

ПРАКТИКУМ ПО ЭКОНОМЕТРИКЕ

**Под редакцией
члена-корреспондента
Российской Академии наук
И.И.Елисеевой**

**Рекомендовано
Учебно-методическим объединением
по образованию в области экономики, статистики,
информационных систем и
математических методов в экономике
в качестве учебного пособия
для экономических вузов**



**МОСКВА
“ФИНАНСЫ И СТАТИСТИКА”
2003**

УДК 330.43(076.5)
ББК 65в6я73
П69

АВТОРЫ:

**И.И. Елисеева, С.В. Курышева,
Н.М. Гордеенко, И.В. Бабаева,
Т.В. Костеева, Б.А. Михайлов**

РЕЦЕНЗЕНТЫ:

**кафедра математической статистики и эконометрики
Московского государственного университета экономики,
статистики и информатики;
В.И. Афанасьев,
доктор экономических наук, профессор**

**Практикум по эконометрике: Учеб. пособие / И. И. Елисеева,
П69 С. В. Курышева, Н. М. Гордеенко и др.; Под ред. И. И. Елисеевой. –
М.: Финансы и статистика, 2003. – 192 с.: ил.**

ISBN 5-279-02313-2

Практикум – дополнение к учебнику «Эконометрика» – обеспечивает методическую поддержку практических занятий. Не требует основательной математической подготовки. Содержит краткие методические указания, решение типовых задач, описание реализации на компьютере с помощью популярных пакетов прикладных программ Excel, Statistica, а также контрольные задания.

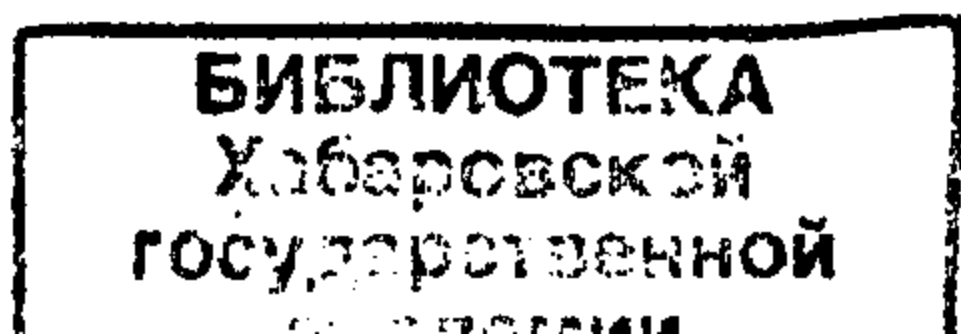
Для преподавателей, аспирантов, студентов экономических вузов, слушателей институтов повышения квалификации.

П $\frac{0702000000 - 154}{010(01) - 2003}$ 231 – 2002

УДК 330.43(076.5)
ББК 65в6я73

ISBN 5-279-02313-2

© Коллектив авторов, 2001



ПРЕДИСЛОВИЕ

Данный практикум представляет собой попытку создания учебного пособия, ориентированного на специфику преподавания эконометрики в экономических вузах. Его структура и содержание базируются на опыте преподавания этой дисциплины в Санкт-Петербургском государственном университете экономики и финансов и изучении зарубежного опыта.

Большое влияние на формирование методики преподавания эконометрики в вузах России, как известно, оказало проведение в 1998–2000 гг. двух международных школ по преподаванию этой дисциплины (руководители – С.А. Айвазян, П.К. Катъшев, А.А. Пересецкий, Я. Магнус), в которых прошла обучение подавляющая часть отечественных преподавателей.

Практикум ориентирован на начальный курс эконометрики.

Изучение этой дисциплины предполагает приобретение студентами опыта построения эконометрических моделей, принятия решений о спецификации и идентификации модели, выбора метода оценки параметров модели, интерпретации результатов, получения прогнозных оценок. Студенты должны также научиться давать статистическую оценку значимости таких искажающих эффектов, как гетероскедастичность остатков зависимой переменной, мультиколлинеарность объясняющих переменных, автокорреляция. В связи с этим курс эконометрики обязательно включает решение задач. Соответственно методическое обеспечение курса должно состоять из учебника и практикума.

Предлагаемый практикум является дополнением к учебнику «Эконометрика», подготовленному тем же коллективом авторов. Практикум охватывает основные темы курса. Главное внимание уделяется построению эконометрических моделей на основе пространственных данных и временных рядов. Все разделы практикума имеют идентичную структуру:

- краткие методические положения, включающие основные понятия, определения, формулы;
- решение типовых задач;
- указания по реализации типовой задачи на компьютере с помощью пакетов прикладных программ (ППП) Excel, Statgraphics или Statistica;
- задачи, предлагаемые студентам для тренировки и для контроля.

Разделы практикума корреспондируют с главами учебника: I раздел практикума соответствует главе 2 учебника, II раздел – главе 3, III раздел – главе 4, IV раздел – главам 5, 6 и 7.

Формулировки практически всех заданий нацелены на применение результатов эконометрического анализа.

Данные, используемые в задачах, охватывают широкий спектр направлений применения эконометрики. Данные могут обновляться и расширяться прежде всего за счет привлечения материалов официальных статистических публикаций, например статистического сборника «Регионы России», содержащего сведения о потреблении, ценах, доходах и т.д. в субъектах Российской Федерации.

Большое число задач составлено таким образом, чтобы обеспечить индивидуализацию работы студента: предусмотрена возможность различных комбинаций объясняющих переменных, выбор различной объясняемой (зависимой) переменной, предлагаются дифференцированные задания. Такая гибкость формулировок заданий позволяет преподавателю учесть вкусы студентов при распределении упражнений, организовать работу в малых группах. Кроме того, каждый раздел практикума содержит упражнения разной степени сложности.

Наличие в практикуме таких рубрик, как «Методические указания», «Решение типовых задач» и «Реализация типовых задач на компьютере», дает возможность студентам освоить материал с минимальными затратами. Эти рубрики полезны и преподавателям для планирования содержания практических занятий, выделения главных понятий, подходов к измерению.

В конце практикума находятся основные статистико-математические таблицы, необходимые для решения задач.

Практикум может быть полезен при освоении не только эконометрики, но и курса «Математическая статистика».

Труд авторов распределен следующим образом: д-р экон. наук И.И. Елисеева – предисловие, разд. 1.1, 1.4 и 3.1; д-р экон. наук С.В. Курышева – разд. 1.1, 1.2, 1.4, 2.1, 2.2, 2.4, 3.1, 3.2, 3.3, 4.2 и 4.4; канд. экон. наук Б.А. Михайлов – разд. 1.2, 1.4, 3.3 и 4.4; канд. экон. наук Н.М. Гордеев – разд. 1.4, 2.4 и 3.3; канд. экон. наук Т.В. Костеева – разд. 1.4, 2.2 и 2.4, 3.2, 3.3, 4.1, 4.2 и 4.4; И.В. Бабаева – разд. 1.3, 2.3, 4.2, 4.3 и 4.4. Работа Т.В. Костеевой выполнена при частичной поддержке гранта института Открытое Общество.

Замечания и пожелания по совершенствованию практикума просим направлять в издательство.

Научный редактор
член-корреспондент РАН

И. И. ЕЛИСЕЕВА

РАЗДЕЛ

ПАРНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

1.1. МЕТОДИЧЕСКИЕ УКАЗАНИЯ

Парная регрессия – уравнение связи двух переменных y и x :

$$y = \hat{f}(x),$$

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, и регрессии, нелинейные по оцениваемым параметрам.

Регрессии, *нелинейные по объясняющим переменным*:

- полиномы разных степеней $y = a + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3 + \varepsilon$;
- равносторонняя гипербола $y = a + \frac{b}{x} + \varepsilon$.

Регрессии, *нелинейные по оцениваемым параметрам*:

- степенная $y = a \cdot x^b \cdot \varepsilon$;
- показательная $y = a \cdot b^x \cdot \varepsilon$;
- экспоненциальная $y = e^{a+b \cdot x} \cdot \varepsilon$.

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, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических \hat{y}_x минимальна, т.е.

$$\sum \left(y - \hat{y}_x \right)^2 \rightarrow \min.$$

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:

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

Тесноту связи изучаемых явлений оценивает *линейный коэффициент парной корреляции* r_{xy} для линейной регрессии ($-1 \leq r_{xy} \leq 1$):

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

и *индекс корреляции* ρ_{xy} — для нелинейной регрессии ($0 \leq \rho_{xy} \leq 1$):

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}}.$$

Оценку качества построенной модели даст коэффициент (индекс) детерминации, а также средняя ошибка аппроксимации.

Средняя ошибка аппроксимации — среднее отклонение расчетных значений от фактических:

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

Допустимый предел значений \bar{A} — не более 8 – 10%.

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

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

Задача *дисперсионного анализа* состоит в анализе дисперсии зависимой переменной:

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

- где $\Sigma(y - \bar{y})^2$ – общая сумма квадратов отклонений;
 $\Sigma(\hat{y}_x - \bar{y})^2$ – сумма квадратов отклонений, обусловленная регрессией («объясненная» или «факторная»);
 $\Sigma(y - \hat{y}_x)^2$ – остаточная сумма квадратов отклонений.

Долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака у характеризует коэффициент (индекс) детерминации R^2 :

$$R^2 = \frac{\Sigma(\hat{y}_x - \bar{y})^2}{\Sigma(y - \bar{y})^2}.$$

Коэффициент детерминации – квадрат коэффициента или индекса корреляции.

F-тест – оценивание качества уравнения регрессии – состоит в проверке гипотезы H_0 о статистической незначимости уравнения регрессии и показателя тесноты связи. Для этого выполняется сравнение фактического $F_{\text{факт}}$ и критического (табличного) $F_{\text{табл}}$ значений *F-критерия Фишера*. $F_{\text{факт}}$ определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы:

$$F_{\text{факт}} = \frac{\Sigma(\hat{y} - \bar{y})^2 / m}{\Sigma(y - \hat{y})^2 / (n - m - 1)} = \frac{r_{xy}^2}{1 - r_{xy}^2} (n - 2),$$

- где n – число единиц совокупности;
 m – число параметров при переменных x .

$F_{\text{табл}}$ – это максимально возможное значение критерия под влиянием случайных факторов при данных степенях свободы и уровне значимости α . Уровень значимости α – вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно α принимается равной 0,05 или 0,01.

Если $F_{\text{табл}} < F_{\text{факт}}$, то H_0 – гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность. Если $F_{\text{табл}} > F_{\text{факт}}$, то гипотеза H_0 не отклоняется и признается статистическая незначимость, ненадежность уравнения регрессии.

Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитываются t -критерий Стьюдента и доверительные интервалы каждого из показателей. Выдвигается гипотеза H_0 о случайной природе показателей, т.е. о незначимом их отличии от нуля. Оценка значимости коэффициентов регрессии и корреляции с помощью t -критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_b = \frac{b}{m_b}; \quad t_a = \frac{a}{m_a}; \quad t_r = \frac{r}{m_r}.$$

Случайные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

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

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

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

Сравнивая фактическое и критическое (табличное) значения t -статистики — $t_{\text{табл}}$ и $t_{\text{факт}}$ — принимаем или отвергаем гипотезу H_0 .

Связь между F -критерием Фишера и t -статистикой Стьюдента выражается равенством

$$t_r^2 = t_b^2 = \sqrt{F}.$$

Если $t_{\text{табл}} < t_{\text{факт}}$, то H_0 отклоняется, т.е. a , b и r_{xy} не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x . Если $t_{\text{табл}} > t_{\text{факт}}$, то гипотеза H_0 не отклоняется и признается случайная природа формирования a , b или r_{xy} .

Для расчета доверительного интервала определяем предельную ошибку Δ для каждого показателя:

$$\Delta_a = t_{\text{табл}} m_a, \quad \Delta_b = t_{\text{табл}} m_b.$$

Формулы для расчета *доверительных интервалов* имеют следующий вид:

$$\begin{aligned} \gamma_a &= a \pm \Delta a; & \gamma_{a_{\min}} &= a - \Delta a; & \gamma_{a_{\max}} &= a + \Delta a; \\ \gamma_b &= b \pm \Delta b; & \gamma_{b_{\min}} &= b - \Delta b; & \gamma_{b_{\max}} &= b + \Delta b. \end{aligned}$$

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

Прогнозное значение y_p определяется путем подстановки в уравнение регрессии $\hat{y}_x = a + b \cdot x$ соответствующего (прогнозного) значения x_p . Вычисляется *средняя стандартная ошибка прогноза* $m_{\hat{y}_p}$:

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

где $\sigma_{\text{ост}} = \sqrt{\frac{\sum (y - \hat{y})^2}{n - m - 1}}$;

и строится *доверительный интервал прогноза*:

$$\gamma_{\hat{y}_p} = \hat{y}_p \pm \Delta \hat{y}_p; \quad \gamma_{\hat{y}_p_{\min}} = \hat{y}_p - \Delta \hat{y}_p; \quad \gamma_{\hat{y}_p_{\max}} = \hat{y}_p + \Delta \hat{y}_p,$$

где $\Delta \hat{y}_p = t_{\text{табл}} \cdot m_{\hat{y}_p}$.

1.2. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Пример 1

По семи территориям Уральского района за 199X г. известны значения двух признаков (табл. 1.1).

Таблица 1.1

Район	Расходы на покупку продовольственных товаров в общих расходах, %, y	Среднедневная заработная плата одного работающего, руб., x
Удмуртская респ.	68,8	45,1
Свердловская обл.	61,2	59,0
Башкортостан	59,9	57,2
Челябинская обл.	56,7	61,8
Пермская обл.	55,0	58,8
Курганская обл.	54,3	47,2
Оренбургская обл.	49,3	55,2

Требуется:

- Для характеристики зависимости y от x рассчитать параметры следующих функций:
 - линейной;
 - степенной;
 - показательной;
 - равносторонней гиперболы.
- Оценить каждую модель через среднюю ошибку аппроксимации \bar{A} и F -критерий Фишера.

Решение

1а. Для расчета параметров a и b линейной регрессии $y = a + b \cdot x$ решаем систему нормальных уравнений относительно a и b :

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

По исходным данным рассчитываем $\sum y$, $\sum x$, $\sum yx$, $\sum x^2$, $\sum y^2$.

Таблица 1.2

	y	x	yx	x ²	y ²	ŷ _x	y - ŷ _x	A _i
1	68,8	45,1	3102,88	2034,01	4733,44	61,3	7,5	10,9
2	61,2	59,0	3610,80	3481,00	3745,44	56,5	4,7	7,7
3	59,9	57,2	3426,28	3271,84	3588,01	57,1	2,8	4,7
4	56,7	61,8	3504,06	3819,24	3214,89	55,5	1,2	2,1
5	55,0	58,8	3234,00	3457,44	3025,00	56,5	-1,5	2,7
6	54,3	47,2	2562,96	2227,84	2948,49	60,5	-6,2	11,4
7	49,3	55,2	2721,36	3047,04	2430,49	57,8	-8,5	17,2
Итого	405,2	384,3	22162,34	21338,41	23685,76	405,2	0,0	56,7
Среднее значение	57,89	54,90	3166,05	3048,34	3383,68	x	x	8,1
σ	5,74	5,86	x	x	x	x	x	x
σ ²	32,92	34,34	x	x	x	x	x	x

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x^2} = \frac{3166,05 - 57,89 \cdot 54,9}{5,86^2} \approx -0,35,$$

$$a = \bar{y} - b \cdot \bar{x} = 57,89 + 0,35 \cdot 54,9 \approx 76,88.$$

Уравнение регрессии: $\hat{y} = 76,88 - 0,35 \cdot x$. С увеличением среднедневной заработной платы на 1 руб. доля расходов на покупку продовольственных товаров снижается в среднем на 0,35 %-ных пункта.

Рассчитаем линейный коэффициент парной корреляции:

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = -0,35 \cdot \frac{5,86}{5,74} = -0,357.$$

Связь умеренная, обратная.

Определим коэффициент детерминации:

$$r_{xy}^2 = (-0,35)^2 = 0,127.$$

Вариация результата на 12,7% объясняется вариацией фактора x.

Подставляя в уравнение регрессии фактические значения x, определим теоретические (расчетные) значения \hat{y}_x . Найдем величину средней ошибки аппроксимации \bar{A} :

$$\bar{A} = \frac{1}{n} \sum A_i = \frac{1}{n} \sum |y - \hat{y}| \cdot 100 \% = \frac{56,7 \cdot 100 \%}{7} = 8,1 \%.$$

В среднем расчетные значения отклоняются от фактических на 8,1%.

Рассчитаем F -критерий:

$$F_{\text{факт}} = \frac{0,127}{0,873} \cdot 5 = 0,7,$$

поскольку $1 \leq F \leq \infty$, следует рассмотреть F^{-1} .

Полученное значение указывает на необходимость принять гипотезу H_0 о случайной природе выявленной зависимости и статистической незначимости параметров уравнения и показателя тесноты связи.

16. Построению степенной модели $y = a \cdot x^b$ предшествует процедура линеаризации переменных. В примере линеаризация производится путем логарифмирования обеих частей уравнения:

$$\lg y = \lg a + b \cdot \lg x;$$

$$Y = C + b \cdot X,$$

где $Y = \lg y$, $X = \lg x$, $C = \lg a$.

Для расчетов используем данные табл. 1.3.

Таблица 1.3

	Y	X	YX	Y^2	X^2	\hat{y}_x	$y - \hat{y}_x$	$(y - \hat{y}_x)^2$	A_i
1	1,8376	1,6542	3,0398	3,3768	2,7364	61,0	7,8	60,8	11,3
2	1,7868	1,7709	3,1642	3,1927	3,1361	56,3	4,9	24,0	8,0
3	1,7774	1,7574	3,1236	3,1592	3,0885	56,8	3,1	9,6	5,2
4	1,7536	1,7910	3,1407	3,0751	3,2077	55,5	1,2	1,4	2,1
5	1,7404	1,7694	3,0795	3,0290	3,1308	56,3	-1,3	1,7	2,4
6	1,7348	1,6739	2,9039	3,0095	2,8019	60,2	-5,9	34,8	10,9
7	1,6928	1,7419	2,9487	2,8656	3,0342	57,4	-8,1	65,6	16,4
Ито- го	12,3234	12,1587	21,4003	21,7078	21,1355	403,5	1,7	197,9	56,3
Сред- нее значе- ние	1,7605	1,7370	3,0572	3,1011	3,0194	x	x	28,27	8,0
σ	0,0425	0,0484	x	x	x	x	x	x	x
σ^2	0,0018	0,0023	x	x	x	x	x	x	x

Рассчитаем C и b :

$$b = \frac{\overline{Y \cdot X} - \bar{Y} \cdot \bar{X}}{\sigma_X^2} = \frac{3,0572 - 1,7605 \cdot 1,7370}{0,0484^2} \approx -0,298;$$

$$C = \bar{Y} - b \cdot \bar{X} = 1,7605 + 0,298 \cdot 1,7370 = 2,278.$$

Получим линейное уравнение: $\hat{Y} = 2,278 - 0,298 \cdot X$.

Выполнив его потенцирование, получим:

$$\hat{y} = 10^{2,278} \cdot x^{-0,298} = 189,7 \cdot x^{-0,298}.$$

Подставляя в данное уравнение фактические значения x , получаем теоретические значения результата \hat{y}_x . По ним рассчитаем показатели: тесноты связи – индекс корреляции ρ_{xy} и среднюю ошибку аппроксимации \bar{A}_i :

$$\rho_{xy} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}} = \sqrt{1 - \frac{28,27}{32,92}} = 0,3758, \quad \bar{A} = 8,0\%.$$

Характеристики степенной модели указывают, что она несколько лучше линейной функции описывает взаимосвязь.

1в. Построению уравнения показательной кривой $y = a \cdot b^x$ предшествует процедура линеаризации переменных при логарифмировании обеих частей уравнения:

$$\lg y = \lg a + x \cdot \lg b;$$

$$Y = C + B \cdot x,$$

где $Y = \lg y$, $C = \lg a$, $B = \lg b$.

Для расчетов используем данные табл. 1.4.

Таблица 1.4

	Y	x	Yx	Y^2	X^2	\hat{y}_x	$y - \hat{y}_x$	$(y - \hat{y}_x)^2$	A_i
1	1,8376	45,1	82,8758	3,3768	2034,01	60,7	8,1	65,61	11,8
2	1,7868	59,0	105,4212	3,1927	3481,00	56,4	4,8	23,04	7,8
3	1,7774	57,2	101,6673	3,1592	3271,84	56,9	3,0	9,00	5,0
4	1,7536	61,8	108,3725	3,0751	3819,24	55,5	1,2	1,44	2,1
5	1,7404	58,8	102,3355	3,0290	3457,44	56,4	-1,4	1,96	2,5
6	1,7348	47,2	81,8826	3,0095	2227,84	60,0	-5,7	32,49	10,5
7	1,6928	55,2	93,4426	2,8656	3047,04	57,5	-8,2	67,24	16,6
Ито- го	12,3234	384,3	675,9974	21,7078	21338,41	403,4	-1,8	200,78	56,3
Сред- нее значе- ние	1,7605	54,9	96,5711	3,1011	3048,34	\bar{x}	\bar{y}	28,68	8,0
σ	0,0425	5,86	\bar{x}	\bar{x}	\bar{x}	\bar{x}	\bar{x}	\bar{x}	\bar{x}
σ^2	0,0018	34,3396	\bar{x}	\bar{x}	\bar{x}	\bar{x}	\bar{x}	\bar{x}	\bar{x}

Значения параметров регрессии A и B составили:

$$B = \frac{\overline{Y \cdot x} - \bar{Y} \cdot \bar{x}}{\sigma_x^2} = \frac{96,5711 - 1,7605 \cdot 54,9}{5,86^2} \approx -0,0023,$$

$$A = \bar{Y} - B \cdot \bar{x} = 1,7605 + 0,0023 \cdot 54,9 = 1,887.$$

Получено линейное уравнение: $\overset{\Lambda}{Y} = 1,887 - 0,0023 \cdot x$.

Произведем потенцирование полученного уравнения и запишем его в обычной форме:

$$\overset{\Lambda}{y} = 10^{1,887} \cdot 10^{-0,0023x} = 77,1 \cdot 0,9947^x.$$

Тесноту связи оценим через индекс корреляции ρ_{xy} :

$$\rho_{xy} = \sqrt{1 - \frac{\sum \left(y - \overset{\Lambda}{y}_x \right)^2}{\sum (y - \bar{y})^2}} = \sqrt{1 - \frac{28,27}{32,92}} = 0,3589.$$

Связь умеренная.

$\bar{A} = 8,0\%$, что говорит о повышенной ошибке аппроксимации, но в допустимых пределах. Показательная функция чуть хуже, чем степенная, она описывает изучаемую зависимость.

1г. Уравнение равносторонней гиперболы $y = a + b \cdot \frac{1}{x}$

линеаризуется при замене: $z = \frac{1}{x}$. Тогда $y = a + b \cdot z$.

Для расчетов используем данные табл. 1.5.

Таблица 1.5

	y	z	yz	z ²	y ²	\hat{y}_x	$y - \hat{y}_x$	$(y - \hat{y}_x)^2$	A_i
1	68,8	0,0222	1,5255	0,000492	4733,44	61,8	7,0	49,00	10,2
2	61,2	0,0169	1,0373	0,000287	3745,44	56,3	4,9	24,01	8,0
3	59,9	0,0175	1,0472	0,000306	3588,01	56,9	3,0	9,00	5,0
4	56,7	0,0162	0,9175	0,000262	3214,89	55,5	1,2	1,44	2,1
5	55	0,0170	0,9354	0,000289	3025,00	56,4	-1,4	1,96	2,5
6	54,3	0,0212	1,1504	0,000449	2948,49	60,8	-6,5	42,25	12,0
7	49,3	0,0181	0,8931	0,000328	2430,49	57,5	-8,2	67,24	16,6
Итого	405,2	0,1291	7,5064	0,002413	23685,76	405,2	0,0	194,90	56,5
Среднее значение	57,9	0,0184	1,0723	0,000345	3383,68	x	x	27,84	8,1
σ	5,74	0,002145	x	X	x	x	x	x	x
σ^2	32,9476	0,000005	x	X	x	x	x	x	x

Значения параметров регрессии a и b составили:

$$a = \bar{y} - b \cdot \bar{z} = 57,89 - 1051,4 \cdot 0,0184 = 38,5;$$

$$b = \frac{\overline{y \cdot z} - \bar{y} \cdot \bar{z}}{\sigma_z^2} = \frac{1,0723 - 57,9 \cdot 0,0184}{0,002145^2} \approx 1051,4.$$

Получено уравнение: $\hat{y} = 38,5 + 1051,4 \cdot \frac{1}{x}$.

$$\text{Индекс корреляции: } \rho_{xy} = \sqrt{1 - \frac{27,84}{32,92}} = 0,3944.$$

$\bar{A} = 8,1\%$. По уравнению равносторонней гиперболы получена наибольшая оценка тесноты связи: $\rho_{xy} = 0,3944$ (по сравнению с линейной, степенной и показательной регрессиями). \bar{A} остается на допустимом уровне:

$$2. \quad F_{\text{факт}} = \frac{\rho_{yx}^2}{1 - \rho_{yx}^2} \cdot \frac{n - m - 1}{m} = \frac{0,1555}{0,8445} \cdot 5 = 0,92,$$

где $F_{\text{табл}} = 6,6 > F_{\text{факт}}$, $\alpha = 0,05$.

Следовательно, принимается гипотеза H_0 о статистически незначимых параметрах этого уравнения. Этот результат можно объяснить сравнительно невысокой теснотой выявленной зависимости и небольшим числом наблюдений.

Пример 2

По территориям региона приводятся данные за 199X г. (табл. 1.6).

Таблица 1.6

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
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

Требуется:

1. Построить линейное уравнение парной регрессии y от x .
2. Рассчитать линейный коэффициент парной корреляции и среднюю ошибку аппроксимации.
3. Оценить статистическую значимость параметров регрессии и корреляции.

4. Выполнить прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума x , составляющем 107% от среднего уровня.

5. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

Решение

1. Для расчета параметров уравнения линейной регрессии строим расчетную таблицу (табл. 1.7).

Таблица 1.7

	x	y	yx	x^2	y^2	\hat{y}_x	$y - \hat{y}_x$	A_i
1	78	133	10374	6084	17689	149	-16	12,0
2	82	148	12136	6724	21904	152	-4	2,7
3	87	134	11658	7569	17956	157	-23	17,2
4	79	154	12166	6241	23716	150	4	2,6
5	89	162	14418	7921	26244	159	3	1,9
6	106	195	20670	11236	38025	174	21	10,8
7	67	139	9313	4489	19321	139	0	0,0
8	88	158	13904	7744	24964	158	0	0,0
9	73	152	11096	5329	23104	144	8	5,3
10	87	162	14094	7569	26244	157	5	3,1
11	76	159	12084	5776	25281	147	12	7,5
12	115	173	19895	13225	29929	183	-10	5,8
Итого	1027	1869	161808	89907	294377	1869	0	68,8
Среднее значение	85,6	155,8	13484,0	7492,3	24531,4	\bar{x}	\bar{y}	5,7
σ	12,95	16,53	\bar{x}	\bar{x}	\bar{x}	\bar{x}	\bar{x}	\bar{x}
σ^2	167,7	273,4	\bar{x}	\bar{x}	\bar{x}	\bar{x}	\bar{x}	\bar{x}

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sum x^2 - (\bar{x})^2} = \frac{13484 - 85,6 \cdot 155,8}{7492,3 - 85,6^2} = \frac{151,8}{164,94} = 0,92;$$

$$a = \bar{y} - b \cdot \bar{x} = 155,8 - 0,92 \cdot 85,6 = 77,0.$$

Получено уравнение регрессии: $y = 77,0 + 0,92 \cdot x$.

С увеличением среднедушевого прожиточного минимума на 1 руб. среднедневная заработная плата возрастает в среднем на 0,92 руб.

2. Тесноту линейной связи оценит коэффициент корреляции:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = 0,92 \cdot \frac{12,95}{16,53} = 0,721; \quad r_{xy}^2 = 0,52.$$

Это означает, что 52% вариации заработной платы (y) объясняется вариацией фактора x – среднедушевого прожиточного минимума. Качество модели определяет средняя ошибка аппроксимации:

$$\bar{A} = \frac{1}{n} \sum A_i = \frac{68,9}{12} = 5,7\%.$$

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

3. Оценку статистической значимости параметров регрессии проведем с помощью t -статистики Стьюдента и путем расчета доверительного интервала каждого из показателей.

Выдвигаем гипотезу H_0 о статистически незначимом отличии показателей от нуля: $a = b = r_{xy} = 0$.

$t_{\text{табл}}$ для числа степеней свободы $df = n - 2 = 12 - 2 = 10$ и $\alpha = 0,05$ составит 2,23.

Определим случайные ошибки $m_a, m_b, m_{r_{xy}}$:

$$m_a = 12,6 \frac{\sqrt{89907}}{12 \cdot 12,95} = 24,3; \quad m_b = \frac{12,6}{12,95 \cdot \sqrt{12}} = 0,281;$$

$$m_{r_{xy}} = \sqrt{\frac{1 - 0,520}{12 - 2}} = 0,219.$$

Тогда

$$t_a = \frac{77}{24,3} = 3,2; \quad t_b = \frac{0,92}{0,281} = 3,3; \quad t_{r_{xy}} = \frac{0,721}{0,219} = 3,3.$$

Фактические значения t -статистики превосходят табличные значения:

$$t_a = 3,2 > t_{\text{табл}} = 2,3; \quad t_b = 3,3 > t_{\text{табл}} = 2,3; \quad t_{r_{xy}} = 3,3 > t_{\text{табл}} = 2,3,$$

поэтому гипотеза H_0 отклоняется, т.е. a, b и r_{xy} не случайно отличаются от нуля, а статистически значимы.

Рассчитаем доверительный интервал для a и b . Для этого определим предельную ошибку для каждого показателя:

$$\Delta_a = 2,23 \cdot 24,3 = 54; \quad \Delta_b = 2,23 \cdot 0,281 = 0,62.$$

Доверительные интервалы:

$$\gamma_a = a \pm \Delta_a = 77 \pm 54 ;$$

$$\gamma_{a_{\min}} = 77 - 54 = 23 ;$$

$$\gamma_{a_{\max}} = 77 + 54 = 131 ;$$

$$\gamma_b = b \pm \Delta_b = 0,92 \pm 0,62 ;$$

$$\gamma_{b_{\min}} = 0,92 - 0,62 = 0,30 ;$$

$$\gamma_{b_{\max}} = 0,92 + 0,62 = 1,54 .$$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью $p = 1 - \alpha = 0,95$ параметры a и b , находясь в указанных границах, не принимают нулевых значений, т.е. не являются статистически незначимыми и существенно отличны от нуля.

4. Полученные оценки уравнения регрессии позволяют использовать его для прогноза. Если прогнозное значение прожиточного минимума составит: $x_p = \bar{x} \cdot 1,07 = 85,6 \cdot 1,07 = 91,6$ тыс. руб., тогда прогнозное значение прожиточного минимума составит:

$$\hat{y}_p = 77 + 0,92 \cdot 91,6 = 161 \text{ тыс. руб.}$$

5. Ошибка прогноза составит:

$$m_{\hat{y}_p} = 12,6 \cdot \sqrt{1 + \frac{1}{12} + \frac{(91,6 - 85,6)^2}{12 \cdot 12,95^2}} = 13,2 \text{ тыс. руб.}$$

Предельная ошибка прогноза, которая в 95% случаев не будет превышена, составит:

$$\Delta_{\hat{y}_p} = t_{\text{табл}} \cdot m_{\hat{y}_p} = 2,23 \cdot 13,2 = 29,4.$$

Доверительный интервал прогноза:

$$\gamma_{\hat{y}_p} = 91,6 \pm 29,4;$$

$$\gamma_{\hat{y}_p_{\min}} = 91,6 - 29,4 = 62,2 \text{ руб.};$$

$$\gamma_{\hat{y}_p_{\max}} = 91,6 + 29,4 = 121 \text{ руб.}$$

Выполненный прогноз среднемесячной заработной платы оказался надежным ($p = 1 - \alpha = 1 - 0,05 = 0,95$), но неточным, так как диапазон верхней и нижней границ доверительного интервала D_γ составляет 1,95 раза:

$$D_\gamma = \frac{\gamma \hat{y}_{\max}}{\gamma \hat{y}_{\min}} = \frac{121}{62} = 1,95.$$

Пример 3

По группе предприятий, производящих однородную продукцию, известно, как зависит себестоимость единицы продукции y от факторов, приведенных в табл. 1.8.

Таблица 1:8

Признак-фактор	Уравнение парной регрессии	Среднее значение фактора
Объем производства, млн руб., x_1	$\hat{y}_{x_1} = 0,62 + 58,74 \cdot \frac{1}{x_1}$	$\bar{x}_1 = 2,64$
Трудоемкость единицы продукции, чел.-час., x_2	$\hat{y}_{x_2} = 9,30 + 9,83 \cdot x_2$	$\bar{x}_2 = 1,38$
Оптовая цена за 1 т энергоносителя, млн руб., x_3	$\hat{y}_{x_3} = 11,75 + x_3^{1,6281}$	$\bar{x}_3 = 1,503$
Доля прибыли, изымаемой государством, %, x_4	$\hat{y}_{x_4} = 14,87 \cdot 1,016^{x_4}$	$\bar{x}_4 = 26,3$

Требуется:

1. Определить с помощью коэффициентов эластичности силу влияния каждого фактора на результат.
2. Ранжировать факторы по силе влияния.

Решение

1. Для уравнения равнобочной гиперболы $\hat{y}_{x_1} = 0,62 + 58,74 \cdot \frac{1}{x_1}$:

$$\bar{\varepsilon}_{yx_1} = f'(x_1) \frac{\bar{x}_1}{\bar{y}} = -\frac{b}{\bar{x}_1^2} \cdot \frac{\bar{x}_1}{a + b/\bar{x}_1} = -\frac{b}{a \cdot \bar{x}_1 + b} = -\frac{58,74}{0,62 \cdot 2,64 + 58,74} = -0,973\%.$$

Для уравнения прямой $\hat{y}_{x_2} = 9,30 + 9,83 \cdot x_2$:

$$\bar{\mathcal{E}}_{yx_2} = f'(x_2) \frac{\bar{x}_2}{\bar{y}} = \frac{b \cdot \bar{x}_2}{a + b \cdot \bar{x}_2} = \frac{9,83 \cdot 1,38}{9,30 + 9,83 \cdot 1,38} = 0,59\% .$$

Для уравнения степенной зависимости $\hat{y}_{x_3} = 11,75 + x_3^{1,6281}$:

$$\bar{\mathcal{E}}_{yx_3} = f'(x_3) \frac{\bar{x}_3}{\bar{y}} = a \cdot b \cdot \bar{x}_3^{b-1} \cdot \frac{\bar{x}_3}{a \cdot \bar{x}_3^b} = b = 1,63\% .$$

Для уравнения показательной зависимости $\hat{y}_{x_4} = 14,87 + 1,016^{x_4}$:

$$\bar{\mathcal{E}}_{yx_4} = f'(x_4) \frac{\bar{x}_4}{\bar{y}} = a \cdot b^{\bar{x}_4} \cdot \ln b \cdot \frac{\bar{x}_4}{a \cdot b^{\bar{x}_4}} = \ln b \cdot \bar{x}_4 = \ln 1,016 \cdot 26,3 = 0,42\%$$

2. Сравнивая значения $\bar{\mathcal{E}}_{yx_i}$, ранжируем x_j по силе их влияния на себестоимость единицы продукции:

а) $\bar{\mathcal{E}}_{yx_3} = 1,63\%$; в) $\bar{\mathcal{E}}_{yx_2} = 0,59\%$;

б) $\bar{\mathcal{E}}_{yx_1} = -0,973\%$; г) $\bar{\mathcal{E}}_{yx_4} = 0,42\%$.

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 -0,97%.

Пример 4

Зависимость потребления продукта A от среднедушевого дохода по данным 20 семей характеризуется следующим образом:

уравнение регрессии $\hat{y}_x = 2 \cdot x^{0,3}$;

индекс корреляции $\rho_{xy} = 0,9$;

остаточная дисперсия $\sigma_{\text{ост}}^2 = 0,06$.

Требуется:

Провести дисперсионный анализ полученных результатов.

Решение

Результаты дисперсионного анализа приведены в табл. 1.9.

Таблица 1.9

Вариация результата y	Число степеней свободы	Сумма квадратов отклонений, S	Дисперсия на одну степень свободы, D	$F_{\text{факт}}$	$F_{\text{табл}}$ $\alpha = 0,05,$ $k_1 = 1; k_2 = 18$
Общая	$df = n - 1 = 19$	6,316	—	—	—
Факторная	$k_1 = m = 1$	5,116	5,116	76,7	4,41
Остаточная	$k_2 = n - m - 1 = 18$	1,200	0,0667	—	—

$$S_{\text{ост}} = \sigma_{\text{ост}}^2 \cdot n = 0,06 \cdot 20 = 1,2;$$

$$S_{\text{общ}} = S_{\text{ост}} : (1 - \rho_{xy}^2) = 1,2 : (1 - 0,81) = 6,316;$$

$$S_{\text{факт}} = 6,316 - 1,2 = 5,116;$$

$$F_{\text{факт}} = \frac{0,9^2}{1 - 0,9^2} \cdot \frac{18}{1} = 76,7.$$

В силу того что $F_{\text{факт}} = 76,7 > F_{\text{табл}} = 4,4$, гипотеза о случайности различий факторной и остаточной дисперсий отклоняется. Эти различия существенны, статистически значимы, уравнение надежно, значимо, показатель тесноты связи надежен и отражает устойчивую зависимость потребления продукта A от среднедушевого дохода.

1.3. РЕАЛИЗАЦИЯ ТИПОВЫХ ЗАДАЧ НА КОМПЬЮТЕРЕ

Решение с помощью ППП Excel

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

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

- 2) выделите область пустых ячеек 5×2 (5 строк, 2 столбца) для вывода результатов регрессионной статистики или область 1×2 – для 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**;

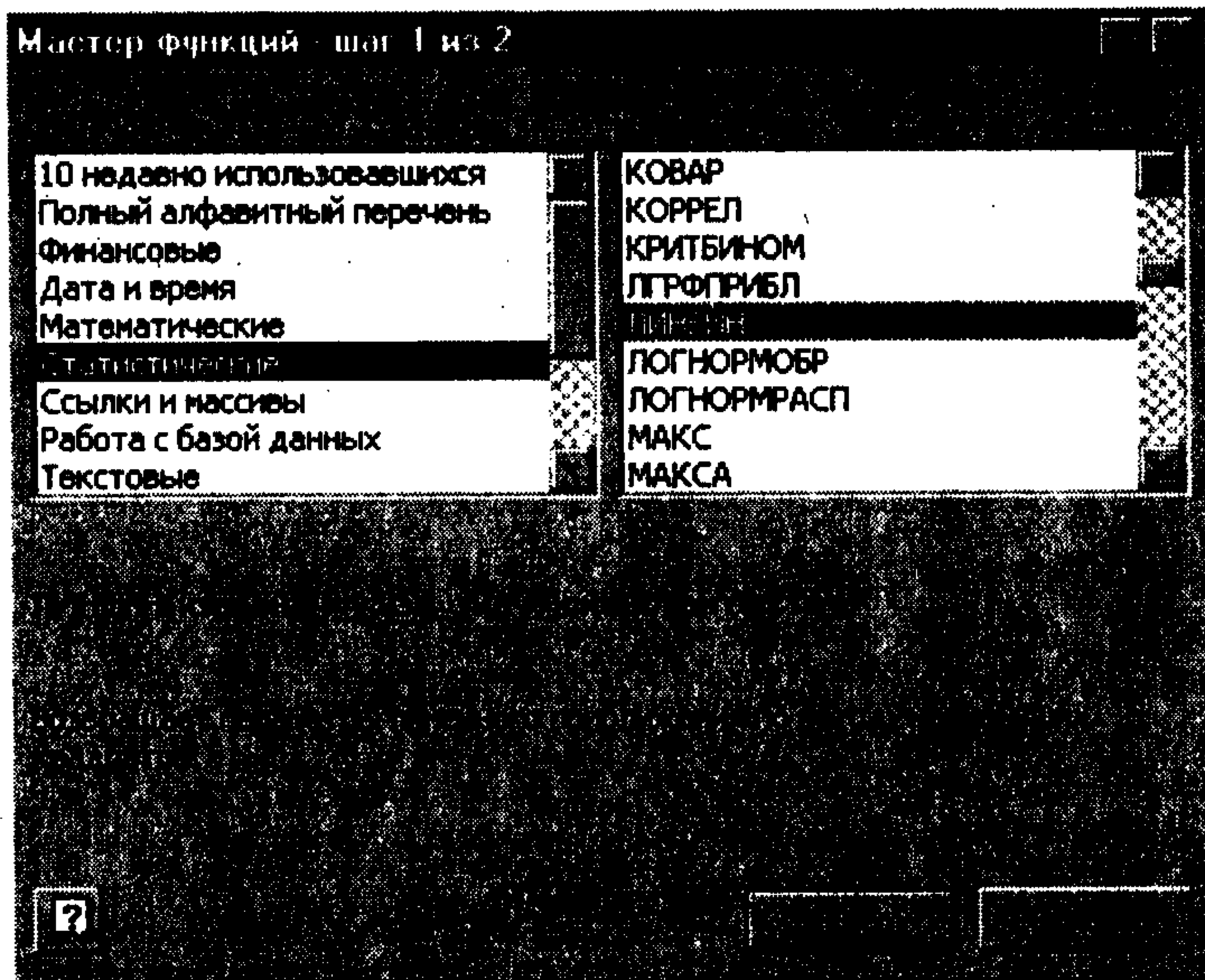


Рис. 1.1. Диалоговое окно «Мастер функций»

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

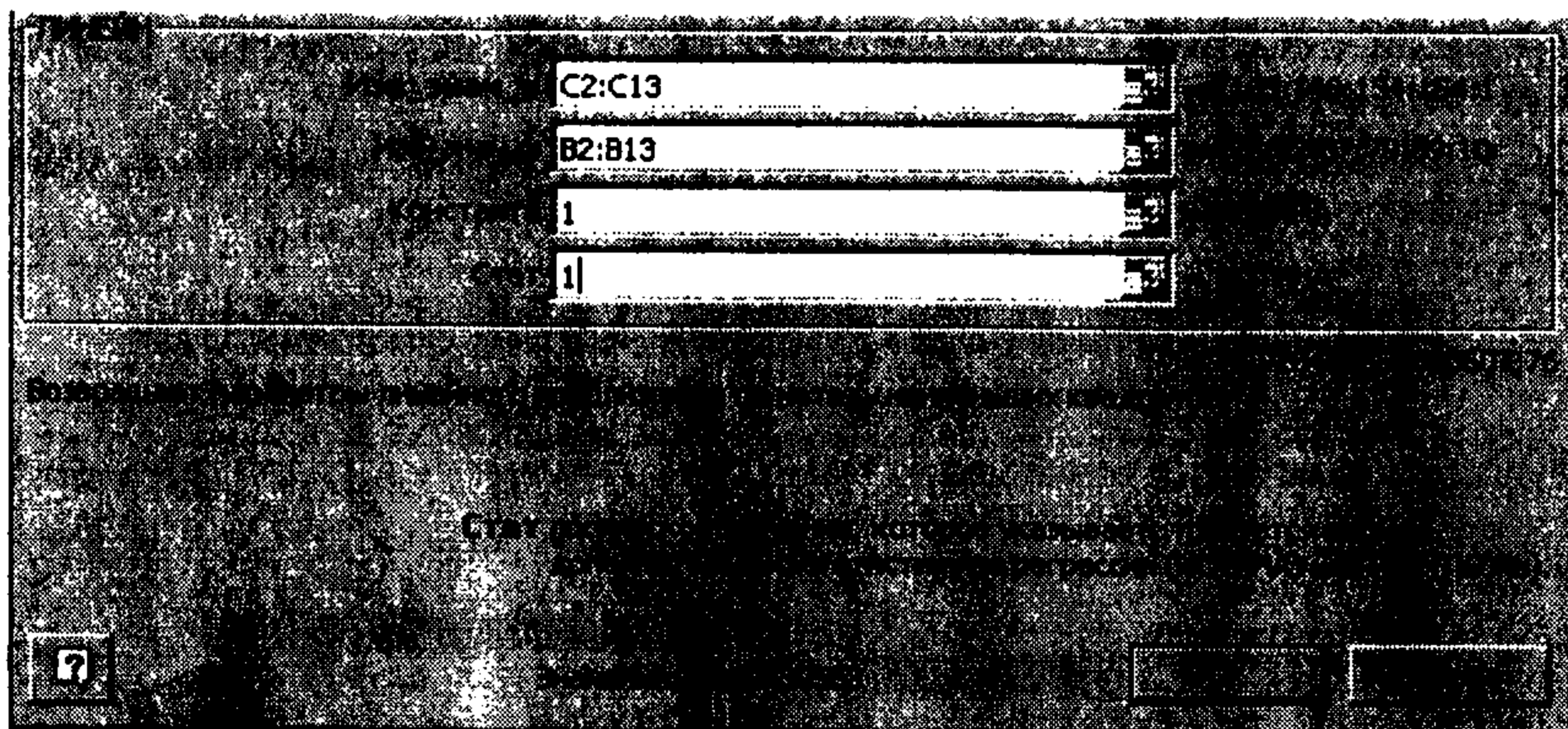


Рис. 1.2. Диалоговое окно ввода аргументов функции **ЛИНЕЙН**

б) 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 <i>b</i>	Значение коэффициента <i>a</i>
Среднеквадратическое отклонение <i>b</i>	Среднеквадратическое отклонение <i>a</i>
Коэффициент детерминации R^2	Среднеквадратическое отклонение <i>y</i>
<i>F</i> -статистика	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов

Для вычисления параметров экспоненциальной кривой $y = \alpha \cdot \beta^x$ в MS Excel применяется встроенная статистическая функция **ЛГРФПРИБЛ**. Порядок вычисления аналогичен применению функции **ЛИНЕЙН**.

Для данных из примера 2 результат вычисления функции **ЛИНЕЙН** представлен на рис. 1.3, функции **ЛГРФПРИБЛ** – на рис. 1.4.

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**

Территории региона	Прожиточный минимум - x	Среднемесячная зарплата - y	Лгрфприбл	
1	78	133	1,005664	95,53277
2	82	148	0,001791	0,154997
3	87	134	0,498671	0,08034
4	79	154	9,946979	10
5	89	162	0,064202	0,064544
6	106	195		
7	67	139		
8	88	158		
9	73	152		
10	87	162		
11	76	159		
12	115	173		

Рис. 1.4. Результат вычисления функции **ЛГРФПРИБЛ**

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

1) проверьте доступ к пакету анализа. В главном меню последовательно выберите **Сервис /Надстройки**. Установите флажок **Пакет анализа** (рис. 1.5);

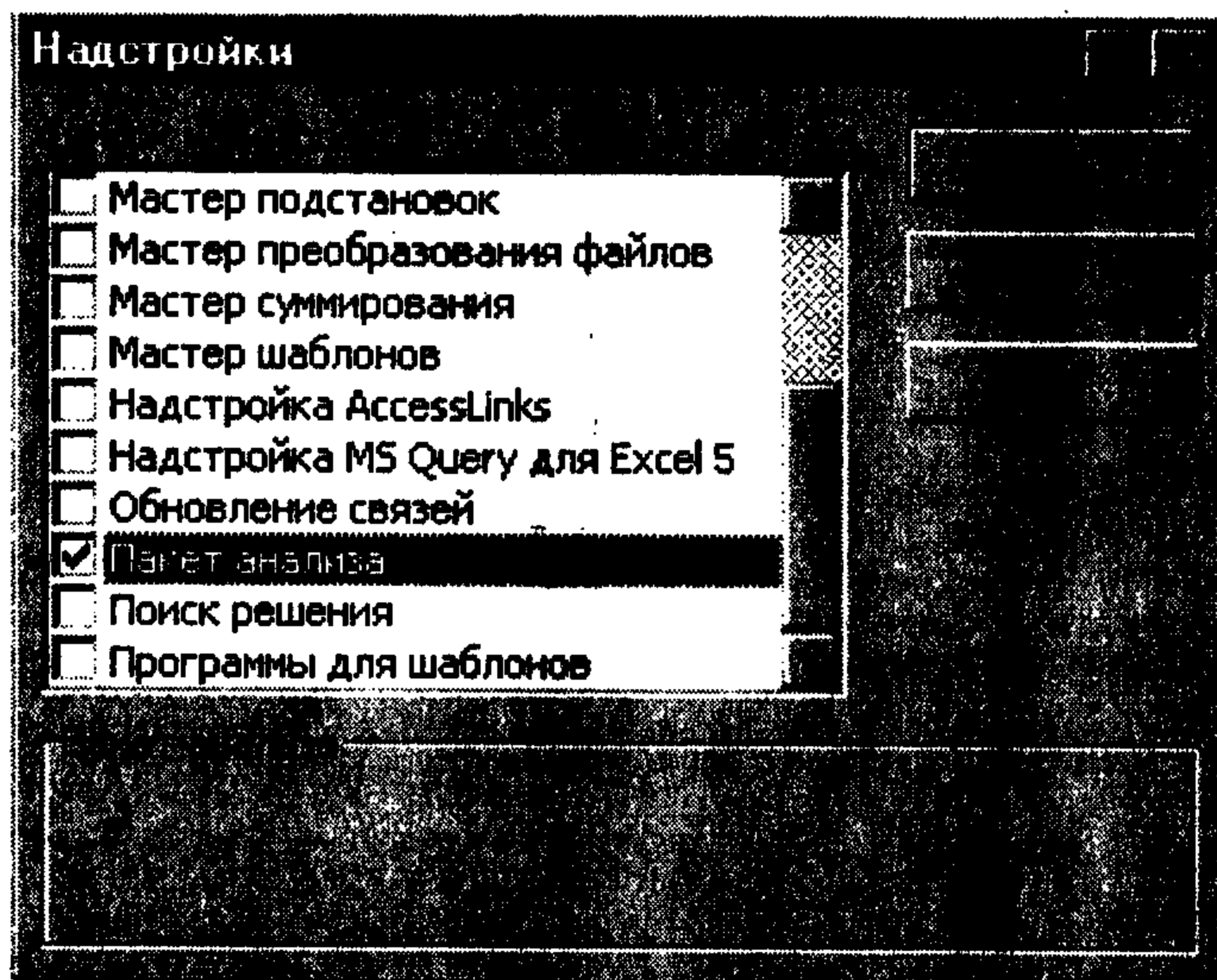


Рис. 1.5. Подключение надстройки **Пакет анализа**

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

3) заполните диалоговое окно ввода данных и параметров вывода (рис. 1.6):

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**.

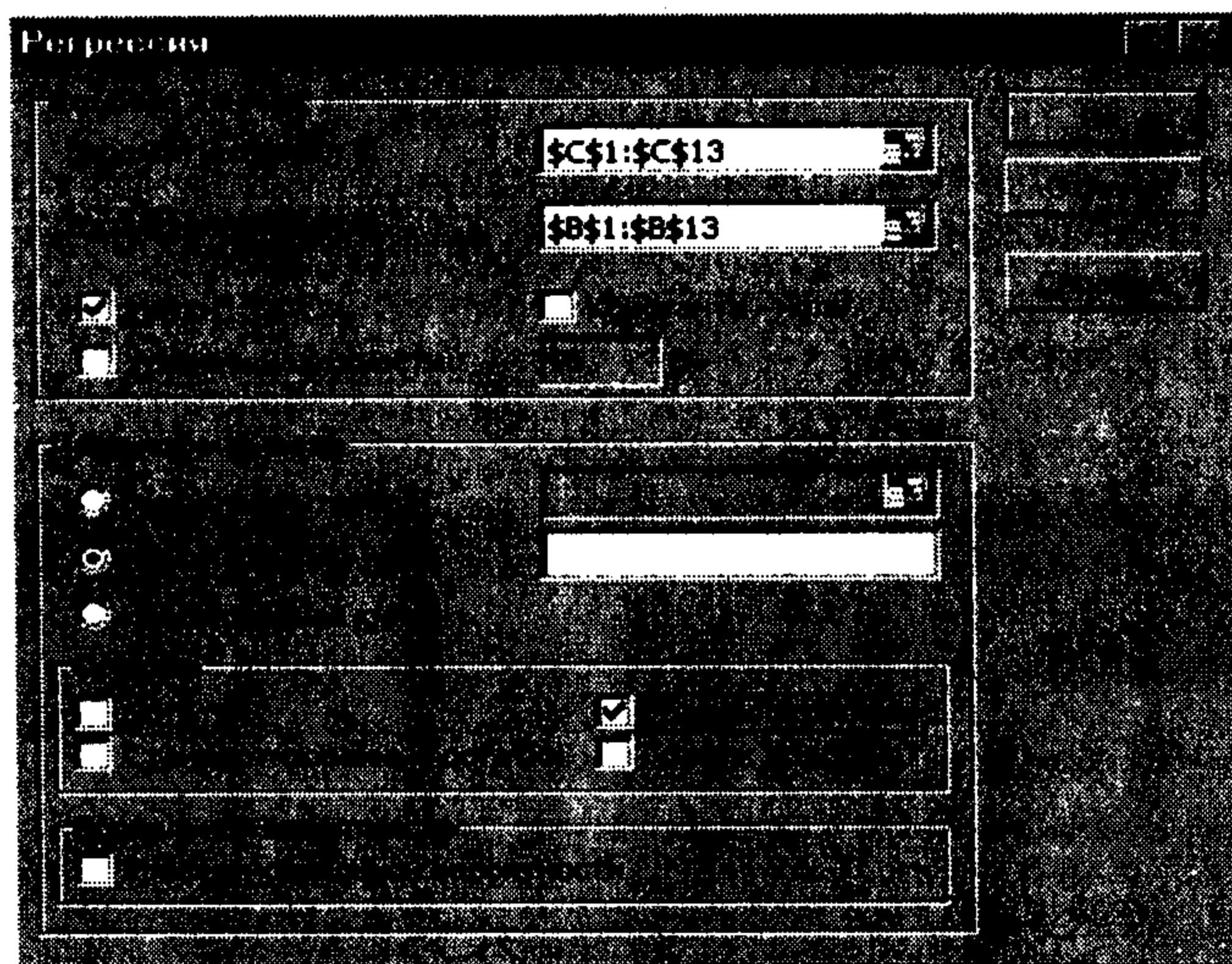


Рис. 1.6. Диалоговое окно ввода параметров инструмента Регрессия

Результаты регрессионного анализа для данных из примера 2 представлены на рис. 1.7.

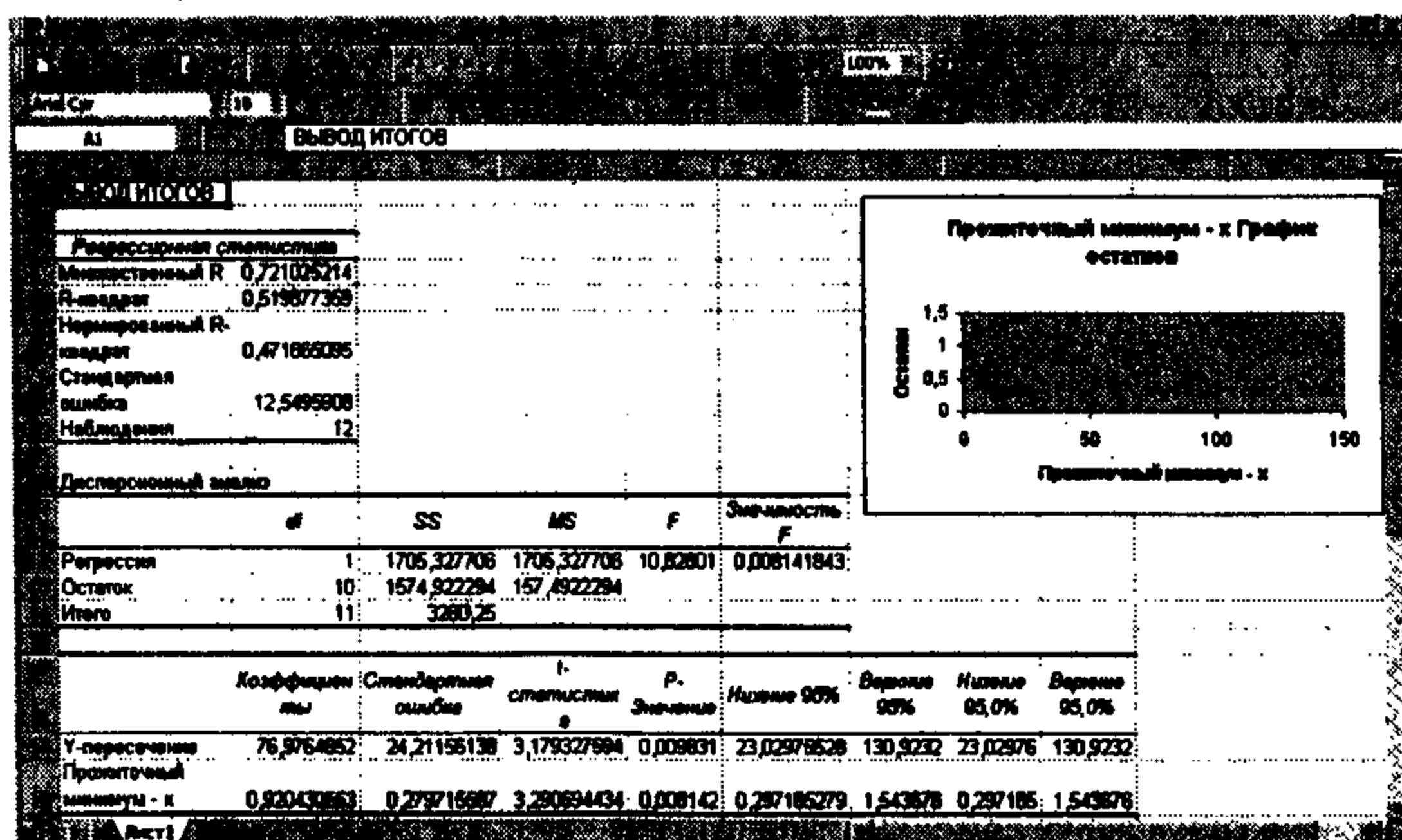


Рис. 1.7. Результат применения инструмента Регрессия

Решение с помощью ППП Statgraphics

Порядок вычислений при использовании функции **Simple Regression** следующий:

- 1) введите исходные данные (рис. 1.8) или откройте существующий файл, содержащий исходные данные;
- 2) в главном меню последовательно выберите **Relate/Simple Regression**;
- 3) заполните диалоговое окно ввода данных. В поле «Y» введите название столбца, содержащего зависимую переменную, в поле «X» – название столбца, содержащего значения факторного признака. Щелкните по кнопке **ОК**;

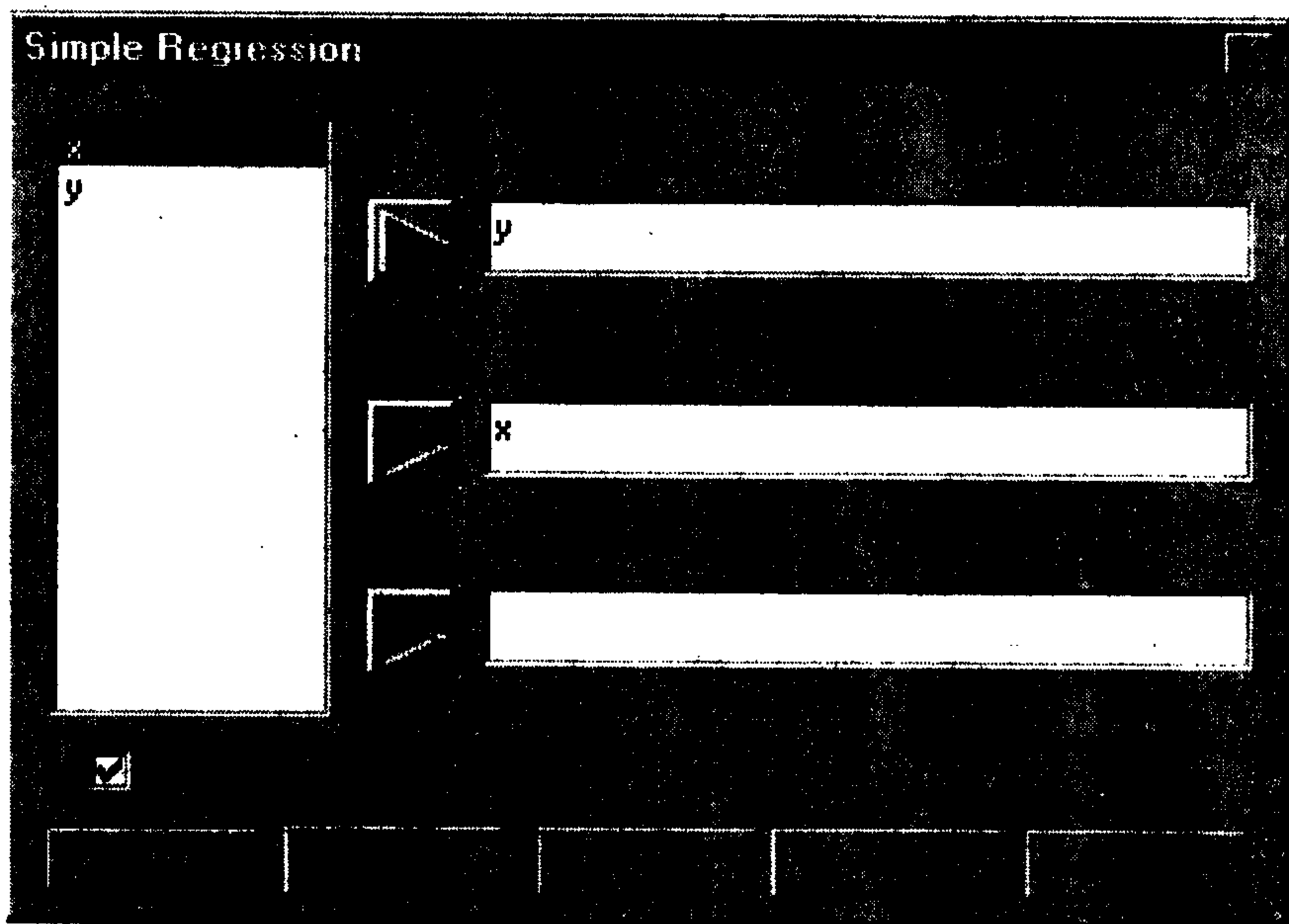


Рис. 1.8. Диалоговое окно ввода данных

- 4) в окне табличных настроек поставьте флажок напротив *Analysis Summary*,

Результаты вычислений появятся в отдельном окне.

Для данных из примера 2 результат применения функции **Simple Regression** представлен на рис. 1.9.

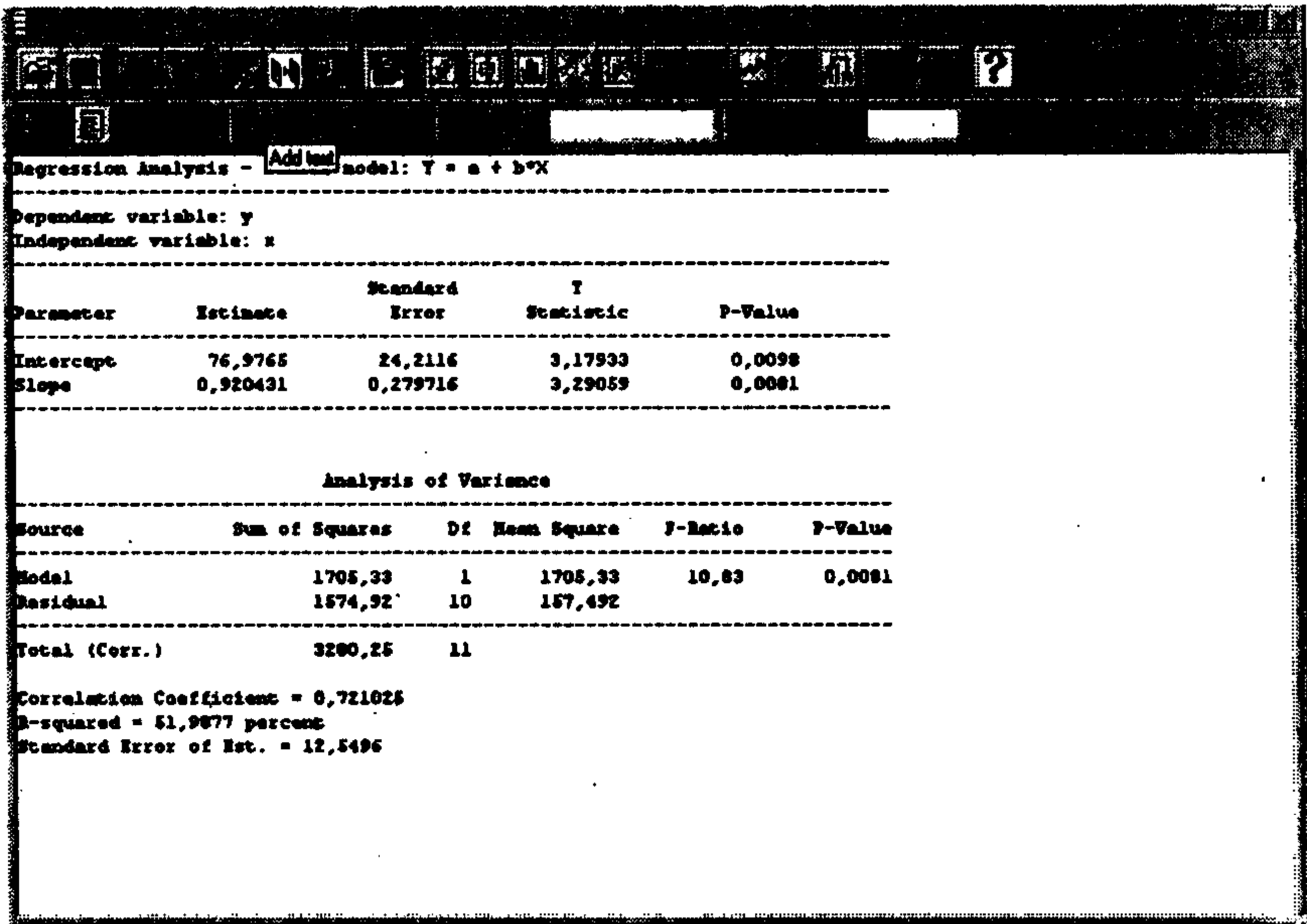


Рис. 1.9. Итоговое окно функции Simple Regression

Как видим, результаты вычислений вручную и с помощью компьютера совпадают.

1.4. КОНТРОЛЬНЫЕ ЗАДАНИЯ

Задача 1

Получены функции:

1. $y = a + bx^3 + \varepsilon,$
2. $y = a + b \ln x + \varepsilon,$
3. $\ln y = a + b \ln x + \varepsilon,$
4. $y = a + bx^c + \varepsilon,$
5. $y^a = b + cx^2 + \varepsilon,$
6. $y = 1 + a(1 - x^b) + \varepsilon,$
7. $y = a + b \frac{x}{10} + \varepsilon,$

Задание

Определите, какие из представленных выше функций линейны по переменным, линейны по параметрам, нелинейны ни по переменным, ни по параметрам.

Задача 2

Исследуя спрос на телевизоры марки N , аналитический отдел компании ABC по данным, собранным по 19 торговым точкам компании, выявил следующую зависимость:

$$\ln y = 10,5 - 0,8 \ln x + \varepsilon,$$

(2,5) (-4,0)

где y – объем продаж телевизоров марки N в отдельной торговой точке;
 x – средняя цена телевизора в данной торговой точке;
в скобках приведены фактические значения t -критерия Стьюдента для параметров уравнения регрессии.

Задание

До проведения этого исследования администрация компании предполагала, что эластичность спроса по цене для телевизоров марки N составляет $-0,9$. Подтвердилось ли предположение администрации результатами исследования?

Задача 3

Для трех видов продукции A , B и C модели зависимости удельных постоянных расходов от объема выпускаемой продукции выглядят следующим образом:

$$\begin{aligned} y_A &= 600, \\ y_B &= 80 + 0,7x, \\ y_C &= 40x^{0,5}. \end{aligned}$$

Задание

1. Определите коэффициенты эластичности по каждому виду продукции и поясните их смысл.
2. Сравните при $x = 1000$ эластичность затрат для продукции B и C .
3. Определите, каким должен быть объем выпускаемой продукции, чтобы коэффициенты эластичности для продукции B и C были равны.

Задача 4

Пусть имеется следующая модель регрессии, характеризующая зависимость y от x :

$$y = 8 - 7x + \varepsilon.$$

Известно также, что $r_{xy} = -0,5$; $n = 20$.

Задание

1. Постройте доверительный интервал для коэффициента регрессии в этой модели:

а) с вероятностью 90%;

б) с вероятностью 99%.

2. Проанализируйте результаты, полученные в п.1, и поясните причины их различий.

Задача 5

Изучается зависимость потребления материалов y от объема производства продукции x . По 20 наблюдениям были получены следующие варианты уравнения регрессии:

$$1. y = 3 + \frac{2}{(6,48)} x + \varepsilon.$$

$$2. \ln y = 2,5 + \frac{0,2}{(6,19)} \cdot \ln x + \varepsilon, \quad r^2 = 0,68.$$

$$3. \ln Y = 1,1 + \frac{0,8}{(6,2)} \cdot \ln X + \varepsilon, \quad r^2 = 0,69.$$

$$4. Y = 3 + \frac{1,5}{(3,0)} \cdot X + \frac{0,1}{(2,65)} \cdot X^2, \quad r^2 = 0,701.$$

В скобках указаны фактические значения t -критерия.

Задание

1. Определите коэффициент детерминации для 1-го уравнения.

2. Запишите функции, характеризующие зависимость y от x во 2-м и 3-м уравнениях.

3. Определите коэффициенты эластичности для каждого из уравнений.

4. Выберите наилучший вариант уравнения регрессии.

Задача 6

По совокупности 30 предприятий торговли изучается зависимость между признаками: x – цена на товар A , тыс. руб.; y – прибыль торгового предприятия, млн руб.

При оценке регрессионной модели были получены следующие промежуточные результаты:

$$\sum (y_j - \hat{y}_x)^2 = 39\,000;$$

$$\sum (y_j - \bar{y})^2 = 120\,000.$$

Задание

1. Поясните, какой показатель корреляции можно определить по этим данным.
2. Постройте таблицу дисперсионного анализа для расчета значения F -критерия Фишера.
3. Сравните фактическое значение F -критерия с табличным. Сделайте выводы.

Задача 7

Зависимость среднемесячной производительности труда от возраста рабочих характеризуется моделью: $y = a + bx + cx^2$. Ее использование привело к результатам, представленным в табл. 1.10.

Таблица 1.10

№ п/п	Производительность труда рабочих, тыс. руб., y		№ п/п	Производительность труда рабочих, тыс. руб., y	
	фактическая	расчетная		фактическая	расчетная
1	12	10	6	11	12
2	8	10	7	12	13
3	13	13	8	9	10
4	15	14	9	11	10
5	16	15	10	9	9

Задание

Оцените качество модели, определив ошибку аппроксимации, индекс корреляции и F -критерий Фишера.

Задача 8

Моделирование прибыли фирмы по уравнению $y = ab^x$ привело к результатам, представленным в табл. 1.11.

Таблица 1.11

№ п/п	Прибыль фирмы, тыс. руб., y		№ п/п	Прибыль фирмы, тыс. руб., y	
	фактическая	расчетная		фактическая	расчетная
1	10	11	5	18	20
2	12	11	6	11	11
3	15	17	7	13	14
4	17	15	8	19	16

Задание

Оцените качество модели. Для этого:

- определите ошибку аппроксимации;
- найдите показатель тесноты связи прибыли с исследуемым в модели фактором;
- рассчитайте F -критерий Фишера. Сделайте выводы.

Задача 9

Изучалась зависимость вида $y = ax^b$. Для преобразованных в логарифмах переменных получены следующие данные:

$$\sum xy = 4,2087; \quad \sum x = 8,2370;$$

$$\sum x^2 = 9,2334; \quad \sum y = 3,9310;$$

$$\sum (Y - \hat{Y}_x)^2 = 0,0014.$$

Задание

- Найдите параметр b .
- Найдите показатель корреляции, предполагая $\sigma_y = 0,08$. Оцените его значимость.
- Оцените его значимость, если известно, что $n = 9$.

Задача 10

Зависимость объема производства y (тыс. ед.) от численности занятых x (чел.) по 15 заводам концерна характеризуется следующим образом:

Уравнение регрессии

$$y = 30 - 0,4x + 0,04x^2$$

Доля остаточной дисперсии в общей 20%

Задание

Определите:

- индекс корреляции;
- значимость уравнения регрессии;
- коэффициент эластичности, предполагая, что численность занятых составляет 30 человек.

Задача 11

По группе 10 заводов, производящих однородную продукцию, получено уравнение регрессии себестоимости единицы продукции y (тыс. руб.) от уровня технической оснащенности x (тыс. руб.):

$$y = 20 + \frac{700}{x}. \text{ Доля остаточной дисперсии в общей составила } 0,19.$$

Задание

Определите:

- коэффициент эластичности, предполагая, что стоимость активных производственных фондов составляет 200 тыс. руб.;
- индекс корреляции;
- F -критерий Фишера. Сделайте выводы.

Задача 12

Зависимость спроса на товар K от его цены характеризуется по 20 наблюдениям уравнением: $\lg y = 1,75 - 0,35 \lg x$. Доля остаточной дисперсии в общей составила 18%.

Задание

- Запишите данное уравнение в виде степенной функции.
- Оцените эластичность спроса на товар в зависимости от его цены.
- Определите индекс корреляции.
- Оцените значимость уравнения регрессии через F -критерий Фишера. Сделайте выводы.

Задача 13

По 20 фермам области получена информация, представленная в табл. 1.12.

Таблица 1.12

Показатель	Среднее значение	Коэффициент вариации
Урожайность, ц/га	27	20
Внесено удобрений на 1 га посева, кг	5	15

Фактическое значение F -критерия Фишера составило 45.

Задание

- Определите линейный коэффициент детерминации.
- Постройте уравнение линейной регрессии.

3. Найдите обобщающий коэффициент эластичности.
4. С вероятностью 0,95 укажите доверительный интервал ожидаемого значения урожайности в предположении роста количества внесенных удобрений на 10% от своего среднего уровня.

Задача 14

Для двух видов продукции А и Б зависимость расходов предприятия y (тыс. руб.) от объема производства x (шт.) характеризуется данными, представленными в табл. 1.13.

Таблица 1.13

Уравнение регрессии	Показатели корреляции	Число наблюдений
$y_A = 160 + 0,8x$	0,85	30
$y_B = 50x^{0,6}$	0,72	25

Задание

1. Поясните смысл величин 0,8 и 0,6 в уравнениях регрессии.
2. Сравните эластичность расходов от объема производства для продукции А и Б при выпуске продукции А в 500 единиц.
1. Определите, каким должен быть выпуск продукции А, чтобы эластичность ее расходов совпадала с эластичностью расходов на продукцию Б.
4. Оцените значимость каждого уравнения регрессии с помощью F -критерия Фишера.

Задача 15

Зависимость объема продаж y (тыс. долл.) от расходов на рекламу x (тыс. долл.) характеризуется по 12 предприятиям концерна следующим образом:

Уравнение регрессии	$y = 10,6 + 0,6x$
Среднее квадратическое отклонение x	$\sigma_x = 4,7$
Среднее квадратическое отклонение y	$\sigma_y = 3,4$

Задание

1. Определите коэффициент корреляции.
2. Постройте таблицу дисперсионного анализа для оценки значимости уравнения регрессии в целом.

3. Найдите стандартную ошибку оценки коэффициента регрессии.
4. Оцените значимость коэффициента регрессии через t -критерий Стьюдента.
5. Определите доверительный интервал для коэффициента регрессии с вероятностью 0,95 и сделайте экономический вывод.

Задача 16

По 20 регионам страны изучается зависимость уровня безработицы y (%) от индекса потребительских цен x (% к предыдущему году). Информация о логарифмах исходных показателей представлена в табл. 1.14.

Таблица 1.14

Показатель	$\ln x$	$\ln y$
Среднее значение	0,6	1,0
Среднее квадратическое отклонение	0,4	0,2

Известно также, что коэффициент корреляции между логарифмами исходных показателей составил $r_{\ln x \ln y} = 0,8$.

Задание

1. Постройте уравнение регрессии зависимости уровня безработицы от индекса потребительских цен в степенной форме.
2. Дайте интерпретацию коэффициента эластичности данной модели регрессии.
3. Определите значение коэффициента детерминации и поясните его смысл.

Задача 17

Изучается зависимость материалоемкости продукции от размера предприятия по 10 однородным заводам (табл. 1.15).

Таблица 1.15

Показатель	Материалоемкость продукции по заводам									
	1	2	3	4	5	6	7	8	9	10
Потреблено материалов на единицу продукции, кг	9	6	5	4	3,7	3,6	3,5	6	7	3,5
Выпуск продукции, тыс. ед.	100	200	300	400	500	600	700	150	120	250

Задание

1. Найдите параметры уравнения $y = a + \frac{b}{x}$.
2. Оцените тесноту связи с помощью индекса корреляции.
3. Охарактеризуйте эластичность изменения материалоемкости продукции.
4. Сделайте вывод о значимости уравнения регрессии.

Задача 18

По территориям Центрального района известны данные за 1995 г. (табл. 1.16).

Таблица 1.16

Район	Доля денежных доходов, направленных на прирост сбережений во вкладах, займах, сертификатах и на покупку валюты, в общей сумме среднедушевого денежного дохода, %, y	Среднемесячная численная заработная плата, тыс. руб., x
Брянская обл.	6,9	289
Владимирская обл.	8,7	334
Ивановская обл.	6,4	300
Калужская обл.	8,4	343
Костромская обл.	6,1	356
Орловская обл.	9,4	289
Рязанская обл.	11,0	341
Смоленская обл.	6,4	327
Тверская обл.	9,3	357
Тульская обл.	8,2	352
Ярославская обл.	8,6	381

Задание

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.
2. Рассчитайте параметры уравнений линейной, степенной, экспоненциальной, полулогарифмической, обратной, гиперболической парной регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. Дайте с помощью среднего (общего) коэффициента эластичности сравнительную оценку силы связи фактора с результатом.

5. Оцените с помощью средней ошибки аппроксимации качество уравнений.

6. Оцените с помощью F -критерия Фишера статистическую надежность результатов регрессионного моделирования. По значениям характеристик, рассчитанных в пп. 4, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.

7. Рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 10% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.

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

Задача 19

По территориям Центрального района известны данные за 1995 г. (табл. 1.17).

Таблица 1.17

Район	Средний размер назначенных ежемесячных пенсий, тыс. руб., y	Прожиточный минимум в среднем на одного пенсионера в месяц, тыс. руб., x
Брянская обл.	240	178
Владимирская обл.	226	202
Ивановская обл.	221	197
Калужская обл.	226	201
Костромская обл.	220	189
г. Москва	250	302
Московская обл.	237	215
Орловская обл.	232	166
Рязанская обл.	215	199
Смоленская обл.	220	180
Тверская обл.	222	181
Тульская обл.	231	186
Ярославская обл.	229	250

Задание

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

2. Рассчитайте параметры уравнений линейной, степенной, экспоненциальной, полулогарифмической, обратной, гиперболической парной регрессии.

3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. Дайте с помощью среднего (общего) коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оцените с помощью средней ошибки аппроксимации качество уравнений.
6. С помощью *F*-критерия Фишера оцените статистическую надежность результатов регрессионного моделирования. По значениям характеристик, рассчитанных в пп. 4, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.
7. Рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 10% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.
8. Оцените полученные результаты, выводы оформите в аналитической записке.

Задача 20

По территориям Центрального и Волго-Вятского районов известны данные за ноябрь 1997 г. (табл. 1.18).

Таблица 1.18

Район	Средняя заработная плата и выплаты социального характера, тыс. руб., <i>y</i>	Прожиточный минимум в среднем на душу населения, тыс. руб., <i>x</i>
Центральный		
Брянская обл.	615	289
Владимирская обл.	727	338
Ивановская обл.	584	287
Калужская обл.	753	324
Костромская обл.	707	307
Орловская обл.	657	304
Рязанская обл.	654	307
Смоленская обл.	693	290
Тверская обл.	704	314
Тульская обл.	780	304
Ярославская обл.	830	341
Волго-Вятский		
Респ. Марий Эл	554	364
Респ. Мордовия	560	342
Чувашская Респ.	545	310
Кировская обл.	672	411
Нижегородская обл.	796	304

Задание

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.
2. Рассчитайте параметры уравнений линейной, степенной, экспоненциальной, полулогарифмической, обратной, гиперболической парной регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. Дайте с помощью среднего (общего) коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оцените с помощью средней ошибки аппроксимации качество уравнений.
6. Оцените с помощью F -критерия Фишера статистическую надежность результатов регрессионного моделирования. По значениям характеристик, рассчитанных в пп. 4, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.
7. Рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 5% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.
8. Оцените полученные результаты, выводы оформите в аналитической записке.

Задача 21

По территориям Волго-Вятского, Центрально-Черноземного и Поволжского районов известны данные за ноябрь 1997 г. (табл. 1.19).

Таблица 1.19

Район	Потребительские расходы в расчете на душу населения, тыс. руб., y	Средняя заработная плата и выплаты социального характера, тыс. руб., x
Волго-Вятский		
Респ. Марий Эл	302	554
Респ. Мордовия	360	560
Чувашская Респ.	310	545
Кировская обл.	415	672
Нижегородская обл.	452	796
Центрально-Черноземный		
Белгородская обл.	502	777
Воронежская обл.	355	632

Район	Потребительские расходы в расчете на душу населения, тыс. руб., у	Средняя заработная плата и выплаты социального характера, тыс. руб., х
Курская обл.	416	688
Липецкая обл.	501	833
Тамбовская обл.	403	577
Поволжский		
Респ. Калмыкия	208	584
Респ. Татарстан	462	949
Астраханская обл.	368	888
Волгоградская обл.	399	831
Пензенская обл.	342	562
Саратовская обл.	354	665
Ульяновская обл.	558	705

Задание

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.
2. Рассчитайте параметры уравнений линейной, степенной, экспоненциальной, полулогарифмической, обратной, гиперболической парной регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. Дайте с помощью среднего (общего) коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оцените с помощью средней ошибки аппроксимации качество уравнений.
6. Оцените с помощью F -критерия Фишера статистическую надежность результатов регрессионного моделирования. По значениям характеристик, рассчитанных в пп. 4, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.
7. Рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 7% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.
8. Оцените полученные результаты, выводы оформите в аналитической записке.

Задача 22

По территориям Северного, Северо-Западного и Центрального районов известны данные за ноябрь 1997 г. (табл. 1.20).

Таблица 1.20

Район	Потребительские расходы на душу населения, тыс. руб., у	Денежные доходы на душу населения, тыс. руб., х
Северный		
Респ. Карелия	596	913
Респ. Коми	417	1095
Архангельская обл.	354	606
Вологодская обл.	526	876
Мурманская обл.	934	1314
Северо-Западный		
Ленинградская обл.	412	593
Новгородская обл.	525	754
Псковская обл.	367	528
Центральный		
Брянская обл.	364	520
Владимирская обл.	336	539
Ивановская обл.	409	540
Калужская обл.	452	682
Костромская обл.	367	537
Московская обл.	328	589
Орловская обл.	460	626
Рязанская обл.	380	521
Смоленская обл.	439	626
Тверская обл.	344	521
Тульская обл.	401	658
Ярославская обл.	514	746

Задание

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.
2. Рассчитайте параметры уравнений линейной, степенной, экспоненциальной, полулогарифмической, обратной, гиперболической парной регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. Рассчитайте средний (общий) коэффициент эластичности.
5. Оцените качество уравнений с помощью средней ошибки аппроксимации.

6. Оцените статистическую надежность результатов регрессионного моделирования с помощью F -критерия Фишера. По значениям характеристик, рассчитанных в пп. 4, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.

7. Рассчитайте ожидаемое значение результата, если прогнозное значение фактора увеличится на 4% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.

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

Задача 23

По территориям Восточно-Сибирского и Дальневосточного районов известны данные за ноябрь 1997 г. (табл. 1.21).

Таблица 1.21

Район	Потребительские расходы на душу населения, тыс. руб., y	Денежные доходы на душу населения, тыс. руб., x
Восточно-Сибирский		
Респ. Бурятия	408	524
Респ. Тыва	249	371
Респ. Хакасия	253	453
Красноярский край	580	1006
Иркутская обл.	651	997
Усть-Ордынский Бурятский авт. округ	139	217
Читинская обл.	322	486
Дальневосточный		
Респ. Саха (Якутия)	899	1989
Еврейская авт. обл.	330	595
Чукотский авт. округ	446	1550
Приморский край	642	937
Хабаровский край	542	761
Амурская обл.	504	767
Камчатская обл.	861	1720
Магаданская обл.	707	1735
Сахалинская обл.	557	1052

Задание

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.
2. Рассчитайте параметры уравнений линейной, степенной, экспоненциальной, полулогарифмической, обратной, гиперболической парной регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. Дайте с помощью среднего (общего) коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оцените качество уравнений с помощью средней ошибки аппроксимации.
6. Оцените статистическую надежность результатов регрессионного моделирования с помощью *F*-критерия Фишера. По значениям характеристик, рассчитанных в пп. 4, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.
7. Рассчитайте ожидаемое значение результата, если значение фактора увеличится на 5% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.
8. Оцените полученные результаты, выводы оформите в аналитической записке.

Задача 24

По территориям Уральского и Западно-Сибирского районов известны данные за ноябрь 1997 г. (табл. 1.22).

Таблица 1.22

Район	Потребительские расходы на душу населения, тыс. руб., <i>y</i>	Денежные доходы на душу населения, тыс. руб., <i>x</i>
Уральский		
Респ. Башкортостан	461	632
Удмуртская Респ.	524	738
Курганская обл.	298	515
Оренбургская обл.	351	640
Пермская обл.	624	942
Свердловская обл.	584	888
Челябинская обл.	425	704

Район	Потребительские расходы на душу населения, тыс. руб., у	Денежные доходы на душу населения, тыс. руб., х
Западно-Сибирский		
Респ. Алтай	277	603
Алтайский край	321	439
Кемеровская обл.	573	985
Новосибирская обл.	576	735
Омская обл.	588	760
Томская обл.	497	830
Тюменская обл.	863	2093

Задание

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.
2. Рассчитайте параметры уравнений линейной, степенной, экспоненциальной, полулогарифмической, обратной, гиперболической парной регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. Дайте с помощью среднего (общего) коэффициента эластичности сравнительную оценку силы связи фактора с результатом.
5. Оцените качество уравнений с помощью средней ошибки аппроксимации.
6. Оцените статистическую надежность результатов регрессионного моделирования с помощью F -критерия Фишера. По значениям характеристик, рассчитанных в пп. 4, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.
7. Рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 8% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.
8. Оцените полученные результаты, выводы оформите в аналитической записке.

Задача 25

По территориям Уральского и Западно-Сибирского районов известны данные за ноябрь 1997 г. (табл. 1.23).

Таблица 1.23

Район	Потребительские расходы на душу населения, тыс. руб., y	Средняя заработная плата и выплаты социального характера, тыс. руб., x
Уральский		
Респ. Башкортостан	461	912
Удмуртская Респ.	524	809
Курганская обл.	298	748
Оренбургская обл.	351	847
Пермская обл.	624	1087
Свердловская обл.	584	1074
Челябинская обл.	425	1008
Западно-Сибирский		
Респ. Алтай	277	682
Алтайский край	321	697
Кемеровская обл.	573	1251
Новосибирская обл.	576	967
Омская обл.	588	898
Томская обл.	497	1263
Тюменская обл.	863	3027

Задание

1. Постройте поле корреляции и сформулируйте гипотезу о форме связи.
2. Рассчитайте параметры уравнений линейной, степенной, экспоненциальной, полулогарифмической, обратной, гиперболической парной регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. Дайте сравнительную оценку силы связи фактора с результатом с помощью среднего (общего) коэффициента эластичности.
5. Оцените качество уравнений с помощью средней ошибки аппроксимации.

6. Оцените статистическую надежность результатов регрессионного моделирования с помощью F -критерия Фишера. По значениям характеристик, рассчитанных в пп. 4, 5 и данном пункте, выберите лучшее уравнение регрессии и дайте его обоснование.

7. Рассчитайте прогнозное значение результата, если прогнозное значение фактора увеличится на 5% от его среднего уровня. Определите доверительный интервал прогноза для уровня значимости $\alpha = 0,05$.

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

Задача 26

Имеются данные по странам за 1994 г., представленные в табл. 1.24.

Таблица 1.24

Страна	Душевой доход*, долл., у	Индекс человеческого развития (ИЧР), x_1	Индекс человеческой бедности (ИЧБ), x_2
Объединенные Арабские Эмираты	1600	0,866	14,9
Таиланд	7100	0,833	11,7
Уругвай	6750	0,883	11,7
Ливия	6130	0,801	18,8
Колумбия	6110	0,848	10,7
Иордания	4190	0,730	10,9
Египет	3850	0,514	34,8
Марокко	3680	0,566	41,7
Перу	3650	0,717	22,8
Шри-Ланка	3280	0,711	20,7
Филиппины	2680	0,672	17,7
Боливия	2600	0,589	22,5
Китай	2600	0,626	17,5
Зимбабве	2200	0,513	17,3
Пакистан	2150	0,445	46,8
Уганда	1370	0,328	41,3
Нигерия	1350	0,393	41,6
Индия	1350	0,446	36,7

* По паритету покупательной способности валют.

Задание

1. Вычислите описательные статистики. Проверьте характер распределения признаков. При необходимости удалите аномальные наблюдения.
2. Постройте матрицу парных коэффициентов корреляции.
3. Постройте парные линейные уравнения регрессии, принимая душевой доход в качестве объясняющей переменной. Постройте графики остатков. Сделайте выводы.
4. Оцените значимость уравнений регрессии в целом и их параметров. Сравните полученные результаты, выберите лучшее уравнение регрессии.

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

2.1. МЕТОДИЧЕСКИЕ УКАЗАНИЯ

Множественная регрессия – уравнение связи с несколькими независимыми переменными:

$$y = f(x_1, x_2, \dots, x_p),$$

где y – зависимая переменная (результативный признак);
 x_1, x_2, \dots, x_p – независимые переменные (факторы).

Для построения уравнения множественной регрессии чаще используются следующие функции:

- линейная – $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \epsilon$;
- степенная – $y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_p^{b_p} \cdot \epsilon$;
- экспонента – $y = e^{a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \epsilon}$;
- гипербола – $y = \frac{1}{a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + \epsilon}$.

Можно использовать и другие функции, приводимые к линейному виду.

Для оценки параметров уравнения множественной регрессии применяют *метод наименьших квадратов* (МНК). Для линейных уравнений и нелинейных уравнений, приводимых к линейным, строится следующая система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии:

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

Для ее решения может быть применён метод определителей:

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

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

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

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

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

где $t_y = \frac{y - \bar{y}}{\sigma_y}, \quad t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$ – стандартизованные переменные;

β_i – стандартизованные коэффициенты регрессии.

К уравнению множественной регрессии в стандартизованном масштабе применим МНК. Стандартизованные коэффициенты регрессии (β -коэффициенты) определяются из следующей системы уравнений:

$$\begin{cases} r_{yx_1} = \beta_1 & + \beta_2 r_{x_2 x_1} + \beta_3 r_{x_3 x_1} + \dots + \beta_p r_{x_p x_1}, \\ r_{yx_2} = \beta_1 r_{x_2 x_1} & + \beta_2 & + \beta_3 r_{x_3 x_2} + \dots + \beta_p r_{x_p x_2}, \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_p} = \beta_1 r_{x_p x_1} & + \beta_2 r_{x_p x_2} + \beta_3 r_{x_3 x_p} + \dots + \beta_p. \end{cases}$$

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

Средние коэффициенты эластичности для линейной регрессии рассчитываются по формуле

$$\bar{\varepsilon}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}}.$$

Для расчета частных коэффициентов эластичности применяется следующая формула:

$$\varepsilon_{yx_i} = b_i \frac{x_i}{\hat{y}_{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p}}.$$

Тесноту совместного влияния факторов на результат оценивает индекс множественной корреляции:

$$R_{yx_1 x_2, \dots, x_p} = \sqrt{1 - \frac{\sigma_{y_{\text{ост}}}^2}{\sigma_y^2}}.$$

Значение индекса множественной корреляции лежит в пределах от 0 до 1 и должно быть больше или равно максимальному парному индексу корреляции:

$$R_{yx_1 x_2, \dots, x_p} \geq r_{yx_i} \quad (i = \overline{1, p}).$$

Индекс множественной корреляции для уравнения в стандартизованном масштабе можно записать в виде

$$R_{yx_1 x_2, \dots, x_p} = \sqrt{\sum \beta_i r_{yx_i}}.$$

При линейной зависимости коэффициент множественной корреляции можно определить через матрицу парных коэффициентов корреляции:

$$R_{yx_1 x_2, \dots, x_p} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}},$$

где

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix} \text{ — определитель матрицы}$$

парных коэффициентов корреляции;

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix} \text{ — определитель матрицы}$$

межфакторной корреляции.

Частные коэффициенты (или индексы) корреляции, измеряющие влияние на y фактора x_i при неизменном уровне других факторов, можно определить по формуле

$$r_{yx_i \cdot x_1x_2 \dots x_{i-1}x_{i+1} \dots x_p} = \sqrt{1 - \frac{1 - R_{yx_1x_2 \dots x_i \dots x_p}^2}{1 - R_{yx_1x_2 \dots x_{i-1}x_{i+1} \dots x_p}^2}}$$

или по рекуррентной формуле:

$$r_{yx_i \cdot x_1x_2 \dots x_p} = \frac{r_{yx_i \cdot x_1x_2 \dots x_{p-1}} - r_{yx_p \cdot x_1x_2 \dots x_{p-1}} r_{x_ix_p \cdot x_1x_2 \dots x_{p-1}}}{\sqrt{(1 - r_{yx_p \cdot x_1x_2 \dots x_{p-1}}^2)(1 - r_{x_ix_p \cdot x_1x_2 \dots x_{p-1}}^2)}}$$

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

Качество построенной модели в целом оценивает коэффициент (индекс) детерминации. Коэффициент множественной детерминации рассчитывается как квадрат индекса множественной корреляции:

$$R_{yx_1x_2, \dots, x_p}^2$$

Скорректированный индекс множественной детерминации содержит поправку на число степеней свободы и рассчитывается по формуле

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

где n – число наблюдений;
 m – число факторов.

Значимость уравнения множественной регрессии в целом оценивается с помощью F -критерия Фишера:

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

Частный F -критерий оценивает статистическую значимость присутствия каждого из факторов в уравнении. В общем виде для фактора x_i частный F -критерий определится как

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

Оценка значимости коэффициентов чистой регрессии с помощью t -критерия Стьюдента сводится к вычислению значения

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

где m_{b_i} – средняя квадратическая ошибка коэффициента регрессии b_i , она может быть определена по следующей формуле:

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

При построении уравнения множественной регрессии может возникнуть проблема мультиколлинеарности факторов, их тесной линейной связанности.

Считается, что две переменные явно коллинеарны, т.е. находятся между собой в линейной зависимости, если $r_{x_i x_j} \geq 0,7$.

По величине парных коэффициентов корреляции обнаруживается лишь явная коллинеарность факторов. Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии *мультиколлинеарности* факторов. Чем сильнее мультиколлинеарность факторов, тем менее надежна оценка распределения суммы объясненной вариации по отдельным факторам с помощью метода наименьших квадратов.

Для оценки мультиколлинеарности факторов может использоваться *определитель матрицы парных коэффициентов корреляции между факторами*.

Если бы факторы не коррелировали между собой, то матрица парных коэффициентов корреляции между факторами была бы единичной матрицей, поскольку все недиагональные элементы $r_{x_i x_j}$ ($x_i \neq x_j$) были бы равны нулю. Так, для включающего три объясняющих переменных уравнения

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + \varepsilon$$

матрица коэффициентов корреляции между факторами имела бы определитель, равный 1:

$$\text{Det } |R| = \begin{vmatrix} r_{x_1 x_1} & r_{x_2 x_1} & r_{x_3 x_1} \\ r_{x_1 x_2} & r_{x_2 x_2} & r_{x_3 x_2} \\ r_{x_1 x_3} & r_{x_2 x_3} & r_{x_3 x_3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1,$$

так как $r_{x_1 x_1} = r_{x_2 x_2} = r_{x_3 x_3} = 1$ и $r_{x_1 x_2} = r_{x_1 x_3} = r_{x_2 x_3} = 0$.

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

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

Чем ближе к 0 определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И наоборот, чем ближе к 1 определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Проверка мультиколлинеарности факторов может быть проведена методом испытания гипотезы о независимости переменных $H_0: \text{Det}|R|=1$. Доказано, что величина $\left[n-1 - \frac{1}{6}(2 \cdot m + 5) \lg \text{Det}R \right]$ имеет приближенное распределение χ^2 с $\left(\frac{1}{2} \cdot n \cdot (n-1) \right)$ степенями свободы. Если фактическое значение χ^2 превосходит табличное (критическое) $\chi_{\text{факт}}^2 > \chi_{\text{табл}(df, \alpha)}^2$, то гипотеза H_0 отклоняется. Это означает, что $\text{Det}|R| \neq 1$, недиагональные ненулевые коэффициенты корреляции указывают на коллинеарность факторов. Мультиколлинеарность считается доказанной.

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

При нарушении гомоскедастичности мы имеем неравенства

$$\sigma_{\varepsilon_i}^2 \neq \sigma_{\varepsilon_j}^2 \neq \sigma^2, \quad j \neq i.$$

При малом объеме выборки для оценки гетероскедастичности может использоваться метод Гольдфельда – Квандта. Основная идея теста Гольдфельда – Квандта состоит в следующем:

- 1) упорядочение n наблюдений по мере возрастания переменной x ;
- 2) исключение из рассмотрения C центральных наблюдений; при этом $(n - C) : 2 > p$, где p – число оцениваемых параметров;
- 3) разделение совокупности из $(n - C)$ наблюдений на две группы (соответственно с малыми и с большими значениями фактора x) и определение по каждой из групп уравнений регрессии;
- 4) определение остаточной суммы квадратов для первой (S_1) и второй (S_2) групп и нахождение их отношения: $R = S_1 : S_2$.

При выполнении нулевой гипотезы о гомоскедастичности отношение R будет удовлетворять F -критерию со степенями свободы $((n - C - 2p) : 2)$ для каждой остаточной суммы квадратов. Чем больше величина R превышает табличное значение F -критерия, тем более нарушена предпосылка о равенстве дисперсий остаточных величин.

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.

Такого вида сконструированные переменные принято в эконометрике называть *фиктивными переменными*. Например, включить в модель фактор «пол» в виде фиктивной переменной можно в следующем виде:

$$z = \begin{cases} 1 - \text{мужской пол,} \\ 0 - \text{женский пол.} \end{cases}$$

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

Пример 1

По 30 территориям России имеются данные, представленные в табл. 2.1.

Таблица 2.1

Признак	Среднее значение	Среднее квадратическое отклонение	Линейный коэффициент парной корреляции
Среднедневной душевой доход, руб., y	86,8	11,44	—
Среднедневная заработная плата одного работающего, руб., x_1	54,9	5,86	$r_{yx_1} = 0,8405$
Средний возраст безработного, лет, x_2	33,5	0,58	$r_{yx_2} = -0,2101$ $r_{x_1x_2} = -0,1160$

Требуется:

1. Построить уравнение множественной регрессии в стандартизованной и естественной форме; рассчитать частные коэффициенты эластичности, сравнить их с β_1 и β_2 , пояснить различия между ними.
2. Рассчитать линейные коэффициенты частной корреляции и коэффициент множественной корреляции, сравнить их с линейными коэффициентами парной корреляции, пояснить различия между ними.
3. Рассчитать общий и частные *F*-критерии Фишера.

Решение

1. Линейное уравнение множественной регрессии y от x_1 и x_2 имеет вид: $y = a + b_1 \cdot x_1 + b_2 \cdot x_2$. Для расчета его параметров применим метод стандартизации переменных и построим искомое уравнение в стандартизованном масштабе: $t_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2}$.

Расчет β -коэффициентов выполним по формулам

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} = \frac{0,8405 - 0,2101 \cdot 0,116}{1 - 0,116^2} = \frac{0,8161}{0,9865} = 0,8273;$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} = \frac{-0,2101 + 0,8405 \cdot 0,116}{1 - 0,116^2} = \frac{-0,1126}{0,9865} = -0,1141.$$

Получим уравнение

$$t_y = 0,8273 t_{x_1} - 0,1141 t_{x_2}.$$

Для построения уравнения в естественной форме рассчитаем b_1 и b_2 , используя формулы для перехода от β_i к b_i :

$$\beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y}; \quad b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}};$$

$$b_1 = \frac{0,8273 \cdot 11,44}{5,86} = 1,6151; \quad b_2 = \frac{-0,1141 \cdot 11,44}{0,58} = -2,2505.$$

Значение a определим из соотношения

$$a = \bar{y} - b_1 \cdot \bar{x}_1 - b_2 \cdot \bar{x}_2 = 86,8 - 1,6151 \cdot 54,9 + 2,2505 \cdot 33,5 = -73,52276,$$

$$y_{x_1 x_2} = -73,52 + 1,62 \cdot x_1 - 2,25 \cdot x_2.$$

Для характеристики относительной силы влияния x_1 и x_2 на y рассчитаем средние коэффициенты эластичности:

$$\bar{\varepsilon}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}};$$

$$\bar{\varepsilon}_{yx_1} = \frac{1,62 \cdot 54,9}{86,8} = 1,0246\%; \quad \bar{\varepsilon}_{yx_2} = \frac{-2,25 \cdot 33,5}{86,8} = -0,8684\%.$$

С увеличением средней заработной платы x_1 на 1% от ее среднего уровня средний душевой доход y возрастает на 1,02% от своего среднего уровня; при повышении среднего возраста безработного x_2 на 1% среднедушевой доход y снижается на 0,87% от своего среднего уровня. Очевидно, что сила влияния средней заработной платы x_1 на средний душевой доход y оказалась большей, чем сила влияния среднего возраста безработного x_2 . К аналогичным выводам о силе связи приходим при сравнении модулей значений β_1 и β_2 :

$$|\beta_1| = |0,8273| > |\beta_2| = |-0,1141|.$$

Различия в силе влияния фактора на результат, полученные при сравнении $\bar{\varepsilon}_{yx_j}$ и β_j , объясняются тем, что коэффициент эластичности исходит из соотношения средних: $\bar{\varepsilon}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}}$, а

β -коэффициент — из соотношения средних квадратических отклонений: $\beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y}$.

2. Линейные коэффициенты частной корреляции здесь рассчитываются по рекуррентной формуле:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1 x_2}^2)}} = \frac{0,8405 - 0,2101 \cdot 0,116}{\sqrt{(1 - 0,2101^2)(1 - 0,116^2)}} = 0,8404;$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1 x_2}^2)}} = \frac{-0,2101 + 0,8405 \cdot 0,116}{\sqrt{(1 - 0,8405^2)(1 - 0,116^2)}} = -0,2092;$$

$$r_{x_1 x_2 \cdot y} = \frac{r_{x_1 x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{yx_2}^2)}} = \frac{-0,116 + 0,8405 \cdot 0,2101}{\sqrt{(1 - 0,8405^2)(1 - 0,2101^2)}} = 0,1144.$$

Если сравнить значения коэффициентов парной и частной корреляции, то приходим к выводу, что из-за слабой межфакторной связи ($r_{x_1 x_2} = -0,116$) коэффициенты парной и частной корреляции отличаются незначительно: выводы о тесноте и направлении связи на основе коэффициентов парной и частной корреляции совпадают:

$$r_{yx_1} = 0,8405; r_{yx_2} = -0,2101; r_{x_1x_2} = 0,1160;$$

$$r_{yx_1 \cdot x_2} = 0,8404; r_{yx_2 \cdot x_1} = -0,2092; r_{x_1x_2 \cdot y} = 0,1144.$$

Расчет линейного коэффициента множественной корреляции выполним с использованием коэффициентов r_{yx_j} и β_j :

$$R_{yx_1x_2} = \sqrt{r_{yx_1} \cdot \beta_1 + r_{yx_2} \beta_2} = \sqrt{0,8405 \cdot 0,8273 + 0,2101 \cdot 0,1141} =$$

$$= \sqrt{0,7193} = 0,8481.$$

Зависимость y от x_1 и x_2 характеризуется как тесная, в которой 72% вариации среднего душевого дохода определяются вариацией учтенных в модели факторов: средней заработной платы и среднего возраста безработного. Прочие факторы, не включенные в модель, составляют соответственно 28% от общей вариации y .

3. Общий F -критерий проверяет гипотезу H_0 о статистической значимости уравнения регрессии и показателя тесноты связи ($R^2 = 0$):

$$F_{\text{факт}} = \frac{R^2_{yx_1x_2}}{1 - R^2_{yx_1x_2}} \cdot \frac{m}{n - m - 1} = \frac{R^2_{yx_1x_2}}{1 - R^2_{yx_1x_2}} \cdot \frac{n - m - 1}{m} =$$

$$= \frac{0,7193}{0,2807} \cdot \frac{27}{2} = 34,6;$$

$$F_{\text{табл}} = 3,4; \alpha = 0,05.$$

Сравнивая $F_{\text{табл}}$ и $F_{\text{факт}}$, приходим к выводу о необходимости отклонить гипотезу H_0 , так как $F_{\text{табл}} = 3,4 < F_{\text{факт}} = 34,6$. С вероятностью $1 - \alpha = 0,95$ делаем заключение о статистической значимости уравнения в целом и показателя тесноты связи $R_{yx_1x_2}$, которые сформировались под неслучайным воздействием факторов x_1 и x_2 .

Частные F -критерии — F_{x_1} и F_{x_2} оценивают статистическую значимость присутствия факторов x_1 и x_2 в уравнении множественной регрессии, оценивают целесообразность включения в уравнение одного фактора после другого фактора, т.е. F_{x_1} оценивает целесообразность включения в уравнение фактора x_1 после того, как в него был включен фактор x_2 . Соответственно F_{x_2} указывает на целесообразность включения в модель фактора x_2 после фактора x_1 :

$$F_{x_1 \text{ факт}} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot \frac{n - m - 1}{1} = \frac{0,8481^2 - 0,2101^2}{1 - 0,8481^2} \cdot \frac{30 - 2 - 1}{1} = 64,9.$$

$$F_{\text{табл}} = 4,21; \alpha = 0,05.$$

Сравнивая $F_{\text{табл}}$ и $F_{\text{факт}}$, приходим к выводу о целесообразности включения в модель фактора x_1 после фактора x_2 , так как $F_{x_1 \text{ факт}} = 64,9 > F_{\text{табл}}$. Гипотезу H_0 о несущественности прироста R^2_y за счет включения дополнительного фактора x_1 отклоняем и приходим к выводу о статистически подтвержденной целесообразности включения фактора x_1 после фактора x_2 .

Целесообразность включения в модель фактора x_2 после фактора x_1 проверяет F_{x_2} :

$$F_{x_2 \text{ факт}} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot \frac{n - m - 1}{1} = \frac{0,8481^2 - 0,8405^2}{1 - 0,8481^2} \cdot \frac{30 - 2 - 1}{1} = 1,234.$$

Низкое значение $F_{x_2 \text{ факт}}$ (немного больше 1) свидетельствует о статистической незначимости прироста $r^2_{yx_1}$ за счет включения в модель фактора x_2 после фактора x_1 . Следовательно, подтверждается нулевая гипотеза H_0 о нецелесообразности включения в модель фактора x_2 (средний возраст безработного). Это означает, что парная регрессионная модель зависимости среднего дохода от средней заработной платы является достаточно статистически значимой, надежной и что нет необходимости улучшать ее, включая дополнительный фактор x_2 (средний возраст безработного).

Пример 2

По 20 территориям России изучаются следующие данные (табл. 2.2): зависимость среднегодового душевого дохода y (тыс. руб.) от доли занятых тяжелым физическим трудом в общей численности занятых x_1 (%) и от доли экономически активного населения в численности всего населения x_2 (%).

Таблица 2.2

Признак	Среднее значение	Среднее квадратическое отклонение	Характеристика тесноты связи	Уравнение связи
у	112,76	31,58	$R_{yx_1x_2} = 0,773$	$\hat{y}_{x_1x_2} = -130,49 + 6,14 \cdot x_1 + 4,13 \cdot x_2$
x_1	5,40	3,34	$r_{yx_1} = 0,746$	$\hat{y}_{x_1} = 74,4 + 7,1 \cdot x_1$
x_2	50,88	1,74	$r_{yx_2} = 0,507$ $r_{x_1x_2} = 0,432$	$\hat{y}_{x_2} = -355,3 + 9,2 \cdot x_2$

Требуется:

1. Составить таблицу дисперсионного анализа для проверки при уровне значимости $\alpha = 0,05$ статистической значимости уравнения множественной регрессии и его показателя тесноты связи.
2. С помощью частных F -критериев Фишера оценить, насколько целесообразно включение в уравнение множественной регрессии фактора x_1 после фактора x_2 и насколько целесообразно включение x_2 после x_1 .
3. Оценить с помощью t -критерия Стьюдента статистическую значимость коэффициентов при переменных x_1 и x_2 множественного уравнения регрессии.

Решение

1. Задача дисперсионного анализа состоит в проверке нулевой гипотезы H_0 о статистической незначимости уравнения регрессии в целом и показателя тесноты связи.

Анализ выполняется при сравнении фактического и табличного (критического) значений F -критерия Фишера $F_{\text{табл}}$ и $F_{\text{факт}}$. $F_{\text{факт}}$ определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы:

$$F_{\text{факт}} = \frac{\sum (\hat{y}_{x_1x_2} - \bar{y})^2}{m} : \frac{\sum (y - \hat{y}_{x_1x_2})^2}{n - m - 1} = \frac{S_{\text{факт}}}{S_{\text{ост}}} \cdot \frac{n - m - 1}{m},$$

- где n — число единиц совокупности;
 m — число факторов в уравнении линейной регрессии;
 \hat{y} — фактическое значение результативного признака;
 $y_{x_1x_2}$ — расчетное значение результативного признака.

Результаты дисперсионного анализа представлены в табл. 2.3.

Таблица 2.3

Вариация результата, у	Число степеней свободы	Сумма квадратов отклонений, S	Дисперсия на одну степень свободы, s ²	F _{факт}	F _{табл} α = 0,05, k ₁ = 2, k ₂ = 17
Общая	df = n - 1 = 19	19945,9	—	—	—
Факторная	k ₁ = m = 2	11918,3	5959,15	12,62	3,59
Остаточная	k ₂ = n - m - 1 = 17	8027,6	472,21	—	—

$$S_{\text{общ}} = \sigma_y^2 \cdot n = (31,58)^2 \cdot 20 = 19945,9;$$

$$S_{\text{факт}} = \sigma_y^2 \cdot n \cdot R_{yx_1x_2}^2 = 19945,9 \cdot (0,773)^2 = 11918,3;$$

$$S_{\text{ост}} = \sigma_y^2 \cdot n \cdot (1 - R_{yx_1x_2}^2) = S_{\text{общ}} - S_{\text{факт}} = 8027,6;$$

$$F_{\text{факт}} = \frac{11918,3}{8027,6} \cdot \frac{17}{2} = 12,62.$$

Сравнивая $F_{\text{табл}}$ и $F_{\text{факт}}$, приходим к выводу о необходимости отклонить гипотезу H_0 и сделать вывод о статистической значимости уравнения регрессии в целом и значения $R_{yx_1x_2}^2$, так как они статистически надежны и сформировались под систематическим действием неслучайных причин. Вероятность того, что допускаются ошибки при отклонении нулевой гипотезы, не превышает 5%, и это является достаточно малой величиной.

2. Частный F -критерий Фишера оценивает статистическую целесообразность включения фактора x_1 в модель после того, как в нее включен фактор x_2 . Частный F -критерий Фишера строится как отношение прироста факторной дисперсии за счет дополнительно включенного фактора (на одну степень свободы) к остаточной дисперсии (на одну степень свободы), подсчитанной по модели с включенными факторами x_1 и x_2 :

$$F_{\text{частн } x_1} = \frac{S_{\text{факт } yx_1x_2} - S_{\text{факт } yx_2}}{S_{\text{ост } yx_1x_2}} \cdot \frac{n - m - 1}{1}.$$

Результаты дисперсионного анализа представлены в табл. 2.4.

Таблица 2.4

Вариация результата, у	Число степеней свободы	Сумма квадратов отклонений, S	Дисперсия на одну степень свободы, s ²	F _{факт}	F _{табл} α = 0,05, k ₁ = 2, k ₂ = 17
Общая	df = n - 1 = 19	19945,9	—	—	—
Факторная	k ₁ = m = 2	11918,3	5959,15	12,62	3,59
В том числе:					
• за счет x ₂	1	5127,1	5127,1	10,86	4,45
• за счет дополнительно включенного x ₁	1	6791,2	6791,2	14,38	4,45
Остаточная	k ₂ = n - m - 1 = 17	8027,6	472,21	—	—

$$S_{\text{общ}} = \sigma_y^2 \cdot n = (31,58)^2 \cdot 20 = 19945,9;$$

$$S_{\text{факт}} = \sigma_y^2 \cdot n \cdot R_{yx_1x_2}^2 = 19945,9 \cdot (0,773)^2 = 11918,3;$$

$$S_{\text{факт } x_2} = \sigma_y^2 \cdot n \cdot r_{yx_2}^2 = 19945,9 \cdot (0,507)^2 = 5127,1;$$

$$S_{\text{факт } x_1} = S_{\text{факт}} - S_{\text{факт } x_2} = 11918,3 - 5127,1 = 6791,2;$$

$$S_{\text{ост}} = \sigma_y^2 \cdot n \cdot (1 - R_{yx_1x_2}^2) = S_{\text{общ}} - S_{\text{факт}} = 8027,6.$$

Включение фактора x₁ после фактора x₂ оказалось статистически значимым и оправданным: прирост факторной дисперсии (в расчете на одну степень свободы) оказался существенным, т.е. следствием дополнительного включения в модель систематически действующего фактора x₁, так как F_{частн x₁} = 14,38 > F_{табл} = 4,45.

Аналогично проверим целесообразность включения в модель дополнительного фактора x₂ после включенного ранее фактора x₁. Расчет выполним с использованием показателей тесноты связи R²_{yx₁x₂} и r²_{yx₁}:

$$F_{\text{частн } x_2} = \frac{R_{yx_1x_2}^2 - r_{yx_1}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{1} = \frac{(0,773)^2 - (0,746)^2}{1 - (0,773)^2} \cdot \frac{17}{1} = 1,73.$$

В силу того что $F_{\text{частн } x_2} = 1,73 < F_{\text{табл}} = 4,45$, приходим к выводу, что включение x_2 после x_1 оказалось бесполезным: прирост факторной дисперсии в расчете на одну степень свободы был несуществен, статистически незначим, т.е. влияние x_2 не является устойчивым, систематическим. Вполне возможно было ограничиться построением линейного уравнения парной регрессии y от x_1 .

3. Оценка с помощью t -критерия Стьюдента значимости коэффициентов b_1 и b_2 связана с сопоставлением их значений с величиной их случайных ошибок: m_{b_1} и m_{b_2} . Расчет значений случайных ошибок достаточно сложен и трудоёмок. Поэтому предлагается более простой способ: расчет значения t -критерия Стьюдента для коэффициентов регрессии линейного уравнения как квадратного корня из соответствующего частного F -критерия Фишера:

$$t_{b_1} = \sqrt{F_{\text{частн } x_1}} = \sqrt{14,38} = 3,79;$$

$$t_{b_2} = \sqrt{F_{\text{частн } x_2}} = \sqrt{1,73} = 1,32.$$

Табличные (критические) значения t -критерия Стьюдента зависят от принятого уровня значимости α (обычно это 0,1; 0,05 или 0,01) и от числа степеней свободы ($n - m - 1$), где n – число единиц совокупности, m – число факторов в уравнении.

В нашем примере при $\alpha = 0,05$; $df = 20 - 3 = 17$; $t_{\text{табл}} = 2,10$. Сравнивая $t_{\text{табл}}$ и $t_{\text{факт}}$, приходим к выводу, что так как $t_{b_1} = 3,79 > 2,11 = t_{\text{табл}}$, коэффициент регрессии b_1 является статистически значимым, надежным, на него можно опираться в анализе и в прогнозе. Так как $t_{b_2} = 1,32 < 2,10 = t_{\text{табл}}$, приходим к заключению, что величина b_2 является статистически незначимой, ненадежной в силу того, что она формируется преимущественно под воздействием случайных факторов. Еще раз подтверждается статистическая значимость влияния x_1 (доли занятых тяжелым физическим трудом) на y (среднедушевой доход) и ненадежность, незначимость влияния x_2 (доли экономически активного населения в численности всего населения).

Пример 3

Зависимость спроса на свинину x_1 от цены на нее x_2 и от цены на говядину x_3 представлена уравнением

$$\lg x_1 = 0,1274 - 0,2143 \cdot \lg x_2 + 2,8254 \cdot \lg x_3.$$

Требуется:

1. Представить данное уравнение в естественной форме (не в логарифмах).
2. Оценить значимость параметров данного уравнения, если известно, что t -критерий для параметра b_2 при x_2 составил 0,827, а для параметра b_3 при x_3 – 1,015.

Решение

1. Представленное степенное уравнение множественной регрессии приводим к естественной форме путём потенцирования обеих частей уравнения:

$$x_1 = 10^{0,1274} \cdot x_2^{-0,2143} \cdot x_3^{2,8254};$$

$$x_1 = 1,3409 \cdot \frac{1}{x_2^{0,2143}} \cdot x_3^{2,8254}.$$

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

$$\bar{\varepsilon}_{x_1 x_2} = -0,2143\%; \quad \bar{\varepsilon}_{x_1 x_3} = 2,8254\%.$$

Спрос на свинину x_1 сильнее связан с ценой на говядину – он увеличивается в среднем на 2,83% при росте цен на 1%. С ценой на свинину спрос на нее связан обратной зависимостью: с ростом цен на 1% потребление снижается в среднем на 0,21%.

2. Табличное значение t -критерия для $\alpha = 0,05$ обычно лежит в интервале 2 – 3 – в зависимости от степеней свободы. В данном примере $t_{b_2} = 0,827$, $t_{b_3} = 1,015$. Это весьма небольшие значения t -критерия, которые свидетельствуют о случайной природе взаимосвязи, о статистической ненадежности всего уравнения, поэтому применять полученное уравнение для прогноза не рекомендуется.

Пример 4

По 20 предприятиям региона (табл. 2.5) 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 2.5

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

Требуется:

1. Оценить показатели вариации каждого признака и сделать вывод о возможностях применения МНК для их изучения.
2. Проанализировать линейные коэффициенты парной и частной корреляции.
3. Написать уравнение множественной регрессии, оценить значимость его параметров, пояснить их экономический смысл.
4. С помощью F -критерия Фишера оценить статистическую надежность уравнения регрессии и R^2_{y, x_1, x_2} . Сравнить значения скорректированного и нескорректированного линейных коэффициентов множественной детерминации.
5. С помощью частных F -критериев Фишера оценить целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1 .
6. Рассчитать средние частные коэффициенты эластичности и дать на их основе сравнительную оценку силы влияния факторов на результат.

2.3. РЕАЛИЗАЦИЯ ТИПОВЫХ ЗАДАЧ НА КОМПЬЮТЕРЕ

1. Решение примера проведем с использованием ППП MS Excel и Statgraphics.

Решение с помощью ППП Excel

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**;

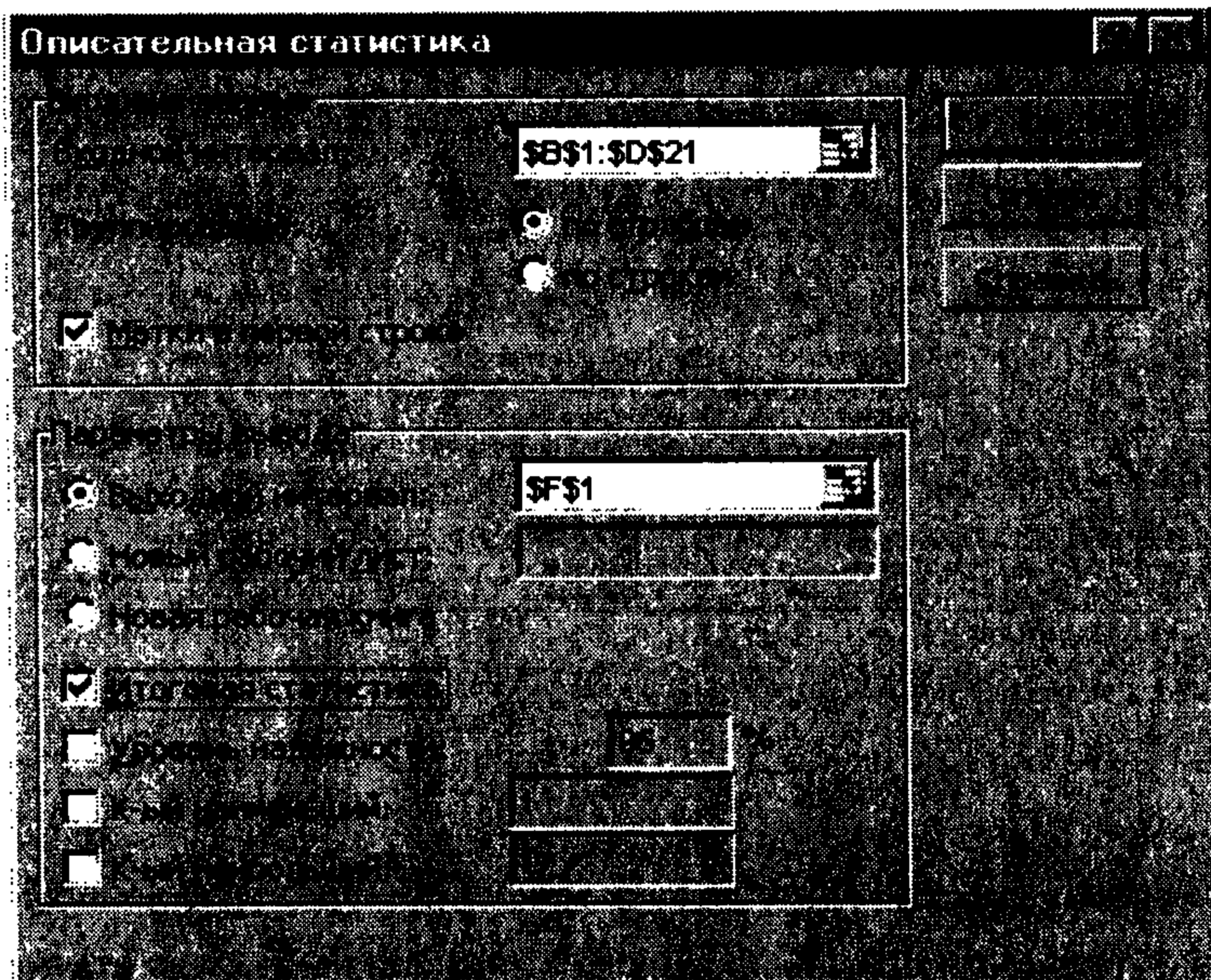


Рис. 2.1. Диалоговое окно ввода параметров инструмента **Описательная статистика**

- 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 указать дополнительно;

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.

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

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

	x_1	x_2	x_3
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
Среднее	9,6	6,19	22,3
Стандартная ошибка	0,549641	0,433523	1,523673
Медиана	9	6,2	20,5
Мода	7	3,9	20
Стандартное отклонение	2,458069	1,938773	6,814072
Дисперсия выборки	6,042105	3,758842	46,43138
Экцесс	-1,196054	-1,331426	-0,536529
Асимметричность	0,445096	0,188101	0,327801
Интервал	7	5,9	26
Минимум	7	3,7	10
Максимум	14	9,6	36
Сумма	192	123,8	446
Счет	20	20	20

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

Решение с помощью ППП Statgraphics

Для проведения многофакторного анализа в ППП Statgraphics используется пункт меню **Multiple Variable Analysis**. Для получения показателей описательной статистики необходимо проделать следующие операции:

- 1) ввести исходные данные или открыть существующий файл, содержащий анализируемые данные;
- 2) в главном меню выбрать **Describe/Numeric Data/Multiple Variable Analysis**;
- 3) заполнить диалоговое окно ввода данных (рис. 2.3). Ввести названия всех столбцов, значения которых вы хотите включить в анализ; щелкнуть по кнопке **ОК**;

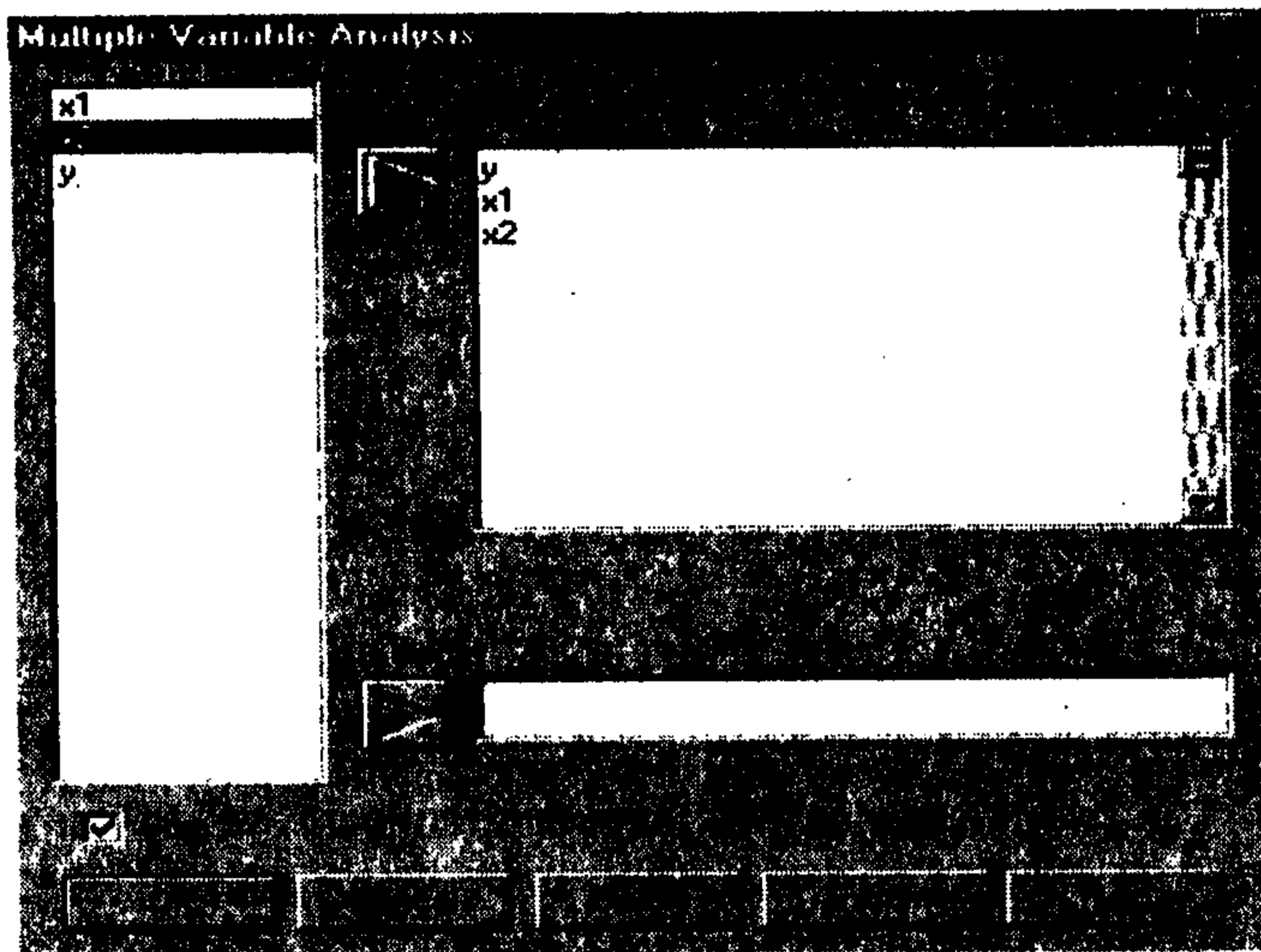


Рис. 2.3. Диалоговое окно ввода данных

4) в окне табличных настроек поставить флажок напротив *Summary Statistics* (рис. 2.4). Итоговая статистика – показатели вариации – появится в отдельном окне.

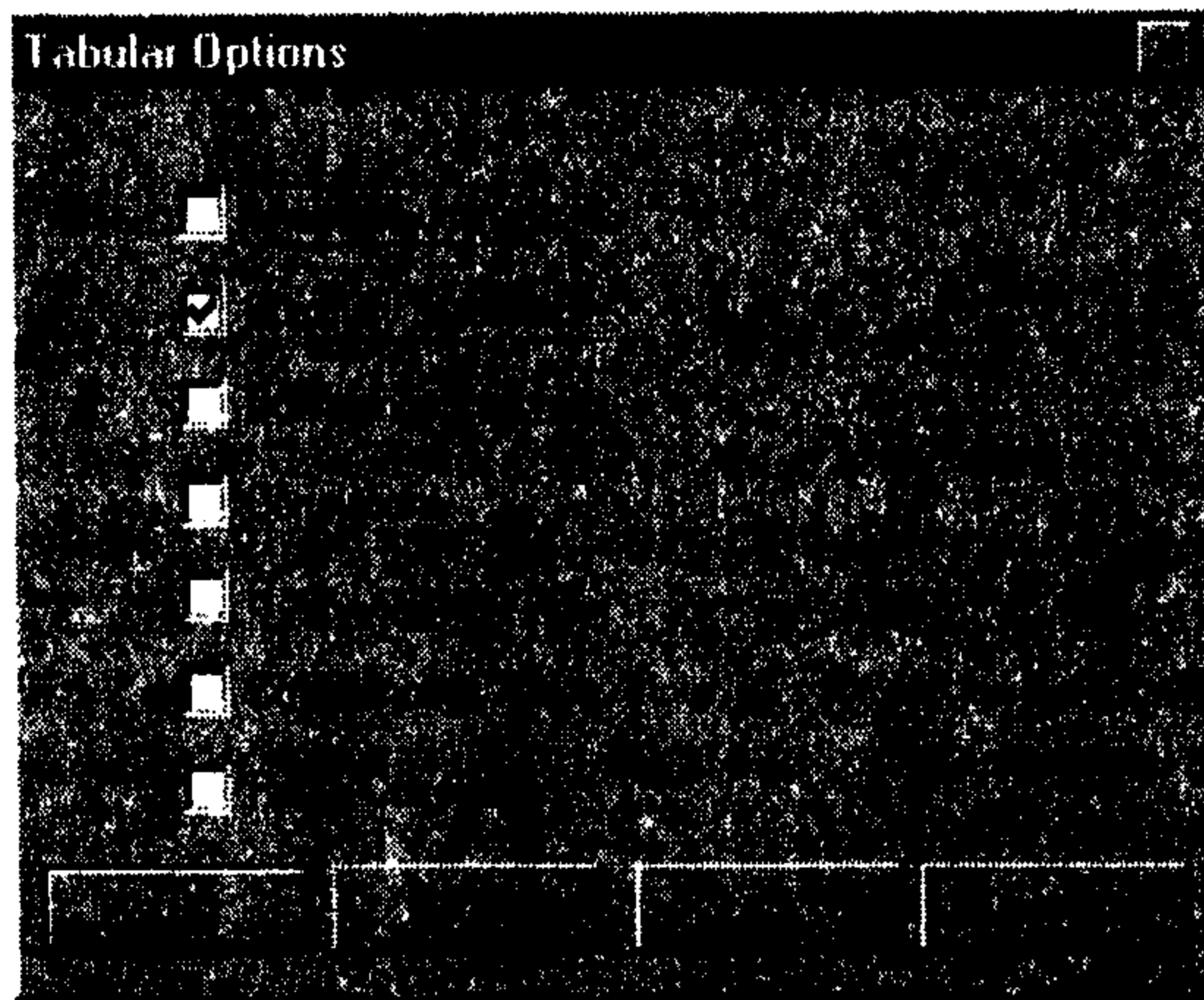


Рис. 2.4. Окно табличных настроек **Multiple Variable Analysis**

Для данных примера 4 результат применения функции **Multiple Variable Analysis** представлен на рис. 2.5.

STATGRAPHICS Plus Untitled Staff.csis [Multiple Variable Analysis]

Summary Statistics

	y	x1	x2
Count	20	20	20
Average	9,6	6,19	22,3
Variance	6,04211	3,75884	46,4316
Standard deviation	2,45807	1,93877	6,81407
Minimum	7,0	3,7	10,0
Maximum	14,0	9,6	36,0
Std. skewness	0,81263	0,343424	0,59848
Std. kurtosis	-1,09184	-1,21542	-0,489782
Sum	192,0	123,8	446,0

Рис. 2.5. Итоговая статистика

Сравнивая значения средних квадратических отклонений и средних величин и определяя коэффициенты вариации:

$$v_y = \frac{\sigma_y}{\bar{y}} \cdot 100\% = \frac{2,45807}{9,6} \cdot 100\% = 25,6\%;$$

$$v_{x_1} = \frac{\sigma_{x_1}}{\bar{x}_1} \cdot 100\% = \frac{1,93877}{6,19} \cdot 100\% = 31,3\%;$$

$$v_{x_{21}} = \frac{\sigma_{x_{21}}}{\bar{x}_2} \cdot 100\% = \frac{6,81407}{22,3} \cdot 100\% = 30,6\%;$$

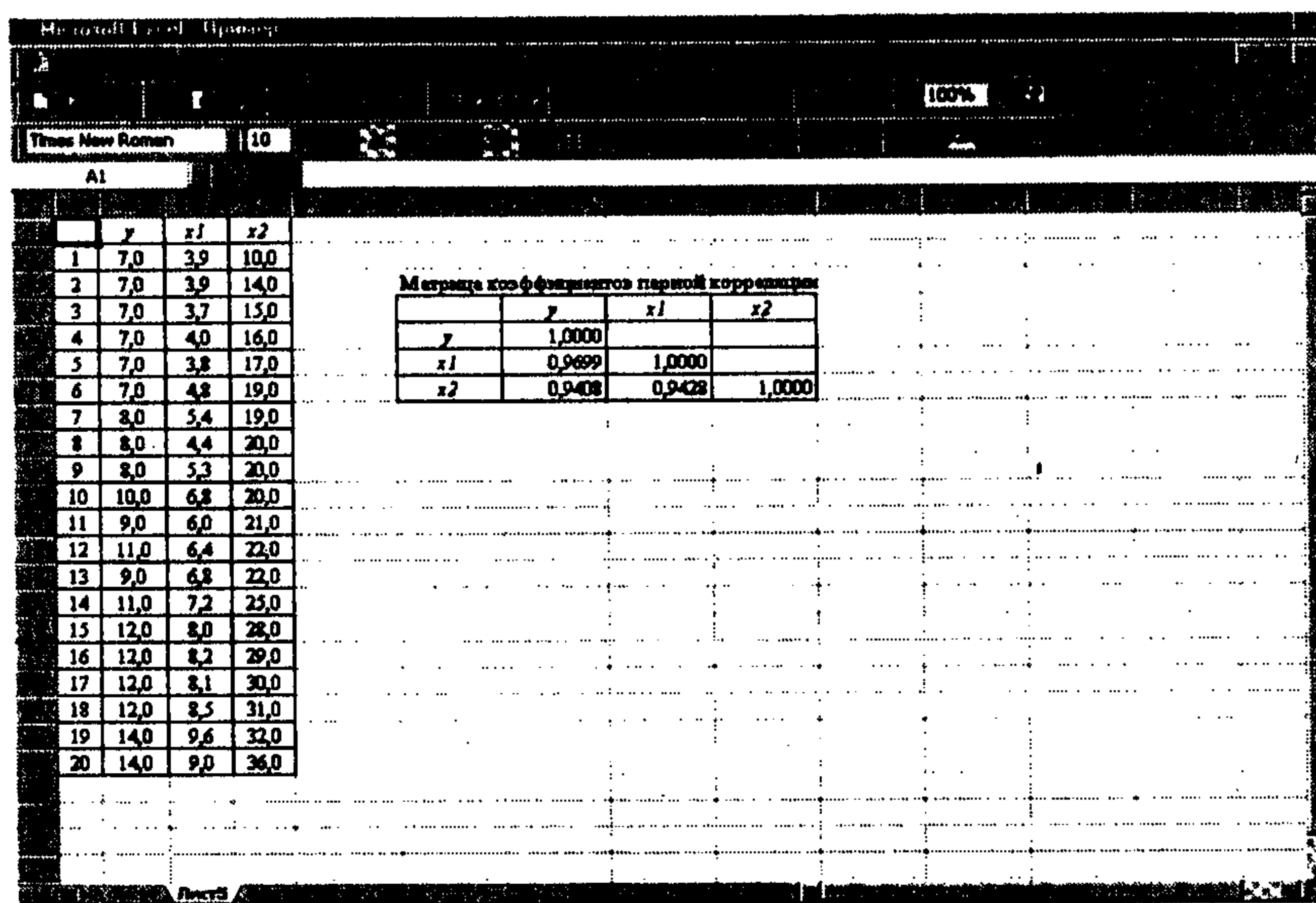
приходим к выводу о повышенном уровне варьирования признаков, хотя и в допустимых пределах, не превышающих 35%. Совокупность предприятий однородна, и для ее изучения могут использоваться метод наименьших квадратов и вероятностные методы оценки статистических гипотез.

2. Значения линейных коэффициентов парной корреляции определяют тесноту попарно связанных переменных, использованных в данном уравнении множественной регрессии. Линейные коэффициенты частной корреляции оценивают тесноту связи значений двух переменных, исключая влияние всех других переменных, представленных в уравнении множественной регрессии.

Решение с помощью ППП Excel

К сожалению, в ППП MS Excel нет специального инструмента для расчета линейных коэффициентов частной корреляции. Матрицу парных коэффициентов корреляции переменных можно рассчитать, используя инструмент 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 (см. рис. 2.1);
- 3) результаты вычислений – матрица коэффициентов парной корреляции – представлены на рис. 2.6.



Матрица коэффициентов парной корреляции

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

Рис. 2.6. Матрица коэффициентов парной корреляции

Решение с помощью ППП Statgraphics

При проведении многофакторного анализа – **Multiple Variable Analysis** – вычисляются линейные коэффициенты парной корреляции и линейные коэффициенты частной корреляции. Последовательность операций описана в п.1 этого примера. Для отображения результатов вычисления на экране необходимо установить флажки напротив *Correlations* и *Partial Correlations* в окне табличных настроек (рис. 2.7).

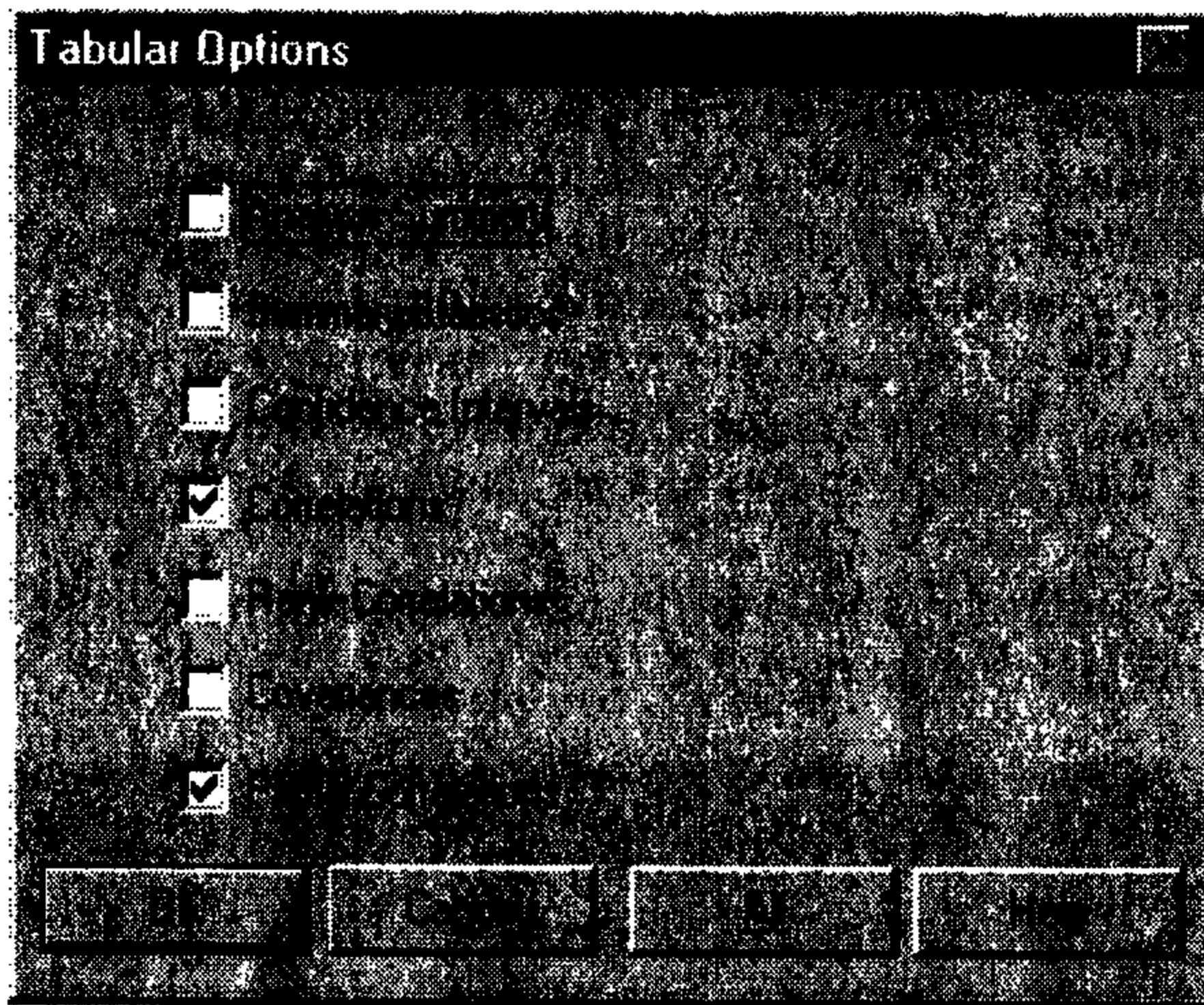


Рис. 2.7. Окно табличных настроек Multiple Variable Analysis

В результате получим матрицы коэффициентов парной и частной корреляции (рис. 2.8).

STATGRAPHICS Plus Untitled StatFolio [Multiple Variable Analysis]

Correlations

	y	x1	x2
y		0,9699 (20) 0,0000	0,9408 (20) 0,0000
x1	0,9699 (20) 0,0000		0,9428 (20) 0,0000
x2	0,9408 (20) 0,0000	0,9428 (20) 0,0000	

Partial Correlations

	y	x1	x2
y		0,7335 (20)	0,3247 (20)
x1	0,7335 (20)		0,3679 (20)
x2	0,3247 (20)	0,3679 (20)	

Рис. 2.8. Матрицы коэффициентов парной и частной корреляции

Значения коэффициентов парной корреляции указывают на весьма тесную связь выработки y как с коэффициентом обновления основных фондов — x_1 , так и с долей рабочих высокой квалификации — x_2 ($r_{yx_1} = 0,9699$ и $r_{yx_2} = 0,9408$). Но в то же время межфакторная связь $r_{x_1x_2} = 0,9428$ весьма тесная и превышает тесноту связи x_2 с y . В связи с этим для улучшения данной модели можно исключить из нее фактор x_2 как малоинформативный, недостаточно статистически надежный.

Коэффициенты частной корреляции дают более точную характеристику тесноты связи двух признаков, чем коэффициенты парной корреляции, так как очищают парную зависимость от взаимодействия данной пары признаков с другими признаками, представленными в модели. Наиболее тесно связаны y и x_1 : $r_{yx_1 \cdot x_2} = 0,7335$; связь y и x_2 гораздо слабее: $r_{yx_2 \cdot x_1} = 0,3247$, а межфакторная зависимость x_1 и x_2 выше, чем парная y и x_2 : $r_{yx_2 \cdot x_1} = 0,3247 < r_{x_1x_2 \cdot y} = 0,3679$. Все это приводит к выводу о необходимости исключить фактор x_2 — доля высококвалифицированных рабочих — из правой части уравнения множественной регрессии.

Если сравнить коэффициенты парной и частной корреляции, то можно увидеть, что из-за высокой межфакторной зависимости коэффициенты парной корреляции дают завышенные оценки тесноты связи:

$$r_{yx_1} = 0,9699; r_{yx_1 \cdot x_2} = 0,7335; r_{yx_2} = 0,9408;$$

$$r_{yx_2 \cdot x_1} = 0,3247.$$

Именно по этой причине рекомендуется при наличии сильной коллинеарности (взаимосвязи) факторов исключать из исследования тот фактор, у которого теснота парной зависимости меньше, чем теснота межфакторной связи.

3. Вычисление параметров линейного уравнения множественной регрессии.

Решение с помощью ППП Excel

Эта операция проводится с помощью инструмента анализа данных **Регрессия**. Она аналогична расчету параметров парной линейной регрессии, описанной в 1-м разделе практикума, только в отличие от парной регрессии в диалоговом окне при заполнении параметра *входной интервал X* следует указать не один столбец, а все столбцы, содержащие значения факторных признаков. Результаты анализа представлены на рис. 2.9.

Дисперсионный анализ					
	df	SS	MS	F	Значимо сть F
Регрессия	2	108,7070945	54,35354726	151,6535	1,45E-11
Остаток	17	6,092905478	0,358406205		
Итого	19	114,8			

	Коэффициенты	Стандартная ошибка	t- статистика	P- Значение	Нижнее 95%	Верхнее 95%	Нижнее 95,0%	Верхнее 95,0%
У-пересечение	1,83530694	0,471064997	3,896080054	0,001162	0,841445	2,829169	0,841445	2,829169
x1	0,945947723	0,212576487	4,449917001	0,000351	0,49745	1,394446	0,49745	1,394446
x2	0,085617787	0,060483309	1,415580577	0,174964	-0,04199	0,213227	-0,04199	0,213227

Рис. 2.9. Результат применения инструмента Регрессия

Решение с помощью ППП Statgraphics

Для вычисления параметров множественной регрессии можно использовать процедуру **Multiple Regression**. Для этого:

- 1) введите исходные данные или откройте существующий файл;
- 2) в главном меню последовательно выберите **Relate / Multiple Regression**;
- 3) заполните диалоговое окно ввода данных. В поле *Depended Variable* введите название столбца, содержащего значения зависимой переменной, в поле *Independed Variable* – названия столбцов, содержащих значения факторов. Щелкните по кнопке **OK**.

Результаты вычисления функции **Multiple Regression** появятся в отдельном окне (рис. 2.10).

По результатам вычислений составим уравнение множественной регрессии вида

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2;$$

$$\hat{y} = 1,8353 + 0,9459 \cdot x_1 + 0,0856 \cdot x_2.$$

Значения случайных ошибок параметров b_0 , b_1 и b_2 с учетом округления:

$$m_{b_0} = 0,4711; m_{b_1} = 0,2126; m_{b_2} = 0,0605.$$

STATISTICAL Pak Untitled Stat Pak (Multiple Regression)

Multiple Regression Analysis

Dependent variable: y

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	1,83531	0,471065	3,89608	0,0012
x1	0,945948	0,212576	4,44992	0,0004
x2	0,0856178	0,0604832	1,41656	0,1750

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	108,707	2	54,3535	151,65	0,0000
Residual	6,09291	17	0,358406		
Total (Corr.)	114,8	19			

R-squared = 94,6926 percent
R-squared (adjusted for d.f.) = 94,0682 percent
Standard Error of Est. = 0,59867
Mean absolute error = 0,426528
Durbin-Watson statistic = 2,39705

Рис. 2.10. Итоговое окно функции Multiple Regression

Они показывают, какое значение данной характеристики сформировалось под влиянием случайных факторов. Эти значения используются для расчета t -критерия Стьюдента:

$$t_{b_0} = 3,90; \quad t_{b_1} = 4,45; \quad t_{b_2} = 1,42.$$

Если значения t -критерия больше 2–3, можно сделать вывод о существенности данного параметра, который формируется под воздействием неслучайных причин. Здесь статистически значимыми являются b_0 и b_1 , а величина b_2 сформировалась под воздействием случайных причин, поэтому фактор x_2 , силу влияния которого оценивает b_2 , можно исключить как несущественно влияющий, неинформативный.

На это же указывает показатель вероятности случайных значений параметров регрессии: если α меньше принятого нами уровня (обычно 0,1; 0,05 или 0,01; это соответствует 10%; 5% или 1% вероятности), делают вывод о неслучайной природе данного значения параметра, т.е. о том, что он статистически значим и надежен. В противном случае принимается гипотеза о случайной природе значения коэффициентов уравнения. Здесь $\alpha_{x_2} = 17,5\% > 5\%$, что позволяет рассматривать x_2 как неинформативный фактор и удалить его для улучшения данного уравнения.

Величина b_0 оценивает агрегированное влияние прочих (кроме учтенных в модели факторов x_1 и x_2) факторов на результат y .

Величины b_1 и b_2 указывают, что с увеличением x_1 и x_2 на единицу их значений результат увеличивается соответственно на 0,9459 и на 0,0856 млн руб. Сравнить эти значения не следует, так как они зависят от единиц измерения каждого признака и потому несопоставимы между собой.

4. Оценку надежности уравнения регрессии в целом и показателя тесноты связи $R_{yx_1x_2}$ дает F -критерий Фишера:

$$F_{\text{факт}} = \frac{\sum (\hat{y}_{x_1x_2} - \bar{y})^2}{m} : \frac{\sum (y - \hat{y}_{x_1x_2})^2}{n - m - 1}.$$

По данным таблиц дисперсионного анализа, представленным на рис. 2.9 и 2.10, $F_{\text{факт}} = 151,65$. Вероятность случайно получить такое значение F -критерия составляет 0,0000, что не превышает допустимый уровень значимости 5%; об этом свидетельствует величина P – значения из этих же таблиц. Следовательно, полученное значение не случайно, оно сформировалось под влиянием существенных факторов, т.е. подтверждается статистическая значимость всего уравнения и показателя тесноты связи $R_{yx_1x_2}^2$.

Значения скорректированного и нескорректированного линейных коэффициентов множественной детерминации приведены на рис. 2.9 и 2.10 в рамках регрессионной статистики.

Нескорректированный коэффициент множественной детерминации $R_{yx_1x_2}^2 = 0,9469$ оценивает долю вариации результата за счет представленных в уравнении факторов в общей вариации результата. Здесь эта доля составляет 94,7% и указывает на весьма высокую степень обусловленности вариации результата вариацией факторов, иными словами – на весьма тесную связь факторов с результатом.

Скорректированный коэффициент множественной детерминации $\hat{R}_{yx_1x_2}^2 = 0,9407$ определяет тесноту связи с учетом степеней свободы общей и остаточной дисперсий. Он дает такую оценку тесноты связи, которая не зависит от числа факторов в модели и потому может сравниваться по разным моделям с разным числом факторов. Оба коэффициента указывают на весьма высокую (более 90%) детерминированность результата y в модели факторами x_1 и x_2 .

5. Информация для оценки с помощью частных F -критериев Фишера целесообразности включения в модель фактора x_1 после фактора x_2 и фактора x_2 после фактора x_1 может быть получена в ППП Statgraphics следующим образом:

- 1) введите исходные данные или откройте существующий файл;
- 2) в главном меню последовательно выберите пункты **Relate / Multiple Regression**;
- 3) заполните диалоговое окно ввода данных. В поле *Depended Variable* введите название столбца, содержащего значения зависимой переменной, в поле *Independed Variable* – названия столбцов, содержащих значения факторов, в том порядке, в котором будет проводиться анализ целесообразности включения факторов в модель. Чтобы оценить статистическую значимость включения в модель фактора x_1 после фактора x_2 , сначала введите фактор x_1 , затем x_2 . Для оценки обратного порядка включения факторов в модель x_2 после x_1 введите x_2 , затем x_1 . Щелкните по кнопке **ОК**;
- 4) в окне табличных настроек поставьте флажок напротив поля *Conditional Sums of Squares*.

Результаты вычисления показаны на рис. 2.11.

The screenshot shows a software window with two ANOVA tables. The top table is titled 'Further ANOVA for Variables in the Order Fitted' and shows the following data:

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
x1	107,989	1	107,989	301,30	0,0000
x2	0,718179	1	0,718179	2,00	0,1750
Model	108,707	2			

The bottom table is also titled 'Further ANOVA for Variables in the Order Fitted' and shows the following data:

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
x2	101,61	1	101,61	289,51	0,0000
x1	7,09707	1	7,09707	19,80	0,0004
Model	108,707	2			

Рис. 2.11. Результаты вычисления частных F -критериев Фишера

Частный F -критерий – $F_{\text{частн } x_2}$ показывает статистическую значимость включения фактора x_2 в модель после того, как в нее включен фактор x_1 .

$F_{\text{частн } x_2} = 2$. Вероятность случайной природы его значения (P -значение = 0,1750) составляет 17,5% против принятого уровня значимости $\alpha = 0,05$, (5%). Следовательно, включение в модель фактора x_2 – доля высококвалифицированных рабочих – после того, как в уравнение включен фактор x_1 – коэффициент обновления основных фондов – статистически нецелесообразно: прирост факторной дисперсии за счет дополнительного признака x_2 оказывается незначи-

мым, несущественным; фактор x_2 включать в уравнение после фактора x_1 не следует.

Если поменять первоначальный порядок включения факторов в модель и рассмотреть вариант включения x_1 после x_2 , то результат расчета частного F -критерия для x_1 будет иным. $F_{\text{частн } x_1} = 19,80$. Вероятность его случайного формирования составила 0,04%, это значительно меньше принятого стандарта $\alpha = 0,05$ (5%). Следовательно, значение частного F -критерия для дополнительно включенного фактора x_1 не случайно, является статистически значимым, надежным, достоверным: прирост факторной дисперсии за счет дополнительного фактора x_1 является существенным. Фактор x_1 должен присутствовать в уравнении, в том числе в варианте, когда он дополнительно включается после фактора x_2 .

Общий вывод состоит в том, что множественная модель с факторами x_1 и x_2 с $R_{yx_1x_2}^2 = 0,9469$ содержит неинформативный фактор x_2 . Если исключить фактор x_2 , то можно ограничиться уравнением парной регрессии:

$$\hat{y}_x = \alpha_0 + \alpha_1 \cdot x = 1,99 + 1,23 \cdot x, \quad r_{yx}^2 = 0,9407,$$

более простым, хорошо детерминированным, пригодным для анализа и для прогноза.

6. Средние частные коэффициенты эластичности $\bar{\mathcal{E}}_{yx_j}$ показывают, на сколько процентов от значения своей средней \bar{y} изменяется результат при изменении фактора x_j на 1% от своей средней \bar{x}_j и при фиксированном воздействии на y всех прочих факторов, включенных в уравнение регрессии. Для линейной зависимости

$$\bar{\mathcal{E}}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}},$$

где b_j – коэффициент регрессии при x_j в уравнении множественной регрессии.

Здесь

$$\bar{\mathcal{E}}_{yx_1} = \frac{0,9459 \cdot 6,19}{9,6} = 0,6099\%,$$

$$\bar{\mathcal{E}}_{yx_2} = \frac{0,0856 \cdot 22,3}{9,6} = 0,1989\%.$$

По значениям частных коэффициентов эластичности можно сделать вывод о более сильном влиянии на результат y признака фактора x_1 , чем признака фактора x_2 : 0,6% против 0,2%.

2.4. КОНТРОЛЬНЫЕ ЗАДАНИЯ

Задача 1

По 19 предприятиям оптовой торговли изучается зависимость объема реализации (y) от размера торговой площади (x_1) и товарных запасов (x_2). Были получены следующие варианты уравнений регрессии:

$$\begin{aligned} 1. \quad y &= 25 + 15x_1 & r^2 &= 0,90 ; \\ 2. \quad y &= 42 + 27x_2 & r^2 &= 0,84 ; \\ 3. \quad y &= 30 + \underset{(2,5)}{10} x_1 + \underset{(4,0)}{8} x_2 & R^2 &= 0,92 ; \\ 4. \quad y &= 21 + \underset{(5,0)}{14} x_1 + \underset{(12,0)}{20} x_2 + \underset{(0,2)}{0,6} x_2^2 & R^2 &= 0,95 . \end{aligned}$$

В скобках указаны значения стандартных ошибок для коэффициентов регрессии.

Задание

1. Проанализируйте тесноту связи результата с каждым из факторов.
2. Выберите наилучшее уравнение регрессии, обоснуйте принятое решение.

Задача 2

Для изучения рынка жилья в городе по данным о 46 коттеджах было построено уравнение множественной регрессии:

$$y = 21,1 - \underset{(1,8)}{6,2} x_1 + \underset{(0,54)}{0,95} x_2 + \underset{(0,83)}{3,57} x_3; \quad R^2 = 0,7,$$

где y – цена объекта, тыс. долл.;
 x_1 – расстояние до центра города, км;
 x_2 – полезная площадь объекта, кв. м;
 x_3 – число этажей в доме, ед.;
 R^2 – коэффициент множественной детерминации.

В скобках указаны значения стандартных ошибок для коэффициентов множественной регрессии.

Задание

1. Проверьте гипотезу о том, что коэффициент регрессии b_1 в генеральной совокупности равен нулю.
2. Проверьте гипотезу о том, что коэффициент регрессии b_2 в генеральной совокупности равен нулю.

3. Проверьте гипотезу о том, что коэффициент регрессии b_3 в генеральной совокупности равен нулю.
4. Проверьте гипотезу о том, что коэффициенты регрессии b_1 , b_2 и b_3 в генеральной совокупности одновременно равны нулю (или что коэффициент детерминации равен нулю).
5. Поясните причины расхождения результатов, полученных в пп. 1, 2 и 3, с результатами, полученными в п. 4.

Задача 3

В результате исследования факторов, определяющих экономический рост, по 73 странам получено следующее уравнение регрессии:

$$\hat{G} = 1,4 - 0,52P + 0,17S + 11,16I - 0,38D - 4,75In, \quad R^2 = 0,60,$$

(-5,9)
(4,34)
(3,91)
(-0,79)
(-2,7)

- где \hat{G} – темпы экономического роста (темпы роста среднедушевого ВВП в % к базисному периоду);
- P – реальный среднедушевой ВВП, %;
- S – бюджетный дефицит, % к ВВП;
- I – объем инвестиций, % к ВВП;
- D – внешний долг, % к ВВП;
- In – уровень инфляции, %.

В скобках указаны фактические значения t -критерия для коэффициентов множественной регрессии.

Задание

1. Проверьте гипотезу о достоверности полученной модели в целом.
2. До получения результатов этого исследования ваш однокурсник заключил с вами пари, что эмпирические результаты по данной модели докажут наличие обратной связи между темпами экономического роста и объемом внешнего долга страны (% к ВВП). Выиграл ли это пари ваш однокурсник?

Задача 4

По 20 предприятиям легкой промышленности получена следующая информация, характеризующая зависимость объема выпуска продукции y (млн руб.) от количества отработанных за год человеко-часов x_1 (тыс. чел.-ч.) и среднегодовой стоимости производственного оборудования x_2 (млн руб.):

Уравнение регрессии	$y = 35 + 0,06x_1 + 2,5x_2$
Множественный коэффициент корреляции	0,9
Сумма квадратов отклонений расчетных значений результата от фактических	3000

Задание

1. Определите коэффициент детерминации в этой модели.
2. Составьте таблицу результатов дисперсионного анализа.
3. Проанализируйте полученные результаты регрессионного анализа.

Задача 5

Анализируется зависимость объема производства продукции предприятиями отрасли черной металлургии от затрат труда и расхода чугуна. Для этого по 20 предприятиям собраны следующие данные: y – объем продукции предприятия в среднем за год (млн руб.), x_1 – среднегодовая списочная численность рабочих предприятия (чел.), x_2 – средние затраты чугуна за год (млн т).

Ниже представлены результаты корреляционного анализа этого массива данных.

Матрицы парных коэффициентов корреляции:

для исходных переменных

	y	x_1	x_2
y	1,00		
x_1	0,78	1,00	
x_2	0,86	0,96	1,00

для натуральных логарифмов
исходных переменных

	$\ln y$	$\ln x_1$	$\ln x_2$
$\ln y$	1,00		
$\ln x_1$	0,86	1,00	
$\ln x_2$	0,90	0,69	1,00

Задание

1. Поясните смысл приведенных выше коэффициентов.
2. Используя эту информацию, опишите ваши предположения относительно:
 - а) знаков коэффициентов регрессии в уравнениях парной линейной регрессии y по x_1 ($y = a + b x_1$) и y по x_2 ($y = a + b x_2$);
 - б) статистической значимости коэффициентов регрессии при переменных x_1 и x_2 в линейном уравнении множественной регрессии и в уравнении множественной регрессии в форме функции Кобба – Дугласа.
3. Определите значения коэффициентов детерминации в уравнениях парной линейной регрессии $y = a + b x_1$ и $y = a + b x_2$. Какое из этих уравнений лучше?
4. Определите частные коэффициенты корреляции для линейного уравнения множественной регрессии.
5. Найдите уравнение множественной линейной регрессии в стандартизованном масштабе и сделайте выводы.

Задача 6

По 25 территориям страны изучается влияние климатических условий на урожайность зерновых y (ц/га). Для этого были отобраны две объясняющие переменные:

x_1 – количество осадков в период вегетации (мм);

x_2 – средняя температура воздуха ($^{\circ}\text{C}$).

Матрица парных коэффициентов корреляции этих показателей имеет следующий вид:

	y	x_1	x_2
y	1,0		
x_1	0,6	1,0	
x_2	-0,5	-0,9	1,0

Задание

1. Определите частные коэффициенты корреляции результата с каждым из факторов. Прокомментируйте различие полученных парных и частных коэффициентов корреляции результатов.

2. Исследователь, анализирующий данную зависимость, намерен определить на основе приведенной выше матрицы, какое уравнение регрессии лучше строить:

- а) парную линейную регрессию y на x_1 ;
- б) парную линейную регрессию y на x_2 ;
- в) множественную линейную регрессию.

Как бы вы ответили на эти вопросы?

3. Постройте уравнение регрессии в стандартизованном масштабе и сделайте выводы.

Задача 7

По 30 наблюдениям матрица парных коэффициентов корреляции оказалась следующей:

	y	x_1	x_2	x_3
y	1,00			
x_1	0,30	1,00		
x_2	0,60	0,10	1,00	
x_3	0,40	0,15	0,80	1,00

Задание

1. Постройте уравнение регрессии в стандартизованном виде и сделайте выводы.

2. Определите показатель множественной корреляции (нескорректированный и скорректированный).

3. Оцените целесообразность включения переменной x_1 в модель после введения в нее переменных x_2 и x_3 .

Задача 8

По 20 предприятиям отрасли были получены следующие результаты регрессионного анализа зависимости объема выпуска продукции y (млн руб.) от численности занятых на предприятии x_1 (чел.) и среднегодовой стоимости основных фондов x_2 (млн руб.):

Коэффициент детерминации	0,81		
Множественный коэффициент корреляции	???		
Уравнение регрессии	$\ln y = ??? + 0,48 \ln x_1 + 0,62 \ln x_2$		
Стандартные ошибки параметров	2	0,06	???
t -критерий для параметров	1,5	???	5

Задание

1. Напишите уравнение регрессии, характеризующее зависимость y от x_1 и x_2 .
2. Восстановите пропущенные характеристики.
3. С вероятностью 0,95 постройте доверительные интервалы для коэффициентов регрессии.
4. Проанализируйте результаты регрессионного анализа.

Задача 9

По 30 предприятиям отрасли были получены следующие результаты регрессионного анализа зависимости объема выпуска продукции y (млн руб.) от численности занятых на предприятии x_1 (чел.) и среднегодовой стоимости основных фондов x_2 (млн руб.):

Коэффициент детерминации	???		
Множественный коэффициент корреляции	0,85		
Уравнение регрессии	$y = ??? + 0,48 x_1 + 20 x_2$		
Стандартные ошибки параметров	2	0,06	???
t -критерий для параметров	1,5	???	4

Задание

1. Восстановите пропущенные характеристики.
2. С вероятностью 0,95 постройте доверительные интервалы для коэффициентов регрессии.
3. Проанализируйте результаты регрессионного анализа.

Задача 10

По данным, полученным от 20 фермерских хозяйств одного из регионов, изучается зависимость объема выпуска продукции растениеводства y (млн руб.) от трех факторов: численности работников

L (чел.), количества минеральных удобрений на 1 га посева M (кг) и количества осадков в период вегетации – R (г). Были получены следующие варианты уравнений регрессии и доверительные интервалы для коэффициентов регрессий (табл. 2.6 и 2.7):

$$1) \hat{y} = -5 + 0,8L + 1,2M, \quad R^2 = 0,75.$$

Таблица 2.6

Граница	Доверительные интервалы для коэффициентов регрессии при факторе	
	L	M
Нижняя	0,4	???
Верхняя	???	1,4
Примечание. Доверительные интервалы построены с вероятностью $P = 0,95$.		

$$2) \hat{y} = 2 + 0,5L + 1,7M - 2R, \quad R^2 = 0,77.$$

Таблица 2.7

Граница	Доверительные интервалы для коэффициентов регрессии при факторе		
	L	M	R
Нижняя	0,1	???	???
Верхняя	???	2,3	1,5
Примечание. Доверительные интервалы построены с вероятностью $P = 0,95$.			

Задание

1. Восстановите пропущенные границы доверительных интервалов в каждом уравнении.
2. Выберите наилучшее уравнение регрессии. Дайте интерпретацию их параметров и доверительных интервалов для коэффициентов регрессии.
3. Каковы ваши предложения относительно значения t -критерия Стьюдента для коэффициента регрессии при факторе R во 2-м уравнении?

Задача 11

По данным, полученным от 20 фермерских хозяйств одного из регионов, изучается зависимость объема выпуска продукции растениеводства y (млн руб.) от четырех факторов: численности работников L (чел.), количества минеральных удобрений на 1 га посева M (кг), количества осадков в период вегетации R (г) и качества почвы Q (баллов). Были получены следующие варианты уравнений регрессии и доверительные интервалы коэффициентов регрессий (табл. 2.8 и 2.9):

$$1) \hat{y} = 2 + 0,5L + 1,7M - 2R, \quad R^2 = 0,77.$$

Таблица 2.8

Граница	Доверительные интервалы для коэффициентов регрессии при факторе		
	L	M	R
Нижняя	0,1	???	???
Верхняя	???	2,3	1,5

Примечание. Доверительные интервалы построены с вероятностью $P = 0,95$.

$$2) \hat{y} = 6,4 + 0,7L + 1,5M - 2R + 0,8Q, \quad R^2 = 0,81.$$

Таблица 2.9

Граница	Доверительные интервалы для коэффициентов регрессии при факторе			
	L	M	R	Q
Нижняя	0,3	-0,2	???	0,4
Верхняя	???	???	-1,2	1,2

Примечание. Доверительные интервалы построены с вероятностью $P = 0,95$.

Задание

1. Восстановите пропущенные границы доверительных интервалов.
2. Выберите наилучшее уравнение регрессии. Дайте интерпретацию его параметров и доверительных интервалов для коэффициентов регрессии на примере одного из факторных признаков.
3. Оцените целесообразность включения в модель $y = f(L, M, R)$ фактора Q .

Задача 12

Производственная функция, полученная по данным за 1990 – 1997 гг., характеризуется уравнением

$$\hat{\lg P} = 0,552 + 0,276 \lg Z + 0,521 \lg K$$

(0,584) (0,065)

$$R^2 = 0,9843, \quad r_{PZ}^2 = 0,7826, \quad r_{PK}^2 = 0,9836,$$

где P – индекс промышленного производства;
 Z – численность рабочих;
 K – капитал.

В скобках указаны значения стандартных ошибок для коэффициентов регрессии.

Задание

1. Дайте интерпретацию параметров уравнения регрессии.
2. Оцените значимость параметров регрессии с помощью t -критерия Стьюдента и сделайте соответствующие выводы о целесообразности включения факторов в модель.
3. Оцените значимость уравнения регрессии в целом с помощью F -критерия Фишера.
4. Найдите величины частных значений F -критерия и сделайте соответствующие выводы.
5. Какова роль факторов, не учтенных в модели, в вариации индекса промышленного производства.

Задача 13

По 30 наблюдениям получены следующие данные:

Уравнение регрессии	$\hat{y} = a + 0,176 x_1 + 0,014 x_2 - 7,75 x_3$
Коэффициент детерминации	0,65
\bar{y}	200
\bar{x}_1	150
\bar{x}_2	20
\bar{x}_3	100

Задание

1. Найдите скорректированный коэффициент корреляции, оцените значимость уравнения регрессии в целом.
2. Определите частные коэффициенты эластичности.
3. Оцените параметр a .

Задача 14

Зависимость потребления электроэнергии y (тыс. кВт · ч) от объемов производства продукции $A - x_1$ (тыс. ед.) и продукции $B - x_2$ (тыс. ед.) характеризуется следующим образом:

Уравнение регрессии в стандартизованном виде	$\hat{t}_y = 0,79 t_{x_1} + 0,56 t_{x_2}$
Коэффициент детерминации	0,95
Коэффициент вариации y , V_y	27%
Коэффициент вариации x_1 , V_{x_1}	45%
Коэффициент вариации x_2 , V_{x_2}	40%

Задание

1. Сделайте выводы о силе влияния факторов на результат.
2. Учитывая значения коэффициентов вариации рассматриваемых признаков, определите частные коэффициенты эластичности, сделайте по ним выводы.
3. Оцените значимость уравнения регрессии, учитывая, что оно построено по 30 наблюдениям.

Задача 15

Имеется информация по 25 наблюдениям (табл. 2.10).

Таблица 2.10

Признак	Среднее значение	Коэффициент вариации, %	Уравнение регрессии
y	35	20	$\hat{y} = 20 + x_1 - 2,0x_2$
x_1	16	30	$\hat{y} = 9 + 1,1x_1$
x_2	8	10	$\hat{y} = 4 - 4,1x_2$

Задание

1. Оцените значимость каждого уравнения регрессии, если известно, что $r_{x_1x_2} = -0,35$.
2. Оцените значимость коэффициентов регрессии уравнения с двумя объясняющими переменными.
3. Определите показатели частной корреляции.
4. Найдите частные коэффициенты эластичности.

Задача 16

Имеется информация по 22 наблюдениям (табл. 2.11).

Таблица 2.11

Признак	Среднее значение	Коэффициент вариации, %	Уравнение регрессии
y	23	20	$\hat{y} = 19 - 2,0x_1 - 0,5x_2$
x_1	6	40	$\hat{y} = 9 - 1,0x_1$
x_2	8	10	$\hat{y} = 4 + 0,6x_2$

Задание

1. Оцените значимость каждого уравнения регрессии, если известно, что $r_{x_1x_2} = -0,5$.
2. Оцените значимость коэффициентов регрессии уравнения с двумя факторами.
3. Найдите скорректированный коэффициент множественной корреляции.
4. Определите показатели частной корреляции.

Задача 17

По совокупности 30 предприятий концерна изучается зависимость прибыли y (тыс. руб.) от выработки продукции на одного работника x_1 (ед.) и индекса цен на продукцию x_2 (%). Данные приведены в табл. 2.12.

Таблица 2.12

Признак	Среднее значение	Среднее квадратическое отклонение	Парный коэффициент корреляции
y	250	38	$r_{yx_1} = 0,68$
x_1	47	12	$r_{yx_2} = 0,63$
x_2	112	21	$r_{x_1x_2} = 0,42$

Задание

1. Постройте линейные уравнения парной регрессии, оцените их значимость с помощью F -критерия Фишера.
2. Найдите уравнение множественной регрессии в стандартизованном и натуральном масштабе.
3. Рассчитайте множественный коэффициент корреляции, общий и частные критерии Фишера и сделайте выводы.

Задача 18

По 30 заводам, выпускающим продукцию А, изучается зависимость потребления электроэнергии y (тыс. кВт · ч) от производства продукции – x_1 (тыс. ед.) и уровня механизации труда – x_2 (%). Данные приведены в табл. 2.13.

Таблица 2.13

Признак	Среднее значение	Среднее квадратическое отклонение	Парный коэффициент корреляции
y	1000	27	$r_{yx_1} = 0,77$
x_1	420	45	$r_{yx_2} = 0,43$
x_2	41,5	18	$r_{x_1x_2} = 0,38$

Задание

1. Постройте уравнение множественной регрессии в стандартизованном и натуральном масштабе.
2. Определите показатели частной и множественной корреляции.
3. Найдите частные коэффициенты эластичности и сравните их с β -коэффициентами.
4. Рассчитайте общий и частные F -критерии Фишера.

Задача 19

Изучается зависимость по 25 предприятиям концерна потребления материалов y (т) от энерговооруженности труда x_1 (кВт · ч на одного рабочего) и объема произведенной продукции x_2 (тыс. ед.). Данные приведены в табл. 2.14.

Таблица 2.14

Признак	Среднее значение	Среднее квадратическое отклонение	Парный коэффициент корреляции
y	12,0	2,0	$r_{yx_1} = 0,52$
x_1	4,3	0,5	$r_{yx_2} = 0,84$
x_2	10,0	1,8	$r_{x_1x_2} = 0,43$

Задание

1. Постройте уравнение множественной регрессии и поясните экономический смысл его параметров.
2. Определите частные коэффициенты эластичности и стандартизованные коэффициенты регрессии.
3. Найдите частные и множественный коэффициенты корреляции.
4. Оцените значимость уравнения регрессии с помощью F -критерия Фишера.

Задача 20

По 50 семьям изучалось потребление мяса – y (кг на душу населения) от дохода – x_1 (руб. на одного члена семьи) и от потребления рыбы – x_2 (кг на душу населения). Результаты оказались следующими:

Уравнение регрессии	$\hat{y} = -180 + 0,2x_1 - 0,4x_2$
Стандартные ошибки параметров	20 0,01 0,25
Множественный коэффициент корреляции	0,85

Задание

1. Используя t -критерий Стьюдента, оцените значимость параметров уравнения.
2. Рассчитайте F -критерий Фишера.
3. Оцените по частным F -критериям Фишера целесообразность включения в модель:
 - а) фактора x_1 после фактора x_2 ;
 - б) фактора x_2 после фактора x_1 .

Задача 21

По 40 предприятиям одной отрасли исследовалась зависимость производительности труда – y от уровня квалификации рабочих – x_1 и энерговооруженности их труда – x_2 . Результаты оказались следующими:

Уравнение регрессии	$\hat{y} = a + 10x_1 + 2x_2$		
Стандартные ошибки параметров	0,5	2	?
t -критерий для параметров	3	?	5
Множественный коэффициент корреляции	0,85		

Задание

1. Определите параметр a и заполните пропущенные значения.
2. Оцените значимость уравнения в целом, используя значение множественного коэффициента корреляции.
3. Какой из факторов оказывает более сильное воздействие на результат?

Задача 22

Изучается влияние стоимости основных и оборотных средств на величину валового дохода торговых предприятий. Для этого по 12 торговым предприятиям были получены данные, приведенные в табл. 2.15.

Таблица 2.15

Номер предприятия	Валовой доход за год, млн руб.	Среднегодовая стоимость, млн руб.	
		основных фондов	оборотных средств
1	203	118	105
2	63	28	56
3	45	17	54
4	113	50	63
5	121	56	28
6	88	102	50
7	110	116	54
8	56	124	42
9	80	114	36
10	237	154	106
11	160	115	88
12	75	98	46

Задание

1. Постройте линейное уравнение множественной регрессии и поясните экономический смысл его параметров.
2. Рассчитайте частные коэффициенты эластичности.
3. Определите стандартизованные коэффициенты регрессии.
4. Сделайте вывод о силе связи результата и факторов.
5. Определите парные и частные коэффициенты корреляции, а также множественный коэффициент корреляции; сделайте выводы.
6. Дайте оценку полученного уравнения на основе коэффициента детерминации и общего F -критерия Фишера.

Задача 23

Имеются данные о деятельности крупнейших компаний США в 1996 г. (табл. 2.16).

Таблица 2.16

№ п/п	Чистый доход, млрд долл. США, y	Оборот капитала, млрд долл. США, x_1	Использованный капитал, млрд долл. США, x_2	Численность служащих, тыс. чел., x_3	Рыночная капитализация компании, млрд долл. США, x_4
1	0,9	31,3	18,9	43,0	40,9
2	1,7	13,4	13,7	64,7	40,5
3	0,7	4,5	18,5	24,0	38,9
4	1,7	10,0	4,8	50,2	38,5
5	2,6	20,0	21,8	106,0	37,3
6	1,3	15,0	5,8	96,6	26,5
7	4,1	137,1	99,0	347,0	37,0
8	1,6	17,9	20,1	85,6	36,8
9	6,9	165,4	60,6	745,0	36,3
10	0,4	2,0	1,4	4,1	35,3
11	1,3	6,8	8,0	26,8	35,3
12	1,9	27,1	18,9	42,7	35,0
13	1,9	13,4	13,2	61,8	26,2
14	1,4	9,8	12,6	212,0	33,1
15	0,4	19,5	12,2	105,0	32,7
16	0,8	6,8	3,2	33,5	32,1
17	1,8	27,0	13,0	142,0	30,5
18	0,9	12,4	6,9	96,0	29,8
19	1,1	17,7	15,0	140,0	25,4
20	1,9	12,7	11,9	59,3	29,3
21	-0,9	21,4	1,6	131,0	29,2
22	1,3	13,5	8,6	70,7	29,2
23	2,0	13,4	11,5	65,4	29,1
24	0,6	4,2	1,9	23,1	27,9
25	0,7	15,5	5,8	80,8	27,2

Задание

1. Рассчитайте параметры линейного уравнения множественной регрессии с полным перечнем факторов.
2. Дайте сравнительную оценку силы связи факторов с результатом с помощью средних (общих) коэффициентов эластичности.
3. Оцените статистическую значимость параметров регрессионной модели с помощью t -критерия; нулевую гипотезу о значимости уравнения и показателей тесноты связи проверьте с помощью F -критерия.
4. Оцените качество уравнения через среднюю ошибку аппроксимации.
5. Рассчитайте матрицы парных и частных коэффициентов корреляции и на их основе и по t -критерию для коэффициентов регрессии отберите информативные факторы в модель. Постройте модель только с информативными факторами и оцените ее параметры.
6. Рассчитайте прогнозное значение результата, если прогнозные значения факторов составляют 80% от их максимальных значений.
7. Рассчитайте ошибки и доверительный интервал прогноза для уровня значимости 5 или 10% ($\alpha = 0,05$; $\alpha = 0,10$).
8. Оцените полученные результаты, выводы оформите в аналитической записке.

Задача 24

Имеются данные о деятельности крупнейших компаний США в 1996 г. (табл. 2.17).

Таблица 2.17

№ п/п	Чистый доход, млрд долл. США, y	Оборот капитала, млрд долл. США, x_1	Использованный капитал, млрд долл. США, x_2	Численность служащих, тыс. чел., x_3
1	6,6	6,9	83,6	222,0
2	3,0	18,0	6,5	32,0
3	6,5	107,9	50,4	82,0
4	3,3	16,7	15,4	45,2
5	0,1	79,6	29,6	299,3
6	3,6	16,2	13,3	41,6
7	1,5	5,9	5,9	17,8
8	5,5	53,1	27,1	151,0
9	2,4	18,8	11,2	82,3
10	3,0	35,3	16,4	103,0
11	4,2	71,9	32,5	225,4
12	2,7	93,6	25,4	675,0
13	1,6	10,0	6,4	43,8
14	2,4	31,5	12,5	102,3
15	3,3	36,7	14,3	105,0
16	1,8	13,8	6,5	49,1
17	2,4	64,8	22,7	50,4
18	1,6	30,4	15,8	480,0
19	1,4	12,1	9,3	71,0
20	0,9	31,3	18,9	43,0

Задание

1. Рассчитайте параметры линейного уравнения множественной регрессии с полным перечнем факторов.
2. Дайте сравнительную оценку силы связи факторов с результатом с помощью средних (общих) коэффициентов эластичности.
3. Оцените статистическую значимость параметров регрессионной модели с помощью t -критерия; нулевую гипотезу о значимости уравнения и показателей тесноты связи проверьте с помощью F -критерия.
4. Оцените качество уравнения через среднюю ошибку аппроксимации.
5. Рассчитайте матрицы парных и частных коэффициентов корреляции и на их основе и по t -критерию для коэффициентов регрессии отберите информативные факторы в модель. Постройте модель только с информативными факторами и оцените ее параметры.
6. Рассчитайте прогнозное значение результата, если прогнозные значения факторов составляют 80% от их максимальных значений.
7. Рассчитайте ошибки и доверительный интервал прогноза для уровня значимости 5 или 10% ($\alpha = 0,05$; $\alpha = 0,10$).
8. Оцените полученные результаты, выводы оформите в аналитической записке.

Задача 25

В табл. 2.18 представлены данные о рынке строящегося жилья в Санкт-Петербурге (по состоянию на декабрь 1996 г.).

Таблица 2.18

№ п/п	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y
1	1	1	39,0	20,0	8,2	0	1	0	15,9
2	3	1	68,4	40,5	10,7	0	1	0	27,0
3	1	1	34,8	16,0	10,7	0	1	12	13,5
4	1	1	39,0	20,0	8,5	0	1	12	15,1
5	2	1	54,7	28,0	10,7	0	1	12	21,1
6	3	1	74,7	46,3	10,7	0	1	12	28,7
7	3	1	71,7	45,9	10,7	0	0	0	27,2
8	3	1	74,5	47,5	10,4	0	0	0	28,3
9	4	1	137,7	87,2	14,6	0	1	0	52,3
10	1	1	40,0	17,7	11,0	1	1	8	22,0
11	2	1	53,0	31,1	10,0	1	1	8	28,0
12	3	1	86,0	48,7	14,0	1	1	8	45,0
13	4	1	98,0	65,8	13,0	1	1	8	51,0

Продолжение

№ п/п	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y
14	2	1	62,6	21,4	11,0	1	1	0	34,4
15	1	1	45,3	20,6	10,4	1	1	8	24,7
16	2	1	56,4	29,7	9,4	1	1	8	30,8
17	1	1	37,0	17,8	8,3	0	1	0	15,9
18	3	1	67,5	43,5	8,3	0	1	0	29,0
19	1	1	37,0	17,8	8,3	0	1	3	15,4
20	3	1	69,0	42,4	8,3	0	1	3	28,6
21	1	1	40,0	20,0	8,3	0	0	0	15,6
22	3	1	69,1	41,3	8,3	0	1	0	27,7
23	2	1	68,1	35,4	13,0	1	1	20	34,1
24	2	1	75,3	41,4	12,1	1	1	20	37,7
25	3	1	83,7	48,5	12,1	1	1	20	41,9
26	1	1	48,7	22,3	12,4	1	1	20	24,4
27	1	1	39,9	18,0	8,1	1	0	0	21,3
28	2	1	68,6	35,5	17,0	1	1	12	36,7
29	1	1	39,0	20,0	9,2	1	0	0	21,5
30	2	1	48,6	31,0	8,0	1	0	0	26,4
31	3	1	98,0	56,0	22,0	1	0	0	53,9
32	2	1	68,5	30,7	8,3	1	1	6	34,2
33	2	1	71,1	36,2	13,3	1	1	6	35,6
34	3	1	68,0	41,0	8,0	1	1	12	34,0
35	1	1	38,0	19,0	7,4	1	1	12	19,0
36	2	1	93,2	49,5	14,0	1	1	12	46,6
37	3	1	117,0	55,2	25,0	1	1	12	58,5
38	1	2	42,0	21,0	10,2	1	0	12	24,2
39	2	2	62,0	35,0	11,0	1	0	12	35,7
40	3	2	89,0	52,3	11,5	1	1	12	51,2
41	4	2	132,0	89,6	11,0	1	1	12	75,9
42	1	2	40,8	19,2	10,1	1	1	6	21,2
43	2	2	59,2	31,9	11,2	1	1	6	30,8
44	3	2	65,4	38,9	9,3	1	1	6	34,0
45	2	2	60,2	36,3	10,9	1	1	12	31,9
46	3	2	82,2	49,7	13,8	1	1	12	43,6
47	3	2	98,4	52,3	15,3	1	1	12	52,2

№ п/п	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y
48	3	3	76,7	44,7	8,0	1	1	0	43,1
49	1	3	38,7	20,0	10,2	1	1	6	25,0
50	2	3	56,4	32,7	10,1	1	1	6	35,2
51	3	3	76,7	44,7	8,0	1	1	6	40,8
52	1	3	38,7	20,0	10,2	1	0	0	18,2
53	1	3	41,5	20,0	10,2	1	1	0	20,1
54	2	3	48,8	28,5	8,0	1	0	0	22,7
55	2	3	57,4	33,5	10,1	1	1	0	27,6
56	3	3	76,7	44,7	8,0	1	1	0	36,0
57	1	4	37,0	17,5	8,3	0	1	7	17,8
58	2	4	54,0	30,5	8,3	0	1	7	25,9
59	3	4	68,0	42,5	8,3	0	1	7	32,6
60	1	4	40,5	16,0	11,0	0	1	3	19,8
61	2	4	61,0	31,0	11,0	0	1	3	29,9
62	3	4	80,0	45,6	11,0	0	1	3	39,2
63	1	3	52,0	21,2	11,2	1	1	18	22,4
64	2	3	78,1	40,0	11,6	1	1	18	35,2
65	3	3	91,6	53,8	16,0	1	0	18	41,2
66	1	4	39,9	19,3	8,4	0	1	6	17,8
67	2	4	56,2	31,4	11,1	0	1	6	25,0
68	3	4	79,1	42,4	15,5	0	1	6	35,2
69	4	4	91,6	55,2	9,4	0	1	6	40,8

Принятые в таблице обозначения:

y – цена квартиры, тыс. долл.;

x_1 – число комнат в квартире;

x_2 – район города (1 – Приморский, Шувалово – Озерки, 2 – Гражданка, 3 – Юго-Запад, 4 – Красносельский);

x_3 – общая площадь квартиры (m^2);

x_4 – жилая площадь квартиры (m^2);

x_5 – площадь кухни (m^2);

x_6 – тип дома (1 – кирпичный, 0 – другой);

x_7 – наличие балкона (1 – есть, 0 – нет);

x_8 – число месяцев до окончания срока строительства.

Задание

1. Определите факторы, формировавшие цену квартир в строящихся домах в Санкт-Петербурге в 1996 г. Сгенерируйте фиктивную переменную z , отражающую местоположение квартиры и позволяющую

разделить всю совокупность квартир на две группы: квартиры на севере города (Приморский район, Шувалово-Озерки, Гражданка) и на юге города (Юго-Запад, Красносельский район).

2. Составьте матрицу парных коэффициентов корреляции:

а) исходных переменных;

б) логарифмов исходных переменных (кроме фиктивных переменных). Вместо переменной x_2 используйте фиктивную переменную z .

3. Постройте уравнение регрессии, характеризующее зависимость цены от всех факторов, в линейной и степенной форме. Установите, какие факторы мультиколлинеарны. В какой модели мультиколлинеарность проявляется сильнее?

4. Постройте модель $y = f(x_3, x_6, x_7, x_8, z)$ в линейной и степенной форме. Какие факторы значимо воздействуют на формирование цены квартиры в этой модели?

5. Существует ли разница в ценах квартир, расположенных в северной и южной частях Санкт-Петербурга? Является ли наличие балкона или лоджии преимуществом квартиры на рынке? Как вы объясните этот факт?

Задача 26

По данным, представленным в табл. 2.19, изучается зависимость индекса человеческого развития y от переменных:

x_1 – ВВП 1997 г., % к 1990 г.;

x_2 – расходы на конечное потребление в текущих ценах, % к ВВП;

x_3 – расходы домашних хозяйств, % к ВВП;

x_4 – валовое накопление, % к ВВП;

x_5 – суточная калорийность питания населения, ккал на душу населения;

x_6 – ожидаемая продолжительность жизни при рождении 1997 г. число лет.

Таблица 2.19

Страна	y	x_1	x_2	x_3	x_4	x_5	x_6
Австрия	0,904	115,0	75,5	56,1	25,2	3343	77,0
Австралия	0,922	123,0	78,5	61,8	21,8	3001	78,2
Белоруссия	0,763	74,0	78,4	59,1	25,7	3101	68,0
Бельгия	0,923	111,0	77,7	63,3	17,8	3543	77,2
Великобритания	0,918	113,0	84,4	64,1	15,9	3237	77,2
Германия	0,906	110,0	75,9	57,0	22,4	3330	77,2
Дания	0,905	119,0	76,0	50,7	20,6	3808	75,7
Индия	0,545	146,0	67,5	57,1	25,2	2415	62,6
Испания	0,894	113,0	78,2	62,0	20,7	3295	78,0

Продолжение

Страна	y	x_1	x_2	x_3	x_4	x_5	x_6
Италия	0,900	108,0	78,1	61,8	17,5	3504	78,2
Канада	0,932	113,0	78,6	58,6	19,7	3056	79,0
Казахстан	0,740	71,0	84,0	71,7	18,5	3007	67,6
Китай	0,701	210,0	59,2	48,0	42,4	2844	69,8
Латвия	0,744	94,0	90,2	63,9	23,0	2861	68,4
Нидерланды	0,921	118,0	72,8	59,1	20,2	3259	77,9
Норвегия	0,927	130,0	67,7	47,5	25,2	3350	78,1
Польша	0,802	127,0	82,6	65,3	22,4	3344	72,5
Россия	0,747	61,0	74,4	53,2	22,7	2704	66,6
США	0,927	117,0	83,3	67,9	18,1	3642	76,7
Украина	0,721	46,0	83,7	61,7	20,1	2753	68,8
Финляндия	0,913	107,0	73,8	52,9	17,3	2916	76,8
Франция	0,918	110,0	79,2	59,9	16,8	3551	78,1
Чехия	0,833	99,2	71,5	51,5	29,9	3177	73,9
Швейцария	0,914	101,0	75,3	61,2	20,3	3280	78,6
Швеция	0,923	105,0	79,0	53,1	14,1	3160	78,5

Задание

1. Постройте матрицу парных коэффициентов корреляции. Рассчитайте коэффициенты множественной детерминации, используя в качестве зависимой переменной каждый фактор. Установите, какие факторы мультиколлинеарны.
2. Постройте уравнение множественной регрессии в линейной форме с полным набором факторов.
3. Оцените статистическую значимость уравнения регрессии и его параметров с помощью критериев Фишера и Стьюдента.
4. Отберите информативные факторы по пп. 1 и 3. Постройте уравнение регрессии со статистически значимыми факторами.

Задача 27

Имеются данные по странам за 1997 г. (табл. 2.20).

Таблица 2.20

Страна	Индекс человеческого развития, y	Ожидаемая продолжительность жизни при рождении в 1997 г., лет, x_1	Суточная калорийность питания населения, ккал на душу, x_2
Австрия	0,904	77,0	3343
Австралия	0,922	78,2	3001
Аргентина	0,827	72,9	3136
Белоруссия	0,763	68,0	3101

Продолжение

Страна	Индекс человеческого развития, у	Ожидаемая продолжительность жизни при рождении в 1997 г., лет, x_1	Суточная калорийность питания населения, ккал на душу, x_2
Бельгия	0,923	77,2	3543
Бразилия	0,739	66,8	2938
Великобритания	0,918	77,2	3237
Венгрия	0,795	70,9	3402
Германия	0,906	77,2	3330
Греция	0,867	78,1	3575
Дания	0,905	75,7	3808
Египет	0,616	66,3	3289
Израиль	0,883	77,8	3272
Индия	0,545	62,6	2415
Испания	0,894	78,0	3295
Италия	0,900	78,2	3504
Канада	0,932	79,0	3056
Казахстан	0,740	67,7	3007
Китай	0,701	69,8	2844
Латвия	0,744	68,4	2861
Нидерланды	0,921	77,9	3259
Норвегия	0,927	78,1	3350
Польша	0,802	72,5	3344
Республика Корея	0,852	72,4	3336
Россия	0,747	66,6	2704
Румыния	0,752	69,9	2943
США	0,927	76,6	3642
Турция	0,728	69,0	3568
Украина	0,721	68,8	2753
Финляндия	0,913	76,8	2916
Франция	0,918	78,1	3551
Чехия	0,833	73,9	3177
Швейцария	0,914	78,6	3280
Швеция	0,923	78,5	3160
ЮАР	0,695	64,1	2933
Япония	0,924	80,0	2905

Задание

1. Постройте матрицу парных коэффициентов корреляции.
2. Постройте парные уравнения регрессии.
3. Оцените статистическую значимость уравнений и их параметров с помощью критериев Фишера и Стьюдента.
4. Постройте уравнение множественной регрессии.
5. Постройте графики остатков. Сделайте выводы.
6. Проведите тестирование ошибок уравнения множественной регрессии на гетероскедастичность, применив тест Гельфельда–Квандта.
7. Оцените статистическую значимость уравнения множественной регрессии. Определите, какое уравнение лучше использовать для прогноза:
 - парную регрессию y на x_1 ;
 - парную регрессию y на x_2 ;
 - множественную регрессию.

Задача 28

Изучается зависимость средней ожидаемой продолжительности жизни от нескольких факторов по данным за 1995 г., представленным в табл. 2.21.

Таблица 2.21

Страна	y	x_1	x_2	x_3	x_4
Мозамбик	47	3,0	2,6	2,4	113
Бурунди	49	2,3	2,6	2,7	98
Чад	48	2,6	2,5	2,5	117
Непал	55	4,3	2,5	2,4	91
Буркина-Фасо	49	2,9	2,8	2,1	99
Мадагаскар	52	2,4	3,1	3,1	89
Бангладеш	58	5,1	1,6	2,1	79
Гаити	57	3,4	2,0	1,7	72
Мали	50	2,0	2,9	2,7	123
Нигерия	53	4,5	2,9	2,8	80
Кения	58	5,1	2,7	2,7	58
Того	56	4,2	3,0	2,8	88
Индия	62	5,2	1,8	2,0	68
Бенин	50	6,5	2,9	2,5	95

Продолжение

Страна	у	x_1	x_2	x_3	x_4
Никарагуа	68	7,4	3,1	4,0	46
Гана	59	7,4	2,8	2,7	73
Ангола	47	4,9	3,1	2,8	124
Пакистан	60	8,3	2,9	3,3	90
Мавритания	51	5,7	2,5	2,7	96
Зимбабве	57	7,5	2,4	2,2	55
Гондурас	67	7,0	3,0	3,8	45
Китай	69	10,8	1,1	1,1	34
Камерун	57	7,8	2,9	3,1	56
Конго	51	7,6	2,9	2,6	90
Шри-Ланка	72	12,1	1,3	2,0	16
Египет	63	14,2	2,0	2,7	56
Индонезия	64	14,1	1,6	2,5	51
Филиппины	66	10,6	2,2	2,7	39
Марокко	65	12,4	2,0	2,6	55
Папуа – Новая Гвинея	57	9,0	2,3	2,3	64
Гватемала	66	12,4	2,9	3,5	44
Эквадор	69	15,6	2,2	3,2	36
Доминиканская Республика	71	14,3	1,9	2,6	37
Ямайка	74	13,1	1,0	1,8	13
Алжир	70	19,6	2,2	4,1	34
Республика Эль- Сальвадор	67	9,7	2,2	3,4	36
Парагвай	68	13,5	2,7	2,9	41
Тунис	69	18,5	1,9	3,0	39
Белоруссия	70	15,6	0,2	0,2	13
Перу	66	14,0	2,0	3,1	47
Таиланд	69	28,0	0,9	1,3	35
Панама	73	22,2	1,7	2,4	23
Турция	67	20,7	1,7	2,1	48
Польша	70	20,0	0,3	0,6	14
Словакия	72	13,4	0,3	0,7	11
Венесуэла	71	29,3	2,3	3,0	23
ЮАР	64	18,6	2,2	2,4	50
Мексика	72	23,7	1,9	2,8	33

Страна	y	x_1	x_2	x_3	x_4
Мавритания	71	49,0	1,3	1,8	16
Бразилия	67	20,0	1,5	1,6	44
Тринидад	72	31,9	0,8	1,8	13
Малайзия	71	33,4	2,4	2,7	12
Чили	72	35,3	1,5	2,1	12
Уругвай	73	24,6	0,6	1,0	18
Аргентина	73	30,8	1,3	2,0	22
Греция	78	43,4	0,6	0,9	8
Республика Корея	72	42,4	0,9	1,9	10
Испания	77	53,8	0,2	1,0	7
Нов. Зеландия	76	60,6	1,4	1,5	7
Ирландия	77	58,1	0,5	1,7	6
Израиль	77	61,1	3,5	3,5	8
Австралия	77	70,2	1,1	1,4	6
Италия	78	73,7	0,2	0,4	7
Канада	78	78,3	1,3	1,0	6
Финляндия	76	65,8	0,5	0,1	5
Гонконг	79	85,1	1,6	1,3	5
Швеция	79	68,7	0,6	0,3	4
Нидерланды	78	73,9	0,7	0,6	6
Бельгия	77	80,3	0,4	0,5	8
Франция	78	78,0	0,5	0,8	6
Сингапур	76	84,4	2,0	1,7	4
Австрия	77	78,8	0,8	0,5	6
США	77	100,0	1,0	1,1	8
Дания	75	78,7	0,3	0,1	6
Япония	80	82,0	0,3	0,6	4
Швейцария	78	95,9	1,0	0,8	6

Принятые в таблице обозначения:

y – средняя ожидаемая продолжительность жизни при рождении, лет;

x_1 – ВВП в паритетах покупательной способности;

x_2 – темпы прироста населения по сравнению с предыдущим годом, %;

x_3 – темпы прироста рабочей силы по сравнению с предыдущим годом, %;

x_4 – коэффициент младенческой смертности, ‰.

Задание

1. Вычислите описательные статистики. Проверьте характер распределения признаков. При необходимости удалите аномальные наблюдения.
2. Постройте матрицу парных коэффициентов корреляции. Установите, какие факторы коллинеарны.
3. Постройте уравнение множественной регрессии, обосновав отбор факторов.
4. Постройте графики остатков. Сделайте выводы.
5. Проведите тестирование ошибок уравнения множественной регрессии на гетероскедастичность, применив тест Гельфельда–Квандта.
6. Оцените статистическую значимость уравнения множественной регрессии. Какие факторы значимо воздействуют на формирование средней ожидаемой продолжительности жизни в этом уравнении?
7. Постройте уравнение множественной регрессии со статистически значимыми факторами.

Задача 29

Имеются данные о продаже квартир на вторичном рынке жилья в Санкт-Петербурге на 01.05.2000 г. (табл. 2.22).

Таблица 2.22

№ п/п	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	13,0	1	1	37,0	21,5	6,5	0	20
2	16,5	1	1	60,0	27,0	22,4	0	10
3	17,0	1	1	60,0	30,0	15,0	0	10
4	15,0	1	1	53,0	26,2	13,0	0	15
5	14,2	1	1	35,0	19,0	9,0	0	8
6	10,5	1	1	30,3	17,5	5,6	1	15
7	23,0	1	1	43,0	25,5	8,5	0	5
8	12,0	1	1	30,0	17,8	5,5	1	10
9	15,6	1	1	35,0	18,0	5,3	1	3
10	12,5	1	1	32,0	17,0	6,0	1	5
11	11,3	1	0	31,0	18,0	5,5	1	10
12	13,0	1	0	33,0	19,6	7,0	0	5
13	21,0	1	0	53,0	26,0	16,0	1	5
14	12,0	1	0	32,2	18,0	6,3	0	20
15	11,0	1	0	31,0	17,3	5,5	1	15
16	11,0	1	0	36,0	19,0	8,0	1	5

№ п/п	y	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇
17	22,5	2	1	48,0	29,0	8,0	1	15
18	26,0	2	1	55,5	35,0	8,0	0	10
19	18,5	2	1	48,0	28,0	8,0	0	10
20	13,2	2	1	44,1	30,0	6,0	1	25
21	25,8	2	1	80,0	51,0	13,0	0	10
22	17,0	2	1	60,0	38,0	10,0	0	12
23	18,0	2	0	50,0	30,0	8,7	1	15
24	21,0	2	0	54,6	32,0	10,0*	1	20
25	14,5	2	0	43,0	27,0	5,5	1	10
26	23,0	2	0	66,0	39,0	12,0	1	5
27	19,5	2	0	53,5	29,5	7,0	1	15
28	14,2	2	0	45,0	29,0	6,0	1	12
29	13,3	2	0	45,0	30,0	5,5	0	5
30	16,1	2	0	50,6	30,8	7,9	0	10
31	13,5	2	0	42,5	28,0	5,2	1	25
32	16,0	2	0	50,1	31,0	6,0	0	10
33	15,5	3	1	68,1	44,4	7,2	0	5
34	38,0	3	1	107,0	58,0	24,0	0	15
35	30,0	3	1	100,0	58,0	20,0	0	15
36	24,0	3	1	71,0	52,0	7,5	1	15
37	32,5	3	1	98,0	51,0	15,0	0	10
38	43,0	3	0	100,0	45,0	35,0	1	25
39	17,8	3	0	58,0	39,0	6,2	0	10
40	28,0	3	0	75,0	40,0	18,0	1	3
41	32,7	3	0	85,0	59,0	9,0	0	5
42	31,0	3	0	66,0	48,0	6,0	0	2
43	33,0	3	0	81,0	52,0	12,0	0	10
44	28,0	3	0	76,4	49,0	10,0	0	5
45	21,5	3	0	55,0	40,5	6,0	1	15
46	15,3	3	0	53,7	37,6	5,5	1	3
47	21,0	3	0	57,0	38,0	6,3	0	7
48	35,5	3	0	62,0	52,0	8,0	0	3
49	22,0	3	0	74,0	47,0	10,0	0	15
50	29,0	3	0	70,0	45,0	9,0	0	2

№ п/п	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
51	16,0	3	0	80,0	54,0	8,0	0	3
52	22,0	3	0	62,0	37,0	10,2	1	5
53	23,0	3	0	69,7	42,0	10,8	0	15
54	19,5	3	0	79,0	50,3	9,1	1	25
55	34,0	3	0	96,4	58,0	12,6	1	5
56	24,5	4	1	90,0	64,0	15,0	0	5
57	27,3	4	1	102,0	66,0	11,8	0	7
58	41,0	4	1	87,0	56,5	12,5	0	10
59	31,0	4	1	114,8	74,0	25,6	0	10
60	35,6	4	1	114,3	74,7	12,0	1	5
61	46,0	4	1	90,0	62,0	8,0	1	5
62	35,0	4	1	116,0	81,0	16,5	0	10
63	42,7	4	1	107,0	75,5	9,5	0	10
64	27,0	4	0	93,0	66,0	10,0	0	15
65	75,0	4	0	176,0	129,0	15,0	0	10
66	38,0	4	0	96,0	69,4	9,0	0	8
67	23,5	4	0	92,0	72,5	9,5	0	10
68	65,0	4	0	176,0	110,0	33,0	1	20
69	23,0	4	0	74,0	49,0	6,5	0	15
70	45,5	4	0	106,0	73,7	9,0	0	10
71	34,0	4	0	88,0	61,7	9,0	0	3
72	23,0	4	0	74,0	45,8	9,0	0	10
73	26,5	4	0	74,7	50,8	8,2	1	10
74	37,0	4	0	115,0	76,0	8,5	0	5
75	30,0	4	0	92,0	62,0	9,0	0	15
76	43,0	4	0	110,0	79,5	10,0	0	5

Принятые в таблице обозначения:

y – цена квартиры, тыс. долл.;

x_1 – число комнат в квартире;

x_2 – район города (1 – центральные, 0 – периферийные);

x_3 – общая площадь квартиры (m^2);

x_4 – жилая площадь квартиры (m^2);

x_5 – площадь кухни (m^2);

x_6 – тип дома (1 – кирпичный, 0 – другой);

x_7 – расстояние от метро, минут пешком.

По этим данным необходимо определить факторы, формировавшие цену квартир на вторичном рынке жилья в Санкт-Петербурге весной 2000 г.

Задание

1. Составьте матрицу парных коэффициентов корреляции.
2. Постройте уравнение регрессии, характеризующее зависимость цены от всех факторов. Установите, какие факторы коллинеарны, определив коэффициенты множественной детерминации для каждого из факторов.
3. Оцените значимость полученного уравнения. Какие факторы значимо воздействуют на формирование цены квартиры в этой модели?
4. Существует ли разница в ценах квартир, расположенных в центральных и в периферийных районах Санкт-Петербурга?
5. Существует ли разница в ценах квартир разных типов домов?
6. Постройте модель формирования цены квартиры за счет значимых факторов.

Задача 30

Изучается зависимость спроса на персональные компьютеры – y от дохода на одного члена семьи – x . Результаты опроса мужчин и женщин представлены на рис. 2.12, а, а результаты опроса всех взрослых в зависимости от жилищных условий приведены на рис. 2.12, б.

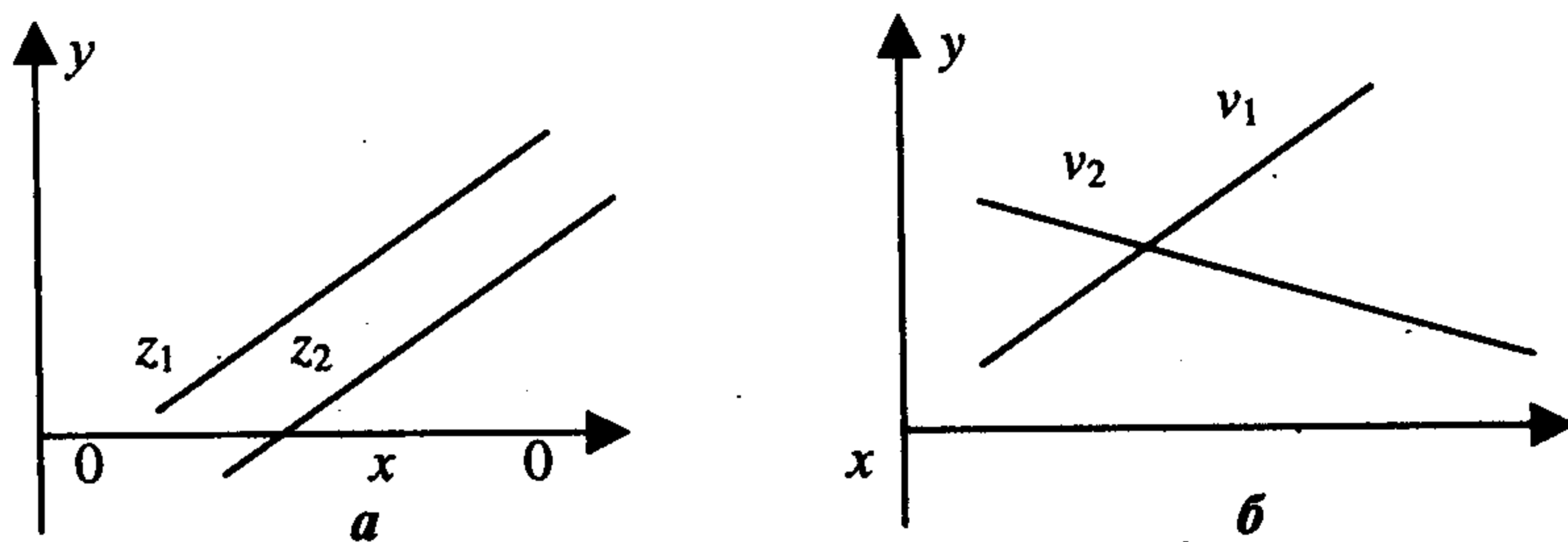


Рис. 2.12. Зависимость спроса на персональные компьютеры – y (тыс. руб.) от дохода на одного члена семьи – x (тыс. руб.) по данным опроса:
а – мужчин (z_1) и женщин (z_2);
б – всего населения при разных жилищных условиях (v_1 – хорошие, v_2 – плохие)

Задание

1. Определите, в каком случае возможно построение уравнения регрессии с включением фиктивной переменной.
2. Напишите общий вид уравнения регрессии с фиктивной переменной.
3. Укажите, как можно ввести в модель фиктивную переменную и как интерпретировать коэффициент регрессии при ней.

РАЗДЕЛ

СИСТЕМА ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ

3.1. МЕТОДИЧЕСКИЕ УКАЗАНИЯ

Сложные экономические процессы описывают 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 :

$$\begin{cases} y_1 = a_{11} \cdot x_1 + a_{12} \cdot x_2 + \dots + a_{1m} \cdot x_m + \epsilon_1, \\ y_2 = a_{21} \cdot x_1 + a_{22} \cdot x_2 + \dots + a_{2m} \cdot x_m + \epsilon_2, \\ \dots \\ y_n = a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + \dots + a_{nm} \cdot x_m + \epsilon_n \end{cases}$$

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:

$$\begin{cases} y_1 = a_{11} \cdot x_1 + a_{12} \cdot x_2 + \dots + a_{1m} \cdot x_m + \epsilon_1, \\ y_2 = b_{21} \cdot y_1 + a_{21} \cdot x_1 + a_{22} \cdot x_2 + \dots + a_{2m} \cdot x_m + \epsilon_2, \\ y_3 = b_{31} \cdot y_1 + b_{32} \cdot y_2 + a_{31} \cdot x_1 + a_{32} \cdot x_2 + \dots + a_{3m} \cdot x_m + \epsilon_3, \\ \dots \\ y_n = b_{n1} \cdot y_1 + b_{n2} \cdot y_2 + \dots + b_{nn-1} \cdot y_{n-1} + a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + \dots \\ + a_{nm} \cdot x_m + \epsilon_n \end{cases}$$

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:

Косвенный МНК состоит в следующем:

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

Двухшаговый МНК заключается в следующем:

- составляют приведенную форму модели и определяют численные значения параметров каждого ее уравнения обычным МНК;
- выявляют эндогенные переменные, находящиеся в правой части структурного уравнения, параметры которого определяют двухшаговым 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.

3.2. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Пример 1

Требуется:

1. Оценить следующую структурную модель на идентификацию:

$$\begin{cases} y_1 = b_{13} \cdot y_3 + a_{11} \cdot x_1 + a_{13} \cdot x_3, \\ y_2 = b_{21} \cdot y_1 + b_{23} \cdot y_3 + a_{22} \cdot x_2, \\ y_3 = b_{32} \cdot y_2 + a_{31} \cdot x_1 + a_{33} \cdot x_3. \end{cases}$$

2. Исходя из приведенной формы модели уравнений

$$\begin{cases} y_1 = 2 \cdot x_1 + 4 \cdot x_2 + 10 \cdot x_3, \\ y_2 = 3 \cdot x_1 - 6 \cdot x_2 + 2 \cdot x_3, \\ y_3 = -5 \cdot x_1 + 8 \cdot x_2 + 5 \cdot x_3, \end{cases}$$

найти структурные коэффициенты модели.

Решение

1. Модель имеет три эндогенные (y_1, y_2, y_3) и три экзогенные (x_1, x_2, x_3) переменные.

Проверим каждое уравнение системы на необходимое (Н) и достаточное (Д) условия идентификации.

Первое уравнение.

Н: эндогенных переменных – 2 (y_1, y_3),
отсутствующих экзогенных – 1 (x_2).

Выполняется необходимое равенство: $2 = 1 + 1$, следовательно, уравнение точно идентифицируемо.

Д: в первом уравнении отсутствуют y_2 и x_2 . Построим матрицу из коэффициентов при них в других уравнениях системы:

Уравнение	Отсутствующие переменные	
	y_2	x_2
Второе	-1	a_{22}
Третье	b_{32}	0

$$\text{Det } A = -1 \cdot 0 - b_{32} \cdot a_{22} \neq 0.$$

Определитель матрицы не равен 0, ранг матрицы равен 2; следовательно, выполняется достаточное условие идентификации, и первое уравнение точно идентифицируемо.

Второе уравнение.

Н: эндогенных переменных – 3 (y_1, y_2, y_3),
отсутствующих экзогенных – 2 (x_1, x_3).

Выполняется необходимое равенство: $3 = 2 + 1$, следовательно, уравнение точно идентифицируемо.

Д: во втором уравнении отсутствуют x_1 и x_3 . Построим матрицу из коэффициентов при них в других уравнениях системы:

Уравнение	Отсутствующие переменные	
	x_1	x_3
Первое	a_{11}	a_{13}
Третье	a_{31}	a_{33}

$$\text{Det } A = a_{11} \cdot a_{33} - a_{31} \cdot a_{13} \neq 0.$$

Определитель матрицы не равен 0, ранг матрицы равен 2, следовательно, выполняется достаточное условие идентификации, и второе уравнение точно идентифицируемо.

Третье уравнение.

Н: эндогенных переменных – 2 (y_2, y_3),
отсутствующих экзогенных – 1 (x_2).

Выполняется необходимое равенство: $2 = 1 + 1$, следовательно, уравнение точно идентифицируемо.

Д: в третьем уравнении отсутствуют y_1 и x_2 . Построим матрицу из коэффициентов при них в других уравнениях системы:

Уравнение	Отсутствующие переменные	
	y_1	x_2
Первое	-1	0
Второе	b_{21}	a_{22}

$$\text{Det } A = -1 \cdot a_{22} - b_{21} \cdot 0 \neq 0.$$

Определитель матрицы не равен 0, ранг матрицы равен 2, следовательно, выполняется достаточное условие идентификации, и третье уравнение точно идентифицируемо.

Следовательно, исследуемая система точно идентифицируема и может быть решена косвенным методом наименьших квадратов.

2. Вычислим структурные коэффициенты модели:

1) из третьего уравнения приведенной формы выразим x_2 (так как его нет в первом уравнении структурной формы):

$$x_2 = \frac{y_3 + 5 \cdot x_1 - 5 \cdot x_3}{8}.$$

Данное выражение содержит переменные y_3 , x_1 и x_3 , которые нужны для первого уравнения структурной формы модели (СФМ). Подставим полученное выражение x_2 в первое уравнение приведенной формы модели (ПФМ):

$$y_1 = 2 \cdot x_1 + 4 \cdot \frac{y_3 + 5 \cdot x_1 - 5 \cdot x_3}{8} + 10 \cdot x_3 \quad \Rightarrow$$

$$y_1 = 0,5 \cdot y_3 + 4,5 \cdot x_1 + 7,5 \cdot x_3 - \text{первое уравнение СФМ};$$

2) во втором уравнении СФМ нет переменных x_1 и x_3 . Структурные параметры второго уравнения СФМ можно будет определить в два этапа:

Первый этап: выразим x_1 в данном случае из первого или третьего уравнения ПФМ. Например, из первого уравнения:

$$x_1 = \frac{y_1 - 4 \cdot x_2 - 10 \cdot x_3}{2} = 0,5 \cdot y_1 - 2 \cdot x_2 - 5 \cdot x_3.$$

Подстановка данного выражения во второе уравнение ПФМ не решило бы задачу до конца, так как в выражении присутствует x_3 , которого нет в СФМ.

Выразим x_3 из третьего уравнения ПФМ:

$$x_3 = \frac{y_3 + 5 \cdot x_1 - 8 \cdot x_2}{5}$$

Подставим его в выражение x_1 :

$$x_1 = 0,5 \cdot y_1 - 2 \cdot x_2 - 5 \cdot \left(\frac{y_3 + 5 \cdot x_1 - 8 \cdot x_2}{5} \right) = 0,5 \cdot y_1 - y_3 + 6 \cdot x_2 - 5 \cdot x_1;$$
$$x_1 = \frac{0,5 \cdot y_1 - y_3 + 6 \cdot x_2}{6}$$

Второй этап: аналогично, чтобы выразить x_3 через искомые y_1 , y_3 и x_2 , заменим в выражении x_3 значение x_1 на полученное из первого уравнения ПФМ:

$$x_3 = \frac{y_3 + 5 \cdot (0,5 \cdot y_1 - 2 \cdot x_2 - 5 \cdot x_3) - 8 \cdot x_2}{5} =$$
$$= 0,2 \cdot y_3 + 0,5 \cdot y_1 - 3,6 \cdot x_2 - 5 \cdot x_3.$$

Следовательно,

$$x_3 = 0,033 \cdot y_3 + 0,083 \cdot y_1 - 0,6 \cdot x_2.$$

Подставим полученные x_1 и x_3 во второе уравнение ПФМ:

$$y_2 = 3 \cdot \frac{0,5 \cdot y_1 - y_3 + 6 \cdot x_2}{6} - 6 \cdot x_2 + 2 \cdot (0,033 \cdot y_3 + 0,083 \cdot y_1 - 0,6 \cdot x_2) \Rightarrow$$
$$y_2 = 0,416 \cdot y_1 - 0,434 \cdot y_3 - 4,2 \cdot x_2 \quad \text{— второе уравнение СФМ.}$$

Это уравнение можно получить из ПФМ иным путем. Суммируя все уравнения, получим

$$\begin{array}{r} y_1 = 2 \cdot x_1 + 4 \cdot x_2 + 10 \cdot x_3, \\ y_2 = 3 \cdot x_1 - 6 \cdot x_2 + 2 \cdot x_3, \\ y_3 = -5 \cdot x_1 + 8 \cdot x_2 + 5 \cdot x_3 \\ \hline y_1 + y_2 + y_3 = 6 \cdot x_2 + 17 \cdot x_3 \end{array}$$

Далее из первого и второго уравнений ПФМ исключим x_1 , домножив первое уравнение на 3, а второе — на (-2) и просуммировав их:

$$\begin{array}{r} 3 \cdot y_1 = 6 \cdot x_1 + 12 \cdot x_2 + 30 \cdot x_3, \\ -2 \cdot y_2 = -6 \cdot x_1 + 12 \cdot x_2 - 4 \cdot x_3. \\ \hline 3 \cdot y_1 - 2 \cdot y_2 = 24 \cdot x_2 + 26 \cdot x_3 \end{array}$$

Затем аналогичным путем из полученных уравнений исключаем x_3 , а именно:

$$\begin{cases} y_1 + y_2 + y_3 = 6 \cdot x_2 + 17 \cdot x_3 \\ 3 \cdot y_1 - 2 \cdot y_2 = 24 \cdot x_2 + 26 \cdot x_3 \end{cases} \begin{array}{l} -26, \\ 17, \end{array}$$

$$\begin{array}{r} -26 \cdot y_1 - 26 \cdot y_2 - 26 \cdot y_3 = -156 \cdot x_2 - 442 \cdot x_3, \\ 51 \cdot y_1 - 34 \cdot y_2 = 408 \cdot x_2 + 442 \cdot x_3, \\ \hline 25 \cdot y_1 - 60 \cdot y_2 - 26 \cdot y_3 = 252 \cdot x_2 \end{array} \Rightarrow$$

$$60 \cdot y_2 = 25 \cdot y_1 - 26 \cdot y_3 - 252 \cdot x_2 \Rightarrow$$

$$y_2 = 0,416 \cdot y_1 - 0,433 \cdot y_3 - 4,2 \cdot x_2;$$

3) из второго уравнения ПФМ выразим x_2 , так как его нет в третьем уравнении СФМ:

$$x_2 = \frac{-y_2 + 3 \cdot x_1 + 2 \cdot x_3}{6} = -0,167 \cdot y_2 + 0,5 \cdot x_1 + 0,333 \cdot x_3.$$

Подставим полученное выражение в третье уравнение ПФМ:

$$y_3 = -5 \cdot x_1 + 8 \cdot (-0,167 \cdot y_2 + 0,5 \cdot x_1 + 0,333 \cdot x_3) + 5 \cdot x_3 \Rightarrow$$

$$y_3 = -1,336 \cdot y_2 - x_1 + 7,664 \cdot x_3 \text{ — третье уравнение СФМ.}$$

Таким образом, СФМ примет вид

$$\begin{cases} y_1 = 0,5 \cdot y_3 + 4,5 \cdot x_1 + 7,5 \cdot x_3, \\ y_2 = 0,416 \cdot y_1 - 0,434 \cdot y_3 - 4,2 \cdot x_2, \\ y_3 = -1,336 \cdot y_2 - x_1 + 7,664 \cdot x_3. \end{cases}$$

Пример 2

Изучается модель вида

$$\begin{cases} y = a_1 + b_1(C + D) + \varepsilon_1, \\ C = a_2 + b_2 \cdot y + b_3 \cdot y_{-1} + \varepsilon_2, \end{cases}$$

где y — валовой национальный доход;
 y_{-1} — валовой национальный доход предшествующего года;
 C — личное потребление;
 D — конечный спрос (помимо личного потребления);
 ε_1 и ε_2 — случайные составляющие.

Информация за девять лет о приростах всех показателей дана в табл. 3.1.

Таблица 3.1

Год	D	y_{-1}	y	C	Год	D	y_{-1}	y	C
1	-6,8	46,7	3,1	7,4	6	44,7	17,8	37,2	8,6
2	22,4	3,1	22,8	30,4	7	23,1	37,2	35,7	30,0
3	-17,3	22,8	7,8	1,3	8	51,2	35,7	46,6	31,4
4	12,0	7,8	21,4	8,7	9	32,3	46,6	56,0	39,1
5	5,9	21,4	17,8	25,8	Σ	167,5	239,1	248,4	182,7

Для данной модели была получена система приведенных уравнений:

$$\begin{cases} y = 8,219 + 0,6688 \cdot D + 0,2610 \cdot y_{-1}, \\ C = 8,636 + 0,3384 \cdot D + 0,2020 \cdot y_{-1}. \end{cases}$$

Требуется:

1. Провести идентификацию модели.
2. Рассчитать параметры первого уравнения структурной модели.

Решение

1. В данной модели две эндогенные переменные (y и C) и две экзогенные переменные (D и y_{-1}). Второе уравнение точно идентифицировано, так как содержит две эндогенные переменные и не содержит одну экзогенную переменную из системы. Иными словами, для второго уравнения имеем по счетному правилу идентификации равенство: $2 = 1 + 1$.

Первое уравнение сверхидентифицировано, так как в нем на параметры при C и D наложено ограничение: они должны быть равны. В этом уравнении содержится одна эндогенная переменная y . Пере-

* Лизер С. Эконометрические методы и задачи: Пер. с англ. — М.: Статистика, 1971. — С. 61.

менная C в данном уравнении не рассматривается как эндогенная, так как она участвует в уравнении не самостоятельно, а вместе с переменной D . В данном уравнении отсутствует одна экзогенная переменная, имеющаяся в системе. По счетному правилу идентификации получаем: $1 + 1 = 2: D + 1 > N$. Это больше, чем число эндогенных переменных в данном уравнении, следовательно, система сверхидентифицирована.

2. Для определения параметров сверхидентифицированной модели используется двухшаговый метод наименьших квадратов.

Шаг 1. На основе системы приведенных уравнений по точно идентифицированному второму уравнению определим теоретические значения эндогенной переменной C . Для этого в приведенное уравнение

$$C = 8,636 + 0,3384 \cdot D + 0,2020 \cdot y_{-1}$$

подставим значения D и y_{-1} , имеющиеся в условии задачи. Получим:

$$\hat{C}_1 = 15,8; \hat{C}_2 = 16,8; \hat{C}_3 = 7,4; \hat{C}_4 = 14,3; \hat{C}_5 = 15,0; \hat{C}_6 = 27,4; \\ \hat{C}_7 = 24,0; \hat{C}_8 = 33,2; \hat{C}_9 = 29,0.$$

Шаг 2. По сверхидентифицированному уравнению структурной формы модели заменяем фактические значения C на теоретические \hat{C} и рассчитываем новую переменную $\hat{C} + D$ (табл. 3.2).

Таблица 3.2

Год	D	\hat{C}	$\hat{C} + D$	Год	D	\hat{C}	$\hat{C} + D$
1	-6,8	15,8	9,0	6	44,7	27,4	72,1
2	22,4	16,8	39,2	7	23,1	24,0	47,1
3	-17,3	7,4	-9,9	8	51,2	33,2	84,4
4	12,0	14,3	26,3	9	32,3	29,0	61,3
5	5,9	15,0	20,9	Σ	167,5	182,9	350,4

Далее к сверхидентифицированному уравнению применяется метод наименьших квадратов. Обозначим новую переменную $\hat{C} + D$ через Z . Решаем уравнение

$$y = a_1 + b_1 \cdot Z.$$

Система нормальных уравнений составит:

$$\begin{cases} \Sigma y = n \cdot a_1 + b_1 \cdot \Sigma Z, \\ \Sigma y \cdot Z = a_1 \cdot \Sigma Z + b_1 \cdot \Sigma Z^2, \end{cases}$$

$$\begin{cases} 248,4 = 9 \cdot a_1 + 350,4 \cdot b_1, \\ 13508,71 = 350,4 \cdot a_1 + 21142,02 \cdot b_1, \end{cases}$$

$$a_1 = 7,678; b_1 = 0,512.$$

Итак, первое уравнение структурной модели будет таким:

$$y = 7,678 + 0,512 \cdot (C + D).$$

Пример 3

Имеются данные за 1990–1994 гг. (табл. 3.3).

Таблица 3.3

Год	Годовое потребление свинины на душу населения, фунтов, y_1	Оптовая цена за фунт, долл., y_2	Доход на душу населения, долл., x_1	Расходы по обработке мяса, % к цене, x_2
1990	60	5,0	1300	60
1991	62	4,0	1300	56
1992	65	4,2	1500	56
1993	62	5,0	1600	63
1994	66	3,8	1800	50

Требуется:

Построить модель вида

$$\begin{cases} y_1 = f(y_2, x_1), \\ y_2 = f(y_1, x_2), \end{cases}$$

рассчитав соответствующие структурные коэффициенты.

Решение

Система одновременных уравнений с двумя эндогенными и двумя экзогенными переменными имеет вид

$$\begin{cases} y_1 = b_{12} \cdot y_2 + a_{11} \cdot x_1 + \epsilon_1, \\ y_2 = b_{21} \cdot y_1 + a_{22} \cdot x_2 + \epsilon_2. \end{cases}$$

В каждом уравнении две эндогенные и одна отсутствующая экзогенная переменная из имеющихся в системе. Для каждого уравнения данной системы действует счетное правило $2 = 1 + 1$. Это означает, что каждое уравнение и система в целом идентифицированы.

Для определения параметров такой системы применяется косвенный метод наименьших квадратов.

С этой целью структурная форма модели преобразуется в приведенную форму:

$$\begin{cases} y_1 = \delta_{11} \cdot x_1 + \delta_{12} \cdot x_2, \\ y_2 = \delta_{21} \cdot x_1 + \delta_{22} \cdot x_2, \end{cases}$$

в которой коэффициенты при x определяются методом наименьших квадратов.

Для нахождения значений δ_{11} и δ_{12} запишем систему нормальных уравнений:

$$\begin{cases} \sum y_1 x_1 = \delta_{11} \cdot \sum x_1^2 + \delta_{12} \cdot \sum x_1 x_2, \\ \sum y_1 x_2 = \delta_{11} \cdot \sum x_1 x_2 + \delta_{12} \cdot \sum x_2^2. \end{cases}$$

При ее решении предполагается, что x и y выражены через отклонения от средних уровней, т. е. матрица исходных данных составит:

	y_1	y_2	x_1	x_2
	-3	0,6	-200	3
	-1	-0,4	-200	-1
	2	-0,2	0	-1
	-1	0,6	100	6
	3	-0,6	300	-7
Σ	0	0,0	0	0

Применительно к ней необходимые суммы оказываются следующими:

$$\sum y_1 x_1 = 1600 ; \sum y_1 x_2 = -37 ; \sum x_1^2 = 180\,000 ;$$

$$\sum x_1 x_2 = -1900 ; \sum x_2^2 = 96.$$

Система нормальных уравнений составит:

$$\begin{cases} 1600 = 180\,000 \cdot \delta_{11} - 1900 \cdot \delta_{12}, \\ -37 = -1900 \cdot \delta_{11} + 96 \cdot \delta_{12}. \end{cases}$$

Решая ее, получим:

$$\delta_{11} = 0,00609 ; \delta_{12} = -0,26481 .$$

Итак, имеем $y_1 = 0,00609 \cdot x_1 - 0,26481 \cdot x_2$.

Аналогично строим систему нормальных уравнений для определения коэффициентов δ_{21} и δ_{22} :

$$\begin{cases} \Sigma y_2 x_1 = \delta_{21} \cdot \Sigma x_1^2 + \delta_{22} \cdot \Sigma x_1 x_2, \\ \Sigma y_2 x_2 = \delta_{21} \cdot \Sigma x_1 x_2 + \delta_{22} \cdot \Sigma x_2^2, \\ \Sigma y_2 x_1 = -160 ; \Sigma y_2 x_2 = 10,2. \end{cases}$$
$$\begin{cases} -160 = 180000 \cdot \delta_{21} - 1900 \cdot \delta_{22}, \\ 10,2 = -1900 \cdot \delta_{21} + 96 \cdot \delta_{22}. \end{cases}$$

Следовательно,

$$\delta_{21} = 0,00029 ; \delta_{22} = 0,11207 ,$$

тогда второе уравнение примет вид

$$y_2 = 0,00029 \cdot x_1 + 0,11207 \cdot x_2.$$

Приведенная форма модели имеет вид

$$\begin{cases} y_1 = 0,00609 \cdot x_1 - 0,26481 \cdot x_2, \\ y_2 = 0,00029 \cdot x_1 + 0,11207 \cdot x_2. \end{cases}$$

Из приведенной формы модели определяем коэффициенты структурной модели:

$$\begin{cases} y_1 = 0,00609 \cdot x_1 - 0,26481 \cdot x_2, \\ x_2 = \frac{y_2 - 0,00029 \cdot x_1}{0,11207}, \end{cases}$$

$$y_1 = 0,00609 \cdot x_1 - 0,26481 \cdot \frac{y_2 - 0,00029 \cdot x_1}{0,11207} = -2,36290 \cdot y_2 + 0,00678 \cdot x_1,$$

$$\begin{cases} y_2 = 0,00029 \cdot x_1 + 0,11207 \cdot x_2, \\ x_1 = \frac{y_1 + 0,26481 \cdot x_2}{0,00609}, \end{cases}$$

$$y_2 = 0,00029 \cdot \frac{y_1 + 0,26481 \cdot x_2}{0,00609} + 0,11207 \cdot x_2 = 0,04762 \cdot y_1 + 0,12468 \cdot x_2.$$

Итак, структурная форма модели имеет вид

$$\begin{cases} y_1 = -2,36290 \cdot y_2 + 0,00678 \cdot x_1 + \varepsilon_1, \\ y_2 = 0,04762 \cdot y_1 + 0,12468 \cdot x_2 + \varepsilon_2. \end{cases}$$

Пример 4

Рассматривается следующая модель:

$$\begin{aligned} C_t &= a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + U_1 && \text{(функция потребления);} \\ I_t &= a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + U_2 && \text{(функция инвестиций);} \\ r_t &= a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + U_3 && \text{(функция денежного рынка);} \\ Y_t &= C_t + I_t + G_t && \text{(тождество дохода),} \end{aligned}$$

где C_t — расходы на потребление в период t ;
 Y_t — совокупный доход в период t ;
 I_t — инвестиции в период t ;
 r_t — процентная ставка в период t ;
 M_t — денежная масса в период t ;
 G_t — государственные расходы в период t ;
 C_{t-1} — расходы на потребление в период $t-1$;
 I_{t-1} — инвестиции в период $t-1$;
 U_1, U_2, U_3 — случайные ошибки.

Требуется:

1. В предположении, что имеются временные ряды данных по всем переменным модели, предложите способ оценки ее параметров.
2. Как изменится ваш ответ на вопрос п. 1, если из модели исключить тождество дохода?

Решение

1. Модель представляет собой систему одновременных уравнений. Для ответа на вопрос о способе оценки параметров модели проверим каждое ее уравнение на идентификацию.

Модель включает четыре эндогенные переменные (C_t , I_t , Y_t и r_t) и четыре predetermined переменные (две экзогенные переменные — M_t и G_t и две лаговые эндогенные переменные — C_{t-1} и I_{t-1}).

Проверим необходимое условие идентификации для уравнений модели.

1 уравнение.

Это уравнение включает две эндогенные переменные (C_t и Y_t) и одну predetermined переменную (C_{t-1}). Следовательно, число predetermined переменных, не входящих в это уравнение, плюс 1, больше числа эндогенных переменных, входящих в уравнение: $3 + 1 > 2$. Уравнение сверхидентифицировано.

II уравнение.

Уравнение II включает две эндогенные переменные (I_t и r_t) и не включает три predetermined переменные. Как и I уравнение, оно сверхидентифицировано.

III уравнение.

Уравнение III тоже включает две эндогенные переменные (Y_t и r_t) и не включает три predetermined переменные. Это уравнение сверхидентифицировано.

IV уравнение.

Уравнение IV представляет собой тождество, параметры которого известны. Необходимости в его идентификации нет.

Проверим для каждого из уравнений достаточное условие идентификации. Для этого составим матрицу коэффициентов при переменных модели:

	C_t	Y_t	C_{t-1}	I_t	r_t	I_{t-1}	M_t	G_t
I уравнение	-1	b_{11}	b_{12}	0	0	0	0	0
II уравнение	0	0	0	-1	b_{21}	b_{22}	0	0
III уравнение	0	b_{31}	0	0	-1	0	b_{32}	0
Тождество	1	-1	0	1	0	0	0	1

В соответствии с достаточным условием идентификации определитель матрицы коэффициентов при переменных, не входящих в исследуемое уравнение, не должен быть равен нулю, а ранг матрицы должен быть равен числу эндогенных переменных модели минус 1, т.е. $4 - 1 = 3$.

I уравнение.

Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

$$A = \begin{pmatrix} -1 & b_{21} & b_{22} & 0 & 0 \\ 0 & -1 & 0 & b_{32} & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Ее ранг равен 3, так как определитель квадратной подматрицы 3×3 этой матрицы не равен нулю:

$$\text{Det } A^* = \begin{vmatrix} -1 & b_{21} & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{vmatrix} \neq 0.$$

Достаточное условие идентификации для I уравнения выполняется.

II уравнение.

Выпишем матрицу коэффициентов при переменных, не входящих в уравнение:

$$A = \begin{pmatrix} -1 & b_{11} & b_{12} & 0 & 0 \\ 0 & b_{31} & 0 & b_{32} & 0 \\ 1 & -1 & 0 & 0 & 1 \end{pmatrix}$$

Ее ранг равен трем, так как определитель квадратной подматрицы 3×3 этой матрицы не равен нулю:

$$\text{Det}A^* = \begin{vmatrix} -1 & 0 & 0 \\ 0 & b_{32} & 0 \\ 1 & 0 & 1 \end{vmatrix} \neq 0.$$

Достаточное условие идентификации для II уравнения выполняется.

III уравнение.

Выпишем матрицу коэффициентов при переменных, не входящих в уравнение:

$$A = \begin{pmatrix} -1 & b_{12} & 0 & 0 & 0 \\ 0 & 0 & -1 & b_{22} & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Ее ранг равен трем, так как определитель квадратной подматрицы 3×3 этой матрицы не равен нулю:

$$\text{Det}A^* = \begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 1 & 1 \end{vmatrix} \neq 0.$$

Достаточное условие идентификации для III уравнения выполняется.

Таким образом, все уравнения модели сверхидентифицированы. Для оценки параметров каждого из уравнений будем применять двухшаговый МНК.

Шаг 1. Запишем приведенную форму модели в общем виде:

$$\begin{aligned} C_t &= A_1 + A_2 \cdot C_{t-1} + A_3 \cdot I_{t-1} + A_4 \cdot M_t + A_5 \cdot G_t + V_1; \\ I_t &= B_1 + B_2 \cdot C_{t-1} + B_3 \cdot I_{t-1} + B_4 \cdot M_t + B_5 \cdot G_t + V_2; \\ Y_t &= D_1 + D_2 \cdot C_{t-1} + D_3 \cdot I_{t-1} + D_4 \cdot M_t + D_5 \cdot G_t + V_3; \\ r_t &= E_1 + E_2 \cdot C_{t-1} + E_3 \cdot I_{t-1} + E_4 \cdot M_t + E_5 \cdot G_t + V_4, \end{aligned}$$

где V_1, V_2, V_3, V_4 – случайные ошибки.

Определим параметры каждого из приведенных выше уравнений в отдельности обычным МНК. Затем найдем расчетные значения эндогенных переменных \hat{Y}_t , \hat{r}_t , используемых в правой части структурной модели, подставляя в каждое уравнение приведенной формы соответствующее значение predetermined переменных.

Шаг 2. В исходных структурных уравнениях заменим эндогенные переменные, выступающие в качестве факторных признаков, их расчетными значениями:

$$C_t = a_1 + b_{11} \cdot \hat{Y}_t + b_{12} \cdot C_{t-1} + U_1^*, \text{ где } U_1^* = U_1 + b_{11} \cdot V_1;$$

$$I_t = a_2 + b_{21} \cdot \hat{r}_t + b_{22} \cdot I_{t-1} + U_2^*, \text{ где } U_2^* = U_2 + b_{21} \cdot V_2;$$

$$r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + U_3, \text{ где } U_3^* = U_3 + b_{31} \cdot V_3.$$

Применяя к каждому из полученных уравнений в отдельности обычный МНК, определим структурные параметры a_1 , b_{11} , b_{12} , a_2 , b_{21} , b_{22} , a_3 , b_{31} и b_{32} .

2. Если из модели исключить тождество дохода, число predetermined переменных модели уменьшится на 1 (из модели будет исключена переменная G_t). Число эндогенных переменных модели также снизится на единицу – переменная Y_t станет экзогенной. В правых частях функции потребления и функции денежного рынка будут находиться только predetermined переменные. Функция инвестиций постулирует зависимость эндогенной переменной I_t от эндогенной переменной r_t (которая зависит только от predetermined переменных) и predetermined переменной I_{t-1} . Таким образом, мы получим рекурсивную систему. Ее параметры можно оценивать обычным МНК, и нет необходимости исследования системы уравнения на идентификацию.

3.3. КОНТРОЛЬНЫЕ ЗАДАНИЯ

Задание к задачам 1–19

1. Применив необходимое и достаточное условие идентификации, определите, идентифицировано ли каждое из уравнений модели.
2. Определите метод оценки параметров модели.
3. Запишите приведенную форму модели.

Задача 1

Модель денежного рынка:

$$\begin{aligned}R_t &= a_1 + b_{11} \cdot M_t + b_{12} Y_t + \varepsilon_1, \\Y_t &= a_2 + b_{21} R_t + b_{22} I_t + \varepsilon_2,\end{aligned}$$

- где R – процентная ставка;
 Y – ВВП;
 M – денежная масса;
 I – внутренние инвестиции;
 t – текущий период.

Задача 2

Модель Менгеса:

$$\begin{aligned}Y_t &= a_1 + b_{11} Y_{t-1} + b_{12} I_t + \varepsilon_1, \\I_t &= a_2 + b_{21} Y_t + b_{22} Q_t + \varepsilon_2, \\C_t &= a_3 + b_{31} Y_t + b_{32} C_{t-1} + b_{33} P_t + \varepsilon_3, \\Q_t &= a_4 + b_{41} Q_{t-1} + b_{42} R_t + \varepsilon_4,\end{aligned}$$

- где Y – национальный доход;
 C – расходы на личное потребление;
 I – чистые инвестиции;
 Q – валовая прибыль экономики;
 P – индекс стоимости жизни;
 R – объем продукции промышленности;
 t – текущий период;
 $t-1$ – предыдущий период.

Задача 3

Одна из версий модифицированной модели Кейнса имеет вид

$$\begin{aligned}C_t &= a_1 + b_{11} Y_t + b_{12} Y_{t-1} + \varepsilon_1, \\I_t &= a_2 + b_{21} Y_t + b_{22} Y_{t-1} + \varepsilon_2, \\Y_t &= C_t + I_t + G_t,\end{aligned}$$

- где C – расходы на потребление;
 Y – доход;
 I – инвестиции;
 G – государственные расходы;
 $t-1$ – текущий период;
 $t-1$ – предыдущий период.

Задача 4

Модель мультипликатора-акселератора:

$$\begin{aligned}C_t &= a_1 + b_{11}R_t + b_{12}C_{t-1} + \varepsilon_1, \\I_t &= a_2 + b_{21}(R_t - R_{t-1}) + \varepsilon_2, \\R_t &= C_t + I_t,\end{aligned}$$

- где C – расходы на потребление;
 R – доход;
 I – инвестиции;
 t – текущий период;
 $t-1$ – предыдущий период.

Задача 5

Конъюнктурная модель имеет вид

$$\begin{aligned}C_t &= a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_1, \\I_t &= a_2 + b_{21}r_t + b_{22}I_{t-1} + \varepsilon_2, \\r_t &= a_3 + b_{31}Y_t + b_{32}M_t + \varepsilon_3, \\Y_t &= C_t + I_t + G_t,\end{aligned}$$

- где C – расходы на потребление;
 Y – ВВП;
 I – инвестиции;
 r – процентная ставка;
 M – денежная масса;
 G – государственные расходы;
 t – текущий период;
 $t-1$ – предыдущий период.

Задача 6

Модель протекционизма Сальватора (упрощенная версия):

$$\begin{aligned}M_t &= a_1 + b_{12}N_t + b_{13}S_t + b_{14}E_{t-1} + b_{15}M_{t-1} + \varepsilon_1, \\N_t &= a_2 + b_{21}M_t + b_{23}S_t + b_{26}Y_t + \varepsilon_2, \\S_t &= a_3 + b_{31}M_t + b_{32}N_t + b_{37}X_t + \varepsilon_3,\end{aligned}$$

- где M – доля импорта в ВВП;
 N – общее число прошений об освобождении от таможенных пошлин;
 S – число удовлетворенных прошений об освобождении от таможенных пошлин;
 E – фиктивная переменная, равная 1 для тех лет, в которые курс доллара на международных валютных рынках был искусственно завышен, и 0 – для всех остальных лет;
 Y – реальный ВВП;
 X – реальный объем чистого экспорта;
 t – текущий период;
 $t-1$ – предыдущий период.

Задача 7

Макроэкономическая модель (упрощенная версия модели Клейна):

$$\begin{aligned}C_t &= a_1 + b_{12}Y_t + b_{13}T_t + \varepsilon_1, \\I_t &= a_2 + b_{21}Y_t + b_{24}K_{t-1} + \varepsilon_2, \\Y_t &= C_t + I_t,\end{aligned}$$

где C – потребление;
 I – инвестиции;
 Y – доход;
 T – налоги;
 K – запас капитала;
 t – текущий период;
 $t-1$ – предыдущий период.

Задача 8

Макроэкономическая модель экономики США (одна из версий):

$$\begin{aligned}C_t &= a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_{1t} && \text{(функция потребления);} \\I_t &= a_2 + b_{21}Y_t + b_{23}r_t + \varepsilon_{2t} && \text{(функция инвестиций);} \\r_t &= a_3 + b_{31}Y_t + b_{34}M_t + b_{35}r_{t-1} + \varepsilon_{3t} && \text{(функция денежного рынка);} \\Y_t &= C_t + I_t + G_t && \text{(тождество дохода),}\end{aligned}$$

где C – потребление;
 Y – ВВП;
 I – инвестиции;
 r – процентная ставка;
 M – денежная масса;
 G – государственные расходы;
 t – текущий период;
 $t-1$ – предыдущий период.

Задача 9

Модель Кейнса (одна из версий):

$$\begin{aligned}C_t &= a_1 + b_{11}Y_t + b_{12}Y_{t-1} + \varepsilon_{1t} && \text{(функция потребления);} \\I_t &= a_2 + b_{21}Y_t + \varepsilon_{2t} && \text{(функция инвестиций);} \\Y_t &= C_t + I_t + G_t && \text{(тождество дохода),}\end{aligned}$$

где C – потребление;
 Y – ВВП;
 I – валовые инвестиции;
 G – государственные расходы;
 t – текущий период;
 $t-1$ – предыдущий период.

Задача 10

Модель денежного и товарного рынков:

$$R_t = a_1 + b_{12}Y_t + b_{14}M_t + \varepsilon_1 \quad (\text{функция денежного рынка});$$

$$Y_t = a_2 + b_{21}R_t + b_{23}I_t + b_{25}G_t + \varepsilon_2 \quad (\text{функция товарного рынка});$$

$$I_t = a_3 + b_{31}R_t + \varepsilon_3 \quad (\text{функция инвестиций}),$$

- где R – процентные ставки;
 Y – реальный ВВП;
 M – денежная масса;
 I – внутренние инвестиции;
 G – реальные государственные расходы.

Задача 11

Для прогнозирования спроса на свою продукцию предприятие использует следующую модель, характеризующую общую экономическую ситуацию в регионе:

$$\begin{aligned} Q_t &= a_1 + b_{11}Y_t + \varepsilon_{1t}, \\ C_t &= a_2 + b_{21}Y_t + \varepsilon_{2t}, \\ I_t &= a_3 + b_{32}(Y_{t-1} - K_{t-1}) + \varepsilon_{3t}, \\ Y_t &= C_t + I_t, \end{aligned}$$

- где Q – реализованная продукция в период t ;
 Y – ВДС региона;
 C – конечное потребление;
 I – инвестиции;
 K – запас капитала;
 t – текущий период;
 $t-1$ – предыдущий период.

Задача 12

Модифицированная модель Кейнса:

$$\begin{aligned} C_t &= a_1 + b_{11}Y_t + \varepsilon_1, \\ I_t &= a_2 + b_{21}Y_t + b_{22}Y_{t-1} + \varepsilon_2, \\ Y_t &= C_t + I_t + G_t, \end{aligned}$$

- где C – расходы на потребление;
 Y – доход;
 I – инвестиции;
 G – государственные расходы;
 t – текущий период;
 $t-1$ – предыдущий период.

Задача 13

Макроэкономическая модель:

$$\begin{aligned}C_t &= a_1 + b_{11}D_t + \varepsilon_{1t}, \\I_t &= a_2 + b_{22}Y_t + b_{23}Y_{t-1} + \varepsilon_{2t}, \\Y_t &= D_t + T_t, \\D_t &= C_t + I_t + G_t,\end{aligned}$$

- где C – расходы на потребление;
 Y – чистый национальный продукт;
 D – чистый национальный доход;
 I – инвестиции;
 T – косвенные налоги;
 G – государственные расходы;
 t – текущий период;
 $t-1$ – предыдущий период.

Задача 14

Дана следующая структурная форма модели:

$$\begin{aligned}C_t &= b_1 + b_2S_t + b_3P_t, \\S_t &= a_1 + a_2R_t + a_3R_{t-1} + a_4t, \\R_t &= S_t + P_t,\end{aligned}$$

- где C_t – личное потребление в период t ;
 S_t – зарплата в период t ;
 P_t – прибыль в период t ;
 R_t – общий доход в период t ;
 R_{t-1} – общий доход в период $t-1$;
 $t-1$ – предыдущий период.

Задача 15

Предложение и спрос на рынке характеризуются следующей моделью:

$$\begin{aligned}q_1 &= a_1 + b_1p + \varepsilon_1, \\q_2 &= a_2 + b_2p + \varepsilon_2, \\q_1 &= q_2,\end{aligned}$$

- где q_1 – спрос на товар;
 q_2 – предложение количества товара;
 p – цена, по которой заключаются сделки.

Задача 16

Гипотетическая модель экономики:

$$C_t = a_1 + b_{11}Y_t + b_{12}J_t + \varepsilon_1,$$

$$J_t = a_2 + b_{21}Y_{t-1} + \varepsilon_2,$$

$$T_t = a_3 + b_{31}Y_t + \varepsilon_3,$$

$$Y_t = C_t + J_t + G_t,$$

- где C – совокупное потребление в период t ;
 Y – совокупный доход в период t ;
 J – инвестиции в период t ;
 T – налоги в период t ;
 G – государственные доходы в период t .

Задача 17

Модель спроса и предложения кейнсианского типа:

$$Q_t^s = a_1 + a_2P_t + a_3P_{t-1} + \varepsilon_1 \quad (\text{предложение}),$$

$$Q_t^d = \beta_1 + \beta_2P_t + \beta_3Y_t + \varepsilon_2 \quad (\text{спрос}),$$

$$Q_t^s = Q_t^d \quad (\text{тождество}),$$

- где Q_t^d – спрос на товар в момент времени t ;
 Q_t^s – предложение товара в момент времени t ;
 P_t – цена товара в момент времени t ;
 Y_t – доход в момент времени t ;
 P_{t-1} – цена товара в предыдущий период.

Задача 18

Модель спроса и предложения на деньги:

$$R_t = a_1 + b_{11}M_t + b_{12}Y_t + \varepsilon_1,$$

$$Y_t = a_2 + b_{21}R_t + \varepsilon_2,$$

- где R – процентные ставки в период t ;
 Y – ВВП в период t ;
 M – денежная масса в период t .

Задача 19

Модель денежного рынка:

$$R_t = a_1 + b_{11}M_t + b_{12}Y_t + \varepsilon_1,$$

$$Y_t = a_2 + b_{21}R_t + b_{22}I_t + \varepsilon_2,$$

$$I_t = a_3 + b_{33}R_t + \varepsilon_3,$$

- где R – процентные ставки;
 Y – ВВП;
 M – денежная масса;
 I – внутренние инвестиции.

Задача 20

Рассматривается следующая модель:

$$S_t = a_1 + b_{11}D_t + b_{12}M_t + b_{13}Un_t + \varepsilon_1,$$

$$C_t = a_2 + b_{21}D_t + b_{22}S_t + b_{23}Un_{t-1} + \varepsilon_2,$$

$$D_t = a_3 + b_{31}S_t + b_{32}C_{t-1} + b_{33}I_t + \varepsilon_3,$$

- где S_t – заработная плата в период t ;
 D_t – чистый национальный доход в период t ;
 M_t – денежная масса в период t ;
 C_t – расходы на потребление в период t ;
 C_{t-1} – расходы на потребление в период $t-1$;
 Un_t – уровень безработицы в период t ;
 Un_{t-1} – уровень безработицы в предыдущий период;
 I_t – инвестиции в период t .

Задание

1. Каким методом вы будете оценивать структурные параметры этой модели?
2. Выпишите приведенную форму модели.
3. Кратко охарактеризуйте методику расчета параметров первого и второго структурного уравнения модели.

Задача 21

Ниже приводятся результаты расчета параметров некоторой эконометрической модели.

Структурная форма модели:

$$Y_1 = -4 + ??? Y_2 - 9,4 X_2 + \varepsilon_1,$$

$$Y_2 = 12,83 - 2,67 Y_1 + ??? X_1 + \varepsilon_2,$$

$$Y_3 = 1,36 - 1,76 Y_1 + 0,828 Y_2 + \varepsilon_3.$$

Приведенная форма модели:

$$Y_1 = 2 + 4 X_1 - 3 X_2 + v_1,$$

$$Y_2 = 7,5 + 5 X_1 + 8 X_2 + v_2,$$

$$Y_3 = 4 + ??? X_1 + ??? X_2 + v_3.$$

Задание

1. Какими методами получены параметры структурной и приведенной форм модели? Обоснуйте возможность применения косвенного МНК для расчета структурных параметров модели.
2. Восстановите пропущенные характеристики.

Задача 22

Эконометрическая модель содержит четыре уравнения, четыре эндогенные переменные (y) и три экзогенные переменные (x). Ниже представлена матрица коэффициентов при переменных в структурной форме этой модели.

Уравнение	y_1	y_2	y_3	y_4	x_1	x_2	x_3
I	-1	0	b_{13}	b_{14}	c_{11}	0	0
II	0	-1	b_{23}	0	c_{21}	0	0
III	0	b_{32}	-1	0	c_{31}	0	c_{33}
IV	b_{41}	b_{42}	b_{43}	-1	0	c_{42}	c_{43}

Задание

Применив необходимое и достаточное условие идентификации, определите, идентифицируемо ли каждое уравнение модели.

Задача 23

Для изучения связи между уровнем инфляции и доходностью обыкновенных акций используется следующая система уравнений регрессии:

$$Rb_t = a_1 + b_{11}Rs_t + b_{12}Rb_{t-1} + b_{13}L_t + b_{14}Y_t + b_{15}N_t + b_{16}I_t + \varepsilon_1,$$

$$Rs_t = a_2 + b_{21}Rb_t + b_{22}Rb_{t-1} + b_{23}L_t + b_{24}Y_t + b_{25}N_t + b_{26}E_t + \varepsilon_2,$$

- где Rb – доходность облигаций;
 Rs – доходность обыкновенных акций;
 L – доход в денежной форме на душу населения;
 Y – доход от всех источников на душу населения;
 N – переменная, характеризующая новые выпуски ценных бумаг за период;
 E – ожидаемая доходность акций на конец периода;
 I – ожидаемый уровень инфляции;
 t – текущий период;
 $t-1$ – предыдущий период.

В этой модели переменные Rb и Rs являются эндогенными.

Задание

1. Определите, является ли данная модель системой одновременных уравнений.
2. Выпишите приведенную форму модели.
3. Каким методом вы будете оценивать структурные параметры этой модели? Обоснуйте ответ.

Задача 24

Имеется следующая модель кейнсианского типа:

$$C_t = a_1 + b_{11}Y_t + b_{12}T_t + \varepsilon_1 \quad (\text{функция потребления});$$

$$I_t = a_2 + b_{21}Y_{t-1} + \varepsilon_2 \quad (\text{функция инвестиций});$$

$$T_t = a_3 + b_{31}Y_t + \varepsilon_3 \quad (\text{функция налогов});$$

$$Y_t = C_t + I_t + G_t \quad (\text{тождество дохода}),$$

- где C – совокупное потребление в период t ;
 Y – совокупный доход в период t ;
 I – инвестиции в период времени t ;
 T – налоги в период времени t ;
 G – государственные расходы в период времени t ;
 Y_{t-1} – совокупный доход в период $t-1$.

В этой модели переменные C , I , T и Y являются эндогенными.

Задание

1. Применяв необходимое и достаточное условие идентификации, определите, идентифицировано ли каждое из уравнений модели. Укажите, каким методом вы будете оценивать структурные параметры каждого уравнения (методику оценки параметров излагать не надо).
2. Напишите приведенную форму модели.
3. Пусть в правую часть функции инвестиций введена дополнительная экзогенная переменная r_t (процентные ставки). Как это изменение повлияет на идентификацию и методы оценки структурных параметров модели?

Задача 25

Изучается зависимость потребления (C) от доходов (Y).

Задание

1. Каким методом вы будете определять параметры функции потребления, если эконометрическая модель имеет следующий вид:

модель А: $C = a + bY + \varepsilon$ (функция потребления);

модель Б: $C = a + bY + \varepsilon$ (функция потребления);
 $Y = C + I$ (тождество дохода);

модель В: $C = a + bY + \varepsilon$ (функция потребления);
 $Y = C + I + G$ (тождество дохода),

где I – инвестиции;

G – госрасходы.

Переменные C, Y – эндогенные.

Дайте развернутый ответ по каждой из моделей А–В, включающий обоснование выбранного вами метода и краткое описание методики расчетов.

2. Предположим, определив по некоторому неизменному по всем моделям массиву исходных данных параметры a и b , для каждой из предложенных моделей вы рассчитали коэффициенты детерминации (назовем их R_A^2, R_B^2, R_V^2 соответственно). Какой из этих коэффициентов будет наиболее высоким? Почему?

Задача 26

Имеется модель, построенная по шести наблюдениям:

$$\begin{aligned} Y_1 &= a_1 + b_{12}Y_2 + \varepsilon_1, \\ Y_2 &= a_2 + b_{21}Y_1 + c_{21}X_1 + \varepsilon_2, \\ Y_3 &= Y_2 + X_2. \end{aligned}$$

Ей соответствует следующая приведенная форма:

$$\begin{aligned} Y_1 &= -1,25 + 22X_1 + 0,67X_2 + v_1, \\ Y_2 &= 2 - 4X_1 + 10X_2 + v_2, \\ Y_3 &= -30 + 12X_1 + 8X_2 + v_3. \end{aligned}$$

Известны также следующие исходные данные:

n	1	2	3	4	5	6
Y_1	3	2	4	1	5	3
X_1	2	3	5	6	10	8
X_2	4	7	3	6	5	5

Задание

1. Определите структурные параметры первого уравнения, если это возможно.
2. Определите структурные параметры второго уравнения, если это возможно.

Задача 27

Имеется следующая модель:

$$Y_t = C_t + I_t + G_t \text{ (тождество дохода);}$$

$$C_t = 0,09 + 0,43YD_{t-1} + 0,23M_t + \epsilon_{1t} \text{ (функция потребления);}$$

$$I_t = 0,08 + 0,40(Y_{t-1} - Y_{t-2}) + 0,48Z_t + 0,1t + \epsilon_{2t} \text{ (функция инвестиций);}$$

$$G_t = 0,13 + 0,67G_{t-1} + \epsilon_{3t} \text{ (уравнение госрасходов),}$$

- где
- Y_t – валовой внутренний продукт в текущем году t ;
 - C_t – расходы на личное потребление в текущем году t ;
 - I_t – валовые внутренние инвестиции в текущем году t ;
 - G_t – государственные расходы плюс чистые иностранные инвестиции в текущем году;
 - YD_{t-1} – располагаемый доход за вычетом налогов в предыдущем году;
 - M_t – денежная масса в текущем году;
 - Z_t – доход от собственности до вычета налогов в текущем году;
 - Y_{t-1} – валовой внутренний продукт в предыдущем году;
 - Y_{t-2} – валовой внутренний продукт два года назад;
 - G_{t-1} – государственные расходы плюс чистые иностранные инвестиции в предыдущем году;
 - t – фактор времени;
 - Y, C, I, G – эндогенные переменные.

Задание

1. Пусть известно следующее:

Переменная	Среднее значение	Среднее квадратическое отклонение
YD_{t-1}	2500	500
M_t	3000	200

Какая из этих переменных оказывает наиболее сильное воздействие на расходы на конечное потребление? Ответ подтвердите соответствующими расчетами ($\sigma_{Ct} = 430$).

2. Дайте интерпретацию параметров функции инвестиций.
3. Параметры каждого уравнения этой модели были найдены обычным методом наименьших квадратов (ОМНК). Изложите ваше мнение относительно возможности применения ОМНК к данной модели.

Задача 28

Имеется следующая модель:

$$\begin{aligned} Y_1 &= a_1 + b_{11}X_1 + b_{12}X_2 + c_{12}Y_2 + \varepsilon_1, \\ Y_2 &= a_2 + b_{22}X_2 + b_{23}X_3 + c_{21}Y_1 + \varepsilon_2, \\ Y_3 &= a_3 + b_{31}X_1 + b_{33}X_3 + \varepsilon_3. \end{aligned}$$

Приведенная форма этой модели имеет вид

$$\begin{aligned} Y_1 &= 6 + 8X_1 + 10X_2 + 4X_3 + v_1, \\ Y_2 &= 16 - 12X_1 - 70X_2 + 8X_3 + v_2, \\ Y_3 &= 10 - 5X_1 - 22X_2 + 5X_3 + v_3. \end{aligned}$$

Задание

1. Определите все возможные структурные коэффициенты на основе приведенной формы модели.
2. Обоснуйте возможность применения выбранного вами метода определения структурных коэффициентов.

Задача 29

Имеется следующая модель:

$$\begin{aligned} Y_1 &= a_1 + b_{12}Y_2 + b_{13}Y_3 + c_{12}X_2 + \varepsilon_1, \\ Y_2 &= a_2 + b_{21}Y_1 + b_{23}Y_3 + c_{21}X_1 + c_{22}X_2 + \varepsilon_2, \\ Y_3 &= a_3 + b_{34}Y_4 + c_{32}X_2 + c_{33}X_3 + \varepsilon_3, \\ Y_4 &= a_4 + b_{42}Y_2 + b_{43}Y_3 + c_{43}X_3 + \varepsilon_4. \end{aligned}$$

Этой модели соответствует следующая приведенная форма:

$$\begin{aligned} Y_1 &= 2 + 3X_1 + 4X_2 - 3X_3 + v_1, \\ Y_2 &= 12 - 6X_1 + 2X_2 + 4X_3 + v_2, \\ Y_3 &= 8 + 5X_1 + 10X_2 + 3X_3 + v_3, \\ Y_4 &= 4 - 3X_1 + 5X_2 - 6X_3 + v_4. \end{aligned}$$

Задание

1. Какие структурные параметры модели можно найти через приведенные коэффициенты. Ответ обоснуйте. В качестве примера найдите параметры какого-либо одного структурного уравнения.

Примечание. Для упрощения расчетов рекомендуется вести их в обыкновенных дробях.

2. Что изменится в вашем ответе на вопрос п.1, если $c_{21} = 0$?

Задача 30

Строится модель вида

$$\begin{aligned} Y_1 &= a_1 + b_2 Y_2 + c_1 X_1 + \varepsilon_1, \\ Y_2 &= a_2 + b_1 Y_1 + c_2 X_2 + \varepsilon_2. \end{aligned}$$

Задание

Определите структурные коэффициенты, учитывая, что

$$\sum Y_1 X_1 = 2600; \sum Y_1 X_2 = 4350; \sum Y_1 = 350; \sum Y_2 = 25; \sum X_1 = 750;$$

$$\sum X_2 = 350; \sum X_1^2 = 1200; \sum X_2^2 = 1800; n = 30; \sum X_1 X_2 = 1500,$$

$$\text{а также } Y_2 = 2X_1 + 3X_2.$$

Задача 31

Имеется следующая гипотетическая структурная модель:

$$Y_1 = b_{12} Y_2 + a_{11} X_1 + a_{12} X_2,$$

$$Y_2 = b_{21} Y_1 + b_{23} Y_3 + a_{22} X_2,$$

$$Y_3 = b_{32} Y_2 + a_{31} X_1 + a_{33} X_3.$$

Приведенная форма исходной модели имеет вид

$$Y_1 = 3X_1 - 6X_2 + 2X_3,$$

$$Y_2 = 2X_1 + 4X_2 + 10X_3,$$

$$Y_3 = -5X_1 + 6X_2 + 5X_3.$$

Задание

1. Проверьте структурную форму модели на идентификацию.

2. Определите структурные коэффициенты модели.

Задача 32

Пусть имеются условные данные, представленные в табл. 3.4.

Таблица 3.4

Период времени	Темп прироста					% безработных, X_1
	заработной платы, Y_1	цен, Y_2	дохода, Y_3	цен на импорт, X_2	экономически активного населения, X_3	
1	2	6	10	2	1	1
2	3	7	12	3	2	2
3	4	8	11	1	5	3
4	5	5	15	4	3	2
5	6	4	14	2	3	3
6	7	9	16	2	4	4
7	8	10	18	3	4	5

Задание

Определите параметры структурной модели следующего вида:

$$Y_1 = b_{12}Y_2 + a_{11}X_1 + a_{12}X_2 + \varepsilon_1,$$

$$Y_2 = b_{21}Y_1 + b_{22}X_2 + a_{23}X_3 + \varepsilon_2,$$

$$Y_3 = b_{31}Y_1 + a_{33}X_3 + \varepsilon_3.$$

Задача 33

В одной из аграрных стран строилась функция потребления за 1988–1997 гг. по данным (в условных денежных единицах), представленным в табл. 3.5.

Таблица 3.5

Показатель	1988 г.	1989 г.	1990 г.	1991 г.	1992 г.	1993 г.	1994 г.	1995 г.	1996 г.	1997 г.
Совокупное потребление	1900	1980	2000	1800	2000	2100	2200	2100	2050	2100
Объем инвестиций	100	200	300	200	100	200	300	200	150	300
Совокупный доход	2000	2180	2300	2000	2100	2300	2500	2300	2200	2400

Задание

1. Постройте функцию потребления, используя модель Кейнса формирования доходов.
2. Дайте интерпретацию результатов приведенной формы модели.

Задача 34

Исследуется зависимость спроса и предложения некоторого товара от его цены, дохода и процентной ставки:

$$Q_t^s = a_1 + a_2 p_t + a_3 R_t + \varepsilon_1,$$

$$Q_t^d = b_1 + b_2 p_t + b_3 Y_t + b_4 Y_{t-1} + \varepsilon_2,$$

$$Q_t^s = Q_t^d = Q_t,$$

где Q_t^s – предложение в момент времени t ;

Q_t^d – спрос в момент времени t ;

p_t – цена товара в момент времени t ;

R_t – процентная ставка в момент времени t ;

Y_t – доход в момент времени t ;

Y_{t-1} – доход предшествующего периода.

Отметим, что в этой модели цена и величина спроса-предложения определяются одновременно, в связи с чем эти переменные должны считаться эндогенными.

Информация за восемь лет о приростах всех показателей представлена в табл. 3.6.

Таблица 3.6

Год	Q_t	R_t	Y_t	Y_{t-1}	p_t
1	40	3,0	15	13	6
2	45	3,0	15	15	6
3	40	2,0	18	15	5
4	50	3,5	20	18	8
5	35	2,5	18	20	5
6	45	4,0	22	18	9
7	50	3,5	21	22	10
8	45	3,5	22	21	9
Σ	350	25,0	151	142	58

Для данной модели была получена система приведенных уравнений:

$$\begin{cases} Q_t = 24,4730 + 5,2374 \cdot R_t + 0,1652 \cdot Y_t - 0,0116 \cdot Y_{t-1}, \\ P_t = -4,4268 + 1,9746 \cdot R_t + 0,1915 \cdot Y_t + 0,1065 \cdot Y_{t-1}. \end{cases}$$

Задание

1. Проведите идентификацию модели.
2. Рассчитайте параметры первого уравнения структурной модели.

РАЗДЕЛ

ВРЕМЕННЫЕ РЯДЫ В ЭКОНОМЕТРИЧЕСКИХ ИССЛЕДОВАНИЯХ

4.1. МЕТОДИЧЕСКИЕ УКАЗАНИЯ

Модели, построенные по данным, характеризующим один объект за ряд последовательных моментов (периодов), 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 временного ряда*.

Аддитивная модель имеет вид: $Y = T + S + E$;

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

Построение аддитивной и мультипликативной моделей сводится к расчету значений T , S и 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 ($T \cdot E$) модели;
- 4) аналитическое выравнивание уровней ($T + E$) или ($T \cdot E$) и расчет значений T с использованием полученного уравнения тренда;
- 5) расчет полученных по модели значений ($T + S$) или ($T \cdot S$);
- 6) расчет абсолютных и/или относительных ошибок.

Автокорреляция уровней ряда – это корреляционная зависимость между последовательными уровнями временного ряда:

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

где $\bar{y}_1 = \frac{\sum_{t=2}^n y_t}{n-1}$; $\bar{y}_2 = \frac{\sum_{t=2}^n y_{t-1}}{n-1}$ – коэффициент автокорреляции уровней ряда первого порядка;

$$r_2 = \frac{\sum_{t=3}^n (y_t - \bar{y}_3) \cdot (y_{t-1} - \bar{y}_4)}{\sqrt{\sum_{t=3}^n (y_t - \bar{y}_3)^2 \cdot \sum_{t=3}^n (y_{t-2} - \bar{y}_4)^2}},$$

где $\bar{y}_3 = \frac{\sum_{t=3}^n y_t}{n-2}$; $\bar{y}_4 = \frac{\sum_{t=3}^n y_{t-1}}{n-2}$ – коэффициент автокорреляции уровней ряда второго порядка.

Формулы для расчета коэффициентов автокорреляции старших порядков легко получить из формулы линейного коэффициента корреляции.

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*. Для этого чаще всего применяются следующие функции:

- линейная $\hat{y}_t = a + b \cdot t$;
- гипербола $\hat{y}_t = a + b/t$;
- экспонента $\hat{y}_t = e^{a+b \cdot t}$;
- степенная функция $\hat{y}_t = a \cdot t^b$;
- парабола второго и более высоких порядков

$$\hat{y}_t = a + b_1 \cdot t + b_2 \cdot t^2 + \dots + b_k \cdot t^k.$$

Параметры трендов определяются обычным МНК, в качестве независимой переменной выступает время $t = 1, 2, \dots, n$, а в качестве зависимой переменной – фактические уровни временного ряда y_t . Критерием отбора наилучшей формы тренда является наибольшее значение скорректированного коэффициента детерминации \bar{R}^2 .

При построении моделей регрессии по временным рядам для устранения тенденции используются следующие методы.

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 $x_t - \hat{x}_t$. Для дальнейшего анализа используют не исходные данные, а отклонения от тренда.

Метод последовательных разностей заключается в следующем: если ряд содержит линейный тренд, тогда исходные данные заменяются первыми разностями:

$$\Delta_t = y_t - y_{t-1} = b + (\varepsilon_t - \varepsilon_{t-1});$$

если параболический тренд – вторыми разностями:

$$\Delta_t^2 = \Delta_t - \Delta_{t-1} = 2 \cdot b_2 + (\varepsilon_t - 2 \cdot \varepsilon_{t-1} + \varepsilon_{t-2}).$$

В случае экспоненциального и степенного тренда метод последовательных разностей применяется к логарифмам исходных данных.

Модель, включающая фактор времени, имеет вид

$$y_t = a + b_1 \cdot x_t + b_2 \cdot t + \varepsilon_t.$$

Параметры a и b этой модели определяются обычным МНК.

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

Для определения автокорреляции остатков используют *критерий Дарбина – Уотсона* и расчет величины:

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

Коэффициент автокорреляции остатков первого порядка определяется по формуле

$$\eta_1^\varepsilon = \frac{\sum_{t=2}^n \varepsilon_t \cdot \varepsilon_{t-1}}{\sum_{t=2}^n \varepsilon_t^2}, \quad -1 \leq \eta_1^\varepsilon \leq 1.$$

Критерий Дарбина – Уотсона и коэффициент автокорреляции остатков первого порядка связаны соотношением

$$d \approx 2(1 - \eta_1^\varepsilon).$$

Эконометрические модели, содержащие не только текущие, но и лаговые значения факторных переменных, называются *моделями с распределенным лагом*.

Модель с распределенным лагом в предположении, что максимальная величина лага конечна, имеет вид

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + \dots + b_p \cdot x_{t-p} + \varepsilon_t.$$

Коэффициент регрессии b_0 при переменной x_t характеризует среднее абсолютное изменение y_t при изменении x_t на 1 ед. своего измерения в некоторый фиксированный момент времени t , без учета воздействия лаговых значений фактора x . Этот коэффициент называют *краткосрочным мультипликатором*.

В момент $(t + 1)$ воздействие факторной переменной x_t на результат y_t составит $(b_0 + b_1)$ условных единиц; в момент времени $(t + 2)$ воздействие можно охарактеризовать суммой $(b_0 + b_1 + b_2)$ и т.д. Эти суммы называют *промежуточными мультипликаторами*. Для максимального лага $(t + l)$ воздействие фактора на результат описывается суммой $(b_0 + b_1 + \dots + b_l = b)$, которая называется *долгосрочным мультипликатором*.

Величины

$$\beta_j = b_j / b, \quad j = \overline{0, l},$$

называются *относительными коэффициентами* модели с распределенным лагом. Если все коэффициенты b_j имеют одинаковые знаки, то для любого j

$$0 < \beta_j < 1 \text{ и } \sum_{j=1}^l \beta_j = 1.$$

Величина *среднего лага* модели множественной регрессии определяется по формуле средней арифметической взвешенной:

$$\bar{l} = \sum_{j=1}^l j \cdot \beta_j$$

и представляет собой средний период, в течение которого будет происходить изменение результата под воздействием изменения фактора в момент t .

Медианный лаг — это период, в течение которого с момента времени t будет реализована половина общего воздействия фактора на результат:

$$\sum_{j=0}^{l_{Me}-1} \beta_j \approx 0,5,$$

где l_{Me} — медианный лаг.

Оценку параметров моделей с распределенными лагами можно проводить согласно одному из двух методов: методу Койка или методу Алмон.

В *распределении Койка* делается предположение, что коэффициенты при лаговых значениях объясняющей переменной убывают в геометрической прогрессии:

$$b_l = b_0 \cdot \lambda^l, \quad l = 0, 1, 2, \dots, \quad 0 < \lambda < 1.$$

Уравнение регрессии преобразуется к виду

$$y_t = a + b_0 \cdot x_t + b_0 \cdot \lambda \cdot x_{t-1} + b_0 \cdot \lambda^2 \cdot x_{t-2} + \dots + \varepsilon_t.$$

После несложных преобразований получаем уравнение, оценки параметров которого приводят к оценкам параметров исходного уравнения.

В *методе Алмон* предполагается, что веса текущих и лаговых значений объясняющих переменных подчиняются полиномиальному распределению:

$$b_j = c_0 + c_1 j + c_2 j^2 + \dots + c_k j^k.$$

Уравнение регрессии примет вид

$$y_t = a + c_0 \cdot z_0 + c_1 \cdot z_1 + c_2 \cdot z_2 + \dots + c_k \cdot z_k + \varepsilon_t,$$

где $z_i = \sum_{j=1}^l j^i \cdot x_{t-j}, \quad i = 1, \dots, k; \quad j = 1, \dots, p.$

Расчет параметров модели с распределенным лагом методом Алмон проводится по следующей схеме:

- 1) устанавливается максимальная величина лага l ;
- 2) определяется степень полинома k , описывающего структуру лага;
- 3) рассчитываются значения переменных z_0, \dots, z_k ;
- 4) определяются параметры уравнения линейной регрессии y_t от z_i ;
- 5) рассчитываются параметры исходной модели с распределенным лагом.

Модели, содержащие в качестве факторов лаговые значения зависимой переменной, называются *моделями авторегрессии*, например:

$$y_t = a + b_0 \cdot x_t + c_1 \cdot y_{t-1} + \varepsilon_t.$$

Как и в модели с распределенным лагом, b_0 в этой модели характеризует краткосрочное изменение y_t под воздействием изменения x_t на 1 ед. Долгосрочный мультипликатор в модели авторегрессии рассчитывается как сумма краткосрочного и промежуточных мультипликаторов:

$$b = b_0 + b_0 \cdot c_1 + b_0 \cdot c_1^2 + b_0 \cdot c_1^3 + \dots = b_0 \cdot (1 + c_1 + c_1^2 + c_1^3 + \dots) = b_0 / (1 - c_1).$$

Отметим, что такая интерпретация коэффициентов модели авторегрессии и расчет долгосрочного мультипликатора основаны на предпосылке о наличии бесконечного лага в воздействии текущего значения зависимой переменной на ее будущие значения.

4.2. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Пример 1

По данным за 18 месяцев построено уравнение регрессии зависимости прибыли предприятия y (млн руб.) от цен на сырье x_1 (тыс. руб. за 1 т) и производительности труда x_2 (ед. продукции на 1 работника):

$$\hat{y} = 200 - 1,5 \cdot x_1 + 4,0 \cdot x_2.$$

При анализе остаточных величин были использованы значения, приведенные в табл. 4.1.

Таблица 4.1

№	y	x_1	x_2
1	210	800	300
2	720	1000	500
3	300	1500	600

$$\sum \varepsilon_t^2 = 10500, \quad \sum (\varepsilon_t - \varepsilon_{t-1})^2 = 40\,000.$$

Требуется:

1. По трем позициям рассчитать $\hat{y}_t, \varepsilon_t, \varepsilon_{t-1}, \varepsilon_t^2, (\varepsilon_t - \varepsilon_{t-1})^2$.
2. Рассчитать критерий Дарбина – Уотсона.
3. Оценить полученный результат при 5%-ном уровне значимости.
4. Указать, пригодно ли уравнение для прогноза.

Решение

1. \hat{y}_t определяется путем подстановки фактических значений x_1 и x_2 в уравнение регрессии:

$$\hat{y}_1 = 200 - 1,5 \cdot 800 + 4,0 \cdot 300 = 200;$$

$$\hat{y}_2 = 200 - 1,5 \cdot 1000 + 4,0 \cdot 500 = 700;$$

$$\hat{y}_3 = 200 - 1,5 \cdot 1500 + 4,0 \cdot 600 = 350.$$

Остатки ε_t рассчитываются по формуле

$$\varepsilon_t = y_t - \hat{y}_t.$$

Следовательно,

$$\varepsilon_1 = 210 - 200 = 10, \quad \varepsilon_2 = 720 - 700 = 20, \quad \varepsilon_3 = 300 - 350 = -50;$$

$$\varepsilon_1^2 = 100, \quad \varepsilon_2^2 = 400, \quad \varepsilon_3^2 = 2500;$$

ε_{t-1} – те же значения, что и ε_t , но со сдвигом на один месяц.

Результаты вычислений оформим в виде табл. 4.2.

Таблица 4.2

№	\hat{y}_t	ε_t	ε_{t-1}	$(\varepsilon_t - \varepsilon_{t-1})$	$(\varepsilon_t - \varepsilon_{t-1})^2$	ε_t^2
1	200	10	–	–	–	100
2	700	20	10	10	100	400
3	350	-50	20	-70	4900	2500
...
Σ					40 000	10 500

2. Критерий Дарбина – Уотсона рассчитывается по формуле

$$d = \frac{\Sigma(\varepsilon_t - \varepsilon_{t-1})^2}{\Sigma\varepsilon_t^2} = \frac{40000}{10500} = 3,81.$$

3. Фактическое значение d сравниваем с табличными значениями при 5%-ном уровне значимости. При $n = 18$ месяцев и $m = 2$ (число факторов) нижнее значение d' равно 1,05, а верхнее – 1,53. Так как фактическое значение d близко к 4, можно считать, что автокорреляция в остатках характеризуется отрицательной величиной. Чтобы проверить значимость отрицательного коэффициента автокорреляции, найдем величину:

$$4 - d = 4 - 3,81 = 0,19,$$

что значительно меньше, чем d' . Это означает наличие в остатках автокорреляции.

4. 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 также, что форма связи неточна, а может быть, в рядах динамики имеется общая тенденция.

Пример 2

Имеются следующие данные о величине дохода на одного члена семьи и расхода на товар A (табл. 4.3).

Таблица 4.3

Показатель	1985 г.	1986 г.	1987 г.	1988 г.	1989 г.	1990 г.
Расходы на товар A , руб.	30	35	39	44	50	53
Доход на одного члена семьи, % к 1985 г.	100	103	105	109	115	118

Требуются:

1. Определить ежегодные абсолютные приросты доходов и расходов и сделать выводы о тенденции развития каждого ряда.
2. Перечислить основные пути устранения тенденции для построения модели спроса на товар A в зависимости от дохода.
3. Построить линейную модель спроса, используя первые разности уровней исходных динамических рядов.
4. Пояснить экономический смысл коэффициента регрессии.
5. Построить линейную модель спроса на товар A , включив в нее фактор времени. Интерпретировать полученные параметры.

Решение

1. Обозначим расходы на товар A через y , а доходы одного члена семьи – через x . Ежегодные абсолютные приросты определяются по формулам

$$\Delta y_t = y_t - y_{t-1}, \quad \Delta x_t = x_t - x_{t-1}.$$

Расчеты можно оформить в виде таблицы (табл. 4.4).

Таблица 4.4

y_t	Δy_t	x_t	Δx_t
30	–	100	–
35	5	103	3
39	4	105	2
44	5	109	4
50	6	115	6
53	3	118	3

Значения Δy не имеют четко выраженной тенденции, они варьируют вокруг среднего уровня, что означает наличие в ряде динамики линейного тренда (линейной тенденции). Аналогичный вывод можно сделать и по ряду x : абсолютные приросты не имеют систематической направленности, они примерно стабильны, а следовательно, ряд характеризуется линейной тенденцией.

2. Так как ряды динамики имеют общую тенденцию к росту, то для построения регрессионной модели спроса на товар A в зависимости от дохода необходимо устранить тенденцию. С этой целью модель может строиться по первым разностям, т.е. $\Delta y = f(\Delta x)$, если ряды динамики характеризуются линейной тенденцией.

Другой возможный путь учета тенденции при построении моделей – найти по каждому ряду уравнение тренда:

$$\hat{y}_t = f(t) \text{ и } \hat{x}_t = f(t)$$

и отклонения от него:

$$dy = y_t - \hat{y}_t; \quad dx = x_t - \hat{x}_t.$$

Далее модель строится по отклонениям от тренда:

$$dy = f(dx).$$

При построении эконометрических моделей чаще используется другой путь учета тенденции – включение в модель фактора времени. Иными словами, модель строится по исходным данным, но в нее в качестве самостоятельного фактора включается время, т.е. $\hat{y}_t = f(x, t)$.

3. Модель имеет вид

$$\Delta \hat{y} = a + b \cdot \Delta x.$$

Для определения параметров a и b применяется МНК. Система нормальных уравнений следующая:

$$\begin{cases} \Sigma \Delta y = n \cdot a + b \cdot \Sigma \Delta x, \\ \Sigma \Delta y \Delta x = a \cdot \Sigma \Delta x + b \cdot \Sigma \Delta^2 x. \end{cases}$$

Применительно к нашим данным имеем

$$\begin{cases} 23 = 5 \cdot a + 18 \cdot b, \\ 88 = 18 \cdot a + 74 \cdot b. \end{cases}$$

Решая эту систему, получим:

$$a = 2,565 \text{ и } b = 0,565,$$

откуда модель имеет вид

$$\Delta \hat{y} = 2,565 + 0,565 \cdot \Delta x.$$

4. Коэффициент регрессии $b = 0,565$ руб. Он означает, что с ростом прироста душевого дохода на 1%-ный пункт расходы на товар A увеличиваются со средним ускорением, равным 0,565 руб.

5. Модель имеет вид

$$\hat{y} = a + b \cdot x + c \cdot t.$$

Применяя МНК, получим систему нормальных уравнений:

$$\begin{cases} \Sigma y = n \cdot a + b \cdot \Sigma x + c \cdot \Sigma t, \\ \Sigma yx = a \cdot \Sigma x + b \cdot \Sigma x^2 + c \cdot \Sigma x \cdot t, \\ \Sigma yt = a \cdot \Sigma t + b \cdot \Sigma x \cdot t + c \cdot \Sigma t^2. \end{cases}$$

Расчеты оформим в виде табл. 4.5.

Таблица 4.5.

t	y	x	Yx	y_t	x_t	x^2	t^2
1	30	100	3000	30	100	10000	1
2	35	103	3605	70	206	10609	4
3	39	105	4095	117	315	11025	9
4	44	109	4796	176	436	11881	16
5	50	115	5750	250	575	13225	25
6	53	118	6254	318	708	13924	36
21	251	650	27500	961	2340	70664	91

Система уравнений примет вид

$$\begin{cases} 251 = 6 \cdot a + 650 \cdot b + 21 \cdot c, \\ 27500 = 650 \cdot a + 70664 \cdot b + 2340 \cdot c, \\ 961 = 21 \cdot a + 2340 \cdot b + 91 \cdot c. \end{cases}$$

Решая ее, получим

$$a = -5,42; b = 0,322; c = 3,516.$$

Уравнение регрессии имеет вид

$$y = -5,42 + 0,322 \cdot x + 3,516 \cdot t.$$

Параметр $b = 0,322$ фиксирует силу связи y и x . Его величина означает, что с ростом дохода на одного члена семьи на 1%-ный пункт при условии неизменной тенденции расходы на товар A возрастают в среднем на 0,322 руб. Параметр $c = 3,516$ характеризует среднегодовой абсолютный прирост расходов на товар A под воздействием прочих факторов при условии неизменного дохода.

Пример 3

По данным за 30 месяцев некоторого временного ряда x_t были получены значения коэффициентов автокорреляции уровней:

$$r_1 = 0,63;$$

$$r_2 = 0,38;$$

$$r_3 = 0,72;$$

$$r_4 = 0,97;$$

$$r_5 = 0,55;$$

$$r_6 = 0,40;$$

$$r_7 = 0,65;$$

r_i — коэффициенты автокорреляции i -го порядка.

Требуется:

1. Охарактеризовать структуру этого ряда, используя графическое изображение.
2. Для прогнозирования значений x_t в будущие периоды предполагается построить уравнение авторегрессии. Выбрать наилучшее уравнение, обосновать выбор. Указать общий вид этого уравнения.

Решение

1. Так как значения всех коэффициентов автокорреляции достаточно высокие, ряд содержит тенденцию. Поскольку наибольшее абсолютное значение имеет коэффициент автокорреляции 4-го порядка r_4 , ряд содержит периодические колебания, цикл этих колебаний равен 4.

График этого ряда можно представить на рис. 4.1.

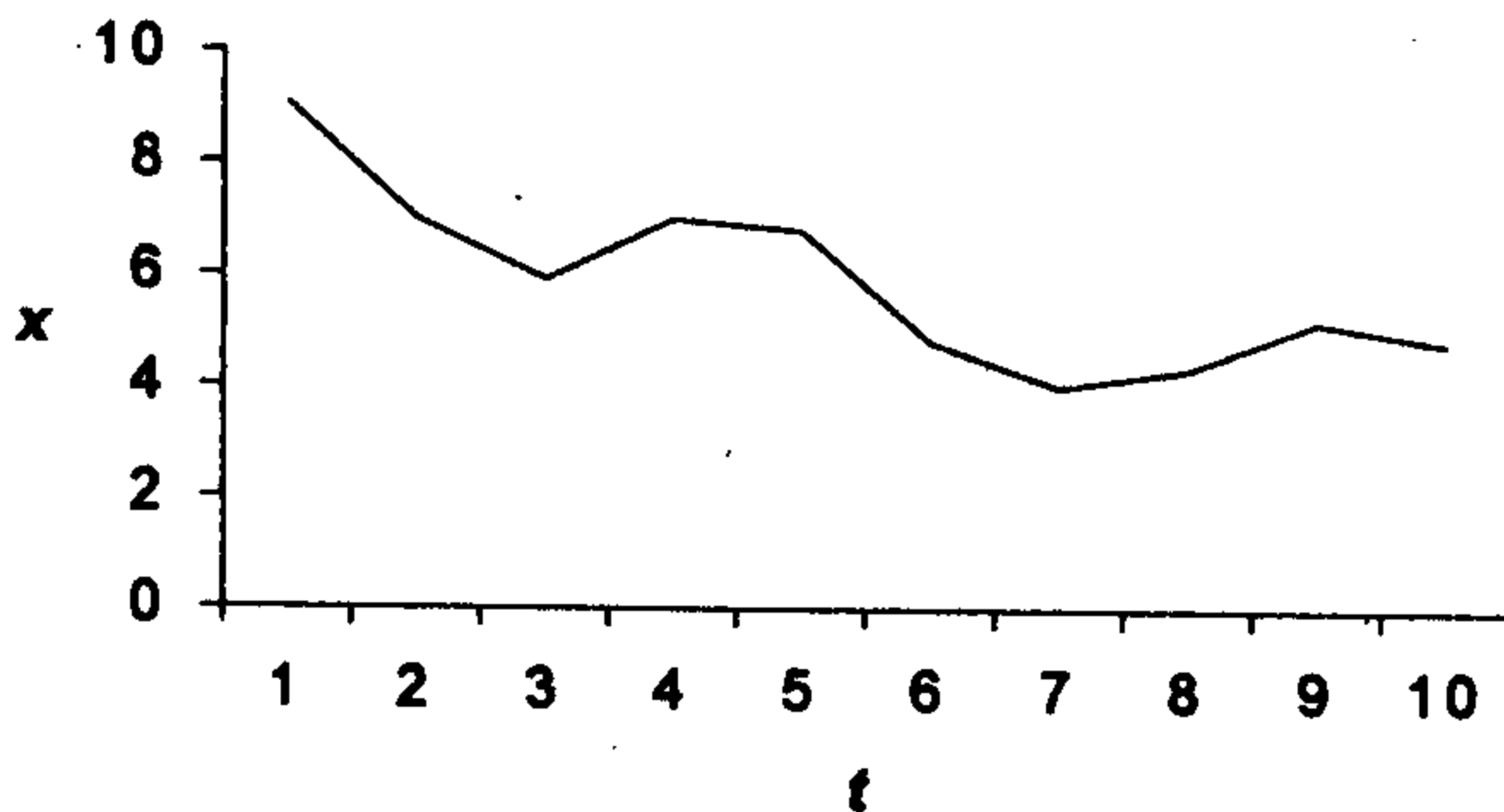
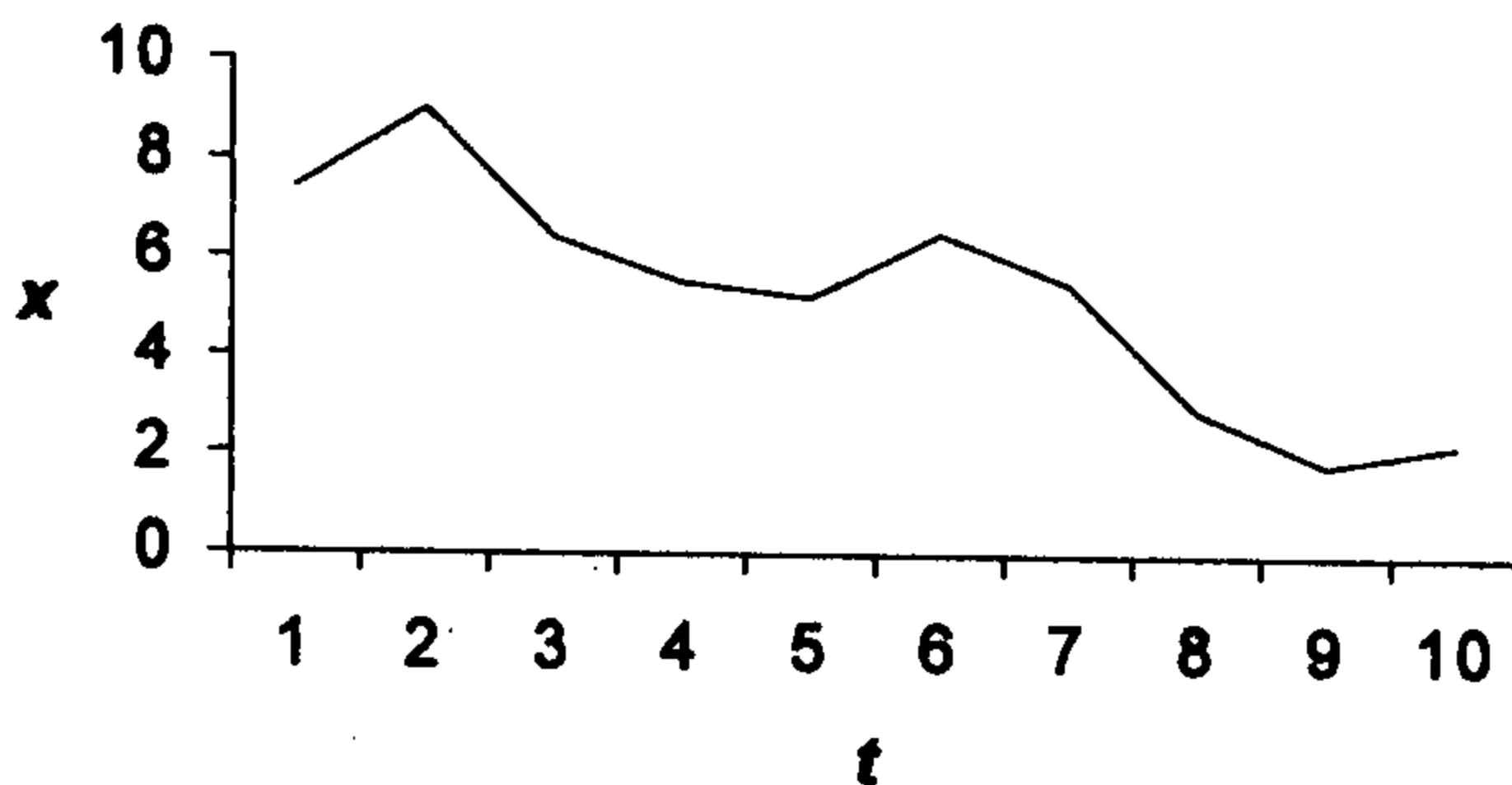


Рис. 4.1. Графики, характеризующие убывающую тенденцию при разных возможных периодических колебаниях

2. Наиболее целесообразно построение уравнения авторегрессии:

$$y_t = a + b \cdot y_{t-4} + u_t$$

так как значение $r_4 = 0,97$ свидетельствует о наличии очень тесной связи между уровнями ряда с лагом в 4 месяца.

Кроме того, возможно построение и множественного уравнения авторегрессии y_t от y_{t-3} и y_{t-4} , так как $r_4 = 0,72$:

$$y_t = a + b_1 \cdot y_{t-3} + b_2 \cdot y_{t-4} + u_t$$

Сравнить полученные уравнения и выбрать наилучшее решение можно с помощью скорректированного коэффициента детерминации. •

Пример 4

На основе помесечных 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 табл. 4.6.

Таблица 4.6

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 ($t = \overline{1;36}$).

Требуется:

1. Определить значение сезонной компоненты за декабрь.
2. На основе построенной модели дать прогноз общего числа браков, заключенных в течение первого квартала следующего года.

Решение

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

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

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 первом квартале следующего года, есть сумма числа браков, заключенных в январе F_{37} , в феврале F_{38} и в 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:

$$S_1 = -1 \text{ — январь;}$$

$$S_2 = 2 \text{ — февраль;}$$

$$S_3 = -0,5 \text{ — март.}$$

Таким образом,

$$F_{37} = T_{37} + S_1 = 3,61 - 1 = 2,61;$$

$$F_{38} = T_{38} + S_2 = 3,64 + 2 = 5,64;$$

$$F_{39} = T_{39} + S_3 = 3,67 - 0,5 = 3,17.$$

Количество браков, заключенных в первом квартале следующего года, составит: $2,61 + 5,64 + 3,17 = 11,42$ тыс., или 11420.

Пример 5

Динамика выпуска продукции Финляндии характеризуется данными (млн долл.), представленными в табл. 4.7.

Таблица 4.7

Год	Выпуск продукции	Год	Выпуск продукции
1961	1054	1979	11172
1962	1104	1980	14150
1963	1149	1981	14004
1964	1291	1982	13088
1965	1427	1983	12518
1966	1505	1984	13471
1967	1513	1985	13617
1968	1635	1986	16356
1969	1987	1987	20037
1970	2306	1988	21748
1971	2367	1989	23298
1972	2913	1990	26570
1973	3837	1991	23080
1974	5490	1992	23981
1975	5502	1993	23446
1976	6342	1994	29658
1977	7665	1995	39573
1978	8570	1996	38435

Требуется:

1. Провести расчет параметров линейного и экспоненциального трендов.
2. Построить графики ряда динамики и трендов.
3. Выбрать наилучший вид тренда на основании графического изображения и значения коэффициента детерминации.

4.3. РЕАЛИЗАЦИЯ ТИПОВЫХ ЗАДАЧ НА КОМПЬЮТЕРЕ

Решение с использованием ППП MS Excel

1. Для определения параметров линейного тренда по методу наименьших квадратов используется статистическая функция **ЛИНЕЙН**, для определения экспоненциального тренда – **ЛГРФПРИБЛ**. Порядок вычисления был рассмотрен в 1-м разделе практикума. В качестве зависимой переменной в данном примере выступает время ($t = 1, 2, \dots, n$). Приведем результаты вычисления функций **ЛИНЕЙН** и **ЛГРФПРИБЛ** (рис. 4.2 и 4.3).

Год, x	Выпуск продукции, y	Линейн	
1961	1054	977,1198198	-1921124,369
1962	1104	60,67808483	120053,2457
1963	1140	0,284084673	3782,03096
1964	1291	259,3175535	34
1965	1427	3709254808	486332921,9
1966	1505		
1967	1513		
1968	1635		
1969	1987		
1970	2306		
1971	2367		
1972	2913		
1973	3837		
1974	5400		
1975	5502		
1976	6342		
1977	7645		
1978	8570		
1979	11172		
1980	14150		
1981	14004		
1982	13088		
1983	12518		

Рис. 4.2. Результат вычисления функции **ЛИНЕЙН**

A1		Год, х	
Год, х	Выпуск продукции, у	Лгрфприбл	
1961	1054	1,004486521	1
1962	1104	9,6069E-05	
1963	1140	0,076463782	1,140451186
1964	1291	2,897809882	35
1965	1427	3,768975299	45,52201173
1966	1505		
1967	1513		
1968	1635		
1969	1987		
1970	2306		
1971	2367		
1972	2913		
1973	3837		
1974	5490		
1975	5502		
1976	6342		
1977	7665		
1978	8570		
1979	11172		
1980	14150		
1981	14004		
1982	13088		
1983	12518		

Рис. 4.3. Результат вычисления функции ЛГРФПРИБЛ

Запишем уравнения линейного и экспоненциального тренда, используя данные рис. 4.2 и 4.3:

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

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

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. 4.4); вид графика выберите в поле рядом со списком типов. Щелкните по кнопке **Далее**;

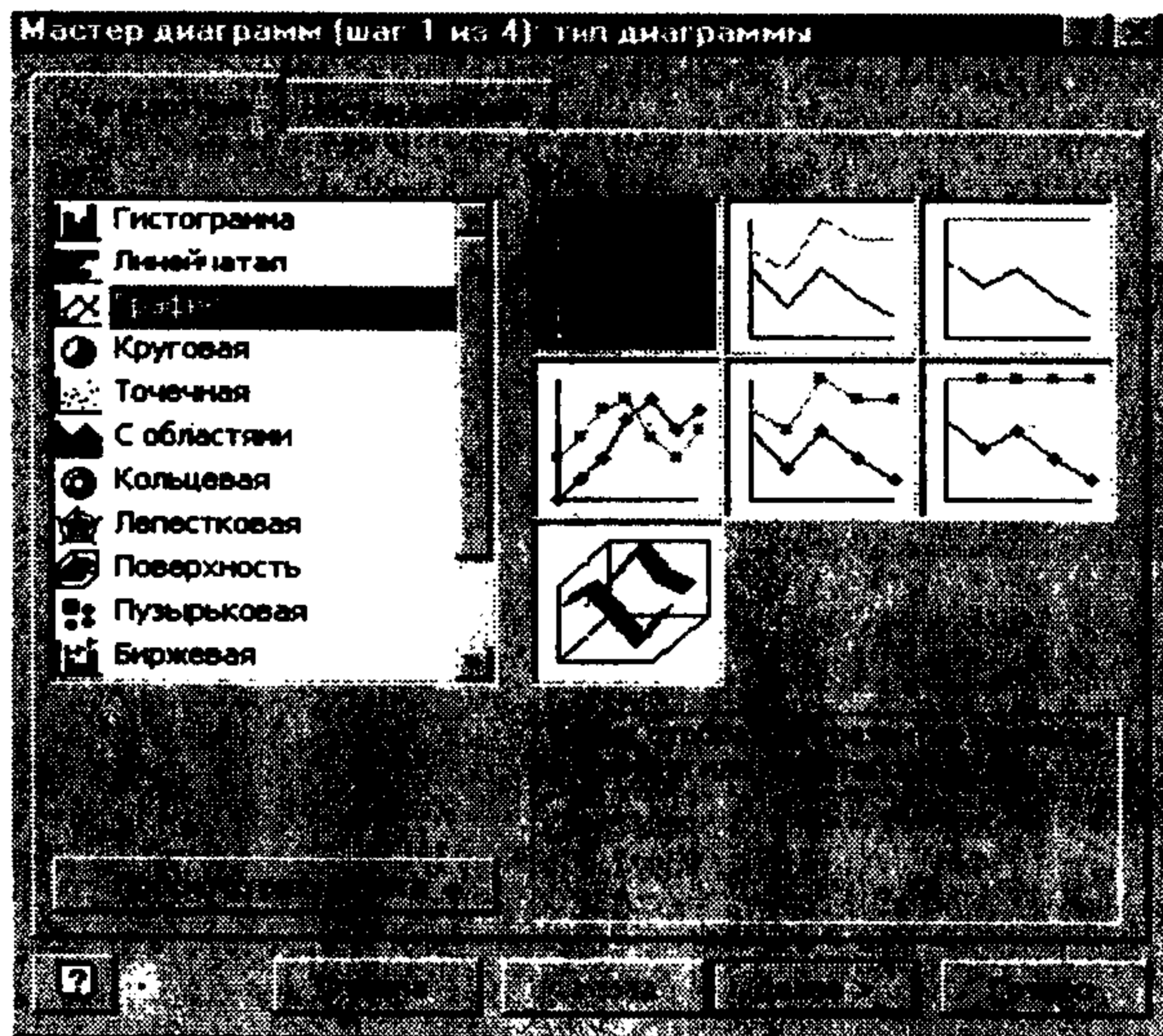


Рис. 4.4. Диалоговое окно Мастера диаграмм: тип диаграммы

4) заполните диапазон данных, как показано на рис. 4.5. Установите флажок размещения данных в столбцах (строках). Щелкните по кнопке **Далее**;

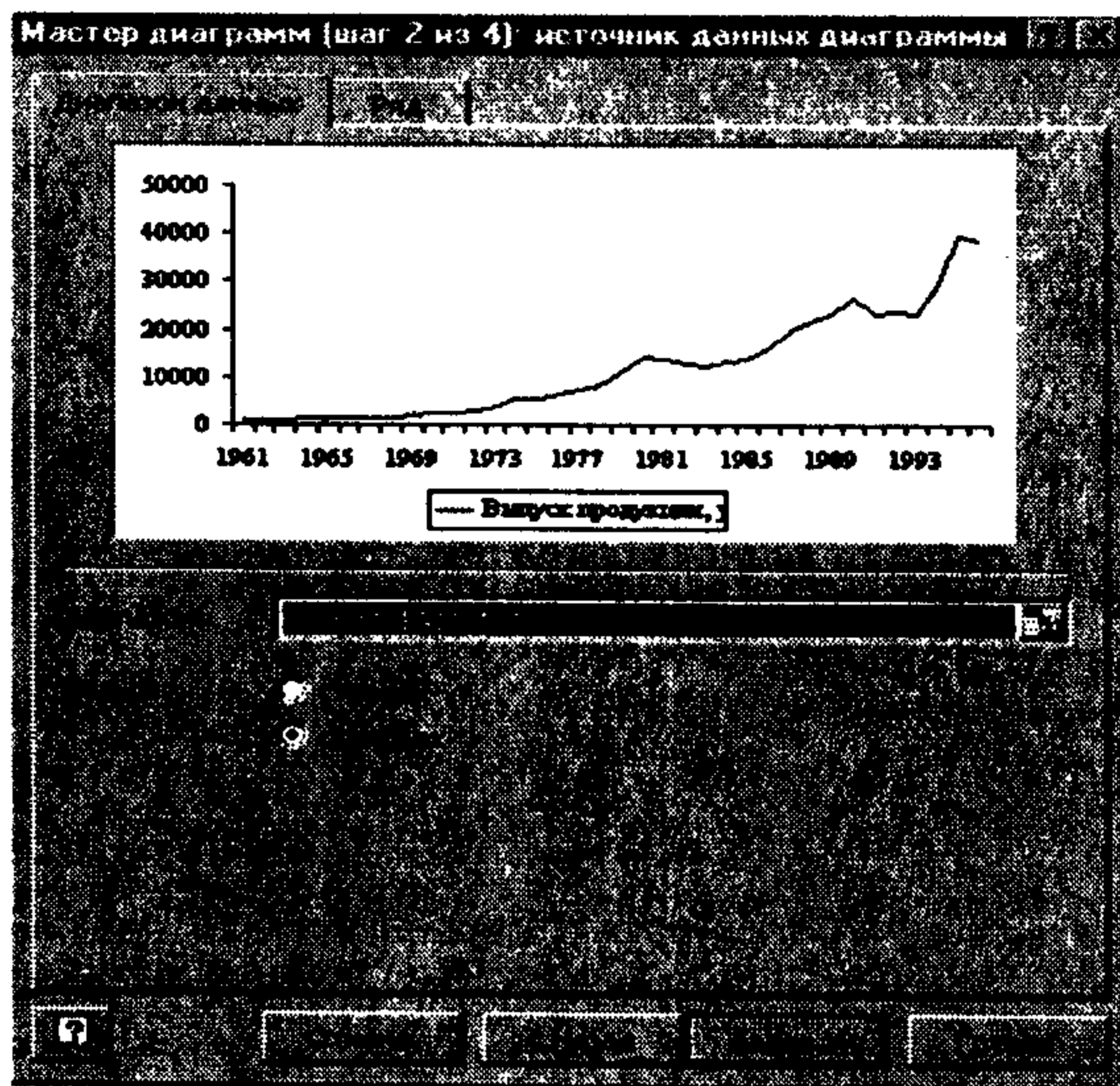


Рис. 4.5. Диалоговое окно Мастера диаграмм: источник данных

5) заполните параметры диаграммы на разных закладках (рис. 4.6): названия диаграммы и осей, значения осей, линии сетки, параметры легенды, таблица и подписи данных. Щелкните по кнопке Далее;

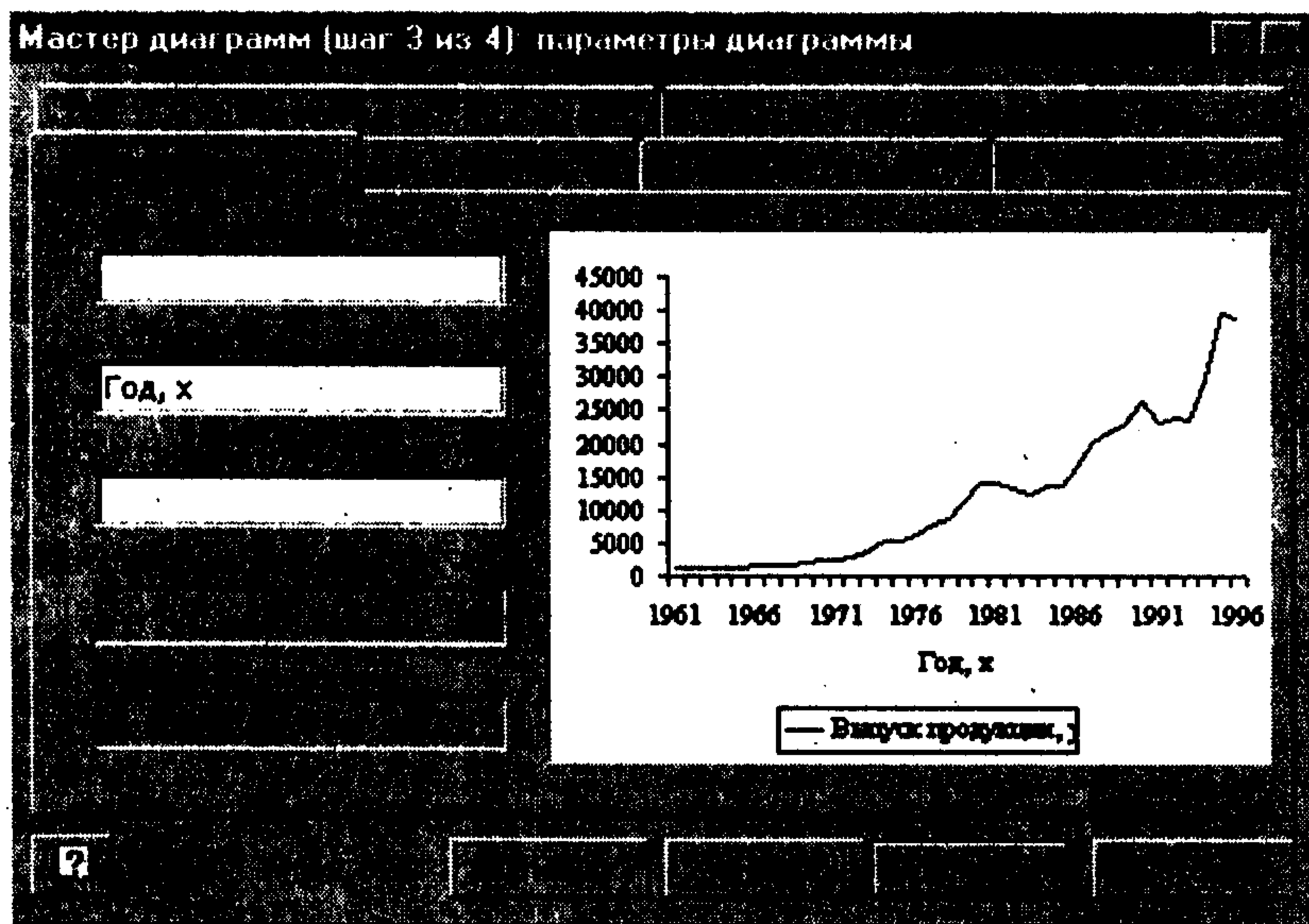


Рис. 4.6. Диалоговое окно Мастера диаграмм: параметры диаграммы

б) укажите место размещения диаграммы на отдельном или на имеющемся листе (рис. 4.7). Щелкните по кнопке Далее. Готовая диаграмма, отражающая динамику уровней изучаемого ряда, представлена на рис. 4.8.

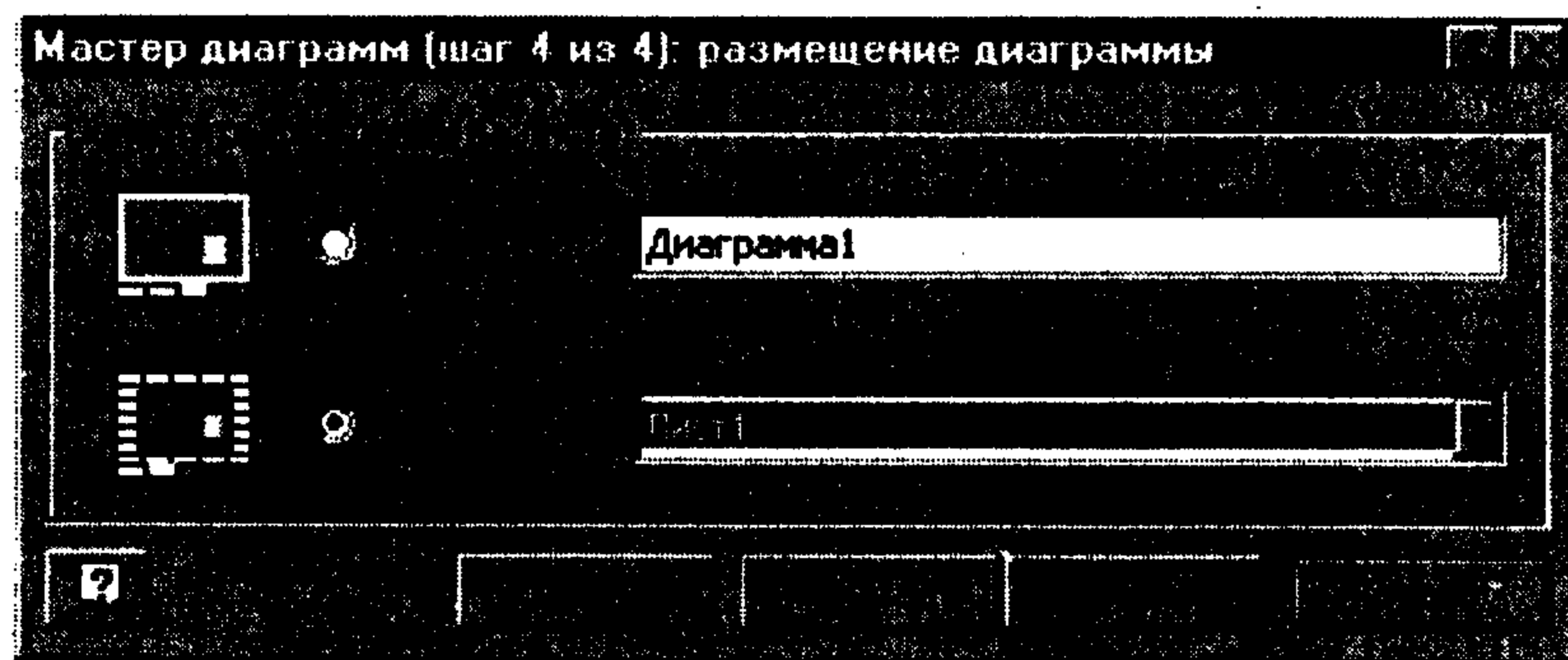


Рис. 4.7. Диалоговое окно Мастера диаграмм: размещение диаграммы

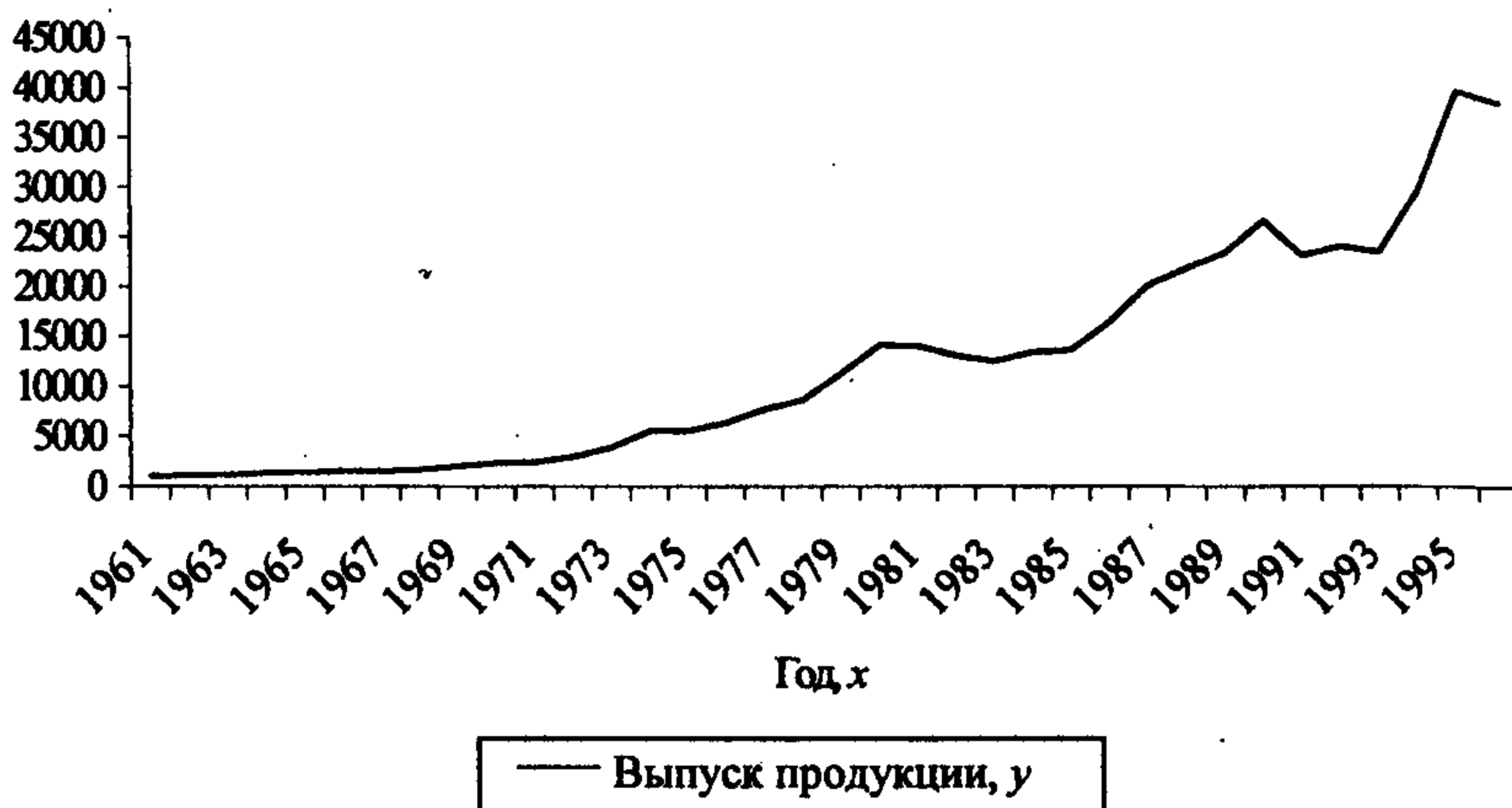


Рис. 4.8. Динамика выпуска продукции

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. 4.9) выберите вид линии тренда и задайте соответствующие параметры. Для полиномиального тренда необходимо задать степень аппроксимирующего полинома, для скользящего среднего – количество точек усреднения.

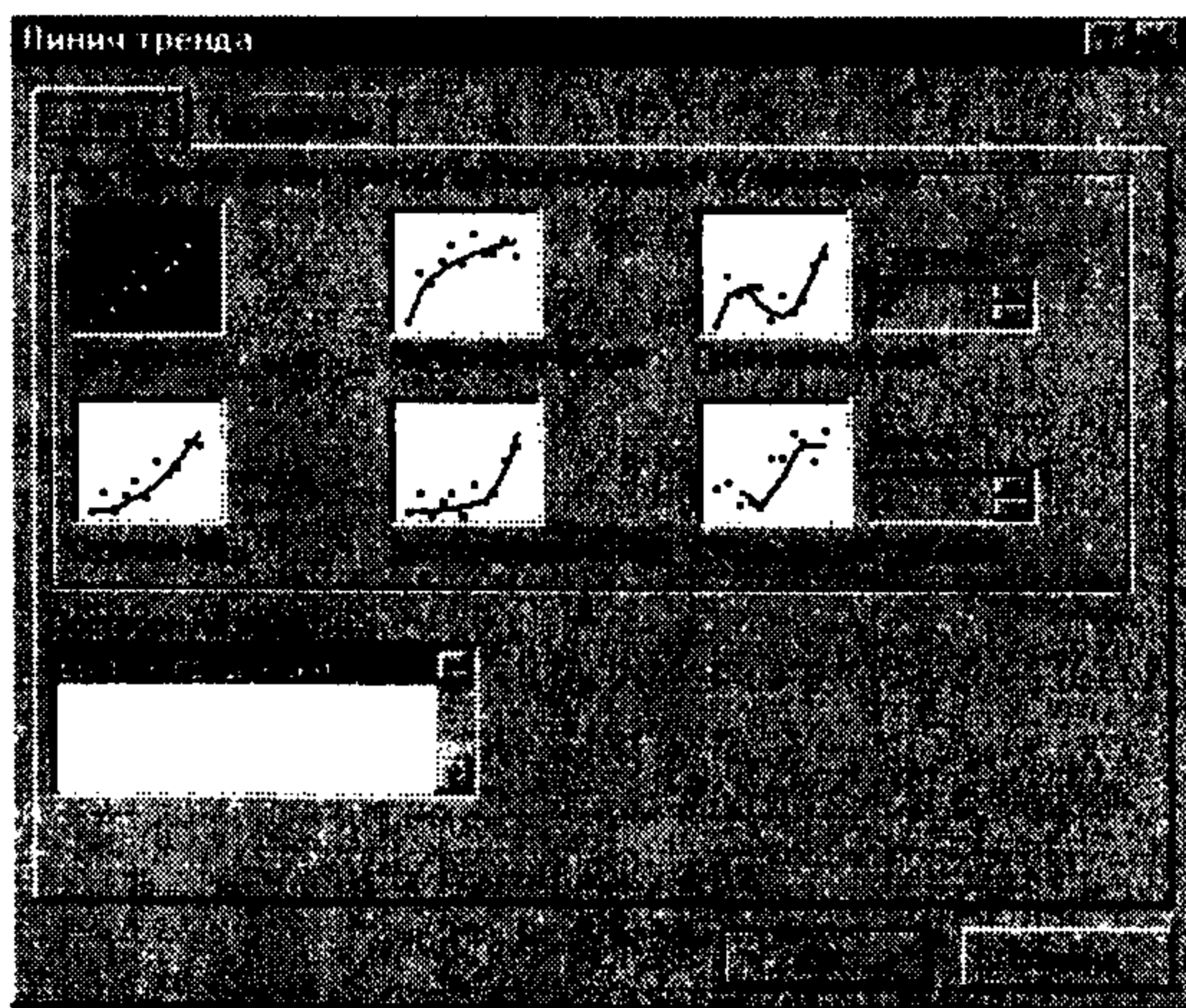


Рис. 4.9. Диалоговое окно типов линий тренда

В качестве дополнительной информации на диаграмме можно отобразить уравнение регрессии и значение среднеквадратического отклонения, установив соответствующие флажки на закладке Параметры (рис. 4.10). Щелкните по кнопке ОК.

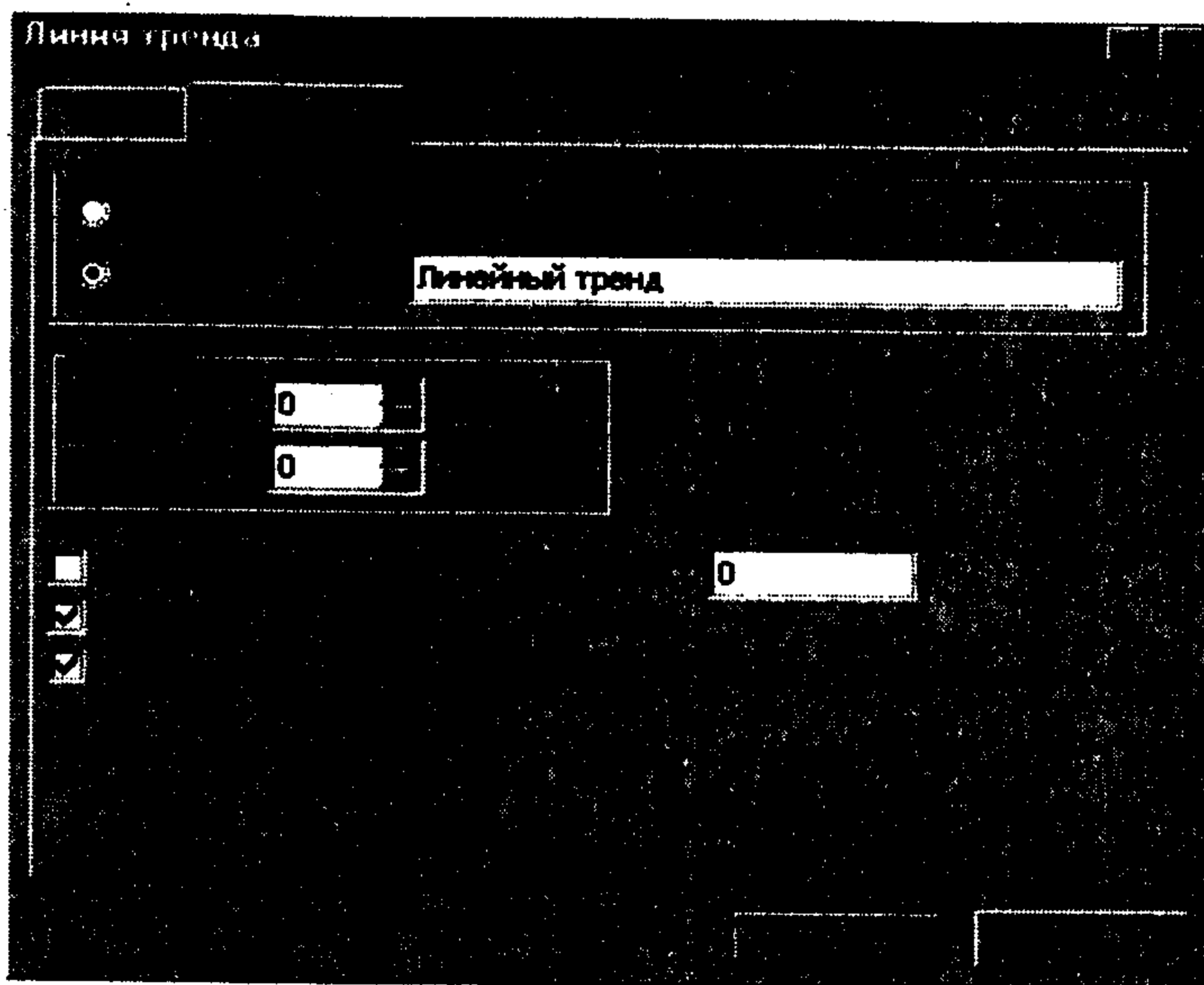


Рис. 4.10. Диалоговое окно параметров линии тренда

На рис. 4.11 – 4.15 представлены различные виды трендов, описывающие исходные данные задачи.

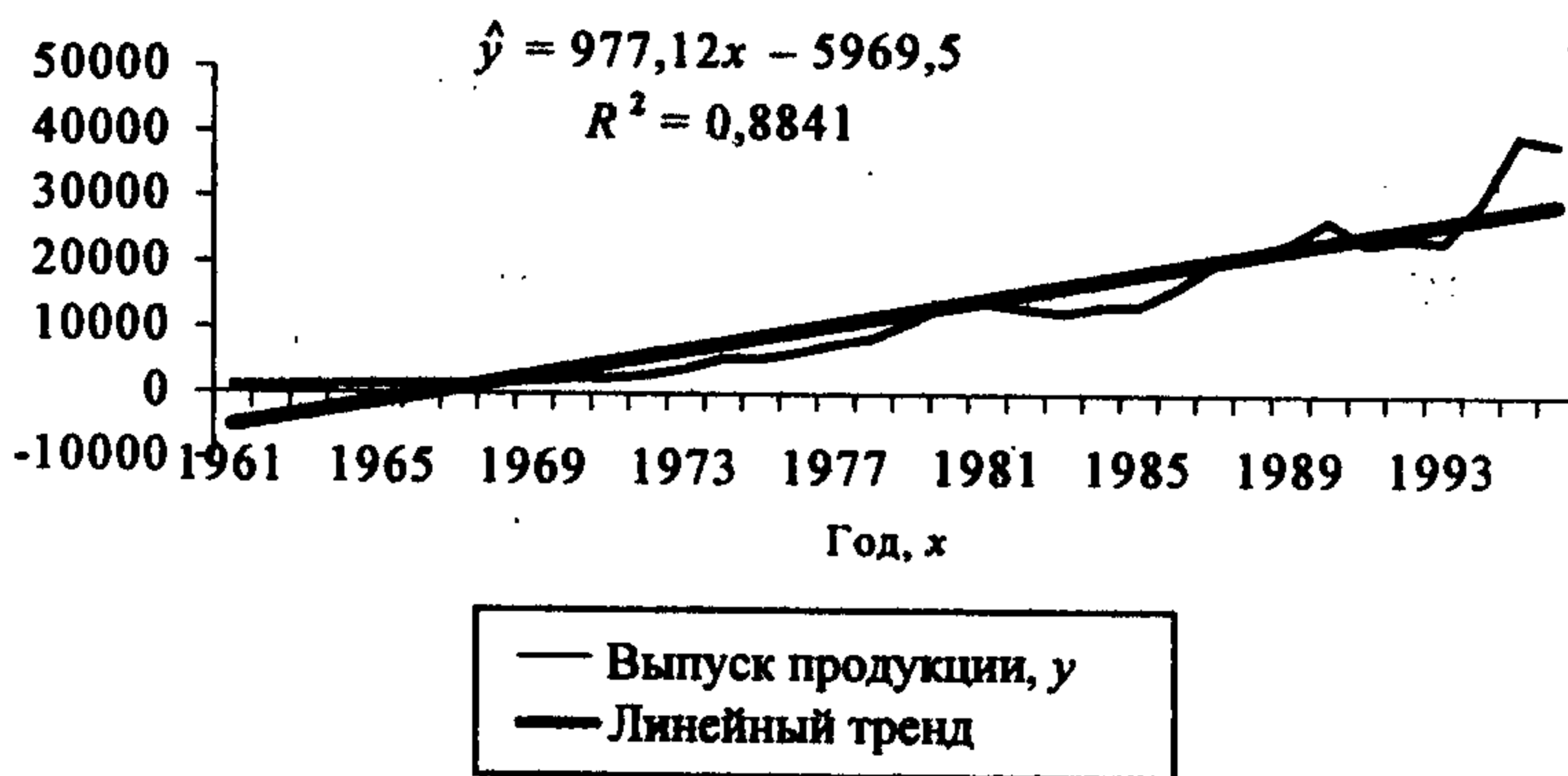
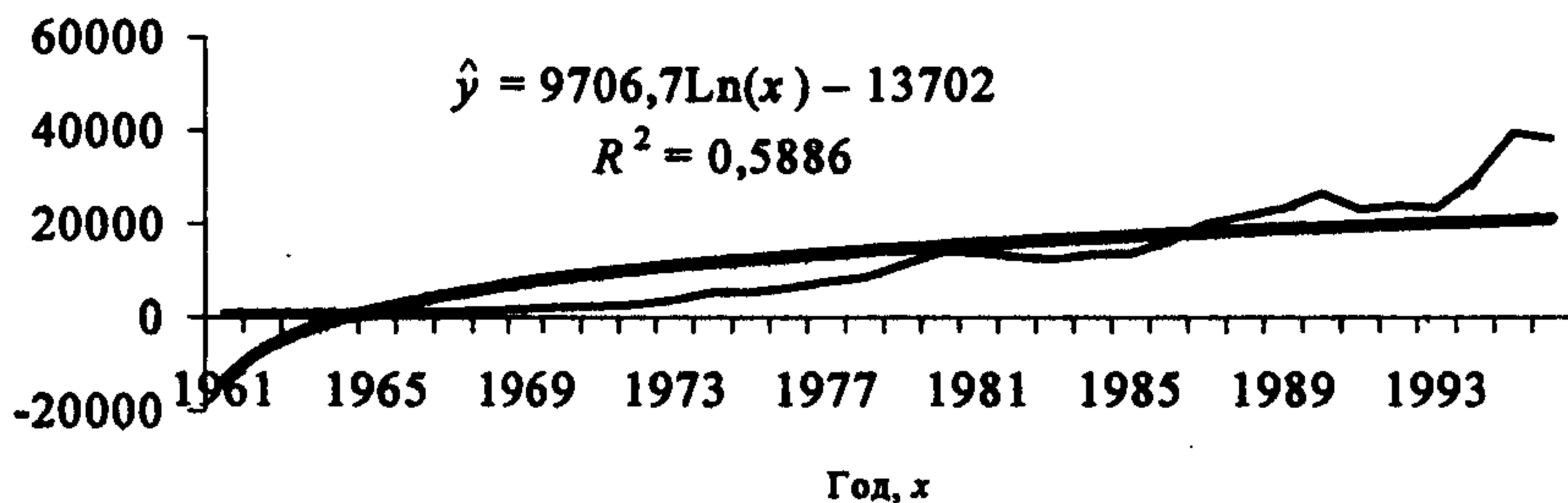
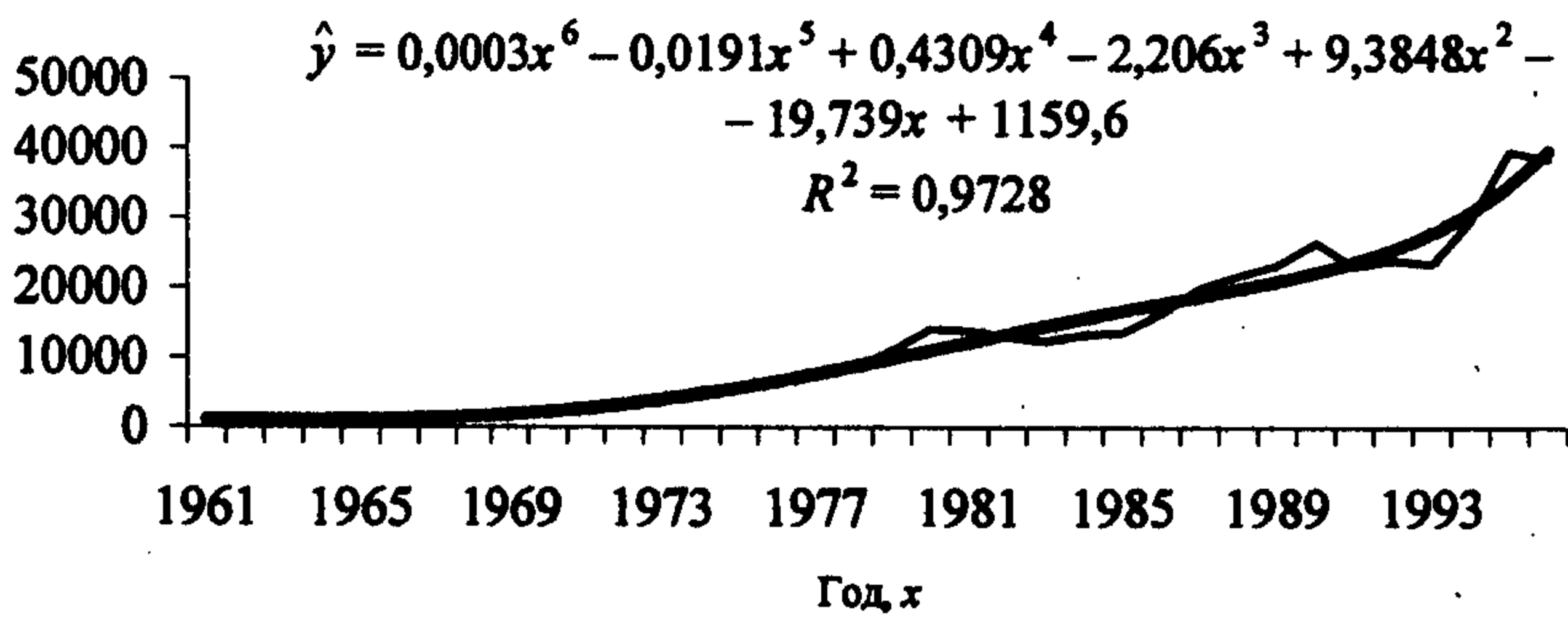


Рис. 4.11. Линейный тренд



— Выпуск продукции, y
 — Логарифмический тренд

Рис. 4.12. Логарифмический тренд



— Выпуск продукции, y
 — Полиномиальный тренд 6-й степени

Рис. 4.13. Полиномиальный тренд

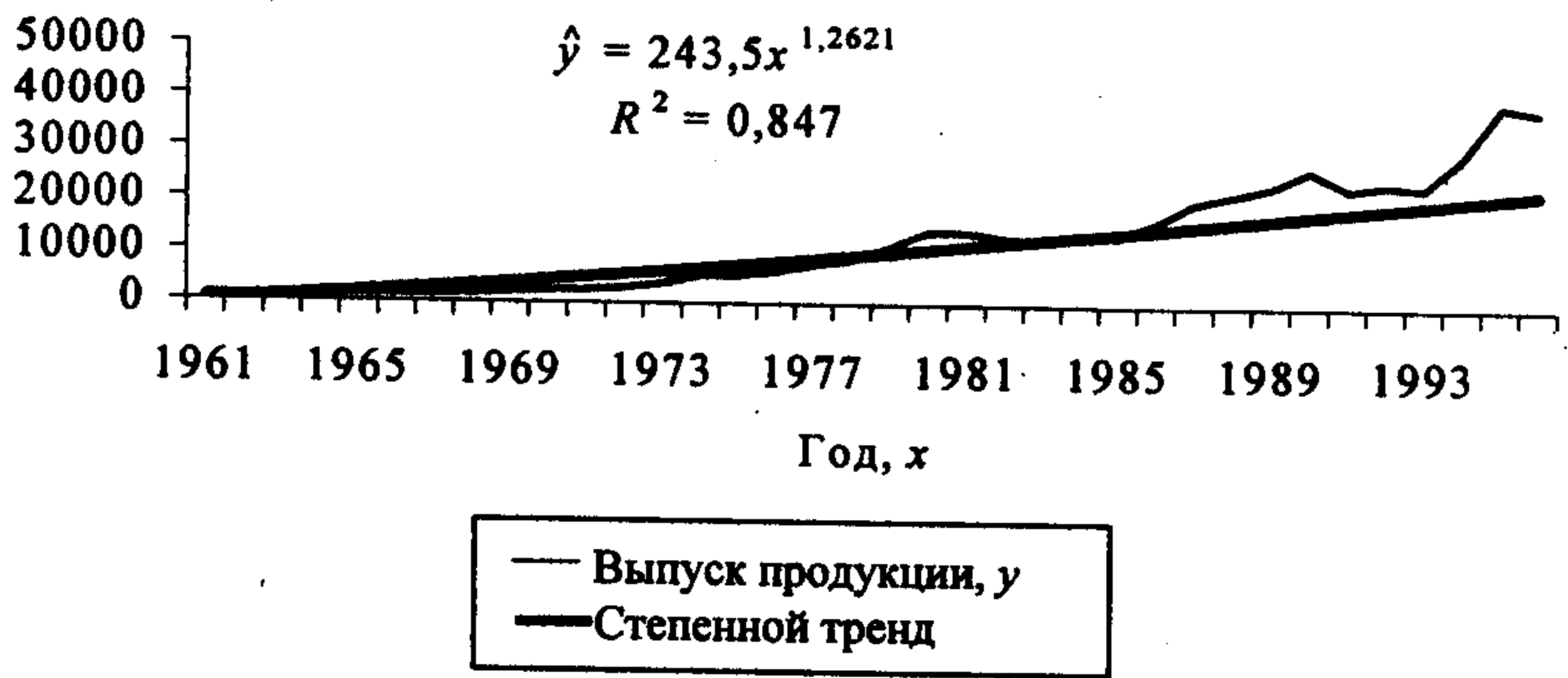


Рис. 4.14. Степенной тренд

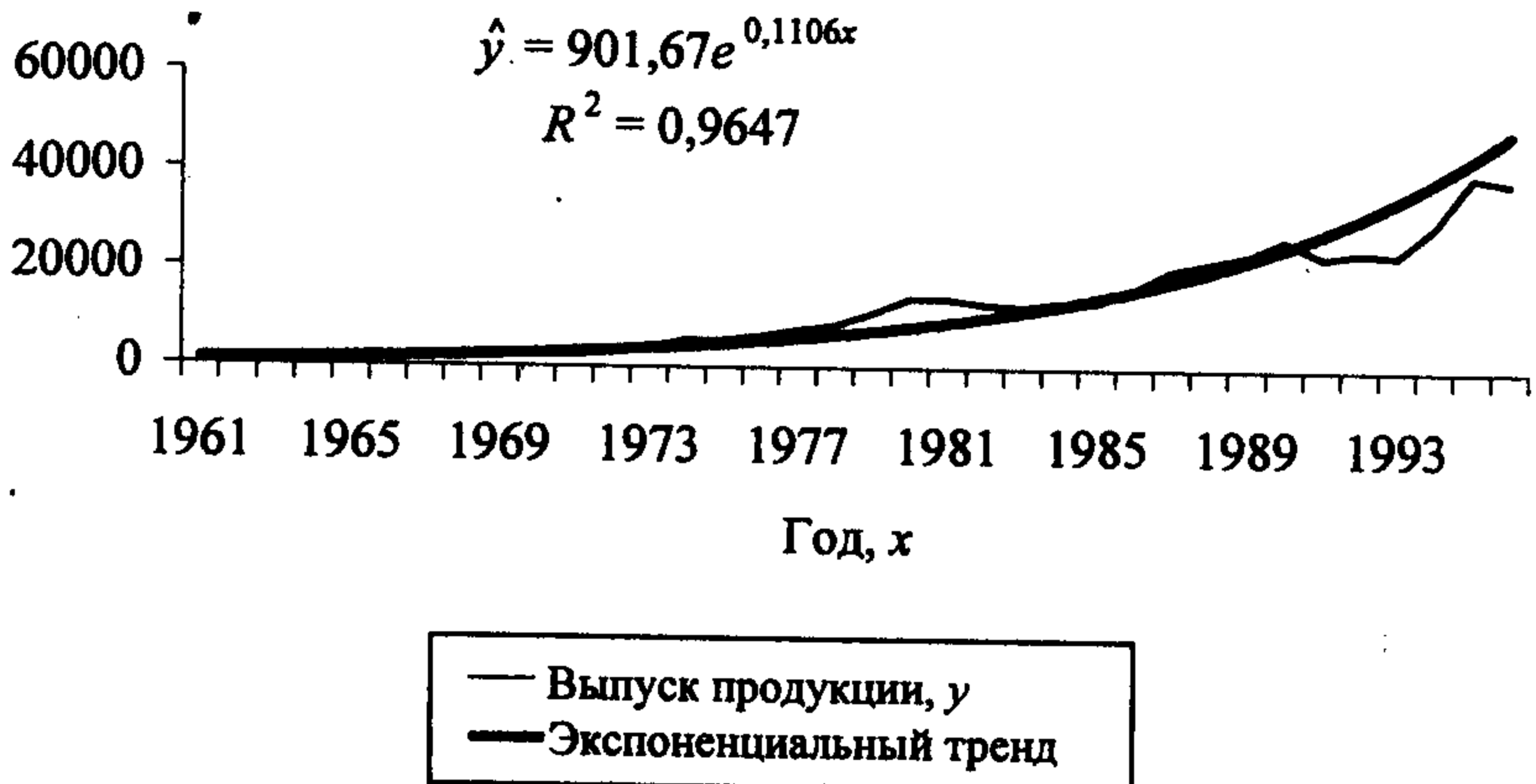


Рис. 4.15. Экспоненциальный тренд

3. Сравним значения R^2 по разным уравнениям трендов:
- полиномиальный 6-й степени — $R^2 = 0,9728$;
 - экспоненциальный — $R^2 = 0,9647$;
 - линейный — $R^2 = 0,8841$;
 - степенной — $R^2 = 0,8470$;
 - логарифмический — $R^2 = 0,5886$.

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

Пример 6

Имеются данные о динамике товарооборота и доходов населения России за 1997 – 1999 гг. (табл. 4.8).

Таблица 4.8

Месяц	Товарооборот, % к предыдущему месяцу	Доходы населения, % к предыдущему месяцу	Месяц	Товарооборот, % к предыдущему месяцу	Доходы населения, % к предыдущему месяцу
Январь	91,5	79,5	Июль	102,3	102,6
Февраль	92,8	100,3	Август	106,8	96,6
Март	104,3	102,9	Сентябрь	96,7	81,5
Апрель	101,5	106,6	Октябрь	92,7	107,8
Май	97,9	92,5	Ноябрь	100,4	69,7
Июнь	98,7	110,1	Декабрь	108,1	122,8
Июль	100,8	96,6	Январь	80,0	63,9
Август	103,7	97,1	Февраль	96,9	107,4
Сентябрь	104,6	98,5	Март	106,0	103,7
Октябрь	100,3	105,7	Апрель	97,6	108,1
Ноябрь	101,5	97,4	Май	100,2	93,9
Декабрь	116,0	129,9	Июнь	100,7	104,1
Январь	82,3	63,9	Июль	100,0	97,2
Февраль	91,6	104,3	Август	106,5	104,6
Март	103,4	101,7	Сентябрь	100,5	98,6
Апрель	100,3	105,5	Октябрь	102,1	104,5
Май	99,2	91,3	Ноябрь	100,5	99,9
Июнь	99,0	102,6	Декабрь	116,0	136,9

Требуется:

1. Оценить параметры модели с распределенными лагами методом Алмон.
2. Постройте таблицу результатов дисперсионного анализа. Оцените значимость построенной модели.

Решение

Решение с использованием ППП Statistica

1. Для построения регрессионной модели с распределенными лагами необходимо априори задать длину максимального лага, для этой задачи выберем длину 3. Тогда уравнение регрессии будет выглядеть следующим образом:

$$\hat{y}_t = a + b_0 \cdot x_t + b_1 \cdot x_{t-1} + b_2 \cdot x_{t-2} + b_3 \cdot x_{t-3} + \varepsilon_t.$$

Для оценки параметров этой модели согласно методу Алмон необходимо задать степень аппроксимирующего полинома. Для решения используем соответствующую процедуру ППП Statistica. Порядок расчетов следующий:

1) введите исходные данные или откройте существующий файл другого формата, содержащий анализируемые данные, в опции **Data Management** в окне переключения модулей (рис. 4.16). Если создаете новый файл данных, в соответствующих ячейках укажите количество строк и столбцов. В нашем случае – 2 столбца, 36 строк;

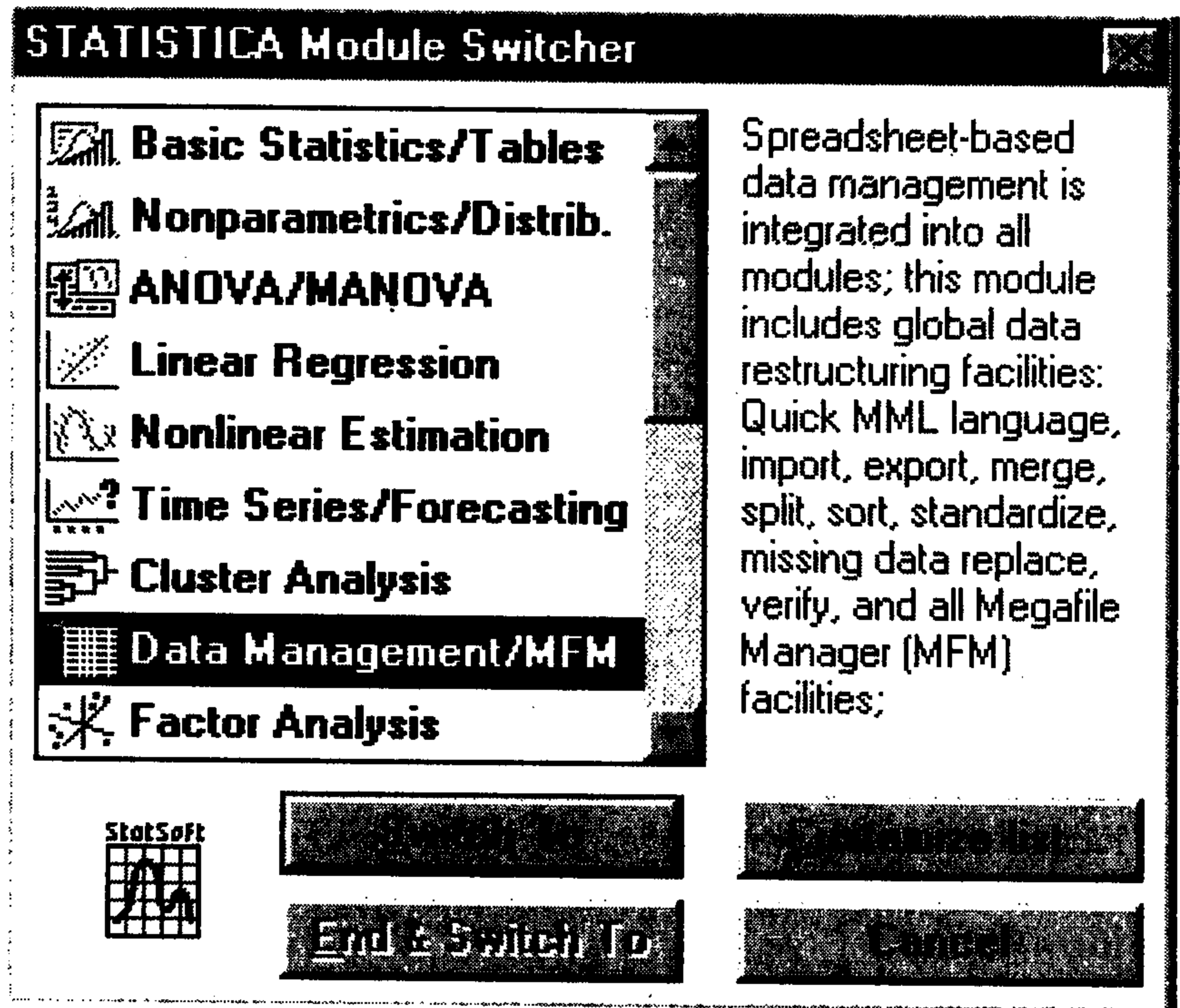


Рис. 4.16. Окно переключения модулей

2) из модуля управления данными перейдите в модуль анализа временных рядов, выбрав в меню пункт **Time Series / Forecasting**;

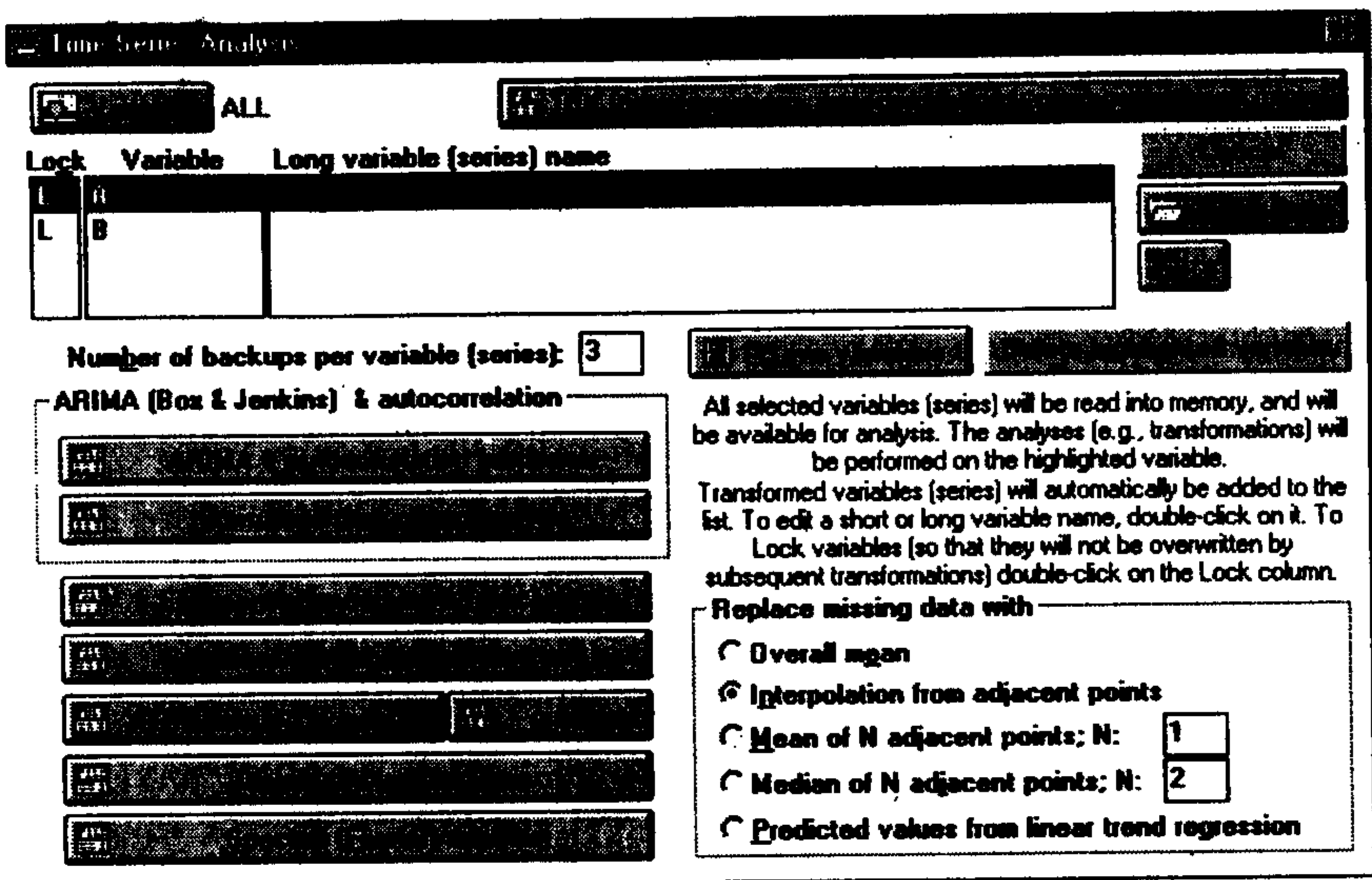


Рис. 4.17. Окно анализа временных рядов

- 3) откройте файл, содержащий данные – **Open Data** (рис. 4.17);
- 4) выделите все переменные, используемые для анализа, – **Variables**. Щелкните по кнопке **ОК** (рис. 4.18).

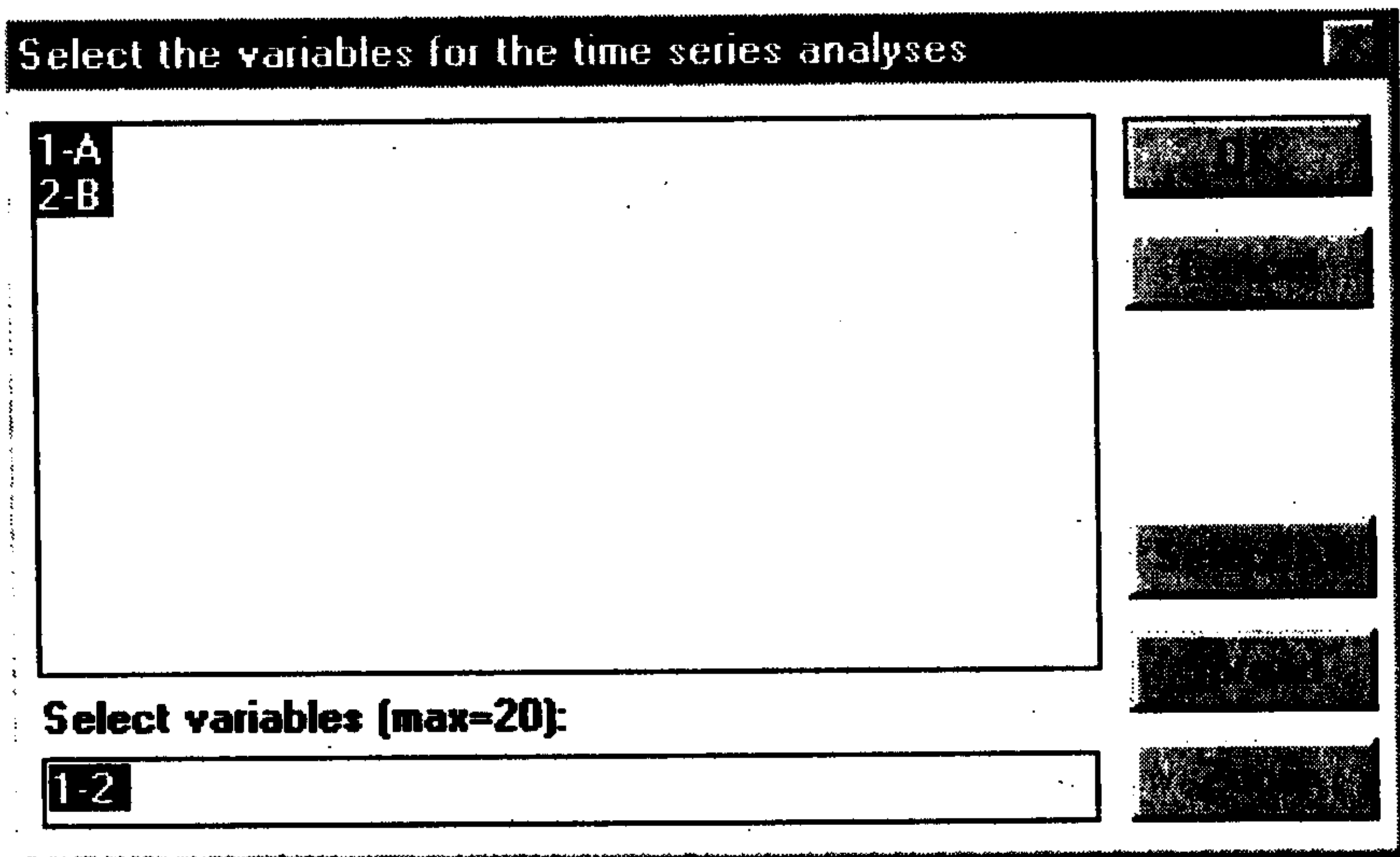


Рис. 4.18. Выбор переменных для анализа

5) щелкните по кнопке **Distributed lags analysis** (см. рис. 4.17);

Рис. 4.19. Окно анализа моделей с распределенными лагами

6) в окне **Distributed Lags Analysis** (рис. 4.19) выделите название зависимой переменной, в появляющемся окне **Independent variable** – название независимой переменной. В ячейке **Lag length** укажите значение максимального лага, в ячейке **Almon polynomial lags** – степень аппроксимирующего полинома. Степень полинома не должна превышать значение максимального лага. Щелкните по кнопке **ОК (Begin analysis)**;

7) результаты расчетов – оценки регрессионных коэффициентов и значимость уравнения – приведены на рис. 4.20 и 4.21.

Dep: A			
Lag: 3 Polyn. order: 2 R=.9989 R-square=.9978 N:33			
Lag			
	.056425825634	8.817606775913	.00000001057
	.248329467570	6.471305334349	.000000440953
	.127333358077	3.402914394983	.001965520839
	.134552413965	2.280682712664	.030099210887

Рис. 4.20. Оценки параметров уравнения с распределенным лагом

TIME SERIES	Dep: A Lag: 3 Polyn. order: 2 R=.9989 R-square=.9978 N:33				
Effect		df			
Regression	333872.9	4	83468.22	3347.999	0.00
	723.0		24.93		
	334595.9				

Рис. 4.21. Результаты дисперсионного анализа

2. Согласно данным таблицы дисперсионного анализа (см. рис. 4.21), полученные значения F -критерия Фишера и коэффициента детерминации R^2 показывают высокий уровень аппроксимации исходных данных.

4.4. КОНТРОЛЬНЫЕ ЗАДАНИЯ

Задача 1

Администрация банка изучает динамику депозитов физических лиц за ряд лет (млн долл. в сопоставимых ценах). Исходные данные представлены ниже:

Время, лет	1	2	3	4	5	6	7	Сумма
Депозиты физических лиц, x ..	2	6	7	3	10	12	13	53

Известно также следующее: $\sum x^2 = 511$.

Задание

1. Постройте уравнение линейного тренда и дайте интерпретацию его параметров.
2. Определите коэффициент детерминации для линейного тренда.
3. Администрация банка предполагает, что среднегодовой абсолютный прирост депозитов физических лиц составляет не менее 2,5 млн долл. Подтверждается ли это предположение результатами, которые вы получили?

Задача 2

Изучается динамика потребления мяса в регионе. Для этого были собраны данные об объемах среднедушевого потребления мяса y_t (кг) за 7 месяцев. Предварительная обработка данных путем логарифмирования привела к получению следующих результатов:

Месяц	1	2	3	4	5	6	7
$\ln y_t$	2,10	2,11	2,13	2,17	2,22	2,28	2,31

Задание

1. Постройте уравнение экспоненциального тренда.
2. Дайте интерпретацию его параметров.

Задача 3

Имеются данные об урожайности зерновых в хозяйствах области:

Год	Урожайность зерновых, ц/га
1	10,2
2	10,7
3	11,7
4	13,1
5	14,9
6	17,2
7	20,0
8	23,2

Задание

1. Обоснуйте выбор типа уравнения тренда.
2. Рассчитайте параметры уравнения тренда.
3. Дайте прогноз урожайности зерновых на следующий год.

Задача 4

Имеются следующие данные об уровне безработицы y_t (%) за 8 месяцев:

Месяц.....	1	2	3	4	5	6	7	8
y_t	8,8	8,6	8,4	8,1	7,9	7,6	7,4	7,0

Задание

1. Определите коэффициенты автокорреляции уровней этого ряда первого и второго порядка.
2. Обоснуйте выбор уравнения тренда и определите его параметры.
3. Интерпретируйте полученные результаты.

Задача 5

Пусть имеется следующий временной ряд:

t	1	2	3	4	5	6	7	8
x_t	20	10

Известно также, что $\sum x_t = 150$, $\sum x_t^2 = 8100$, $\sum_{t=2}^n x_t x_{t-1} = 7350$.

Задание

1. Определите коэффициент автокорреляции уровней этого ряда первого порядка.
2. Установите, включает ли исследуемый временной ряд тенденцию.

Задача 6

Экспорт, импорт, внешнеторговый оборот Австрии и Бельгии за 1961 – 1995 гг. характеризуются данными, представленными в табл. 4.9.

Таблица 4.9

Год	Австралия, млн шиллингов			Бельгия, млн франков		
	Экспорт	Импорт	Внешне- торговый оборот	Экспорт	Импорт	Внешне- торговый оборот
1961	44	43	87	202	209	411
1962	47	46	93	219	221	440
1963	51	51	102	239	248	487
1964	56	56	112	278	283	561
1965	62	63	125	306	305	611
1966	67	71	138	328	337	665
1967	72	74	146	352	351	703
1968	79	80	159	402	400	802
1969	95	91	186	483	474	957
1970	117	131	248	562	533	1095
1971	129	126	255	609	581	1190
1972	146	144	290	683	633	1316
1973	166	164	330	846	811	1657
1974	204	206	410	1116	1109	2225
1975	209	205	414	1065	1061	2126
1976	236	247	483	1266	1261	2527
1977	257	278	535	1474	1499	2973
1978	281	280	561	1540	1570	3110
1979	328	332	660	1798	1866	3664
1980	366	386	752	2026	2125	4151
1981	405	419	824	2286	2357	4643
1982	431	412	843	2640	2694	5334
1983	450	434	884	2924	2864	5788
1984	498	496	994	3337	3277	6614
1985	549	547	1096	3479	3379	6858
1986	523	510	1033	3367	3187	6554
1987	527	520	1047	3477	3334	6811
1988	590	584	1174	3900	3719	7619
1989	669	661	1330	4498	4320	8818

Год	Австралия, млн шиллингов			Бельгия, млн франков		
	Экспорт	Импорт	Внешне- торговый оборот	Экспорт	Импорт	Внешне- торговый оборот
1990	737	720	1457	4660	4506	9166
1991	775	758	1533	4846	4658	9504
1992	792	772	1564	4980	4713	9693
1993	787	773	1560	5012	4674	9686
1994	835	842	1677	5491	5108	10599
1995	887	911	1798	5764	5377	11141

Задание

1. По каждому ряду постройте график динамики.
2. Проведите расчет параметров трендов разной формы.
3. Оцените качество каждого тренда через среднюю ошибку аппроксимации, линейный коэффициент автокорреляции отклонений.
4. Оцените статистическую значимость трендов через F -критерий, значимость параметров тренда – через t -критерий.
5. Выберите лучшую форму тренда и выполните по ней точечный прогноз на 1998 г.
6. Оцените ошибку прогноза и постройте доверительный интервал прогноза для уровня значимости 0,05.

Задача 7

Имеются поквартальные данные по розничному товарообороту России в 1995 – 1999 гг. (табл. 4.10).

Таблица 4.10

Номер квартала	Товарооборот, % к предыдущему периоду	Номер квартала	Товарооборот, % к предыдущему периоду
1	100,0	11	98,8
2	93,9	12	101,9
3	96,5	13	113,1
4	101,8	14	98,4
5	107,8	15	97,3
6	96,3	16	102,1
7	95,7	17	97,6
8	98,2	18	83,7
9	104,0	19	84,3
10	99,0	20	88,4

Задание

1. Постройте график временного ряда.
2. Постройте мультипликативную модель временного ряда.
3. Оцените качество модели через показатели средней абсолютной ошибки и среднего относительного отклонения.

Задача 8

Имеются данные об объеме экспорта из Российской Федерации (млрд долл., цены Фондовой Общероссийской биржи (ФОБ)) за 1994 – 1999 гг. (табл. 4.11).

Таблица 4.11

Номер квартала	Экспорт, млрд долл., цены ФОБ	Номер квартала	Экспорт, млрд долл., цены ФОБ
1	4087	13	6975
2	4737	14	6891
3	5768	15	7527
4	6005	16	7971
5	5639	17	5875
6	6745	18	6140
7	6311	19	6248
8	7107	20	6041
9	5741	21	4626
10	7087	22	6501
11	7310	23	6284
12	8600	24	6707

Задание

1. Постройте график временного ряда.
2. Постройте аддитивную и мультипликативную модели временного ряда.
3. Оцените качество каждой модели через показатели средней абсолютной ошибки и среднего относительного отклонения. Выберите лучшую модель.

Задача 9

Для прогнозирования объема продаж компании ABC (млн руб.) на основе поквартальных данных за 1993–1997 гг. была построена аддитивная модель временного ряда объема продаж. Уравнение, моделирующее динамику трендовой компоненты этой модели, имеет вид: $T = 100 + 2 \cdot t$ (при построении тренда для моделирования переменной времени использовались натуральные числа, начиная с 1). Показатели за 1996 г., полученные в ходе построения аддитивной модели, представлены в табл. 4.12.

Таблица 4.12

Время года	Фактический объем продаж в 1996 г.	Компонента, полученная по аддитивной модели		
		трендовая	сезонная	случайная
Зима	100			+4
Весна			10	+5
Лето	150		25	
Осень				

Задание

Определите недостающие в таблице данные, учитывая, что объем продаж компании ABC за 1996 г. в целом составил 490 млн руб.

Задача 10

Имеются данные о разрешениях на строительство нового частного жилья, выданных в США в 1990–1994 гг., % к уровню 1987 г. (табл. 4.13).

Таблица 4.13

Месяц	1990 г.	1991 г.	1992 г.	1993 г.	1994 г.
Январь	72,9	61,4	71,2	78,3	86,4
Февраль	113,4	51,0	69,9	76,4	87,5
Март	86,2	55,3	74,3	74,5	80,2
Апрель	80,8	59,1	70,2	68,5	84,3
Май	73,7	59,5	68,4	71,6	86,8
Июнь	69,2	64,3	68,5	72,1	86,9
Июль	71,9	62,5	68,6	73,3	85,2
Август	69,9	63,1	70,6	76,2	85,0
Сентябрь	69,4	61,2	69,7	79,8	87,5
Октябрь	63,3	63,2	72,3	81,2	90,0
Ноябрь	60,0	64,3	73,5	83,5	88,4
Декабрь	61,0	63,9	72,5	88,0	85,7

Задание

1. Рассчитайте трендовую и сезонную компоненты.
2. Постройте аддитивную модель этого ряда.
3. Постройте автокорреляционную функцию временного ряда количества разрешений на строительство частного нового жилья. Охарактеризуйте структуру этого ряда.

Задача 11

В табл. 4.14 приводятся данные об объемах продаж в перерабатывающей промышленности и торговле, в сопоставимых ценах 1987 г., млрд долл.

Таблица 4.14

Месяц	1990 г.	1991 г.	1992 г.	1993 г.	1994 г.
Январь	472,5	477,9	510,9	541,0	578,2
Февраль	482,1	467,5	484,7	512,3	539,4
Март	489,5	470,9	486,6	512,6	545,3
Апрель	493,6	469,1	488,4	511,5	551,9
Май	488,0	478,1	489,5	511,9	549,7
Июнь	490,6	480,6	486,6	513,9	550,1
Июль	492,5	479,3	491,8	520,0	554,0
Август	488,1	484,2	495,2	515,9	550,0
Сентябрь	493,1	484,9	491,8	524,2	565,6
Октябрь	484,5	485,6	496,1	527,1	564,7
Ноябрь	483,0	486,1	498,8	529,8	566,9
Декабрь	476,9	484,7	501,5	534,9	572,7

Задание

1. Рассчитайте трендовую и сезонную компоненты.
2. Постройте мультипликативную модель этого ряда.
3. Постройте автокорреляционную функцию временного ряда объема продаж в перерабатывающей промышленности и торговле. Охарактеризуйте структуру этого ряда.

Задача 12

На основе помесечных данных о потреблении электроэнергии в регионе (млн кВт · ч) за последние 3 года была построена аддитивная модель временного ряда. Скорректированные значения сезонной компоненты за соответствующие месяцы приводятся ниже:

январь	+25	май	-32	сентябрь	+2
февраль	+10	июнь	-38	октябрь	+15
март	+6	июль	-25	ноябрь	+27
апрель	-4	август	-18	декабрь	?

Уравнение тренда выглядит следующим образом:

$$T = 300 + 1,5 \cdot t$$

(при расчете параметров тренда для моделирования переменной времени использовались натуральные числа $t = 1 : 36$).

Задание

1. Определите значение сезонной компоненты за декабрь.
2. На основе построенной модели дайте точечный прогноз ожидаемого потребления электроэнергии в течение первого квартала следующего года.

Задача 13

На основе поквартальных данных об уровне безработицы в летнем курортном городе (% от экономически активного населения) за последние 5 лет была построена мультипликативная модель временного ряда. Скорректированные значения сезонной компоненты за каждый квартал приводятся ниже:

I квартал.....1,4	III квартал....0,7
II квартал....0,8	IV квартал.....-

Уравнение тренда выглядит следующим образом:

$$T = 9,2 - 0,3t$$

(при расчете параметров тренда для нумерации кварталов использовались натуральные числа $t = 1 + 20$).

Задание

1. Определите значения сезонной компоненты за IV квартал.
2. На основе построенной модели дайте точечные прогнозы уровня безработицы на I и II квартал следующего года.

Задача 14

В целях прогнозирования объема экспорта страны на будущие периоды были собраны данные за 30 лет по следующим показателям: y_t – объем экспорта (млрд долл., в сопоставимых ценах); x_t – индекс физического объема промышленного производства (в % к предыдущему году). Ниже представлены результаты предварительной обработки исходных данных.

1. Уравнения линейных трендов:

а) для ряда Y_t

$$Y_t = 3,1 + 1,35t + \varepsilon_t \quad R^2 = 0,91 \quad d = 2,31;$$

б) для ряда X_t

$$X_t = -8,4 + 4,8t + \varepsilon_t \quad R^2 = 0,89 \quad d = 2,08.$$

2. Уравнение регрессии по уровням временных рядов:

$$Y_t = -10,5 + 0,5X_t + \varepsilon_t \quad R^2 = 0,95 \quad d = 2,21.$$

3. Уравнение регрессии по первым разностям уровней временных рядов:

$$\Delta Y_t = 1,4 + 0,03\Delta X_t + \varepsilon_t \quad R^2 = 0,86 \quad d = 2,25.$$

4. Уравнение регрессии по вторым разностям уровней временных рядов:

$$\Delta^2 Y_t = 0,7 + 0,012\Delta^2 X_t + \varepsilon_t \quad R^2 = 0,47 \quad d = 2,69.$$

5. Уравнение регрессии по уровням временных рядов с включением фактора времени:

$$Y_t = 4,23 + 0,24X_t + 0,78t + \varepsilon_t \quad R^2 = 0,97 \quad d = 0,9.$$

Задание

1. Сформулируйте свои предположения относительно величины коэффициента автокорреляции первого порядка в каждом из рядов. Ответ обоснуйте.

2. Выберите наилучшее уравнение регрессии, которое можно использовать для прогнозирования объема экспорта, и дайте интерпретацию его параметров.

3. Пусть известна информация за последние три года (табл. 4.15).

Таблица 4.15

Год (номер периода t)	28	29	30	31
y_t	38	742	43	???
x_t	120	126	121	124

Используя выбранное вами в п. 2 уравнение, дайте точечный прогноз ожидаемого значения y_t на ближайший год (период 31).

Задача 15

Изучается зависимость объема продаж бензина (y_t) от динамики потребительских цен (x_t). Полученные за последние 6 кварталов данные представлены в табл. 4.16.

Таблица 4.16

Показатель	1 кв.	2 кв.	3 кв.	4 кв.	5 кв.	6 кв.
Индекс потребительских цен, % к кварталу 1	100	104	112	117	121	126
Средний за день объем продаж бензина в течение квартала, тыс. л	89	83	80	77	75	72

Известно также, что $\sum x_t = 680$, $\sum y_t = 476$, $\sum x_t y_t = 53648$, $\sum x_t^2 = 77566$.

Задание

1. Постройте модель зависимости объема продаж бензина от индекса потребительских цен с включением фактора времени.
2. Дайте интерпретацию параметров полученной вами модели.

Задача 16

Годовое потребление товара A и доходы населения (тыс. руб.) за 1989–1997 гг. приведены в табл. 4.17.

Таблица 4.17

Показатель	1989 г.	1990 г.	1991 г.	1992 г.	1993 г.	1994 г.	1995 г.	1996 г.	1997 г.
Потребление	46	50	54	59	62	67	75	86	100
Доходы	53	57	64	70	73	82	95	110	127

Задание

1. Определите уравнение регрессии, включив в него фактор времени, если известно, что $\Sigma Y = 599$, $\Sigma X = 731$, $\Sigma YX = 52179$, $\Sigma X^2 = 64361$, $\Sigma Y^2 = 42367$.
2. Интерпретируйте полученные результаты.

Задача 17

В табл. 4.18 приводятся данные об уровне дивидендов, выплачиваемых по обыкновенным акциям (в процентах), и среднегодовой стоимости основных фондов компании (X , млн руб.) в сопоставимых ценах за последние девять лет.

Таблица 4.18

Показатель	1	2	3	4	5	6	7	8	9
Среднегодовая стоимость основных фондов	72	75	77	77	79	80	78	79	80
Дивиденды по обыкновенным акциям	4,2	3,0	2,4	2,0	1,9	1,7	1,8	1,6	1,7

Задание

1. Определите параметры уравнения регрессии по первым разностям и дайте их интерпретацию. В качестве зависимой переменной используйте показатель дивидендов по обыкновенным акциям.
2. В чем состоит причина построения уравнения регрессии по первым разностям, а не по исходным уровням рядов?

Задача 18

В табл. 4.19 приводятся данные о потреблении и личных доходах населения за 1985–1991 гг.

Таблица 4.19

Показатель	1985 г.	1986 г.	1987 г.	1988 г.	1989 г.	1990 г.	1991 г.
Потребление, тыс. долл.	300	310	325	340	350	370	385
Личные доходы, тыс. долл.	335	340	360	378	400	417	430

Задание

1. Постройте уравнение линейной регрессии, используя метод первых разностей.
2. Охарактеризуйте тесноту связи между рядами по их уровням, по первым разностям. Сделайте выводы.

Задача 19

По данным за 30 лет изучается зависимость рентабельности продукции компании y_t (%) от численности занятых ручным трудом x_t (чел.). Были получены следующие варианты уравнений регрессии:

а) по уровням временных рядов:

$$\hat{Y}_t = 2 - 0,5X_t + \varepsilon_t \quad R^2 = 0,9025 \quad d = 0,8;$$

б) по первым разностям уровней:

$$\Delta \hat{Y}_t = 3 + 0,10\Delta X_t + \varepsilon_t \quad R^2 = 0,49 \quad d = 1,2;$$

в) по вторым разностям уровней:

$$\Delta^2 \hat{Y}_t = 15 - 0,062\Delta^2 X_t + \varepsilon_t \quad R^2 = 0,7225 \quad d = 2,1;$$

г) по уровням рядов с включением фактора времени:

$$\hat{Y}_t = -7 - 0,02X_t + 0,3 \cdot t + \varepsilon_t \quad R^2 = 0,95 \quad d = 2,2.$$

$$t_\phi = \quad -(3,1) \quad (3,7)$$

В табл. 4.20 приведены известные коэффициенты автокорреляции первого порядка.

Таблица 4.20

Ряд	По уровням ряда	По первым разностям уровней ряда	По вторым разностям уровней ряда
x_t	0,99	0,80	0,05
y_t	0,86	0,86	0,10

Задание

1. Определите коэффициенты корреляции по уровням временных рядов, по первым разностям временных рядов и по вторым разностям временных рядов. Охарактеризуйте тесноту связи между временными рядами рентабельности продукции и численности занятых ручным трудом. Обоснуйте ваш выбор одной из мер тесноты связи.
2. Исследуйте полученные уравнения регрессии на автокорреляцию в остатках.
3. Выберите наилучшее уравнение регрессии и дайте интерпретацию его параметров.

Задача 20

Имеются данные за десять лет (1987–1996 гг.) о производительности труда и электровооруженности труда на одном из предприятий промышленности области (табл. 4.21).

Таблица 4.21

Показатель	1987 г.	1988 г.	1989 г.	1990 г.	1991 г.	1992 г.	1993 г.	1994 г.	1995 г.	1996 г.
Среднегодовая выработка продукции на 1 рабочего, усл. ед., у	28,7	31,7	31,7	32,6	33,9	31,2	33,3	42,6	46,0	49,9
Электровооруженность, кВт·ч/чел.-ч, х	3,33	3,39	3,50	3,63	3,81	3,84	3,88	4,07	4,12	4,17

Результаты аналитического выравнивания привели к следующим уравнениям трендов для каждого из рядов:

а) для временного ряда производительности труда:

$$\hat{y}_t = 33,19 + 1,04 \cdot t + 0,09 \cdot t^2 \quad (t = -9, -7, -5, -3, -1, 1, 3, 5, 7, 9);$$

б) для временного ряда электровооруженности:

$$\hat{x}_t = 3,774 + 0,049 \cdot t \quad (t = -9, -7, -5, -3, -1, 1, 3, 5, 7, 9).$$

Задание

1. Определите коэффициент корреляции между временными рядами, используя непосредственно исходные уровни, первые разности для электровооруженности и вторые разности для производительности труда, отклонения от основной тенденции.
2. Объясните различия полученных результатов.
3. Рассчитайте коэффициент автокорреляции внутри каждого временного ряда.

Задача 21

На основе данных за последние 20 лет изучается зависимость между уровнем дивидендов по обыкновенным акциям y (%) от прибыли компании x (тыс. долл.). Имеется следующая информация:

1) результаты аналитического выравнивания рядов:

тренд в форме параболы второго порядка

а) для ряда Y_t : $\hat{Y}_t = 0,8 + 0,3t - 0,05t^2$, ($R^2 = 0,95$);

б) для ряда X_t : $\hat{Y}_t = 3 - 0,65t - 0,01t^2$, ($R^2 = 0,85$);

линейные тренды

а) для ряда: Y_t : $\hat{Y}_t = 2 + 0,05t$ ($R^2 = 0,38$);

б) для ряда X_t : $\hat{Y}_t = 0,65 + 0,8t$ ($R^2 = 0,24$);

(R^2 – коэффициент детерминации);

2) коэффициенты корреляции:

по исходным данным уровням рядов – 0,98;

по отклонениям от трендов в форме параболы второго порядка – 0,78;

по отклонениям от линейных трендов – 0,45;

по первым разностям – 0,42;

по вторым разностям – 0,84.

Задание

1. Есть ли взаимосвязь между исследуемыми временными рядами? Если есть, укажите ее количественную характеристику (характеристики). Ответ обоснуйте.

2. Поясните причины различий полученных мер тесноты связи.

Задача 22

Администрация торговой фирмы интересуется, есть ли взаимосвязь между объемом продаж и удельным весом женщин среди работников компании. Для этого были собраны данные за последние девять лет (табл. 4.22).

Таблица 4.22

Показатель	1	2	3	4	5	6	7	8	9
Объем продаж, тыс. долл., y_t	378	385	393	403	414	428	444	462	481
Удельный вес женщин среди работников ком- пании, % , x_t	25	24	27	30	31	29	31	33	34

Известны также следующие данные: $\Sigma y_t = 3788$, $\Sigma y_t^2 = 1604488$,
 $\Sigma x_t = 264$, $\Sigma x_t^2 = 78388$, $\Sigma x_t y_t = 112001$.

Уравнения трендов для каждого из рядов составили:

а) для ряда x_t

$$\hat{x}_t = 23,5 + 1,17 \cdot t;$$

б) для ряда y_t

$$\hat{y}_t = 374,14 + 3,33 \cdot t + 0,95 \cdot t^2.$$

Задание

1. Определите коэффициент корреляции между изучаемыми рядами по их уровням.
2. Определите коэффициент корреляции между изучаемыми рядами по отклонениям от указанных выше линейного и параболического трендов соответственно.
3. Выбрав одну из полученных мер в пп. 1 и 2, охарактеризуйте тесноту связи между временными рядами объемов продаж и долей женщин среди работников компании. Обоснуйте ваш выбор.

Задача 23

Имеются данные об экспорте и импорте Германии, млрд долл. США, за 1985 – 1996 гг. (табл. 4.23).

Таблица 4.23

Год	Экспорт	Импорт	Год	Экспорт	Импорт
1985	184	158	1991	403	390
1986	243	191	1992	422	402
1987	294	228	1993	382	346
1988	323	280	1994	430	385
1989	341	270	1995	524	464
1990	410	346	1996	521	456

Задание

1. Постройте график одновременного движения экспорта и импорта Германии.
2. Постройте по каждому ряду тренды и выберите лучший из них.
3. Постройте уравнение регрессии и оцените тесноту и силу связи двух рядов (по отклонениям от тренда и по множественной регрессионной модели с включением в нее фактора времени).

4. Выполните прогноз уровней одного ряда исходя из его связи с уровнями другого ряда.

5. Прогнозные значения уровней ряда и доверительный интервал прогноза нанесите на график.

Задача 24

В табл. 4.24 указаны остатки регрессии.

Таблица 4.24

Год	Остатки	Год	Остатки	Год	Остатки
1980	-0,7	1984	0	1988	0,0
1981	0	1985	0,3	1989	0,3
1982	-0,2	1986	-0,1	1990	0,3
1983	0,9	1987	-0,1	1991	-0,1

Задание

1. Оцените автокорреляцию остатков.
2. Примените критерий Дарбина – Уотсона и сделайте выводы относительно рассматриваемой регрессии.

Задача 25

Рассмотрите следующие модели регрессии, описывающие динамику заработной платы:

$$\text{модель А } W_t = 8,56 + 0,36 P_t + 0,74 P_{t-1} + 0,24 P_{t-1} - 2,53 Un_t + \varepsilon_t$$

$(t_{\text{факт}})$ (2,3) (3,7) (2,8) (-4,1)

$$R^2 = 0,9 \quad d = 1,7;$$

$$\text{модель Б } W_t = 9,01 + 0,32 P_t - 2,70 Un_t + 0,2 W_{t-1} + \varepsilon_t$$

$(t_{\text{факт}})$ (3,5) (-4,7) (2,7)

$$R^2 = 0,85 \quad d = 2,1,$$

где W_t – средняя заработная плата в году t ;

P_t – индекс цен в году t (в процентах по сравнению с базисным периодом);

Un_t – уровень безработицы в году t .

Исходные данные по W_t , P_t и Un_t были собраны за 30 лет (данные годовичные).

Задание

1. Используя модель А, охарактеризуйте силу связи между изменением цен и уровнем средней заработной платы.
2. Используя модель Б, охарактеризуйте силу связи между изменением цен и уровнем средней заработной платы.

3. Что вы можете сказать относительно автокорреляции в остатках по моделям А и Б? Ответ обоснуйте.

4. Какая из двух моделей лучше? Ответ обоснуйте.

Задача 26

Изучается зависимость объема ВВП Y_t (млрд долл.) от уровня прибыли в экономике x_t (млрд долл.) по данным за 30 лет. Была получена следующая модель:

$$Y_t = -5 + \underset{(2,2)}{1,5} X_t + \underset{(2,3)}{2} X_{t-1} + \underset{(2,5)}{4} X_{t-2} + \underset{(2,3)}{2,5} X_{t-3} + \underset{(2,4)}{2} X_{t-4} + \varepsilon_t,$$

$$R^2 = 0,9 \quad d = 2,65$$

В скобках указаны значения t -критерия для коэффициентов регрессии.

Задание

1. Проанализируйте полученные результаты регрессионного анализа: определите краткосрочный и долгосрочный мультипликаторы, охарактеризуйте структуру лага.
2. Перечислите основные эконометрические проблемы, возникающие при построении моделей с распределенным лагом.

Задача 27

По результатам изучения зависимости удельных постоянных затрат (коп.) от инвестиций в НИОКР (млн руб.) по некоторому виду продукции администрация компании получила следующую модель по данным за последние 38 лет:

$$Y_t = 231 - 0,2X_{t-1} - 0,15X_{t-2} - 0,5X_{t-3} + u_t, \quad R^2 = 0,87.$$

Задание

1. Каковы ваши предположения относительно структуры лага в этой модели?
2. Дайте интерпретацию параметров этой модели.

Задача 28

Предположим, по данным о динамике показателей сбережений населения и дохода в городе была получена модель авторегрессии, описывающая зависимость сбережений в среднем на душу населения за год S_t (млн руб.) от среднедушевого совокупного годового дохода Y_t (млн руб.) и сбережений предшествующего года S_{t-1} :

$$S_t = -53 + 0,12Y_t + 0,03S_{t-1} + \varepsilon_t.$$

Задание

Определите краткосрочную и долгосрочную склонность к накоплению.

Задача 29

Исследуя зависимость капитальных расходов от капитальных ассигнований, Ш. Алмон получила следующую модель¹:

$$\hat{E}_t = 0,048A_t + 0,099A_{t-1} + 0,141A_{t-2} + 0,165A_{t-3} + 0,167A_{t-4} + \\ + 0,146A_{t-5} + 0,105A_{t-6} + 0,053A_{t-7} - 283S_{1t} + 13S_{2t} - 50S_{3t} + 320S_{4t},$$

(0,023) (0,016) (0,013) (0,023) (0,023)
(0,013) (0,016) (0,024)

$$n = 36 \quad \bar{R}^2 = 0,92 \quad d = 0,89$$

(в скобках указаны стандартные ошибки для коэффициентов регрессии),

где E_t – капитальные расходы в квартале t (млн долл.);
 A_t – капитальные ассигнования в квартале t (млн долл.);
 S_{kt} – фиктивная переменная, равная 1 в квартале k и равная 0 – в остальных кварталах, $k = 1 + 4$ (Алмон построила уравнение с константой и тремя фиктивными переменными, а затем определила коэффициент регрессии при четвертой фиктивной переменной таким образом, чтобы сумма всех четырех коэффициентов и константы была равна 0)².

Задание

1. Охарактеризуйте структуру лага графически.
2. Рассчитайте относительные коэффициенты в этой модели и дайте количественную характеристику структуры лага. Определите средний и медианный лаг.
3. Выпишите краткосрочный, промежуточные и долгосрочный мультипликаторы в данной модели. Поясните смысл этих показателей.

Задача 30

В табл. 4.25 приводятся данные об уровне производительности труда (выпуск продукции в среднем за 1 ч, % к уровню 1982 г.) по экономике США (X) и среднечасовой заработной плате в экономике США (Y), в сопоставимых ценах 1982 г., долл., в 1960–1990 гг.

¹ Almon S. The distributed lags between capital appropriations and expenditures // *Econometric*. – 1965. – С. 183.

² Там же. – С. 183.

Таблица 4.25

Год	X	Y	Год	X	Y	Год	X	Y
1960	65,6	6,79	1970	87,0	8,03	1980	98,6	7,78
1961	68,1	6,88	1971	90,2	8,21	1981	99,9	7,69
1962	70,4	7,07	1972	92,6	8,53	1982	100,0	7,68
1963	73,3	7,17	1973	95,0	8,55	1983	102,2	7,79
1964	76,5	7,33	1974	93,3	8,28	1984	104,6	7,80
1965	78,6	7,52	1975	95,5	8,12	1985	106,1	7,77
1966	81,0	7,62	1976	98,3	8,24	1986	108,3	7,81
1967	83,0	7,72	1977	99,8	8,36	1987	109,4	7,73
1968	85,4	7,89	1978	100,4	8,40	1988	110,4	7,69
1969	85,9	7,98	1979	99,3	8,17	1989	109,5	7,64
						1990	109,7	7,53

Задание

1. Оцените обычным МНК параметры модели с распределенным лагом, характеризующей зависимость заработной платы от производительности труда, при величине лага 2, 3 и 4. Проанализируйте полученные результаты.

2. Оцените параметры этой же модели при величине лага 3 и 4 в предположении полиномиальной структуры лага (в качестве функции, описывающей структуру лага, выберите полином второй степени). Проанализируйте полученные результаты. Сравните их с результатами, полученными вами в п.1. Сделайте выводы.

Задача 31

Имеются данные о динамике оборота розничной торговли и потребительских цен региона за 1998 – 1999 гг. (табл. 4.26).

Таблица 4.26

Месяц	Оборот розничной торговли, % к предыдущему месяцу	Индекс потребительских цен, % к предыдущему месяцу
Январь	70,8	101,7
Февраль	98,7	101,1
Март	97,9	100,4
Апрель	99,6	100,1
Май	96,1	100,0
Июнь	103,4	100,1
Июль	95,5	100,0
Август	102,9	105,8
Сентябрь	77,6	145,0
Октябрь	102,3	99,8
Ноябрь	102,9	102,7

Месяц	Оборот розничной торговли, % к предыдущему месяцу	Индекс потребительских цен, % к предыдущему месяцу
Декабрь	123,1	109,4
Январь	74,3	110,0
Февраль	92,9	106,4
Март	106,0	103,2
Апрель	99,8	103,2
Май	105,2	102,9
Июнь	99,7	100,8
Июль	99,7	101,6
Август	107,9	101,5
Сентябрь	98,8	101,4
Октябрь	104,6	101,7
Ноябрь	106,4	101,7
Декабрь	122,7	101,2

Задание

1. Постройте автокорреляционную функцию каждого временного ряда. Охарактеризуйте структуру рядов.
2. Используя метод Алмон, оцените параметры модели с распределенным лагом. Длину лага выберите не более 4, степень аппроксимирующего полинома – не более 3. Оцените качество построенной модели.
3. Используя метод Койка, оцените параметры модели с распределенным лагом. Длину лага выберите не более 4.
4. Сравните результаты, полученные в п. 2 и 3.

Задача 32

Динамика объема платных услуг населению региона по кварталам 1996 – 1999 гг. характеризуется данными, представленными в табл. 4.27.

Таблица 4.27

Квартал	Объем платных услуг населению, млн руб.	Квартал	Объем платных услуг населению, млн руб.
1	2428	9	3528
2	2010	10	3838
3	2981	11	3916
4	3074	12	4142
5	2893	13	4441
6	3198	14	5583
7	3250	15	6230
8	3495	16	6497

Задание

1. Постройте автокорреляционную функцию временного ряда.
2. Охарактеризуйте структуру этого ряда.

Задача 33

Динамика выпуска продукции за 1986–1997 гг. представлена в табл. 4.28.

Таблица 4.28

Год	Выпуск продукции, ед.	Год	Выпуск продукции, ед.	Год	Выпуск продукции, ед.
1986	25	1990	30	1994	40
1987	27	1991	35	1995	42
1988	30	1992	33	1996	45
1989	29	1993	40	1997	44
Σ	111		138		171

Задание

1. Постройте уравнение авторегрессии с лагом в 2 года.
 2. Измерьте автокорреляцию остатков и сделайте выводы.
- В расчетах используйте следующие данные:

$$\Sigma y_t y_{t-2} = 12486, \quad \Sigma y_{t-2}^2 = 11273.$$

Задача 34

Динамика цен на товар *A* по кварталам характеризуется следующими данными:

<i>t</i>	1	2	3	4	5	...	20	21	22	23	24
<i>y_t</i>	2	3	3	6	4	...	15	12	13	13	14

Получены коэффициенты автокорреляции уровней временного ряда:

$$r_1 = 0,87025;$$

$$r_2 = 0,76579;$$

$$r_3 = 0,79343;$$

$$r_4 = 0,82278;$$

$$r_5 = 0,77790;$$

$$r_6 = 0,67833;$$

r_i – коэффициенты автокорреляции *i*-го порядка.

Задание

1. Постройте два лучших уравнения авторегрессии первого порядка. Оцените значимость полученных уравнений.
2. Постройте уравнение авторегрессии второго порядка. Для оценки параметров регрессии используйте МНК.
3. Постройте прогноз y_t на 25-й квартал по уравнению авторегрессии второго порядка.

В расчетах используйте следующие данные:

$$\text{при } n = 24, \Sigma y_t = 186, \Sigma y_t^2 = 1794;$$

$$\text{при лаге } (t - 4) n = 20, \Sigma y_t y_{t-1} = 1617, \Sigma y_t y_{t-4} = 1359,$$

$$\Sigma y_{t-1} y_{t-4} = 1271, \Sigma y_{t-1}^2 = 1576, \Sigma y_{t-4}^2 = 1116.$$

Задания для задач 35 – 42.

1. Найдите коэффициенты автокорреляции разного порядка и выберите величину лага.
2. Постройте авторегрессионную функцию.
3. Рассчитайте прогнозные значения на три года вперед.

Задача 35

В табл. 4.29 приводятся сведения об уровне среднегодовых цен на какао-бобы из Бразилии, амер. центы за фунт.

Таблица 4.29

Год	Цена	Год	Цена	Год	Цена	Год	Цена
1970	29,4	1977	183,5	1984	105,3	1991	47,5
1971	23,5	1978	153,5	1985	94,9	1992	45,0
1972	26,2	1979	140,7	1986	92,0	1993	44,5
1973	48,5	1980	107,1	1987	83,9	1994	55,9
1974	73,4	1981	87,5	1988	72,7	1995	60,5
1975	56,6	1982	68,3	1989	56,9	1996	64,1
1976	77,0	1983	83,1	1990	49,1	1997	71,0

Задача 36

В табл. 4.30 приводятся сведения об уровне среднегодовых цен на рис из Таиланда на рынках Бангкока, амер. доллары за метрическую тонну.

Таблица 4.30

Год	Цена	Год	Цена	Год	Цена	Год	Цена
1970	143	1977	272	1984	252	1991	287
1971	130	1978	369	1985	217	1992	291
1972	150	1979	334	1986	210	1993	237
1973	296	1980	434	1987	229	1994	269
1974	542	1981	483	1988	302	1995	321
1975	363	1982	293	1989	320	1996	338
1976	254	1983	277	1990	270	1997	303

Задача 37

В табл. 4.31 приводятся сведения об уровне среднегодовых цен на говядину из США на рынках Нью-Йорка, амер. центы за фунт.

Таблица 4.31

Год	Цена	Год	Цена	Год	Цена	Год	Цена
1970	41	1977	51	1984	97	1991	90
1971	42	1978	71	1985	89	1992	90
1972	49	1979	92	1986	77	1993	93
1973	64	1980	87	1987	81	1994	87
1974	53	1981	86	1988	82	1995	84
1975	44	1982	99	1989	87	1996	85
1976	52	1983	96	1990	94	1997	86

Задача 38

В табл. 4.32 приводятся сведения об уровне среднегодовых цен на каучук из Малайзии на рынках Сингапура, амер. центы за фунт.

Таблица 4.32

Год	Цена	Год	Цена	Год	Цена	Год	Цена
1970	18,5	1977	36,9	1984	43,4	1991	37,5
1971	15,1	1978	44,7	1985	34,4	1992	39,1
1972	15,1	1979	57,3	1986	36,6	1993	37,7
1973	30,8	1980	64,6	1987	44,7	1994	51,1
1974	34,1	1981	50,9	1988	53,7	1995	71,7
1975	25,4	1982	38,9	1989	44,0	1996	63,6
1976	35,1	1983	48,3	1990	39,2	1997	46,2

Задача 39

В табл. 4.33 приводятся сведения об уровне среднегодовых цен на каучук, поступивший на рынки Нью-Йорка из всех источников, амер. центы за фунт.

Таблица 4.33

Год	Цена	Год	Цена	Год	Цена	Год	Цена
1970	21,1	1977	41,5	1984	49,6	1991	47,6
1971	18,0	1978	49,9	1985	41,8	1992	46,6
1972	18,1	1979	64,2	1986	41,2	1993	47,3
1973	35,1	1980	73,4	1987	44,1	1994	48,9
1974	39,7	1981	56,9	1988	48,8	1995	56,7
1975	29,8	1982	45,3	1989	48,7	1996	54,8
1976	39,5	1983	56,1	1990	50,2	1997	53,5

Задача 40

В табл.4.34 приводятся сведения об уровне среднегодовых цен на мировых рынках на шерсть из Новой Зеландии, амер. центы за килограмм.

Таблица 4.34

Год	Цена	Год	Цена	Год	Цена	Год	Цена
1970	73,8	1977	256,4	1984	230,7	1991	249,3
1971	72,6	1978	249,6	1985	234,9	1992	242,9
1972	106,9	1979	300,4	1986	248,5	1993	234,3
1973	237,5	1980	316,7	1987	333,0	1994	287,9
1974	214,7	1981	274,6	1988	403,2	1995	356,2
1975	147,6	1982	239,7	1989	386,3	1996	348,3
1976	202,9	1983	221,9	1990	341,5		

Задача 41

В табл. 4.35 приводятся сведения об уровне среднегодовых цен на мировых рынках на невытую шерсть из Австралии, амер. центы за килограмм.

Таблица 4.35

Год	Цена	Год	Цена	Год	Цена	Год	Цена
1970	98,2	1977	227,0	1984	282,0	1991	307,5
1971	79,7	1978	234,8	1985	258,5	1992	302,6
1972	117,8	1979	259,6	1986	259,5	1993	240,4
1973	305,1	1980	302,5	1987	343,2	1994	323,2
1974	251,9	1981	328,5	1988	567,1	1995	395,8
1975	182,4	1982	306,5	1989	520,9	1996	325,7
1976	197,9	1983	269,3	1990	446,6	1997	358,5

Задача 42

В табл. 4.36 приводятся сведения об уровне среднегодовых цен на рис из Таиланда на рынках Бангкока, амер. доллары за метрическую тонну.

Таблица 4.36

Год	Цена	Год	Цена	Год	Цена	Год	Цена
1970	143,0	1977	272,4	1984	252,3	1991	287,1
1971	130,3	1978	368,5	1985	217,4	1992	291,0
1972	149,9	1979	334,3	1986	210,2	1993	237,3
1973	296,6	1980	433,7	1987	229,8	1994	269,5
1974	541,5	1981	482,8	1988	301,5	1995	320,8
1975	363,2	1982	293,4	1989	320,3	1996	338,1
1976	254,1	1983	276,8	1990	270,2	1997	302,7

ПРИЛОЖЕНИЯ

СТАТИСТИКО-МАТЕМАТИЧЕСКИЕ ТАБЛИЦЫ

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

$k_1 \backslash k_2$	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	254,32
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

$k_1 \backslash k_2$	1	2	3	4	5	6	8	12	24	∞
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

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

Число степеней свободы d.f.	α			Число степеней свободы d.f.	α		
	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	4,6041	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. Критические значения корреляции для уровневой значимости 0,05 и 0,01

d.f.	$\alpha = 0,05$	$\alpha = 0,01$	d.f.	$\alpha = 0,05$	$\alpha = 0,01$
1	0,996917	0,9998766	17	0,4555	0,5751
2	0,95000	0,99000	18	0,4438	0,5614
3	0,8783	0,95873	19	0,4329	0,5487
4	0,8114	0,91720	20	0,4227	0,5368
5	0,7545	0,8745	25	0,3809	0,4869
6	0,7067	0,8343	30	0,3494	0,4487
7	0,6664	0,7977	35	0,3246	0,4182
8	0,6319	0,7646	40	0,3044	0,3932
9	0,6021	0,7348	45	0,2875	0,3721
10	0,5760	0,7079	50	0,2732	0,3541
11	0,5529	0,6835	60	0,2500	0,3248
12	0,5324	0,6614	70	0,2319	0,3017
13	0,5139	0,6411	80	0,2172	0,2830
14	0,4973	0,6226	90	0,2050	0,2673
15	0,4821	0,6055	100	0,1946	0,2540
16	0,4683	0,5897			

Для простой корреляции d.f. на 2 меньше, чем число пар вариантов; в случае частной корреляции необходимо также вычесть число исключаемых переменных.

4. Значения статистик Дарбина – Уотсона $d_L d_U$ при 5%-ном уровне значимости

n	$k^1=1$		$k^1=2$		$k^1=3$		$k^1=4$		$k^1=5$	
	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				

n	k ¹ =1		k ¹ =2		k ¹ =3		k ¹ =4		k ¹ =5	
	d _L	d _U	d _L	d _U	d _L	d _U	d _L	d _U	d _L	d _U
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	0,86	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	1,90
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

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	3
-------------------	---

I раздел

ПАРНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ.....	5
---	----------

1.1. МЕТОДИЧЕСКИЕ УКАЗАНИЯ	5
1.2. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ	10
1.3. РЕАЛИЗАЦИЯ ТИПОВЫХ ЗАДАЧ НА КОМПЬЮТЕРЕ	22
1.4. КОНТРОЛЬНЫЕ ЗАДАНИЯ	29

II раздел

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ.....	49
--	-----------

2.1. МЕТОДИЧЕСКИЕ УКАЗАНИЯ	49
2.2. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ	56
2.3. РЕАЛИЗАЦИЯ ТИПОВЫХ ЗАДАЧ НА КОМПЬЮТЕРЕ	66
2.4. КОНТРОЛЬНЫЕ ЗАДАНИЯ	79

III раздел

СИСТЕМА ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ.....	106
--	------------

3.1. МЕТОДИЧЕСКИЕ УКАЗАНИЯ	106
3.2. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ	108
3.3. КОНТРОЛЬНЫЕ ЗАДАНИЯ	121

IV раздел

ВРЕМЕННЫЕ РЯДЫ В ЭКОНОМЕТРИЧЕСКИХ ИССЛЕДОВАНИЯХ.....	137
---	------------

4.1. МЕТОДИЧЕСКИЕ УКАЗАНИЯ	137
4.2. РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ	142
4.3. РЕАЛИЗАЦИЯ ТИПОВЫХ ЗАДАЧ НА КОМПЬЮТЕРЕ	151
4.4. КОНТРОЛЬНЫЕ ЗАДАНИЯ	163

ПРИЛОЖЕНИЯ.

СТАТИСТИКО-МАТЕМАТИЧЕСКИЕ ТАБЛИЦЫ	187
--	------------

1. Таблица значений F -критерия Фишера при уровне значимости $\alpha = 0,05$	187
2. Критические значения t -критерия Стьюдента при уровне значимости 0,10, 0,05, 0,01 (двухсторонний).....	188
3. Критические значения корреляции для уровневой значимости 0,05 и 0,01.....	189
4. Значения статистик Дарбина – Уотсона $d_L d_U$ при 5%-ном уровне значимости	189

Учебное издание

**Елисеева Ирина Ильинична
Курышева Светлана Владимировна
Гордеевко Нелли Михайловна и др.**

ПРАКТИКУМ ПО ЭКОНОМЕТРИКЕ

Заведующая редакцией *Л. А. Табакова*
Редактор *Л. А. Табакова*
Младший редактор *Н. А. Федорова*
Художественный редактор *Ю. И. Артюхов*
Технический редактор *И. В. Завгородняя*
Корректоры *Г. Б. Абудеева, Н. П. Сперанская*
Компьютерная верстка *И. В. Витте*
Обложка художника *Е. К. Самойлова*

ИБ № 4210

Подписано в печать 21.10.2002. Формат 60×88/16
Гарнитура «Times». Печать офсетная
Усл. п. л. 11,76. Уч.-изд. л. 11,2.
Доп. тираж 7000 экз. Заказ 3654. «С» 154

Издательство «Финансы и статистика»
101000, Москва, ул. Покровка, 7
Телефон (095) 925-35-02. Факс (095) 925-09-57
E-mail: mail@finstat.ru <http://www.finstat.ru>

ГУП «Великолукская городская типография»
Комитета по средствам массовой информации Псковской области
182100, Великие Луки, ул. Полиграфистов, 78/12
Тел./факс: (811-53) 3-62-95
E-mail: VTL@MART.RU