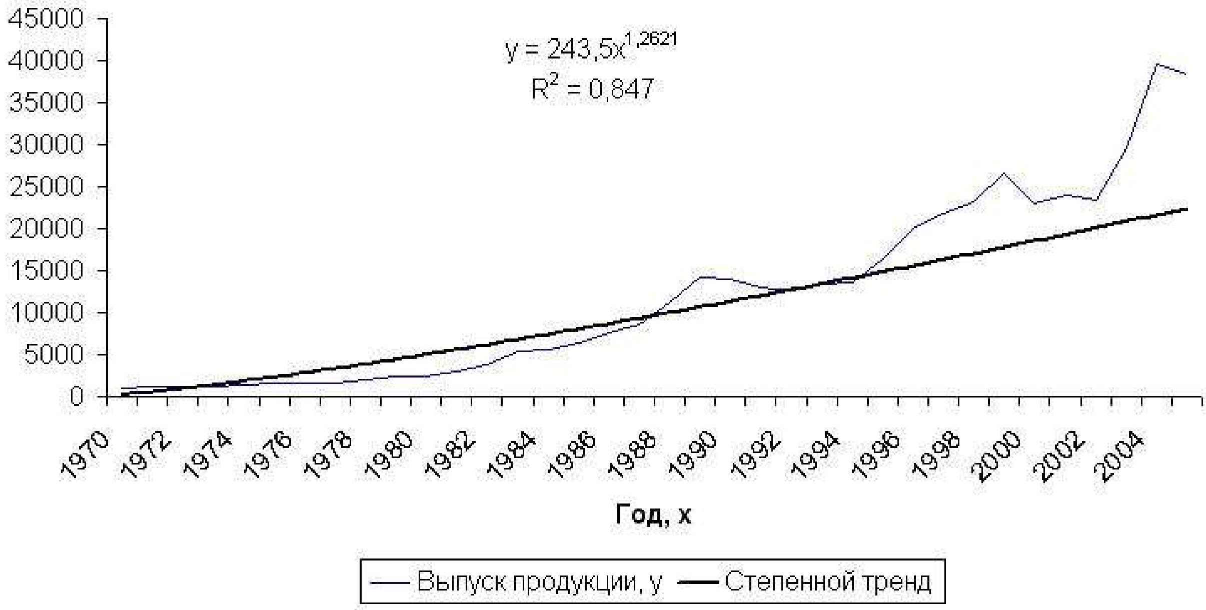


Рис.3.13Степенной тренд

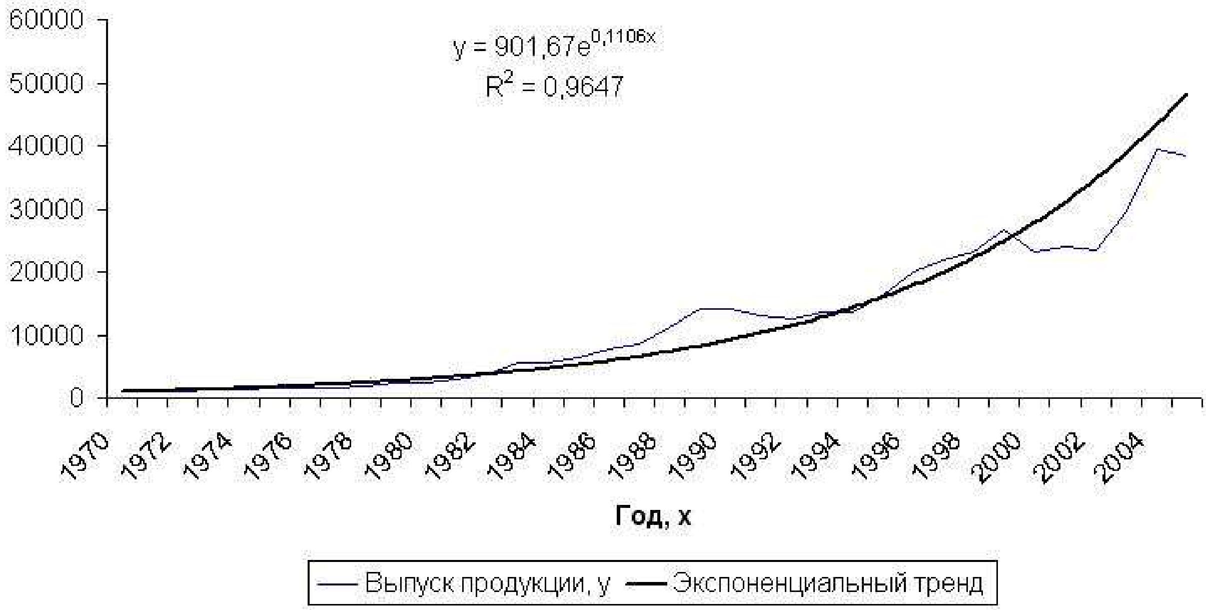


Рис.3.14 Экспоненциальный тренд

3. Сравним значения r_{xy}^2 (или R^2) по разным уравнениям трендов: полиномиальный 6-й степени - r_{xy}^2=0,9728; экспоненциальный - r_{xy}^2=0,9647; линейный - r_{xy}^2=0,8841; степенной - r_{xy}^2=0,8470; логарифмический - r_{xy}^2=0,5886.

Исходные данные лучше всего описывает полином 6-й степени. Следовательно, в рассматриваемом примере для прогнозных значений следует использовать полиномиальное уравнение.

3.4 Контрольные вопросы

1. Каковы основные элементы временного ряда?
2. В чем состоит задача эконометрического анализа временного ряда?
3. Перечислите основные виды трендов.
4. Что представляют собой параметры линейного и экспоненциального трендов?
5. Что такое аддитивная модель временного ряда? Перечислите этапы ее построения.
6. Как строится мультипликативная модель временного ряда?
7. Что такое скорректированная сезонная компонента и для чего она применяется?
8. Как выбрать наиболее предпочтительный тренд?
9. Пояснить особенности применения аддитивных и мультипликативных моделей.
10. Поясните расчет сезонной компоненты в аддитивных и мультипликативных моделях временных рядов.

3.5 Пример варианта промежуточного тестирования

1. Прогнозное значение уровня временного ряда в аддитивной модели равно:

а) разности трендового значения и значения сезонной компоненты;
б) трендовому значению;
в) случайному значению;
г) сумме трендового значения, случайного значения и значения сезонной компоненты.
д) сумме трендового и случайного значения.

2. На основе помесячных данных о числе раскрытых преступлений за последние два года была построена аддитивная модель временного ряда. Скорректированное значение сезонной компоненты за январь – S=-2, уравнение тренда: $\hat{y}_t = 10{,}5 + 0{,}1 \cdot t$. На основе модели число раскрытых преступлений на январь следующего года составит:

а) 12,6;
б) 10,5;
в) 12,5;
г) 11;
д) 15.

3. Для описания темпов роста заработной платы были рассмотрены следующие виды трендов: экспоненциальный, полиномиальный 8 степени, линейный, степенной и логарифмический. Значения коэффициентов детерминации для каждого тренда составляют соответственно:

а) 0,99;
б) 0,95;
в) 0,25;
г) 0,85;
д) 0,55.

Какой тренд следует выбрать?

4. Чему равно скорректированное значение сезонной компоненты за четвертый квартал, если значения первых трех кварталов составляют соответственно: +25, -15, +10?

а) +20;
б) +50;
в) -20;
г) -50;
д) 0.

5. Чему должна быть равна сумма значений сезонной компоненты внутри одного цикла:

а) 1;
б) 0;
в) -1;
г) 100%
д) нет постоянного значения.

6. Модель временного ряда – это:

а) модель циклического процесса;
б) модель, построенная по данным, характеризующим один объект за ряд последовательных моментов времени;
в) модель поведения процесса в пространстве и времени;
г) значения характеристик объекта в начальный момент времени;
д) модель, построенная по данным, характеризующим ряд объектов в определенный момент времени.

7. Модель, в которой временной ряд представлен произведением трендовой, циклической и случайной компонент, называется:

а) мультипликативной;
б) аддитивной;
в) множественной;
г) линейной;
д) нелинейной.

4 Система экономических уравнений

4.1 Методические указания

Сложные экономические процессы описываются с помощью *системы взаимосвязанных (одновременных) уравнений*.

Различают несколько видов систем уравнений:

— *система независимых уравнений* – когда каждая зависимая переменная *y* рассматривается как функция одного и того же набора факторов x;

$$\begin{cases} y_1 = a_{11} \cdot x_1 + a_{12} \cdot x_2 + ... + a_{1m} \cdot x_m \\ y_2 = a_{21} \cdot x_1 + a_{22} \cdot x_2 + ... + a_{2m} \cdot x_m \\ .... \\ y_n = a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + ... + a_{nm} \cdot x_m \end{cases}$$

Для решения этой системы и нахождения ее параметров используется метод наименьших квадратов.

— *система рекурсивных уравнений* – когда зависимая переменная *y* одного уравнения выступает в виде фактора *x* в другом уравнении.

$$\begin{cases} y_1 = a_{11} \cdot x_1 + a_{12} \cdot x_2 + ... + a_{1m} \cdot x_m \\ y_2 = b_{21} \cdot y_1 + a_{21} \cdot x_1 + a_{22} x_2 + ... + a_{2m} \cdot x_m \\ y_3 = b_{31} \cdot y_1 + b_{32} \cdot y_2 + a_{31} \cdot x_1 + a_{32} x_2 + ... + a_{3m} \cdot x_m \\ .... \\ y_n = b_{n1} \cdot y_1 + b_{n2} \cdot y_2 + ... + a_{n1} \cdot x_1 + a_{n2} x_2 + ... + a_{nm} \cdot x_m \end{cases}$$

Для решения этой системы и нахождения ее параметров используется метод наименьших квадратов.

— *система взаимосвязанных (совместных) уравнений* – когда одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других – в правую.

$$\begin{cases} y_1 = b_{12} \cdot y_2 + b_{13} \cdot y_3 + ... + a_{11} \cdot x_1 + a_{12}x_2 + ... + a_{1m} \cdot x_m \\ y_2 = b_{21} \cdot y_1 + b_{23} \cdot y_3 + ... + a_{21} \cdot x_1 + a_{22}x_2 + ... + a_{2m} \cdot x_m \\ .... \\ y_n = b_{n1} \cdot y_1 + b_{n2} \cdot y_2 + ... + a_{n1} \cdot x_1 + a_{n2}x_2 + ... + a_{nm} \cdot x_m \end{cases}$$

Такая система уравнений называется *структурной формой модели*.

Система совместных, одновременных уравнений (или структурная форма модели) обычно содержит эндогенные и экзогенные переменные.

Эндогенные переменные (у) – взаимозависимые переменные, которые определяются внутри модели (системы).

Экзогенные переменные (х) – независимые переменные, которые определяются вне системы.

Классификация переменных на эндогенные и экзогенные зависит от теоретической концепции принятой модели. Экономические переменные могут выступать в одних моделях как эндогенные, а в других как экзогенные переменные. Внеэкономические переменные (например, климатические условия, социальное положение, пол, возрастная категория) входят в систему только как экзогенные переменные. В качестве экзогенных переменных могут рассматриваться значения эндогенных переменных за предшествующий период времени (*лаговые переменные*).

Коэффициенты a и b при переменных – *структурные коэффициенты модели*.

Система линейных функций эндогенных переменных от всех предопределенных переменных системы – *приведенная форма модели*:

$$\begin{cases} y_1 = \delta_{11} \cdot x_1 + \delta_{12} \cdot x_2 + ... + \delta_{1m} \cdot x_m \\ y_2 = \delta_{21} \cdot x_1 + \delta_{22} \cdot x_2 + ... + \delta_{2m} \cdot x_m \\ .... \\ y_n = \delta_{n1} \cdot x_1 + \delta_{n2} \cdot x_2 + ... + \delta_{nm} \cdot x_m \end{cases}$$

где δ – коэффициенты приведенной формы модели.

При переходе от приведенной формы модели к структурной возникает проблема идентификации.

Идентификация – это единственность соответствия между приведенной и структурной формами модели.

С позиции идентифицируемости структурные модели можно подразделить на три вида:

— идентифицируемые;

— неидентифицируемые;

— сверхидентифицируемые.

Модель идентифицируема, если все ее структурные коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, т.е. если число параметров структурной модели равно числу параметров приведенной формы модели. В этом случае структурные коэффициенты модели оцениваются через параметры приведенной формы модели и модель идентифицируема.

Модель неидентифицируема, если число приведенных коэффициентов меньше числа структурных коэффициентов, и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели.

Модель сверхидентифицируема, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. В этой модели число структурных коэффициентов меньше числа коэффициентов приведенной формы.

Необходимое условие идентификации – выполнение счетного правила:

$D + 1 = H$ – уравнение идентифицируемо;

$D + 1 < H$ – уравнение неидентифицируемо;

$D + 1 > H$ – уравнение сверхидентифицируемо;

где H – число эндогенных переменных в уравнении,

D – число предопределенных переменных, отсутствующих в уравнении, но присутствующих в системе.

Достаточное условие идентификации – определитель матрицы, составленной из коэффициентов при переменных, отсутствующих в исследуемом уравнении, не равен *0* и ранг этой матрицы не менее числа эндогенных переменных системы без единицы.

Для решения идентифицируемого уравнения применяется *косвенный метод наименьших квадратов*, для решения сверхидентифицированных – *двухшаговый метод наименьших квадратов*.

Косвенный метод наименьших квадратов (КМНК) применяется в случае точно идентифицируемой структурной модели. Процедура применения КМНК предполагает выполнение следующих этапов работы:

- составляют приведенную форму модели и определяют численный значения параметров каждого его уравнения обычным МНК;
- путем алгебраических преобразований переходят от приведенной формы к уравнениям структурной формы модели, получая тем самым численные оценки структурных параметров.

Если система сверхидентифицируема, то КМНК не используется, так как он не дает однозначных оценок для параметров структурной модели. В этом случае могут использоваться разные методы оценивания, среди которых наиболее распространенным и простым является двухшаговый метод наименьших квадратов (ДМНК):

- составляют приведенную форму модели и определяют численный значения параметров каждого его уравнения обычным МНК;
- выявляют эндогенные переменные, находящиеся в правой части структурного уравнения, параметры которого определяют косвенным МНК, и находят расчетные значения по соответствующим уравнениям приведенной формы модели;
- обычным МНК определяют параметры структурного уравнения, используя в качестве исходных данных фактические значения предопределенных переменных и расчетные значения эндогенных

переменных, стоящих в правой части данного структурного уравнения.

4.2 Контрольные вопросы

1) Назовите возможные способы построения систем уравнений. В чем их отличия?
2) В чем заключаются проблемы идентификации модели?
3) Каковы необходимые условия идентификации?
4) Каковы достаточные условия идентификации?
5) Что такое эндогенные переменные?
6) В чем состоит косвенный метод наименьших квадратов?
7) Что такое двухшаговый метод наименьших квадратов? В каком случае он применяется?
8) Что такое лаговые переменные?
9) Что такое экзогенные переменные?
10) Что такое структурная форма модели? Для чего она применяется?

4.3 Примерный вариант итогового тестирования

1. Труды каких ученых XIX века явились существенным толчком в развитии статистической обработки результатов и применении парной корреляции в экономических исследованиях (например, при изучении связи между уровнем бедности и формами помощи бедным, уровнем брачности и благосостоянием)?

а) Ф. Гальтон, К. Пирсон, Ф. Эджворт.
б) Дж. Кларк, Г. Мур, А. Маршалл.
в) К. Жюгляр, С. Кузнец, Н. Кондратьев
г) В. Петти, Г. Кинг, Ч. Давенант
д) У. Персон, У. Митчелл, В. Леонтьев

2. На основе помесячных данных о числе раскрытых преступлений за последние два года была построена аддитивная модель временного ряда. Скорректированное значение сезонной компоненты за январь – S=-2, уравнение тренда: $\hat{y}_t = 10{,}5 + 0{,}1 \cdot t$. На основе модели число раскрытых преступлений на январь следующего года составит:

а) 12,6;
б) 10,5;
в) 12,5;
г) 11;
д) 15.

3. Эндогенные переменные – это:

а) независимые заданные параметры;
б) независимые переменные;
в) независимые переменные, которые определяются вне системы;
г) взаимозависимые переменные;
д) взаимозависимые переменные, которые определяются внутри модели.

4. Идентификация модели – это единственность соответствия между:

а) приведенной и рекурсивной формами модели;
б) системой рекурсивных и независимых уравнений;
в) системой совместных и независимых уравнений;
г) приведенной и структурной формами модели;
д) системой совместных и рекурсивных уравнений.

4. На какие виды можно разделить структурные модели с точки зрения идентифицируемости?

а) идентифицируемые и неидентифицируемые;
б) идентифицируемые, неидентифицируемые и сверхидентифицируемые;
в) идентифицируемые и квазиидентифицируемые;
г) сверхидентифицируемые и неидентифицируемые;
д) идентифицируемые, неидентифицируемые и квазиидентифицируемые.

6. В каких пределах лежат значения линейного коэффициента парной корреляции для линейной регрессии:

а) $r_{xy} \le 1$;
б) $0 \le r_{xy} \le 1$;
в) $-1 \le r_{xy} \le 0$;
г) $r_{xy} \ge 1$;
д) $-1 \le r_{xy} \le 1$.

7. Величина F-критерия для проверки гипотезы H_0:$D_{факт}$=$D_{ост}$ это:

а) отношение факторной дисперсии к остаточной;
б) произведение факторной и остаточной дисперсий;
в) отношение общей дисперсии к остаточной;
г) произведение общей и остаточной дисперсий;
д) отношение общей дисперсии к произведению факторной и остаточной.

8. При получении уравнения линейной регрессии, описывающей зависимость расходов на покупку бытовой техники в общих расходах от уровня заработной платы был рассчитан линейный коэффициент парной корреляции $r_{xy}=0,3$. Чему равен коэффициент детерминации:

а) 0,3;
б) 0,7;
в) 1,3;
г) 1,7;
д) 0,09.

9. Величина у складывается из двух параметров $y=\hat{y}_x+\varepsilon$, где $\hat{y}$ - это расчетное (теоретическое) значение. Что означает параметр ε?

а) "Возмущение", включающее в себя влияние неучтенных в модели факторов, случайные ошибки и особенности измерений;
б) "Колебание", отображающее величину среднего разброса результативного фактора;
в) "Распределение", показывающее разброс величины результативного фактора;
г) "Коллапс", равный минимальному значению из возможных значений признак-фактора;
д) "Комбинирование", показывающее величину разброса неучтенных факторов и погрешности измерений.

10. Какие величины скоррелированы максимально тесно?

а) Уровень благосостояния и "уровень брачности" (людей, состоящих в браке). r_{xy}=0,01;
б) Количество студентов в университете и доход от студентов-контрактников. $r_{xy}=0,8$;
в) Уровень бедности и форма бедности. $r_{xy}=0,3$;
г) Выплаты зарплаты персоналу фирмы и общий доход фирмы. $r_{xy}=-0,7$;
д) Средний размер пенсий и прожиточный минимум. $r_{xy}=-0,1$.

Литература

1) Эконометрика: Учебник / Под ред. И. И. Елисеевой. – М.: Финансы и статистика, 2001. – 344 с.: ил.
2) Эконометрия / В.И. Суслов и др. – Новосибирский государственный университет, 2005. – 744с.
3) Шалабанов А.К., Роганов Д.А. Эконометрика. Учебно-методическое пособие. ТИСБИ, Казань, 2004. – 198с.
4) Доугерти К. Введение в эконометрику. – М.: Финансы и статистика, 1999

Приложение 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	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28

k_1 / k_2	1	2	3	4	5	6	8	12	24	∞
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
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

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

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

Приложение 2

Индивидуальные задания для решения практических задач

Исходные данные к задаче №1 Парная линейная регрессия и корреляция

Вар. 1		**Вар. 2**		**Вар. 3**		**Вар. 4**		**Вар. 5**	
y	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*
11,8	3,4	28,7	8,5	8,2	3,8	16,2	12,5	47,8	7,1
25,8	7,3	22,5	6,1	0,7	0,6	16,9	13,9	37,5	5,1
13,8	4,0	30,5	9,2	7,4	3,5	13,9	7,9	50,8	7,7
8,9	2,6	21,8	5,8	6,5	3,1	14,1	8,2	36,4	4,8
21,8	6,2	12,7	2,2	3,4	1,8	16,5	13,1	21,2	1,8
32,1	9,0	28,2	8,4	5,1	2,5	18,3	16,7	47,0	7,0
26,1	7,4	36,7	11,8	8,2	3,8	13,7	7,4	61,1	9,8
39,1	10,9	31,4	9,6	5,3	2,6	14,2	8,5	52,3	8,0
37,6	10,6	28,3	8,4	1,6	1,0	15,9	11,8	47,1	7,0
25,4	7,2	21,7	5,8	2,6	1,4	15,6	11,3	36,2	4,8

Вар. 6		**Вар. 7**		**Вар. 8**		**Вар. 9**		**Вар. 10**	
y	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*
14,7	3,1	4,1	0,8	5,8	1,0	6,3	7,6	0,9	0,9
32,2	6,6	3,7	1,2	5,0	1,4	0,5	1,1	0,7	2,4
17,2	3,6	3,3	1,6	5,3	1,3	5,7	6,9	0,8	2,0
11,1	2,4	4,8	0,1	4,6	1,6	5,0	6,1	0,7	2,3
27,3	5,6	3,1	1,8	4,3	1,8	2,6	3,5	0,6	3,0
40,1	8,2	3,1	1,7	7,3	0,3	3,9	4,9	0,9	0,3
32,6	6,7	3,7	1,2	2,6	2,7	6,3	7,6	0,8	1,5
48,9	9,9	4,4	0,5	3,4	2,2	4,1	5,1	0,6	3,0
47,0	9,6	3,2	1,7	3,2	2,3	1,2	1,9	0,9	0,8
31,8	6,5	3,9	1,0	3,1	2,4	2,0	2,8	0,6	3,1

Исходные данные к задаче №2 Множественная регрессия и корреляция (варианты 1-5)

Вар. 1			**Вар. 2**			**Вар. 3**			**Вар. 4**			**Вар. 5**		
y	X_1	X_2	*y*	X_1	X_2	*y*	X_1	X_2	*y*	X_1	X_2	*y*	X_1	X_2
441,76	28,5	20,28	965,42	38	8,11	15,91	0,07	5,02	123,92	2,48	7,02	3,45	0,68	0,45
136,75	6,93	46,77	264,44	9,24	18,7	36,47	3,73	7,09	192,92	4,49	5,46	4,15	0,05	1,15
305,17	18,2	45,84	639,51	24,27	18,33	25,84	2,22	4,28	178,23	4,1	4,44	4,52	4,46	1,52
233,05	14,57	20,51	501,43	19,43	8,2	36,64	4,65	3,04	149,8	3,22	6,94	4,05	0,99	1,05
432,36	26,73	44,68	923,18	35,65	17,87	23,73	2,33	1,89	179,35	4,08	6,21	4,28	1,49	1,28
344,96	22,86	2,89	765,56	30,48	1,15	22,59	2,09	1,93	153,48	3,3	7,94	3,41	3,3	0,41
217,76	13,69	17,64	469,96	18,25	7,05	38,66	4,23	6,82	99,41	1,7	9,56	3,75	1,42	0,75
431,75	18,51	10,65	747,65	30,74	16,26	26,45	2,62	3,02	85,92	1,44	5,22	4,56	2,34	1,56
369,97	24,64	0,49	823,24	32,85	0,19	21,02	1,01	5,41	98,51	1,7	8,8	4,36	0,32	1,36
204,97	13,34	6,94	450,92	17,78	2,77	28,46	3,36	1,5	125,19	2,67	1,72	3,92	4,94	0,92
284,04	18,46	10,13	623,82	24,61	4,05	38,62	4,19	6,97	167,04	3,63	9,79	4,13	2,91	1,13
363,32	23,75	9,93	800,19	31,67	3,97	14,65	0,09	3,78	86,84	1,54	2,71	4,58	2,11	1,58
435,94	27,65	30,18	943,84	36,87	12,07	36,49	3,4	8,6	179,26	4,19	2,52	3,11	2,57	0,11
341,71	22,14	13,6	749,96	29,52	5,44	28,87	1,89	8,53	138,17	2,84	8,75	4,2	1,66	1,2
331,72	12,15	9,94	804,52	32,05	3,97	37,32	4,15	5,93	102,19	1,85	7,37	3,1	2,16	0,1

Исходные данные к задаче №2 Множественная регрессия и корреляция (варианты 6-10)

Вар. 6			**Вар. 7**			**Вар. 8**			**Вар. 9**			**Вар. 10**		
y	X_1	X_2	y	X_1	X_2	y	X_1	X_2	y	X_1	X_2	y	X_1	X_2
4,17	2,07	0,82	15,36	0,59	2,75	86,74	1,98	6,32	386,17	34,20	6,49	28,64	0,13	4,52
3,51	1,52	1,48	11,26	0,09	2,53	135,04	3,59	4,91	105,78	8,32	14,96	65,65	7,09	6,38
4,46	4,37	0,58	11,83	1,38	1,1	124,76	3,28	4,00	255,80	21,84	14,66	46,51	4,22	3,85
4,12	0,07	0,87	15,11	1,3	1,86	104,86	2,58	6,25	200,57	17,49	6,56	65,95	8,84	2,74
3,13	3,83	1,86	20,76	1,96	2,19	125,55	3,26	5,59	369,27	32,09	14,30	42,71	4,43	1,70
3,63	4,85	1,36	15,12	1,31	1,85	107,44	2,64	7,15	306,22	27,43	0,92	40,66	3,97	1,74
4,57	4,95	0,42	16,37	0,8	2,71	69,59	1,36	8,60	187,98	16,43	5,64	69,59	8,04	6,14
3,32	3,94	1,67	16,45	0,81	2,72	60,14	1,15	4,70	299,06	27,67	13,01	47,61	4,98	2,72
3,74	2,19	1,25	14,97	1,25	1,89	68,96	1,36	7,92	329,30	29,57	0,15	37,84	1,92	4,87
4,73	2,49	0,26	10,31	1,46	0,7	87,63	2,14	1,55	180,37	16,00	2,22	51,23	6,38	1,35
4,58	1,06	0,41	10,74	0,75	1,64	116,93	2,90	8,81	249,53	22,15	3,24	69,52	7,96	6,27
3,78	3,21	1,21	12,09	0,01	2,79	60,79	1,23	2,44	320,08	28,50	3,18	26,37	0,17	3,40
3,74	1,6	1,25	11,91	0,11	2,65	125,48	3,35	2,27	377,54	33,18	9,66	65,68	6,46	7,74
4,25	4,8	0,74	16,87	1,5	1,96	96,72	2,27	7,88	299,98	26,57	4,35	51,97	3,59	7,68
3,84	3,63	1,15	13,4	1,58	1,17	71,53	1,48	6,63	321,81	28,85	3,18	67,18	7,89	5,34

Исходные данные к задаче №3 Временные ряды

Вар. 1	**Вар. 2**	**Вар. 3**	**Вар. 4**	**Вар. 5**	**Вар. 6**	**Вар. 7**	**Вар. 8**	**Вар. 9**	**Вар. 10**
5216	2395	5817	6329	7458	2456	10945	3912	1964	2676
6011	2960	7120	7180	8113	2554	11042	4508	2427	3275
8715	3200	7056	7762	8561	2621	11108	6536	2624	3246
11009	4100	6655	9436	9849	2813	11299	8257	3362	3061
12990	4325	6406	10301	10515	2912	11397	9743	3547	2947
13689	4631	6822	10669	10797	2954	11439	10267	3797	3138
13950	5175	6886	11352	11323	3033	11516	10463	4244	3168
14497	5268	8085	11259	11250	3022	11506	10873	4320	3719
11635	4995	9695	10916	11059	2983	11467	8726	4096	4460
14690	4856	10444	11010	12415	2993	11477	11018	3982	4804
17500	5965	11122	12772	14236	3196	11678	13125	4891	5116
19980	7230	12558	15140	15083	3468	11947	14985	5929	5777
21600	7650	11026	16245	15850	3594	12072	16200	6273	5072
25960	8257	18760	17098	16188	3692	12169	19470	6771	8630
26308	9753	11200	19343	15742	3950	12425	19731	7997	5152
27570	8200	11500	17238	17469	3704	12185	20678	6724	5290
27390	8560	16500	17668	15923	3759	12238	20543	7019	7590
29675	8200	19200	17320	23947	3719	12196	22256	6724	8832
37980	14300	25600	27510	22958	5585	13076	28485	11726	11776
40153	10500	21350	25060	23905	4925	13380	30115	8610	9821