ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
Государственное образовательное учреждение
высшего профессионального образования
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ

Е. Г. Семенова, М. С. Смирнова

ОСНОВЫ ЭКОНОМЕТРИЧЕСКОГО АНАЛИЗА

Учебное пособие

Допущено УМО по образованию
в области прикладной математики и управления качеством
в качестве учебного пособия для студентов
высших учебных заведений, обучающихся по специальности
220501 – «Управление качеством»

Санкт-Петербург
2006

УДК [519.2+330.4] (075)
ББК 65в6я7
С30

Рецензенты:

доктор технических наук, профессор *И. Г. Мироненко*;
доктор технических наук, профессор *В. М. Балашов*

Утверждено редакционно-издательским советом университета
в качестве учебного пособия

Семенова Е. Г. , Смирнова М. С.

С30 Основы эконометрического анализа: учеб. пособие / Е. Г. Семенова, М. С. Смирнова; ГУАП. – СПб., 2006. – 72 с.: ил
ISBN 5-8088-0195-8

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

Предназначено для студентов и аспирантов соответствующих экономических и управленческих направлений обучения.

УДК [519.2+330.4] (075)
ББК 65в6я7

ISBN 5-8088-0195-8

ПРЕДИСЛОВИЕ

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

– методология принятия решений о спецификации и идентификации моделей;

– ознакомление с методами и приемами интерпретации результатов эконометрического моделирования;

– изучение принципов выбора метода оценки параметров моделей;

– выработка устойчивых практических навыков разработки прогнозных оценок.

В результате усвоения материала дисциплины студент должен знать:

– терминологию, основные понятия и определения;

– методологические основы эконометрического моделирования;

– принципы и методы построения эконометрических моделей на основе пространственных данных и временных рядов;

– принципы решения типовых задач с учетом мультиколлинеарости и автокорреляции;

– возможности реализации типовых задач на компьютере с помощью пакета прикладных программ Excel.

Введение

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

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

Существуют различные варианты определения эконометрики:

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

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

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

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

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

1. ПАРНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

1.1. Методические указания

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

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

Парная регрессия – уравнение связи двух переменных y и x:

$$y = f(x),$$

где y – зависимая переменная (результативный признак), x – независимая, объясняющая переменная (признак-фактор).

Различают линейные и нелинейные регрессии.

Линейная регрессия: $y = a + bx + \varepsilon$.

Нелинейные регрессии делятся на два класса:

– регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам;

– регрессии, нелинейные по оцениваемым параметрам.

Регрессии, нелинейные по объясняющим переменным:

– полиномы разных степеней $y = a + b_1x + b_2x^2 + b_3x^3 + \varepsilon$;

– равносторонняя гипербола $y = a + \dfrac{b}{x} + \varepsilon$.

Регрессии, нелинейные по оцениваемым параметрам:

– степенная $y = ax^b + \varepsilon$;

– показательная $y = ab^x + \varepsilon$;

– экспоненциальная $y = e^{a+bx} + \varepsilon$.

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

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

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

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

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

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

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

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

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

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

Можно воспользоваться готовыми формулами, которые вытекают из этой системы:

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

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

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

Индекс корреляции ρ_{xy} – для нелинейной регрессии $0 \le \rho_{xy} \le 1$, причем, чем ближе к единице, тем теснее связь рассматриваемых признаков, тем более надежно найденное уравнение регрессии:

$$\rho_{xy} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}} = \sqrt{1 - \frac{\sum\left(y - \hat{y}_x\right)^2}{\sum\left(y - \overline{y}\right)^2}}\ .$$

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

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

$$r^2_{xy} = \frac{\sigma^2_{y\ \text{объясн}}}{\sigma^2_{y\ \text{общ}}} = \frac{\sum(\hat{y}_x - \overline{y})^2}{\sum(y - \overline{y})^2}.$$

Средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических.

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

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

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

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

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

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

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

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

где $\sum(y-\overline{y})^2$ – общая сумма квадратов отклонений; $\sum\left(\hat{y}_x - \overline{y}\right)^2$ – сумма квадратов отклонений, обусловленная регрессией (объясненная, или факторная); $\sum\left(y - \hat{y}_x\right)^2$ – остаточная сумма квадратов отклонений (необъясненная).

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

При расчете объясненной суммы квадратов $\sum\left(\hat{y}_x - \overline{y}\right)^2$ используются теоретические (расчетные) результативного признака $\hat{y}_x$, найденные по линии регрессии:

$$\hat{y}_x = a + bx.$$

Сумма квадратов отклонений, обусловленных линейной регрессией, составляет:

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

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

Следовательно, имеем два равенства:

$$\sum(y-\bar{y})^2=\sum\left(\hat{y}_x-\bar{y}\right)^2+\sum\left(y-\hat{y}_x\right)^2, \quad n-1=1+(n-2).$$

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

$$D_{\text{общ}}=\frac{\sum(y-\bar{y})^2}{n-1}; \quad D_{\text{факт}}=\frac{\sum(\hat{y}_x-\bar{y})^2}{1}; \quad D_{\text{ост}}=\frac{\sum(y-\hat{y}_x)^2}{n-2}.$$

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

$$F=\frac{D_{\text{факт}}}{D_{\text{ост}}}.$$

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

Табличное значение F-критерия – это максимальная величина отношения дисперсий, которая может иметь место при случайном их расхождении для данного уровня вероятности наличия нулевой гипотезы. Вычисленное значение F-критерия признается достоверным (отличным от единицы), если оно больше табличного. В этом случае нулевая гипотеза об отсутствии связи признаков отклоняется и делается вывод о существенности этой связи: $F_{\text{факт}} > F_{\text{табл}}$. H_0 отклоняется.

Если же величина окажется меньше табличной $F_{\text{факт}} < F_{\text{табл}}$, то вероятность нулевой гипотезы выше заданного уровня (например, 0,05) и она не может быть отклонена без серьезного риска сделать неправильный вывод о наличии связи. В этом случае уравнение регрессии считается статистически незначимым, H_0 не отклоняется:

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

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

Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитываются t-критерии Стьюдента. С этой целью по каждому из параметров определяется его стандартная ошибка: m_b и m_a.

Стандартная ошибка коэффициента регрессии определяется по следующей формуле:

$$m_b = \sqrt{\frac{\sum(y-\hat{y}_x)^2 \big/ (n-2)}{\sum(x-\overline{x})^2}}.$$

Для оценки существующей коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т. е. определяется фактическое значение t-критерия Стьюдента: $t_b = \frac{b}{m_b}$, которое затем сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы $(n-2)$. Если фактическое значение t-критерия превышает табличное, то гипотезу о несущественности коэффициента регрессии можно отклонить.

Можно доказать равенство $t_b^2 = F$:

$$t_b^2 = \frac{b^2}{m_b^2} = b^2 \Bigg/ \frac{\sum(y-\hat{y}_x)^2 \big/ (n-2)}{\sum(x-\overline{x})^2} = \frac{b^2\sum(x-\overline{x})^2}{\sum(y-\hat{y}_x)^2 \big/ (n-2)} =$$

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

Стандартная ошибка параметра a определяется по следующей формуле:

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

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции:

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}.$$

Фактическое значение t-критерия определяется как:

$$t_r = \frac{r}{\sqrt{1 - r^2}} \sqrt{n - 2}.$$

Данная формула свидетельствует, что в парной линейной регрессии $t_r^2 = F$, так как $F = \frac{r^2}{1 - r^2}(n - 2)$, кроме того, $t_b^2 = F$, следовательно, $t_r^2 = t_b^2$.

Таким образом, проверка гипотез о значимости коэффициентов регрессии и корреляции равносильна проверке гипотезы о существенности линейного уравнения регрессии.

После того как произведена оценка параметров модели, рассчитывая разности фактических и теоретических значений результативного признака y, можно определить оценки случайной составляющей $y - \hat{y}_x$. Поскольку они не являются реальными случайными остатками, их можно считать некоторой выборочной реализацией неизвестного остатка заданного уравнения, т. е. ε_i.

При изменении спецификации модели, добавлении в нее новых наблюдений выборочные оценки остатков ε_i могут меняться. Поэтому в задачу регрессионного анализа входит не только построение самой модели, но и исследование случайных отклонений ε_i, т. е. остаточных величин.

При использовании критериев Фишера и Стьюдента делаются предположения относительно поведения остатков ε_i – остатки представляют собой независимые случайные величины, и их среднее значение равно 0; они имеют одинаковую (постоянную) дисперсию и подчиняются нормальному распределению.

Дисперсия остатков должна быть гомоскедастичной, т. е. для каждого значения фактора x_j остатки ε_i должны иметь одинаковую дис-

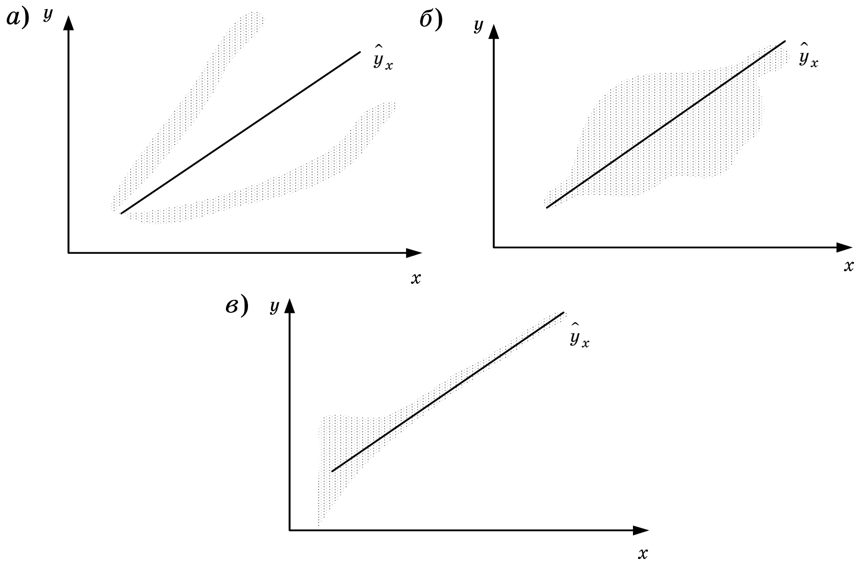


Рис. 1. *Примеры гетероскедастичности*

персию. Если это условие не соблюдается, то имеет место *гетероскедастичность*. Наличие гетероскедастичности можно наглядно видеть из поля корреляции (рис. 1): *а* – дисперсия остатков растет по мере увеличения x; *б* – дисперсия остатков достигает максимальной величины при средних значениях переменной x и уменьшается при минимальных и максимальных значениях x; *в* – максимальная дисперсия остатков при малых значениях x, и дисперсия остатков однородна по мере увеличения значений x.

1.2. Решение типовых задач

Задача 1.1. По семи территориям Уральского района за 200x год известны значения двух признаков (табл. 1.1).

Требуется:

– для характеристики зависимости y от x рассчитать параметры;

– оценить модель через ошибку аппроксимации $\bar{A}$ и F-критерий.

Решение. Для расчета параметров a и b линейной регрессии $y = a + bx$ решаем систему нормальных уравнений относительно a и b:

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

Таблица 1.1

Район	Расходы на покупку продовольственных товаров в общих расходах, y, %	Среднедневная заработная плата одного работающего, x, у. е.
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

По исходным данным рассчитываем $\sum y$, $\sum x$, $\sum yx$, $\sum x^2$, $\sum y^2$ (табл. 1.2)

$$b = \frac{\overline{yx} - \bar{y}\,\bar{x}}{\sigma_x^2} = \frac{3166{,}05 - 57{,}89 \cdot 54{,}9}{34{,}33} = -0{,}35,$$

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

Таблица 1.2

№ п/п	y	x	yx	x^2	y^2	$\hat{y}_x$	$y - \hat{y}_x$	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	2562,06	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	384,2	22162,34	21338,41	23685,76	405,2	0,0	56,7
Среднее значение	57,89	54,9	3166,05	3048,34	3383,68	–	–	8,1
σ	5,7	5,9	–	–	–	–	–	–
σ^2	32,43	34,33	–	–	–	–	–	–

Уравнение регрессии: $y = 76,88 - 0,35x$.

С увеличением среднедневной заработной платы на одну условную единицу (у. е.) доля расходов на покупку продовольственных товаров снижается в среднем на $0,35\%$ пункта.

Рассчитаем линейный коэффициент парной корреляции:

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} = -0,35\frac{5,9}{5,7} = -0,362.$$

Связь умеренная, обратная.

Определим коэффициент детерминации:

$$r_{xy}^2 = (-0,362)^2 = 0,131.$$

Вариация результата на $13,1\%$ объясняется вариацией фактора x.

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

Найдем величину средней ошибки аппроксимации $\bar{A}$:

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

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

Рассчитаем F-критерий:

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

Полученное значение указывает на необходимость применять гипотезу H_0 о случайной природе выявленной зависимости, т. к. $F_{\text{факт}} < F_{\text{табл}} = 6,61$ (табличные значения F-критерия приведены в Приложении).

Задача 1.2. По территориям региона приводятся данные (табл. 1.3).

Таблица 1.3

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, у. е.	Среднедневная заработная плата, у. е.
1	78	133
2	82	148
3	87	134

Окончание табл. 1.3

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, у. е.	Среднедневная заработная плата, у. е.
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

Таблица 1.4

№ п/п	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	–	–	5,7
σ	12,95	16,53	–	–	–	–	–	–
σ^2	167,7	273,4	–	–	–	–	–	–

Требуется:

– построить линейное уравнение парной регрессии y от x;

– рассчитать линейный коэффициент парной корреляции и среднюю ошибку аппроксимации.

Решение. Для расчета параметров уравнения линейной регрессии строим расчетную табл. 1.4.

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

$$a = \bar{y} - b\bar{x} = 155{,}89 - 0{,}92 \cdot 85{,}6 = 77{,}0.$$

Получено уравнение регрессии: $y = 77{,}0 + 0{,}92x$.

С увеличением среднедушевого прожиточного минимума на одну у. е. среднедневная заработная плата возрастет в среднем на 0,92 у.е.

Тесноту линейной связи оценит коэффициент корреляции:

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

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 хорошее, так как $\bar{A}$ не превышает $8–10\%$.

Задача 1.3. По группе предприятий, производящих однородную продукцию, известно, как зависит себестоимость единицы продукции y от факторов, приведенных в табл. 1.5.

Таблица 1.5

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

Требуется:

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

– ранжировать факторы по силе влияния.

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

$$\bar{Э}_{yx_1} = f'(x_1)\frac{\bar{x}_1}{\bar{y}} = -\frac{b}{\bar{x}_1^2} \cdot \frac{\bar{x}_1}{a + b/\bar{x}_1} = -\frac{b}{a\bar{x}_1 + b} =$$

$$= -\frac{58{,}85}{0{,}65 \cdot 2{,}55 + 58{,}85} = -0{,}973\% .$$

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

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

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

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

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

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

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

а) $\bar{Э}_{yx_3} = 1{,}55\%$, б) $\bar{Э}_{yx_1} = -0{,}973\%$, в) $\bar{Э}_{yx_2} = 0{,}62\%$, г) $\bar{Э}_{yx_4} = 0{,}26\%$.

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

1.3. Решение с помощью ППП Excel

Задача 1.4. По территориям региона приводятся данные (табл. 1.6).

Таблица 1.6

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

1) Встроенная статистическая функция ЛИНЕЙН определяет параметры линейной регрессии $y = a + bx$. 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 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;

5. Заполните аргументы функции (рис. 1.2):

Известные_значения_y – диапазон, содержащий данные результативного признака;

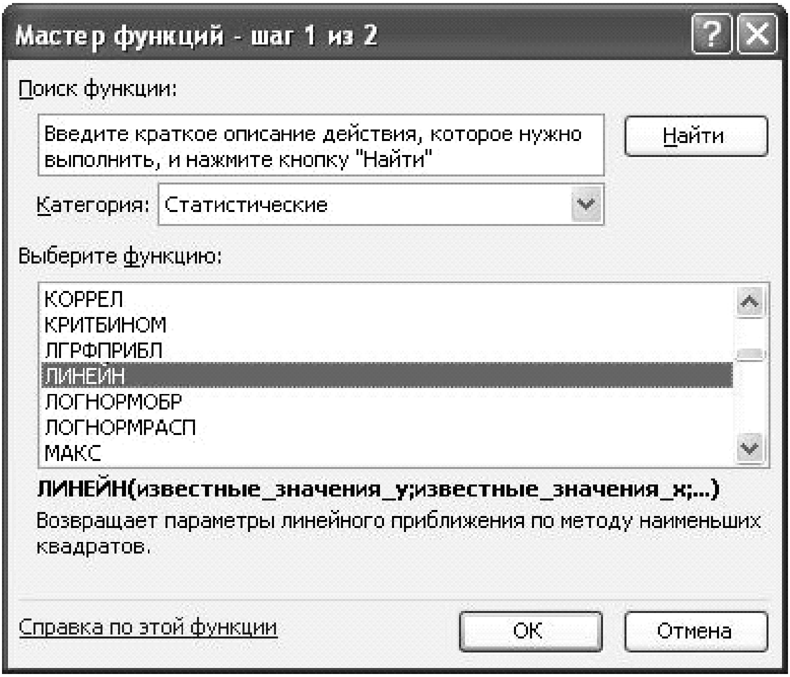


Рис. 1.1. Диалоговое окно Мастер функций

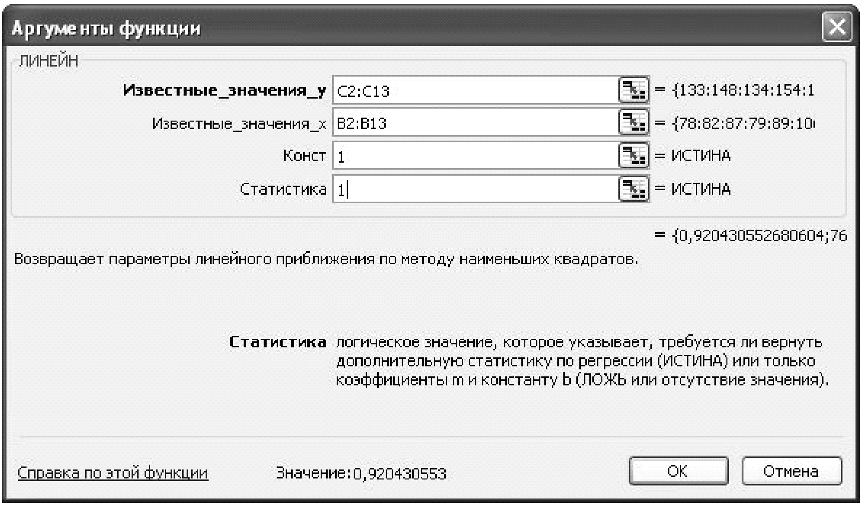


***Рис. 1.2.** Диалоговое окно ввода аргументов функции ЛИНЕЙН*

Известные_значения_x – диапазон, содержащий данные факторов независимого признака;

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

Статистика – логическое значение, которое указывает, выводить дополнительную информацию по регрессионному анализу или нет. Если Статистика = 1, то дополнительная информация выводится, если Статистика = 0, то выводятся только оценки параметров уравнения. Щелкните по кнопке OK;

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

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

Значение коэффициента b	Значение коэффициента a
Среднеквадратическое отклонение b	Среднеквадратическое отклонение a
Коэффициент детерминации	Среднеквадратическое отклонение y
F-статистика	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов

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

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

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

2. В главном меню выберите Сервис/Анализ данных/Регрессия. Щелкните по кнопке OK.

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

Входной интервал Y – диапазон, содержащий данные результативного признака;

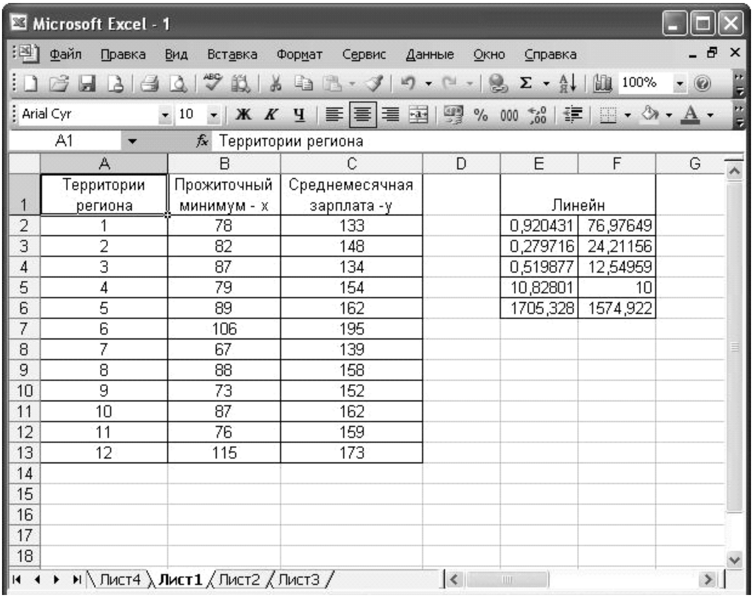


Рис. 1.3. *Результат вычисления функции ЛИНЕЙН*

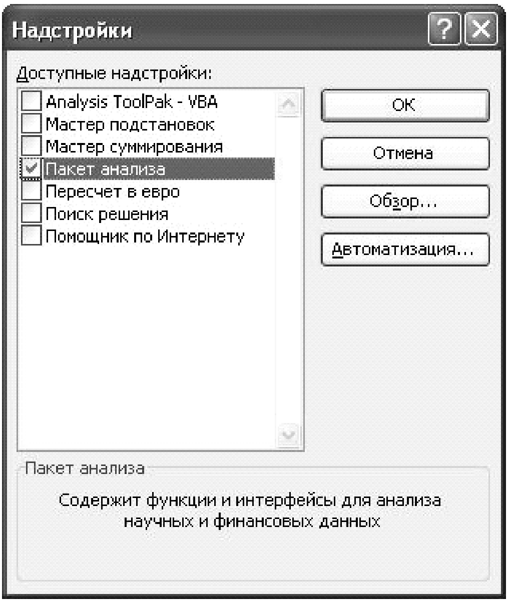


Рис. 1.4. *Подключение надстройки Пакет анализа*

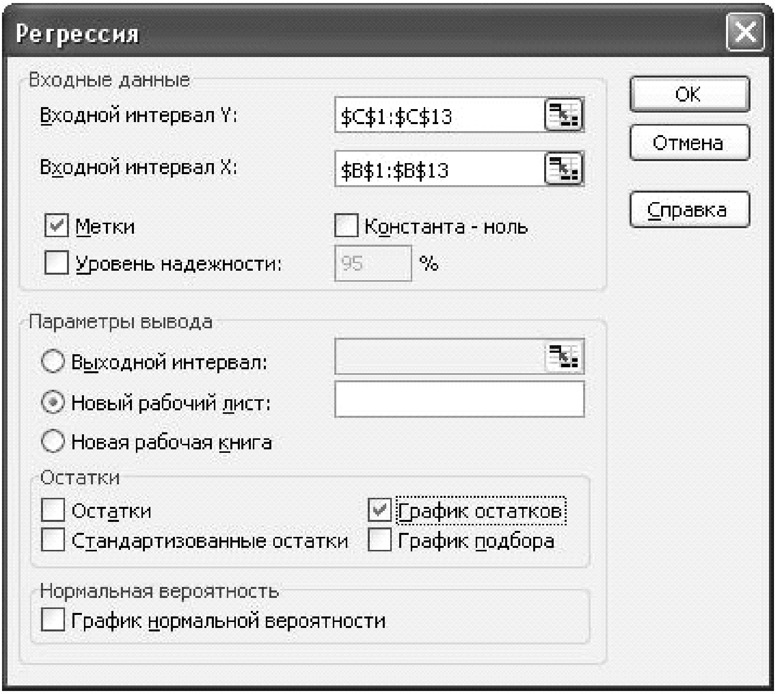


Рис. 1.5. *Диалоговое окно ввода параметров инструмента Регрессия*

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

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

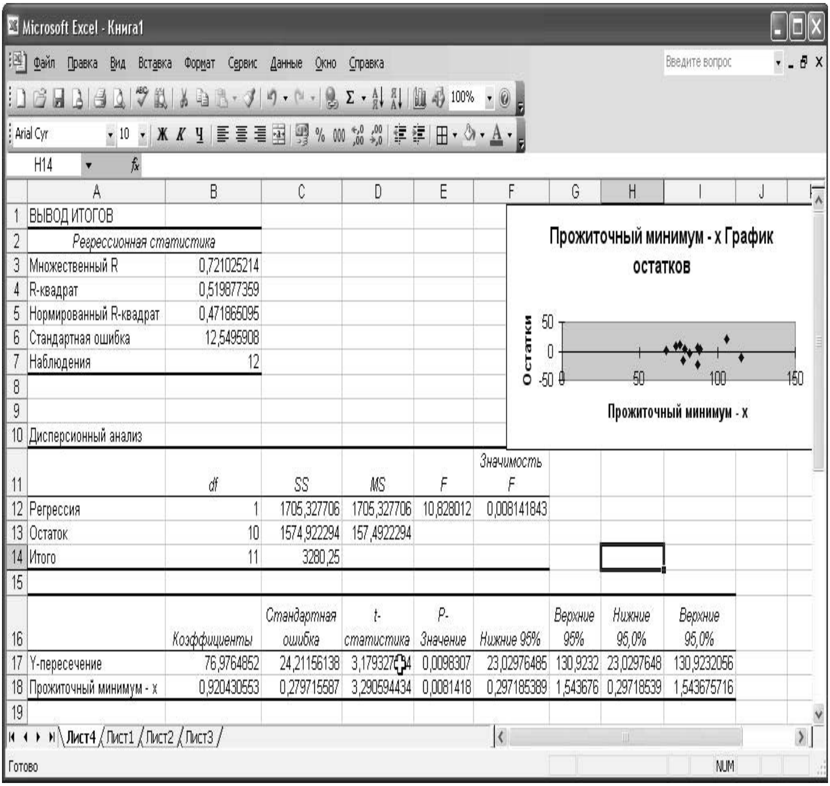
Константа – ноль – флажок, указывающий на наличие или отсутствие свободного члена в уравнении;

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 OK.

Результаты регрессионного анализа для данных из задачи 1.4 приведены на рис. 1.6.



ВЫВОД ИТОГОВ	
Регрессионная статистика	
Множественный R	0,721025214
R-квадрат	0,519877359
Нормированный R-квадрат	0,471865095
Стандартная ошибка	12,5495908
Наблюдения	12

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	1705,327706	1705,327706	10,828012	0,008141843
Остаток	10	1574,922294	157,4922294		
Итого	11	3280,25			

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

Рис. 1.6. *результат применения инструмента Регрессия*

Контрольные задания

Задача 1. Для трех видов продукции *A, B, C* модели зависимости удельных постоянных расходов от объема выпускаемой продукции выглядят следующим образом:

$$Y_A = 500,\ Y_B = 85 + 0{,}75x,\ Y_C = 40x^{0,5}.$$

Задание:

– определите коэффициенты эластичности по каждому виду продукции и поясните их смысл;

– сравните при $x = 1000$ эластичность затрат для продукции *B* и *C*;

– определите, каким должен быть объем выпускаемой продукции, чтобы коэффициенты эластичности для продукции *B* и *C* были равны.

Задача 2. Исследуя спрос на продукцию марки *N*, аналитический отдел компании АБС, по данным, собранным по 19 торговым точкам компании, выявил следующую зависимость: $\ln y = 15{,}0 - 0{,}85 \ln x$, где

y – объем продаж телевизоров марки N в отдельной торговой точке; x – средняя цена телевизора в данной торговой точке.

Задание.

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

Задача 3.

По территориям Центрального района известны данные за 200х г. (табл. 1.7).

Таблица 1.7

Район, обл.	Средний размер назначенных ежемесячных пенсий, y, р.	Прожиточный минимум в среднем на одного пенсионера в месяц, x, р.
Брянская	4400	3780
Владимирская	4260	3020
Ивановская	4210	3970
Калужская	4260	3010
Костромская	4200	2890
Московская	4500	4020
Орловская	4370	3150
Рязанская	4320	3660
Смоленская	4150	3990
Тверская	4200	3800
Тульская	4220	3810
Ярославская	4310	3860

Задание:

– рассчитайте параметры уравнения линейной парной регрессии;

– оцените тесноту связи с помощью показателей корреляции и детерминации;

– дайте с помощью среднего коэффициента эластичности сравнительную оценку силы связи фактора с результатом;

– оцените с помощью средней ошибки аппроксимации качество уравнения.

Задача 4. По территориям Уральского и Западно-Сибирского районов известны данные за 200х г. (табл. 1.8)

Таблица 1.8

Район	Потребительские расходы на душу населения, y, р.	Денежные доходы на душу населения, x, р.
Уральский		
Респ. Башкортостан	6610	8320
Удмуртская респ.	7240	9380
Курганская обл.	4980	7150
Пермская обл.	5510	8400
Свердловская обл.	7840	9880
Челябинская обл.	6250	8040
Западно-Сибирский район		
Респ. Алтай	4770	8030
Кемеровская обл.	7730	6390
Новосибирская обл.	7760	9850
Омская обл.	7880	9600
Томская обл.	6970	10300
Тюменская обл.	10630	22930

Задание:

– рассчитайте параметры уравнения линейной парной регрессии;

– оцените тесноту связи с помощью показателей корреляции и детерминации;

– дайте с помощью среднего коэффициента эластичности сравнительную оценку силы связи фактора с результатом;

– оцените с помощью средней ошибки аппроксимации качество уравнения.

1.5. Контрольные вопросы

1. Что такое коэффициент регрессии? Каковы способы его оценивания?
2. В чем смысл коэффициента детерминации?
3. Для чего применяется критерий Фишера? В чем его суть?
4. Для чего применяется анализ дисперсии в критерии Фишера?
5. Для чего необходим расчет дисперсии на одну степень свободы?
6. Перечислите виды нелинейных моделей.

7. Как определяются коэффициенты эластичности по различным видам регрессионных моделей?

8. Как определяется средняя ошибка аппроксимации? В чем ее смысл?

9. Что представляют собой остаточные величины? Какие требования к ним предъявляются?

2. 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, ..., x_p – независимые переменные (факторы).

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

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

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

Для ее решения может быть применен метод определителей:

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

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

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

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

где $t_y = \dfrac{y - \overline{y}}{\sigma_y}$, $t_{x_i} = \dfrac{x_i - \overline{x}_i}{\sigma_{x_i}}$ – стандартизованные переменные; β_i – стандартизованные коэффициенты регрессии.

Связь коэффициентов множественной регрессии b_i со стандартизованными коэффициентами β_i описывается следующим соотношением:

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

Параметр a определяется как

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

Для расчета *частных коэффициентов эластичности* применяется следующая формула:

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

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

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

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

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

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

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

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

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

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

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

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

Исследования остатков ε_i предполагают проверку наличия следующих пяти предпосылок МНК:

1) случайный характер остатков;

2) нулевая средняя величина остатков, не зависящая от x_i;

3) гомоскедастичность – дисперсия каждого отклонения ε_i, одинакова для всех значений x;

4) отсутствие автокорреляции остатков – значения остатков ε_i распределены независимо друг от друга;

5) остатки подчиняются нормальному распределению.

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

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

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

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

2.2. Решение типовых задач

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

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

Требуется:

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

– оценить значимость параметров данного уравнения, если известно, что t-критерий для параметра b_2 при x_2 составил 0,8, а для параметра b_3 при x_3 – 1,1.

Решение.

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

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

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

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

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

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

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

Задача 2.2. Имеются данные о ценах и дивидендах по обыкновенным акциям, а также о доходности компании (табл. 2.1).

Таблица 2.1

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

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

Решение.

Необходимо построить расчетную табл. 2.2.

Таблица 2.2

№ п/п	y	x_1	x_2	x_2*x_2	x_1*x_1	$y*x_1$	$y*x_2$	x_1*x_2
1	25	15,2	2,6	6,76	231,04	380,0	65,0	39,52
2	20	13,9	2,1	4,41	193,21	278,0	42,0	29,19
3	15	15,8	1,5	2,25	249,64	237,0	22,5	23,70
4	34	12,8	3,1	9,61	163,84	435,2	105,4	39,68
5	20	6,9	2,5	6,25	47,61	138,0	50,0	17,25
6	33	14,6	3,1	9,61	213,16	481,8	102,3	45,26
7	28	15,4	2,9	8,41	237,16	431,2	81,2	44,66
8	30	17,3	2,8	7,84	299,29	519,0	84,0	48,44
9	23	13,7	2,4	5,76	187,69	315,1	55,2	32,88
10	24	12,7	2,4	5,76	161,29	304,8	57,6	30,48
11	25	15,3	2,6	6,76	234,09	382,5	65,0	39,78
12	26	15,2	2,8	7,84	231,04	395,2	72,8	42,56
13	26	12,0	2,7	7,29	144,0	312,0	70,2	32,40
14	20	15,3	1,9	3,61	234,09	306,0	38,0	29,07
15	20	13,7	1,9	3,61	187,69	274,0	38,0	26,03
16	13	13,3	1,6	2,56	176,89	172,9	20,8	21,28
17	21	15,1	2,4	5,76	228,01	317,1	50,4	36,24
18	31	15,0	3,0	9,0	225,0	465,0	93,0	45,0
19	26	11,2	3,1	9,61	125,44	291,2	80,6	34,72
20	11	12,1	2,0	4,0	146,41	133,1	22,0	24,20
Итого	471	276,5	49,4	126,7	3916,59	6569,1	1216	682,34
Среднее значение	23,55	–	–	–	–	325,455	60,8	34,117
σ	6,07	2,168	0,484					

По данным табл. 2.2 строится система нормальных уравнений с тремя неизвестными:

$$\begin{cases}\sum y = na + b_1\sum x_1 + b_2\sum x_2,\\ \sum yx_1 = a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2,\\ \sum yx_2 = a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2,\end{cases}$$

$$\begin{cases}471 = 20a + 276{,}5b_1 + 49{,}4b_2,\\ 6569{,}1 = 276{,}5a + 3916{,}59b_1 + 682{,}34b_2,\\ 1216 = 49{,}4a + 682{,}34b_1 + 126{,}7b_2.\end{cases}$$

Из этой системы находятся коэффициенты a, b_1, b_2:

$$a = -13{,}925;\ b_1 = 0{,}686;\ b_2 = 11{,}331.$$

Таким образом, уравнение множественной регрессии имеет следующий вид:

$$\hat{b} = -13{,}925 + 0{,}686x_1 + 11{,}331x_2.$$

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

2.3. Решение с помощью ППП Excel

Задача 2.3. По 20 предприятиям региона (табл. 2.3) 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 (%).

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 OK;

Таблица 2.3

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

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

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

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

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

Выходной интервал – достаточно указывать левую верхнюю ячейку будущего диапазона;

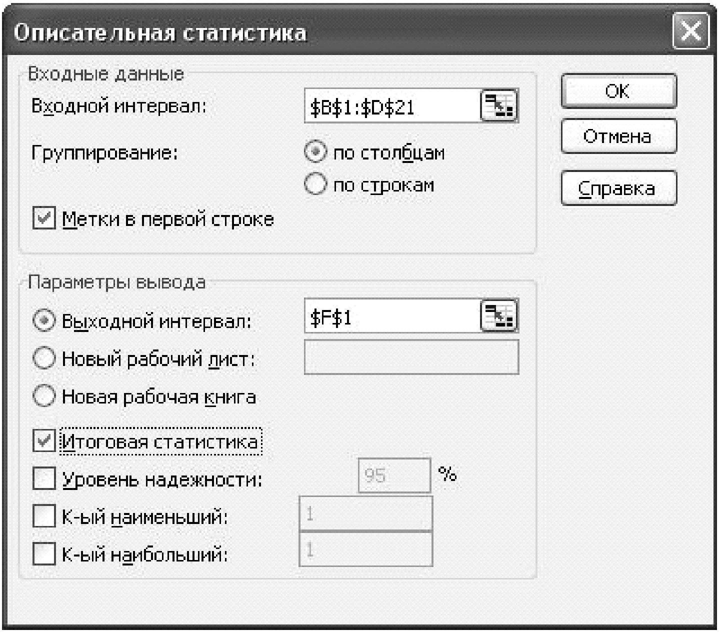


Рис. 2.1. Диалоговое окно ввода параметров инструмента Описательная статистика

Новый рабочий лист – можно задать произвольное имя нового листа.

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

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

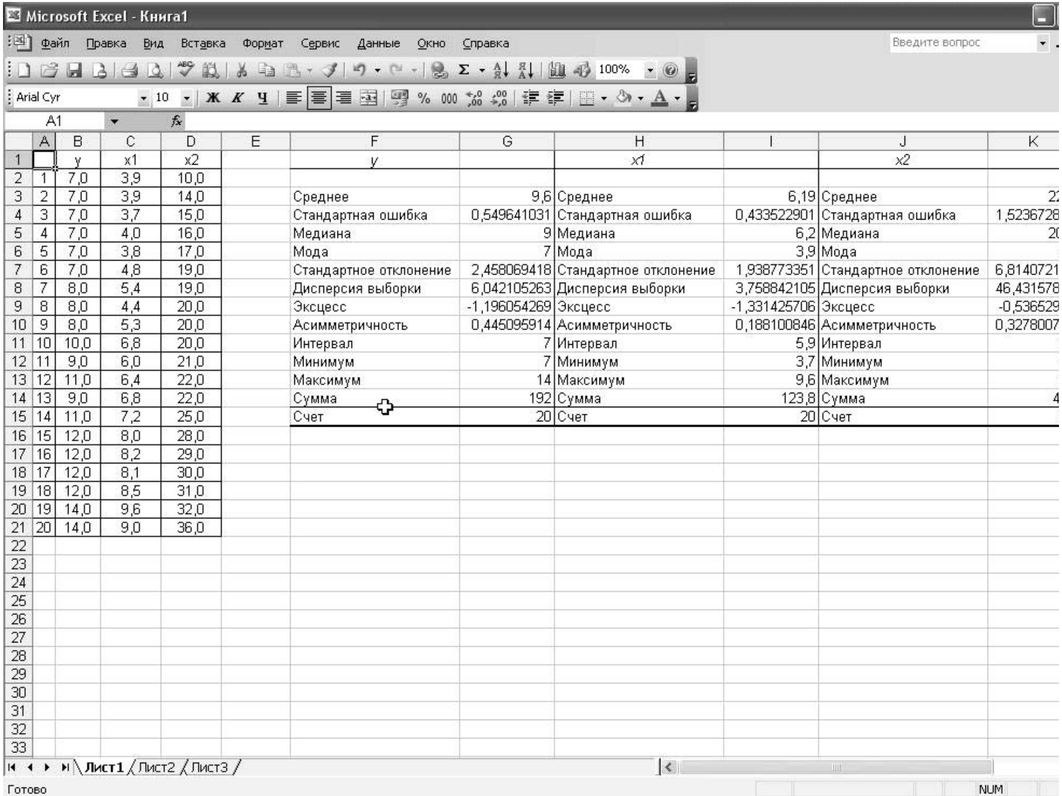
Матрица парных коэффициентов корреляции переменных рассчитывается с использованием инструмента анализа данных Корреляция. Для этого:

1) в главном меню последовательно выберите пункты Сервис / Анализ данных / Корреляция; щелкните по кнопке OK;

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

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

Для вычисления параметров линейного уравнения множественной регрессии используется инструмент анализа данных Регрессия.



	A	B	C	D	E	F	G	H	I	J	K
1		y	x1	x2		y		x1		x2	
2	1	7,0	3,9	10,0							
3	2	7,0	3,9	14,0		Среднее	9,6	Среднее	6,19	Среднее	22
4	3	7,0	3,7	15,0		Стандартная ошибка	0,549641031	Стандартная ошибка	0,433522901	Стандартная ошибка	1,5236728
5	4	7,0	4,0	16,0		Медиана	9	Медиана	6,2	Медиана	20
6	5	7,0	3,8	17,0		Мода	7	Мода	3,9	Мода	
7	6	7,0	4,8	19,0		Стандартное отклонение	2,458069418	Стандартное отклонение	1,938773351	Стандартное отклонение	6,8140721
8	7	8,0	5,4	19,0		Дисперсия выборки	6,042105263	Дисперсия выборки	3,758842105	Дисперсия выборки	46,431578
9	8	8,0	4,4	20,0		Эксцесс	-1,196054269	Эксцесс	-1,331425706	Эксцесс	-0,536529
10	9	8,0	5,3	20,0		Асимметричность	0,445095914	Асимметричность	0,188100846	Асимметричность	0,3278007
11	10	10,0	6,8	20,0		Интервал	7	Интервал	5,9	Интервал	
12	11	9,0	6,0	21,0		Минимум	7	Минимум	3,7	Минимум	
13	12	11,0	6,4	22,0		Максимум	14	Максимум	9,6	Максимум	
14	13	9,0	6,8	22,0		Сумма	192	Сумма	123,8	Сумма	4
15	14	11,0	7,2	25,0		Счет	20	Счет	20	Счет	
16	15	12,0	8,0	28,0							
17	16	12,0	8,2	29,0							
18	17	12,0	8,1	30,0							
19	18	12,0	8,5	31,0							
20	19	14,0	9,6	32,0							
21	20	14,0	9,0	36,0							

Рис. 2.2. *Результат применения инструмента Описательная статистика*

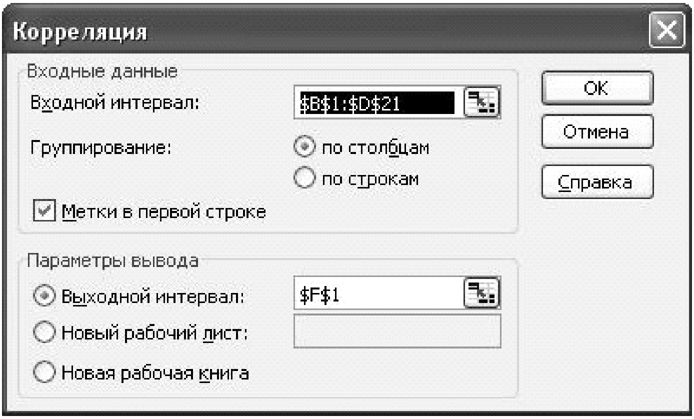
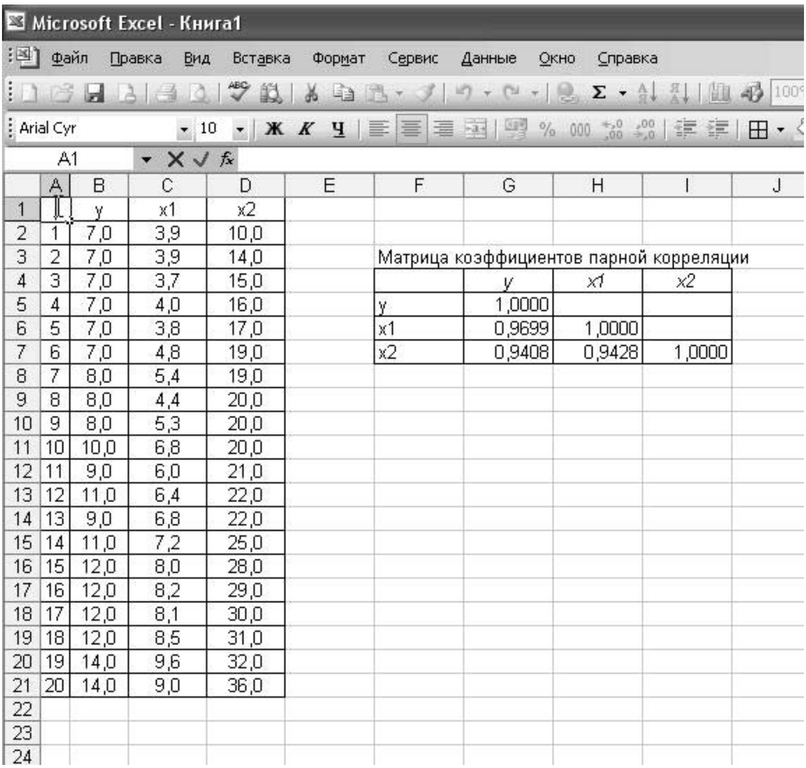


Рис. 2.3. *Диалоговое окно ввода параметров инструмента Корреляция*



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

Матрица коэффициентов парной корреляции

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

Рис. 2.4. *Матрица коэффициентов парной корреляции*

Она аналогична расчету параметров парной линейной регрессии, описанной выше, только в отличие от парной регрессии в диалоговом окне при заполнении параметра *входной интервал X* следует указы-

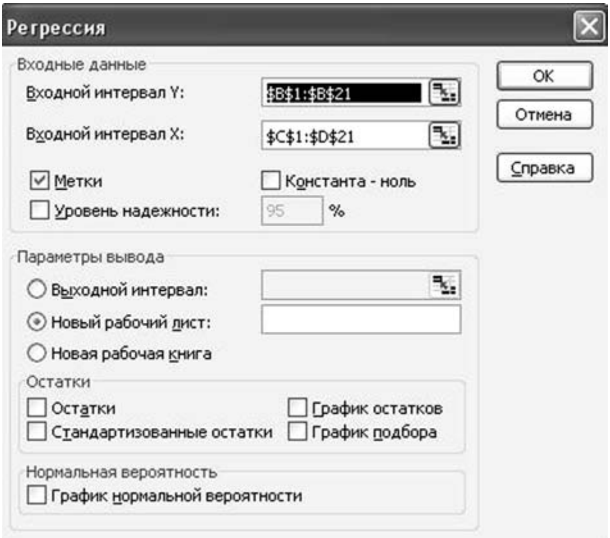
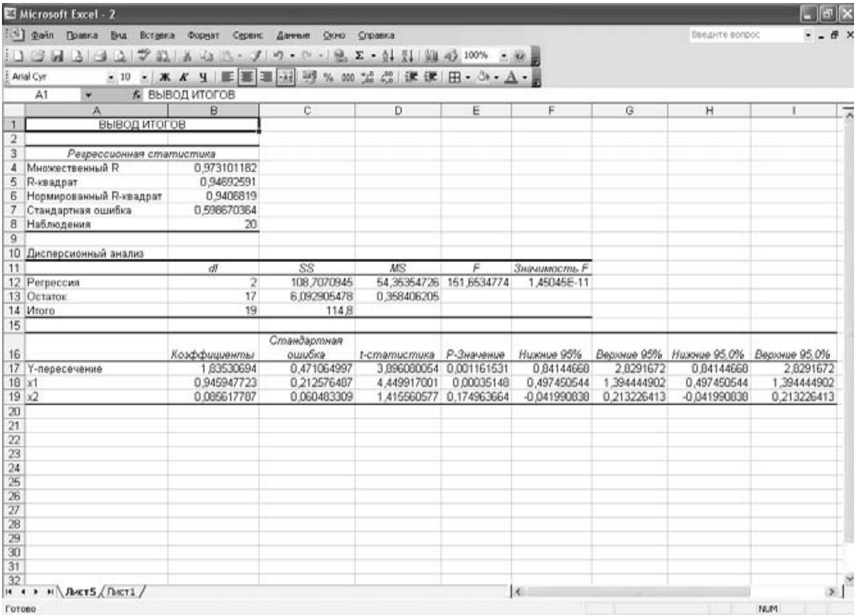


Рис. 2.5. *Диалоговое окно ввода параметров инструмента Регрессия*



	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
2									
3	*Регрессионная статистика*								
4	Множественный R	0,973101182							
5	R-квадрат	0,94692591							
6	Нормированный R-квадрат	0,9406819							
7	Стандартная ошибка	0,598670364							
8	Наблюдения	20							
9									
10	Дисперсионный анализ								
11		*df*	*SS*	*MS*	*F*	*Значимость F*			
12	Регрессия	2	108,7070945	54,35354726	161,6534774	1,45046E-11			
13	Остаток	17	6,092905478	0,358406205					
14	Итого	19	114,8						
15									
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
17	Y-пересечение	1,83530694	0,471064997	3,896080054	0,001161531	0,84144668	2,8291672	0,84144668	2,8291672
18	x1	0,945947723	0,212576487	4,449917001	0,00035148	0,497450544	1,394444902	0,497450544	1,394444902
19	x2	0,085617787	0,060483309	1,415560577	0,174963664	-0,041990838	0,213226413	-0,041990838	0,213226413

Рис. 2.6. *Результат применения инструмента* Регрессия

вать не один столбец, а все столбцы, содержащие значения факторных признаков (рис. 2.5).

Результаты анализа приведены на рис. 2.6.

2.4. Контрольные задания

Задача 1. Зависимость спроса на товар К от его цены характеризуется по 20 наблюдениям уравнением:

$$\lg y = 5{,}25 - 0{,}85 \lg x.$$

Задание.

– записать данное уравнение в виде степенной функции;

– оценить эластичность спроса на товар в зависимости от его цены.

Задача 2. Изучается влияние стоимости основных и оборотных средств на величину валового дохода торговых предприятий (табл. 2.4).

Таблица 2.4

№ п/п	y, валовый доход за год	x_1, среднегодовая стоимость основных фондов	x_2, среднегодовая стоимость оборотных средств
1	303	118	205
2	163	128	156
3	145	117	154
4	213	150	163
5	221	156	128
6	188	202	150
7	210	216	154
8	156	224	142
9	180	214	136
10	237	254	206
11	160	215	188
12	175	198	146

Задание:

– используя ППП Excel, получить таблицу статистических характеристик;

– построить линейное уравнение множественной регрессии и пояснить экономический смысл его параметров;

– рассчитать частные коэффициенты эластичности;

– определить частные и парные коэффициенты корреляции, а также множественный коэффициент корреляции;

· дать оценку полученного уравнения на основе коэффициента детерминации и F-критерия Фишера.

Контрольные вопросы

1. Что представляет собой множественная регрессия?
2. Какие функции применяются для построения множественной регрессии?
3. Чему равен определитель матрицы, если между всеми факторами существует полная линейная связь?
4. Какие факторы включаются в модель множественной регрессии?
5. Как определяется множественный коэффициент корреляции?
6. Что такое фиктивные переменные? Как они входят в уравнение множественной регрессии?
7. Каким критериям должны отвечать оценки параметров регрессии?
8. Для чего применяется анализ остатков при наличии регрессионной модели?

3. ВРЕМЕННЫЕ РЯДЫ В ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ

3.1. Методические указания

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

Временной ряд – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов. Каждый уровень временного ряда формируется из трендовой (T), циклической (S) и случайной (E) компонент. Модели, в которых временной ряд представлен как сумма перечисленных компонент – аддитивные модели, как произведение – мультипликативные модели. Основная задача эконометрического исследования отдельного временного ряда – выявление и придание количественного выражения каждой из перечисленных выше компонент с тем, чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух и более временных рядов.

Аддитивная модель имеет следующий вид:

$$Y = T + S + E.$$

Мультипликативная модель:

$$Y = TSE.$$

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

Построение модели включает следующие шаги:

1) выравнивание исходного ряда методом скользящей средней;

2) расчет значений сезонной компоненты S;

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

4) аналитическое выравнивание уровней ($T + E$) или ($T \cdot E$) и расчет значений Т с использованием полученного уравнения тренда;

5) расчет полученных по модели значений $(T + S)$ или $(T \cdot S)$;

6) расчет абсолютных или относительных ошибок.

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

3.2. Решение типовых задач

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

Таблица 3.1

Год	Квартал	t	Количество возбужденных дел, y_t
2002	I	1	375
	II	2	371
	III	3	869
	IV	4	1015
2003	I	5	357
	II	6	471
	III	7	992
	IV	8	1020
2004	I	9	390
	II	10	355
	III	11	992
	IV	12	905
2005	I	13	461
	II	14	454
	III	15	920
	IV	16	927

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

Шаг 1. Проводится выравнивание исходных уровней ряда методом скользящей средней (табл. 3.2). Для этого:

Таблица 3.2

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

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

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

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

Шаг 2. Находятся оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними. Эти оценки используются для расчета значений сезонной компоненты S. Для этого находятся средние за каждый квартал (по всем годам) оценки сезонной компоненты S_i. В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю (табл. 3.3).

Таблица 3.3

Показатели	Год	№ квартала, i			
		I	II	III	IV
	2002	–	–	213,75	349,5
	2003	–336,75	–238,375	277,875	316,25
	2004	–299,25	–319,875	322,625	214,375
	2005	–233	–233,75	–	–
Всего за i-й квартал		–869	–792	814,25	880,125
Средняя оценка сезонной компоненты для i-го квартала, $\bar{S}_i$		–289,667	–264	271,417	293,375
Скорректированная сезонная компонента, S_i		–292,448	–266,781	268,636	290,593

Для данной модели имеем:

$$-289{,}667-264+271{,}417+293{,}375=11{,}125.$$

Корректирующий коэффициент $k=11{,}125/4=2{,}781.$

Расчет скорректированных значений сезонной компоненты

$$(\,S_i=\bar{S}_i-k\,).$$

Проверка равенства нулю суммы значений сезонной компоненты:

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

Шаг 3. Исключается влияние сезонной компоненты, путем вычитания ее значения из каждого уровня исходного временного ряда. Получаются величины $T + E = Y - S$. Эти значения рассчитываются за каждый момент времени и содержат только тенденцию и случайную компоненту (табл. 3.4).

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

$$T=671{,}777+0{,}9233t.$$

Таблица 3.4

t	y_t	S_i	$y_t - S_i$	T	$T+S$	$E = y_t - (T+S)$	E_2
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	375	–292,448	667,448	672,700	380,252	–5,252	27,584
2	371	–266,781	637,781	673,624	406,843	–35,843	1284,721
3	869	268,636	600,364	674,547	943,183	–74,183	5503,117
4	1015	290,593	724,407	675,470	966,063	48,937	2394,830
5	357	–292,448	649,448	676,394	383,946	–26,946	726,087
6	471	–266,781	737,781	677,317	410,536	60,464	3655,895
7	992	268,636	723,364	678,240	946,876	45,124	2036,175
8	1020	290,593	729,407	679,163	969,756	50,244	2524,460
9	390	–292,448	682,448	680,087	387,639	2,361	5,574
10	355	–266,781	621,781	681,010	414,229	–59,229	3508,074
11	992	268,636	723,364	681,933	950,569	41,431	1716,528
12	905	290,593	614,407	682,857	973,450	–68,450	4685,403
13	461	–292,448	753,448	683,780	391,332	69,668	4853,630
14	454	–266,781	720,781	684,703	417,922	36,078	1301,622
15	920	268,636	651,364	685,627	954,263	–34,263	1173,953
16	927	290,593	636,407	686,550	977,143	–50,143	2514,320

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

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

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

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

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

Шаг 6. Прогнозирование по аддитивной модели. Необходимо дать прогноз об общем объеме правонарушений на I-й и II-й кварталы 2006 года. Прогнозное значение F_t уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда

$$T = 671{,}777 + 0{,}9233t.$$

Получим

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

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

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

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

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

То есть в первые два квартала 2006 года следует ожидать порядка 395 и 422 правонарушений соответственно.

Задача 3.2. На основе помесячных данных о числе браков (тыс.) в регионе за последние три года была построена аддитивная модель временного ряда. Скорректированные значения сезонной компоненты за соответствующие месяцы приведены в табл. 3.5.

Таблица 3.5

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:

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

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

Требуется:

– определить значение сезонной компоненты за декабрь;

– на основе постоянной модели дать прогноз, заключенных в течение 1-го квартала следующего года.

Решение.

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

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

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

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

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

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

Соответствующие значения сезонных компонент составят:

$S_1 = -1$	(январь)
$S_2 = 2$	(февраль)
$S_3 = -0,5$	(март)

Таким образом,

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

Количество браков, заключенных в I-м квартале следующего года, составит:

$$2,61 + 5,64 + 3,17 = 11,42 \text{ тыс.}$$

3.3. Решение с помощью ППП Excel

Задача 3.3. Динамика выпуска продукции Швеции характеризуется данными (млн дол.), приведенными в табл. 3.6.

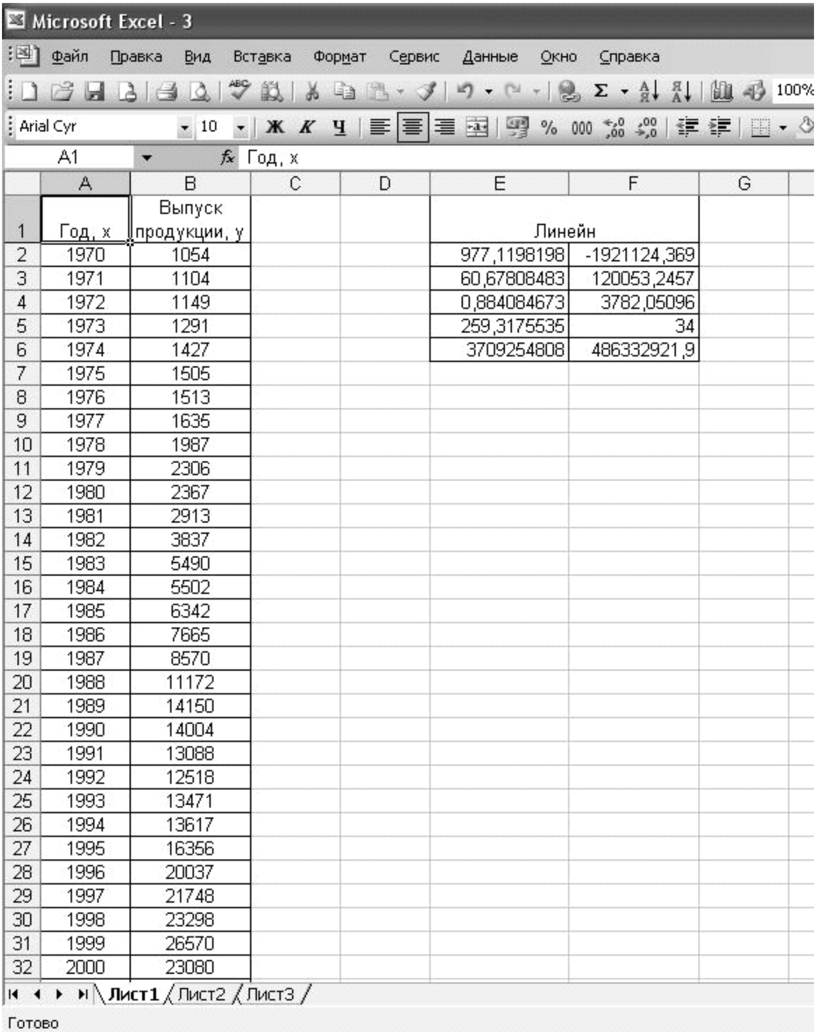
Таблица 3.6

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

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

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

$$\hat{y}_t = -1921124{,}37 + 977{,}12t, \ \hat{y}_t = -1{,}0045^t.$$



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

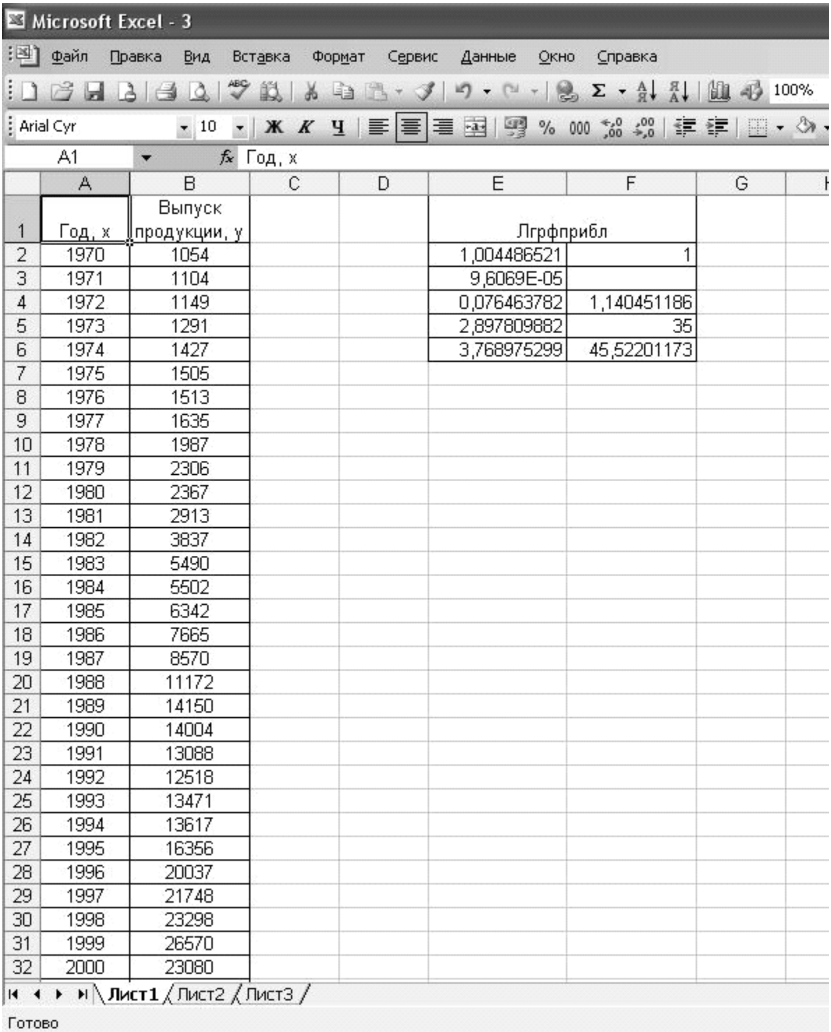
Рис. 3.1. *Результат вычисления функции ЛИНЕЙН*

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:



	A	B	C	D	E	F
1	Год, x	Выпуск продукции, y			Лгрфприбл	
2	1970	1054			1,004486521	1
3	1971	1104			9,6069E-05	
4	1972	1149			0,076463782	1,140451186
5	1973	1291			2,897809882	35
6	1974	1427			3,768975299	45,52201173
7	1975	1505				
8	1976	1513				
9	1977	1635				
10	1978	1987				
11	1979	2306				
12	1980	2367				
13	1981	2913				
14	1982	3837				
15	1983	5490				
16	1984	5502				
17	1985	6342				
18	1986	7665				
19	1987	8570				
20	1988	11172				
21	1989	14150				
22	1990	14004				
23	1991	13088				
24	1992	12518				
25	1993	13471				
26	1994	13617				
27	1995	16356				
28	1996	20037				
29	1997	21748				
30	1998	23298				
31	1999	26570				
32	2000	23080				

Рис. 3.2. *Результат вычисления функции ЛГРФПРИБЛ*

а) в главном меню выберите Вставка / Диаграмма;

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

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

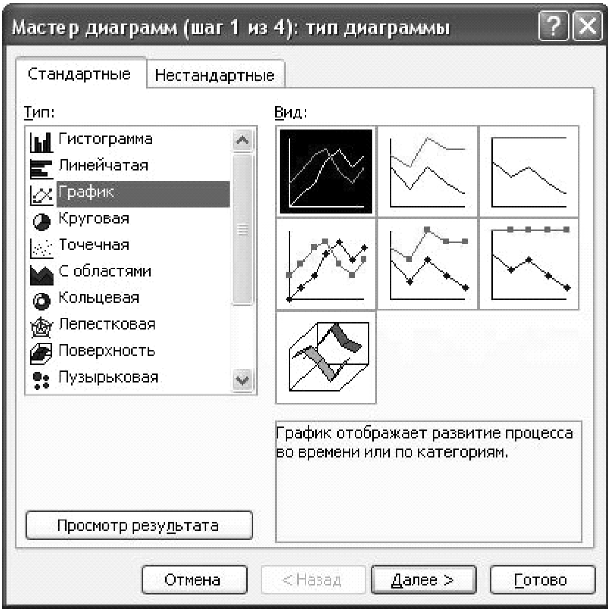


Рис. 3.3. *Диалоговое окно Мастера диаграмм: тип диаграммы*

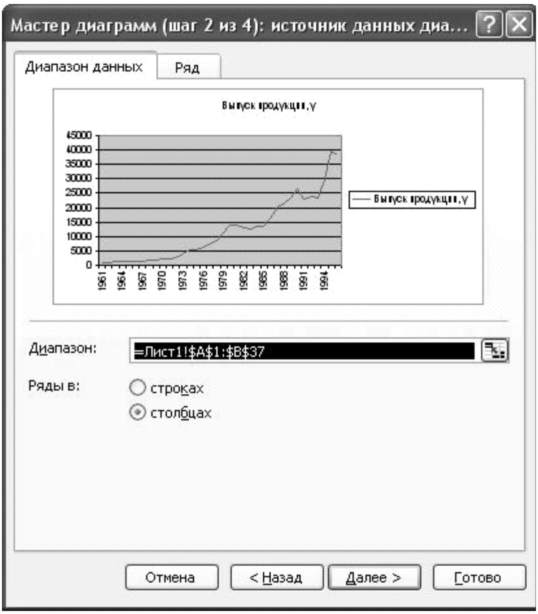


Рис. 3.4. *Диалоговое окно Мастера диаграмм: источник данных*

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

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

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

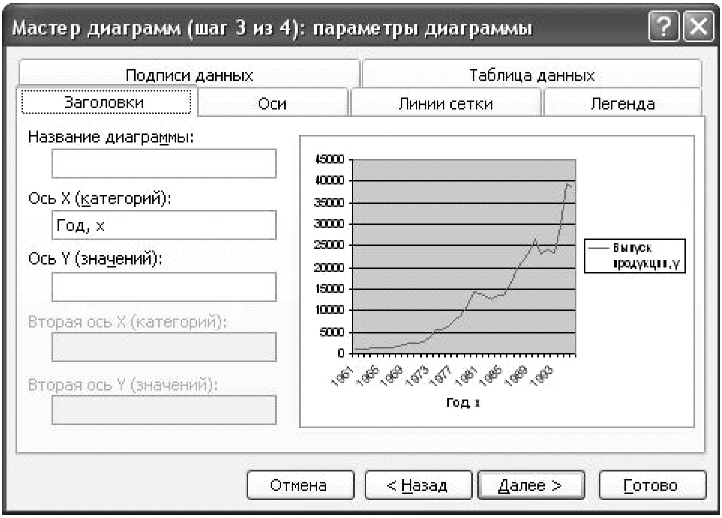


Рис. 3.5. *Диалоговое окно Мастера диаграмм: параметры диаграммы*

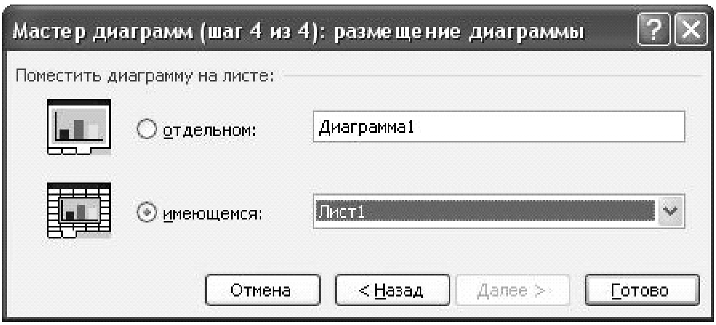


Рис. 3.6. *Диалоговое окно Мастера диаграмм: размещение диаграммы*

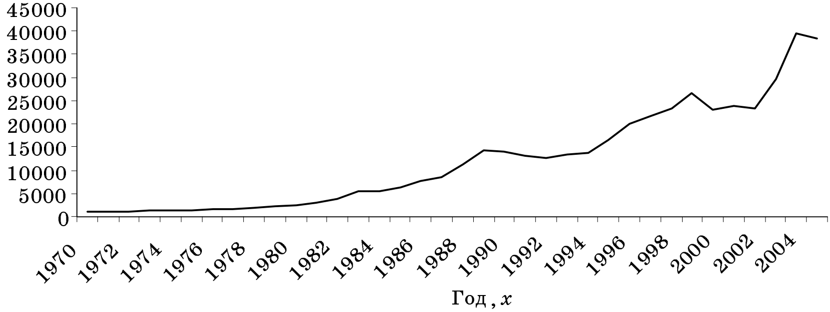


Рис. 3.7. *Динамика выпуска продукции :*

—— – выпуск продукции, y

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

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

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

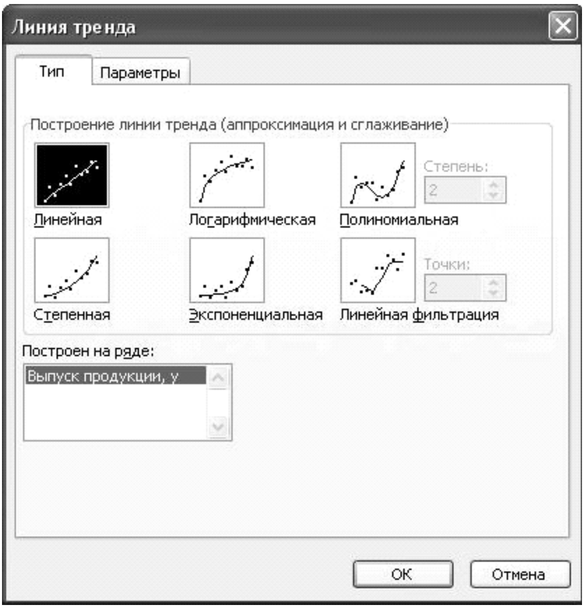


Рис. 3.8. *Диалоговое окно типов линий тренда*

В качестве дополнительной информации на диаграмме можно отобразить уравнение регрессии и значение среднеквадратического отклонения, установив соответствующие флажки на закладке Параметры (рис. 3.9). Щелкните по кнопке OK.

На рис. 3.10–3.14 приведены различные виды трендов, описывающие исходные данные задачи.

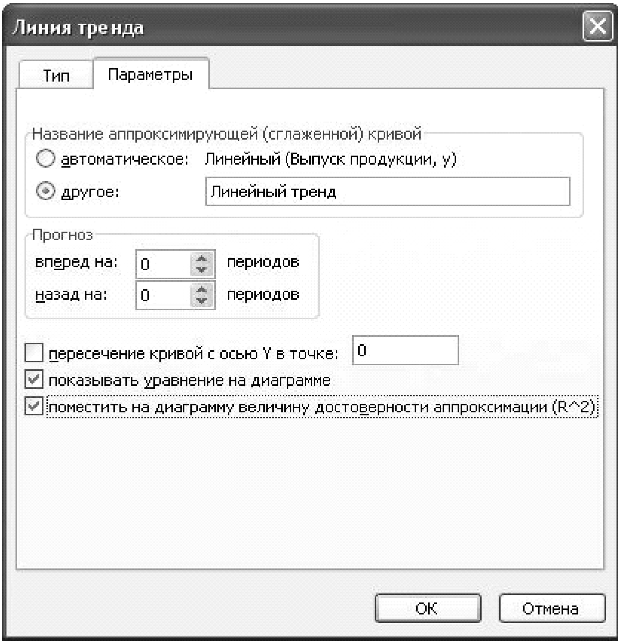


Рис. 3.9. *Диалоговое окно параметров линии тренда*

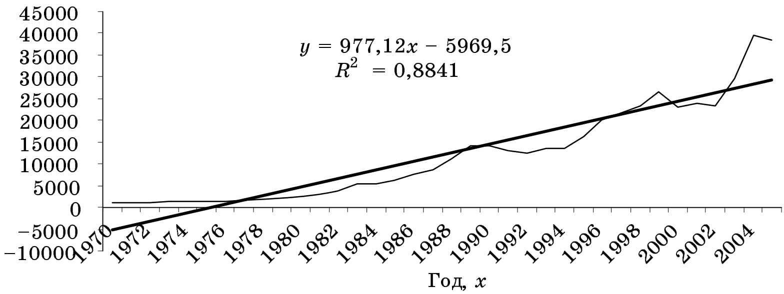


Рис. 3.10. *Линейный тренд:*

— – выпуск продукции, y; — – линейный тренд

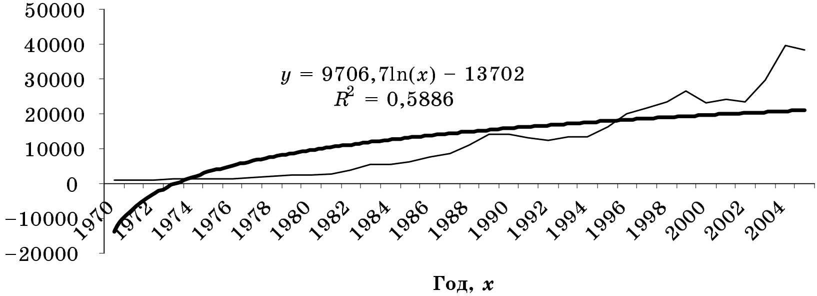


Рис. 3.11. *Логарифмический тренд:*

— – выпуск продукции, y; ▬ – логарифмический тренд

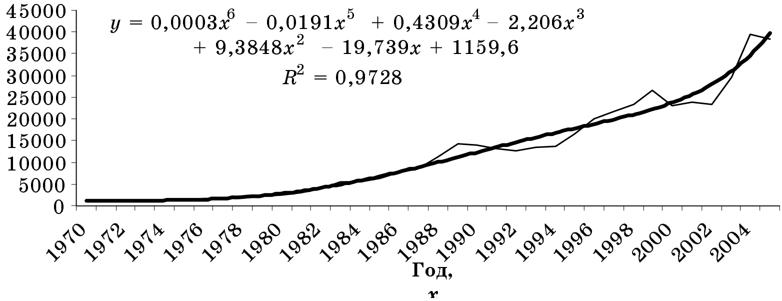


Рис. 3.12. *Полиномиальный тренд:*

— – выпуск продукции, y; ▬ – полиномиальный тренд 6-й степени

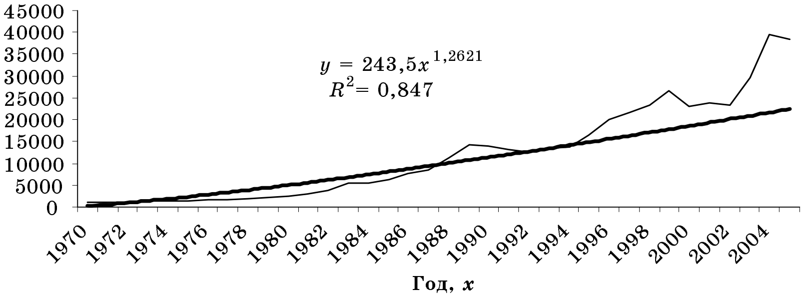


Рис. 3.13. *Степенной тренд:*

— – выпуск продукции, y; ▬ – степенной тренд

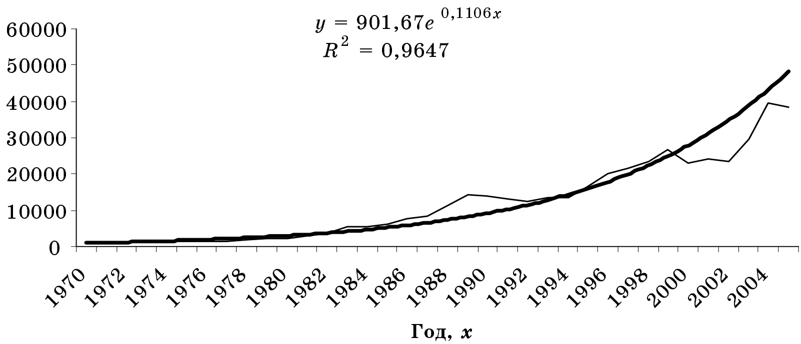


Рис. 3.14. *Экспоненциальный тренд:*

— – выпуск продукции, y; — – экспоненциальный тренд

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

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

3.4. Контрольные задания

Задача 1. На основе помесячных данных о потреблении электроэнергии за последние три года была построена аддитивная модель временного ряда (табл. 3.7).

Таблица 3.7

Месяц	Скорректированные значения сезонной компоненты	Месяц	Скорректированные значения сезонной компоненты
Январь	+25	Июль	–25
Февраль	+10	Август	–18
Март	+6	Сентябрь	+2
Апрель	–4	Октябрь	+15
Май	–32	Ноябрь	+27
Июнь	–38	Декабрь	?

Уравнение тренда выглядит следующим образом: $\hat{y}_t = 300 + 1,5t$.

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

Задание:

– определить значение сезонной компоненты за декабрь;

– дать точечный прогноз ожидаемого потребления электроэнергии в течение I-го квартала следующего года.

Задача 2. Имеются помесячные данные о темпах роста заработной платы за 10 месяцев 2000 г. в процентах к уровню декабря 1999 г. (табл. 3.8).

Таблица 3.8

Месяц	Темпы роста номинальной месячной заработной платы
Январь	82,9
Февраль	87,3
Март	99,4
Апрель	104,8
Май	107,2
Июнь	121,6
Июль	118,6
Август	114,1
Сентябрь	123,0
Октябрь	127,3

Используя ППП Excel, выбрать наилучший тип тренда и определить его параметры.

Контрольные задания

1. Каковы основные элементы временного ряда?
2. В чем состоит задача эконометрического анализа временного ряда?
3. Перечислите основные виды трендов.
4. Что представляют собой параметры линейного и экспоненциального трендов?
5. Что такое аддитивная модель временного ряда? Перечислите этапы ее построения.
6. Как строится мультипликативная модель временного ряда?
7. Что такое скорректированная сезонная компонента и для чего она применяется?

4. СИСТЕМА ЭКОНОМИЧЕСКИХ УРАВНЕНИЙ

4.1. Методические указания

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

Различают несколько видов систем уравнений

Система независимых уравнений – когда каждая зависимая переменная y рассматривается как функция одного и того же набора факторов x:

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

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

Система рекурсивных уравнений – когда зависимая переменная y одного уравнения выступает в виде фактора x в другом уравнении:

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

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

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

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

Такая система уравнений называется *структурной формой модели*.

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

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

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

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

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

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

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

где d – коэффициенты приведенной формы модели.

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

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

С позиции идентифицируемости структурные модели можно подразделить на три вида:

– идентифицируемые;
– неидентифицируемые;
– сверхидентифицируемые.

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

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

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

Необходимое условие идентификации – выполнение счетного правила:

$D + 1 = H$ – уравнение идентифицируемо;
$D + 1 < H$ – уравнение неидентифицируемо;
$D + 1 > H$ – уравнение сверхидентифицируемо,

где H – число эндогенных переменных в уравнении; D – число предопределенных переменных, отсутствующих в уравнении, но присутствующих в системе.

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

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

Косвенный метод наименьших квадратов применяется в случае точно идентифицируемой структурной модели. При этом предполагается выполнение следующих этапов работы:

1) составляют приведенную форму модели и определяют численный значения параметров каждого его уравнения обычным МНК;

2) путем алгебраических преобразований переходят от приведенной формы к уравнениям структурной формы модели, получая тем самым численные оценки структурных параметров.

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

– составляют приведенную форму модели и определяют численные значения параметров каждого его уравнения обычным МНК;

– выявляют эндогенные переменные, находящиеся в правой части структурного уравнения, параметры которого определяют косвенным МНК, и находят расчетные значения по соответствующим уравнениям приведенной формы модели;

– обычным МНК определяют параметры структурного уравнения, используя в качестве исходных данных фактические значения предопределенных переменных и расчетные значения эндогенных переменных, стоящих в правой части данного структурного уравнения.

4.2. Решение типовой задачи

Задача 4.1. Изучается модель вида

$$y = a_1 + b_1(C + D);$$
$$C = a_2 + b_2 y + b_3 y_{-1},$$

где y – валовой национальный доход; y_{-1} – валовой национальный доход предшествующего года; C – личное потребление; D – конечный спрос (помимо личного потребления).

Информация за девять лет о приростах всех показателей дана в табл. 4.1.

Для данной модели была получена система приведенных уравнений:

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

Требуется:

– провести идентификацию модели;

– рассчитать параметры первого уравнения структурной модели.

Таблица 4.1

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

Решение. 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 на параметры при C и D наложено ограничение: они должны быть равны. В этом уравнении содержится одна эндогенная переменная y. Переменная 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: $1 + 1 = 2$: $D + 1 > H$. Это больше, чем число эндогенных переменных в уравнении. Следовательно, система сверхидентифицирована.

Для определения параметров сверидентифицированной 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 $C = 8{,}636 + 0{,}3384D + 0{,}2020y_{-1}$ подставим значения D и y_{-1}. Получим:

$$\widehat{C}_1 = 15{,}8;\ \ \widehat{C}_2 = 16{,}8;\ \ \widehat{C}_3 = 7{,}4;\ \ \widehat{C}_4 = 14{,}3;\ \ \widehat{C}_5 = 15{,}0;\ \ \widehat{C}_6 = 27{,}4;$$

$$\widehat{C}_7 = 24{,}0;\ \ \widehat{C}_8 = 33{,}2;\ \ \widehat{C}_9 = 29{,}0.$$

Шаг 2. По сверхидентифицированному уравнению структурной формы модели заменяем фактические значения C на теоретические $\widehat{C}$ и рассчитываем новую переменную $\widehat{C}+D$ (табл. 4.2).

Таблица 4.2

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

Далее к сверхидентифицированному уравнению применяем метод наименьших квадратов. Обозначим новую переменную $\widehat{C}+D$ через Z. Решаем следующее уравнение:

$$y = a_1 + b_1 Z.$$

Система нормальных уравнений составит:

$$\begin{cases} \sum y = na_1 + b_1 \sum Z, \\ \sum yZ = a_1 \sum Z + b_1 \sum Z^2, \end{cases} \qquad \begin{cases} 248{,}4 = 9a_1 + 350{,}4b_1, \\ 13508{,}71 = 350{,}4a_1 + 21142{,}02b_1. \end{cases}$$

$$a_1 = 7{,}678; \quad b_1 = 0{,}512;$$

Первое уравнение структурной модели будет иметь следующий вид:

$$y = 7{,}678 + 0{,}512(C + D).$$

4.3. Контрольные задания

Задача 1. Модель Кейнса (упрощенная версия), описывающая макроэкономическое равновесие, зависимость объема производства и уровня занятости от размеров совокупного спроса, при условии, что отсутствует изменение заработной платы и цен, представлена следующими уравнениями:

$$C_t = a_1 + b_{11}y_t + b_{12}y_{t-1} \text{ – функция потребления;}$$

$$I_t = a_2 + b_{21}y_t \text{ – функция инвестиций;}$$

$$Y_t = C_t + I_t + G_t \text{ – тождество доходов,}$$

где C – потребление; Y – ВВП; I – валовые инвестиции; G – государственные расходы; t – текущий период; t–1 – предыдущий период.

Задание:

– применив необходимое и достаточное условие идентификации, определите, идентифицировано ли каждое из уравнений;

– определите метод оценки параметров модели;

– запишите приведенную форму модели.

Задача 2. Модель спроса и предложения на деньги

$$R_t = a_1 + b_{11}M_t + b_{12}Y_t;$$
$$Y_t = a_2 + b_{21}R_t,$$

где R – процентные ставки в период t; Y – ВВП в период t; M – денежная масса в период t.

Задание:

– применив необходимое и достаточное условие идентификации, определите, идентифицировано ли каждое из уравнений;

– определите метод оценки параметров модели;

– запишите приведенную форму модели.

Задача 3. Макроэкономическая модель (упрощенная модель Клейна):

$$C_t = a_1 + b_{12}Y_t + b_{13}T_t;$$
$$I_t = a_2 + b_{21}Y_t + b_{24}K_{t-1};$$
$$Y_t = C_t + I_t,$$

где C – потребление; I – инвестиции; Y – доход; T – налоги; K – запас капитала; t – текущий период; t–1 – предыдущий период.

Задание:

– применив необходимое и достаточное условие идентификации, определите, идентифицировано ли каждое из уравнений;

– определите метод оценки параметров модели;

– запишите приведенную форму модели.

Контрольные вопросы

1. Назовите возможные способы построения систем уравнений. В чем их отличия?
2. В чем заключаются проблемы идентификации модели?
3. Каковы необходимые условия идентификации?
4. Каковы достаточные условия идентификации?
5. Что такое эндогенные и экзогенные переменные?
6. В чем состоит косвенный метод наименьших квадратов?
7. Что такое двухшаговый метод наименьших квадратов? В каком случае он применяется?

Библиографический список

1. Эконометрика: учебник / Под ред. И. И. Елисеевой. – М.: Финансы и статистика, 2001. – 344 с.: ил.

2. Эконометрия / В. И. Суслов и др. – Новосибирский государственный университет, 2005. – 744 с.

3. Шалабанов А. К., Роганов Д. А. Эконометрика. Учебно-методическое пособие. ТИСБИ, Казань, 2004. – 198 с.

4. Доугерти К. Введение в эконометрику. – М.: Финансы и статистика, 1999

Таблица 1. *Значения F-критерия Фишера при уровне значимости $a = 0{,}05$*

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

Окончание табл.1

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
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

Таблица 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	4,5041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

СОДЕРЖАНИЕ

Предисловие 3
Введение 4
1. Парная регрессия и корреляция 6
1.1. Методические указания 6
1.2. Решение типовых задач 13
1.3. Решение с помощью ППП Excel 19
1.4. Контрольные задания 24
Контрольные вопросы 26
2. Множественная регрессия и корреляция 28
2.1. Методические указания 28
2.2. Решение типовых задач 31
2.3. Решение с помощью ППП Excel 34
2.4. Контрольные задания 40
Контрольные вопросы 41
3. Временные ряды в экономических исследованиях 42
3.1. Методические указания 42
3.2. Решение типовых задач 43
3.3. Решение с помощью ППП Excel 49
3.4. Контрольные задания 57
Контрольные задания 58
4. Система экономических уравнений 59
4.1. Методические указания 59
4.2. Решение типовой задачи 62
4.3. Контрольные задания 64
Контрольные вопросы 66
Библиографический список 67
Приложение 68

Учебное издание

Семенова Елена Георгиевна
Смирнова Марина Сергеевна

ОСНОВЫ ЭКОНОМЕТРИЧЕСКОГО АНАЛИЗА

Учебное пособие

Редактор *А. Г. Ларионова*
Вестальщик *А. Н. Колешко*

Сдано в набор 24.09.06. Подписано к печати 10.10.06. Формат 60×84 1/16.
Гарнитура SchoolBookC. Бумага офсетная. Печать офсетная.
Усл. печ. л. 4,38. Уч. -изд. л. 4,51. Тираж 300 экз. Заказ № 503.

Редакционно-издательский центр ГУАП
190000, Санкт-Петербург, Б. Морская ул., 67