Министерство образования и науки Российской Федерации
Волгоградский государственный архитектурно-строительный университет

ЭКОНОМЕТРИКА (ПРОДВИНУТЫЙ УРОВЕНЬ)

Методические указания к практическим работам

Составители Т. В. Соловьева, С. В. Харланова

Волгоград
ВолгГАСУ
2016

УДК 330.43(076.5)
ББК 65В631я73
Э40

Э40 **Эконометрика** (продвинутый уровень): методические указания к практическим работам / сост. Т. В. Соловьева, С. В. Харланова ; М-во образования и науки Росс. Федерации, Волгогр. гос. архит.-строит. ун-т. — Электронные текстовые и графические данные (1,6 Мбайт). — Волгоград : ВолгГАСУ, 2016. — Электронное издание сетевого распространения. — Систем. требования: PC 486 DX-33; Microsoft Windows XP; Internet Explorer 6.0; Adobe Reader 6.0. — Официальный сайт Волгоградского государственного архитектурно-строительного университета. Режим доступа: http://www.vgasu.ru/publishing/on-line/ — Загл. с титул. экрана.

Предназначены для студентов-магистров, аспирантов всех специальностей очной и заочной форм обучения. Приведены необходимые теоретические сведения и формулы, примеры решения типовых задач.

УДК 330.43 (076.5)
ББК 65В631я73

Оглавление

1. Понятие эконометрики и основные этапы эконометрического моделирования

Эконометрика — одна из базовых дисциплин экономического образования во всем мире.

Слово «эконометрика» произошло от двух слов: «экономика» и «метрика» (от греч. «метрон» — правило определения расстояния между двумя точками в пространстве, «метрия» — измерение).

Существуют различные варианты определения эконометрики. Это расширенные варианты, при которых к эконометрике относят все, что связано с измерениями в экономике, а также узко инструментально ориентированные, при которых понимают определенный набор математико-статистических средств, позволяющих верифицировать модельные соотношения (проверка их истинности и адекватности) между анализируемыми экономическими показателями.

Объединяя эти два определения, можно сказать, что *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 закономерностям, обусловленным экономической теорией.

Эконометрика ставит своей целью количественно охарактеризовать те экономические закономерности, которые экономическая теория выявляет и определяет лишь в общем. Эконометрика представляет собой сочетание трех наук: экономической теории; математической и экономической статистики; математики.

На современном этапе развития науки неотъемлемым фактором развития эконометрики является развитие компьютерных технологий и специальных пакетов прикладных программ.

Анализ экономических процессов и явлений осуществляется с помощью математических моделей, построенных на эмпирических данных. Практически все эконометрические методы и приемы изучения закономерностей позаимствованы из математической статистики. Специфика применения методов математической статистики в эконометрике заключается в том, что практически все экономические показатели являются величинами случайными, а не результатами контролируемого эксперимента. Кроме того, зачастую экономические данные содержат ошибки измерения. Поэтому существуют определенные усовершенствования и дополнения методов, которые в математической статистике не используются и в эконометрике разрабатываются специальные методы, позволяющие устранить или снизить влияние ошибок на результаты экспериментов.

С помощью эконометрики решается очень широкий круг задач, которые можно классифицировать по трем признакам:

1. *по конечным прикладным целям*

а) прогноз социально-экономических показателей, определяющих состояние и развитие изучаемой системы;

б) моделирование возможных вариантов социально-экономического развития системы для определения тех параметров, которые оказывают наиболее мощное влияние на состояние системы в целом:

2. *по уровню иерархии*

а) задачи, решаемые на макроуровне (страны в целом):

б) задачи, решаемые на мезоуровне (уровень отраслей, регионов);

в) задачи, решаемые на микроуровне (уровень предприятия, фирмы, семьи):

3. *по области решения проблем изучаемой экономической системы*

а) рынок;

б) инвестиционная, социальная, финансовая политика;

в) ценообразование;

г) распределительные отношения;

д) спрос и потребление;

е) отдельно выделенный комплекс проблем.

1.1. Основные виды эконометрических моделей

Выделяют три основных класса эконометрических моделей.

1. *Регрессионные модели с одним уравнением*.

В подобных моделях зависимая или результативная переменная представляется в виде функции факторных или независимых переменных. Регрессионные модели делятся на парные (с одним факторным признаком) и множественные регрессии. В зависимости от вида функции модели делятся на линейные и нелинейные регрессии.

2. *Модель временных рядов*.

Модель представляет собой зависимость результативного признака от переменной времени или переменных, относящихся к другим моментам времени. Модели временных рядов делятся на модели, построенные по стационарным и нестационарным временным рядам. Стационарные временные ряды характеризуются постоянными во времени величинами: средней, дисперсией и автокорреляцией. В противном случае временной ряд является нестационарным

3. *Системы одновременных уравнений*.

Данные модели описываются системами взаимозависимых регрессионных уравнений. Системы могут состоять из тождеств и регрессионных уравнений, каждое из которых может содержать не только факторные переменные, но и результативные переменные из других уравнений системы. Для тождеств характерно то, что их вид и значения параметров известны.

Регрессионные уравнения, из которых состоит система, называются поведенческими уравнениями. В поведенческих уравнениях значения параметров являются неизвестными и подлежат оцениванию.

1.2. Эконометрическое моделирование

Для описания сущности эконометрической модели весь процесс моделирования разбивают на несколько основных этапов.

1. *Постановочный*. На данном этапе определяются конечные цели и задачи исследования и набор участвующих в модели факторных и результативных экономических переменных.

Включение в эконометрическую модель той или иной переменной должно быть теоретически обосновано. Число переменных не должно быть слишком большим. Факторные переменные не должны быть связаны функциональной или тесной корреляционной зависимостью. Присутствие в модели мультиколлинеарности может привести к негативным последствиям всего процесса моделирования.

2. *Априорный*. На этом этапе проводятся теоретический анализ сущности изучаемого процесса, а также формирование и формализация известной до моделирования (априорной) информации.

3. *Параметризация*. Осуществляется выбор общего вида модели и выявление состава и формы, входящих в неё связей, т.е. происходит непосредственно моделирование.

Основная задача этапа моделирования заключается в выборе наиболее оптимального вида функции зависимости результативной переменной от факторных признаков. Если возникает возможность выбора между линейной и нелинейной формой зависимости, то предпочтение всегда отдается линейной форме, как наиболее простой и надежной.

Помимо этого, на этапе моделирования решается задача спецификации модели путем аппроксимации математической формой выявленных связей и соотношений между переменными; определения зависимых и независимых переменных; формулировки исходных предпосылок и ограничений модели.

Успех эконометрического моделирования во многом зависит от правильного решения проблемы спецификации модели.

4. *Информационный***.** Происходит сбор необходимой статистической базы данных, т.е. наблюдаемых (эмпирических) значений экономических переменных, анализ качества собранной информации.

5. *Идентификация модели*. Осуществляется статистический анализ модели и производится статистическое оценивание неизвестных параметров.

6. *Верификация модели*. Проверяются достоверность и адекватность модели, т.е. определяется, насколько успешно решены задачи спецификации и идентификации модели, какова точность расчетов, полученных на её основе. Кроме того, построенная модель должна быть адекватна реальному экономическому процессу. Если качество модели оказывается неудовлетворительным, то снова возвращаются ко второму этапу моделирования.

7. *Интерпретация результатов моделирования*. Среди наиболее известных эконометрических моделей можно выделить модели потребительского и сберегательного потребления; модели взаимосвязи риска и доходности ценных бумаг; модели предложения труда; макроэкономические модели (модель роста); модели инвестиций; маркетинговые модели; модели валютных курсов и валютных кризисов и многие другие.

1.3. Классификация видов эконометрических переменных и типов данных

В эконометрических исследованиях, как правило, используются два типа выборочных данных: пространственные данные (cross-sectional data); временные данные (time-series data).

Под пространственными данными понимается совокупность экономической информации, относящейся к разным объектам, полученной за один и тот же период или момент времени. Пространственные данные представляют собой выборочную совокупность из некоторой генеральной совокупности. Например, совокупность различной информации по какому-либо предприятию (численность рабочих, объем производства, размер заработной платы, размер основных фондов и т.д.) есть пространственные данные.

Под временными данными понимается совокупность экономической информации, характеризующей один и тот же объект, но за разные периоды времени. Отдельно взятый временной ряд можно считать выборкой из бесконечного ряда значений показателей во времени. Например, данные о динамике индекса потребительских цен, ежедневные обменные курсы валют представляют собой временные данные. Временная информация упорядочена во времени в отличие от пространственных данных.

Существуют определенные отличия временного ряда от пространственной выборки. Элементы временного ряда не являются статистически независимыми, в отличие от случайной пространственной выборки, т.е. они подвержены явлению автокорреляции (зависимости между прошлыми и текущими наблюдениями временного ряда). Кроме того, элементы временного ряда не являются одинаково распределенными величинами.

Экономические переменные, участвующие в любой эконометрической модели, делятся на четыре вида:

1) экзогенные (независимые) — переменные, значения которых задаются извне;

2) эндогенные (зависимые) переменные, значения которых определяются внутри модели;

3) лаговые — экзогенные или эндогенные переменные в эконометрической модели, взятые в предыдущий момент времени и находящиеся в уравнении с переменными, относящимися к текущему моменту времени;

4) преопределенные (объясняющие переменные) — лаговые и текущие экзогенные переменные, а также лаговые эндогенные переменные.

Любая эконометрическая модель предназначена для объяснения значений одной или нескольких текущих эндогенных переменных в зависимости от значений предопределенных переменных.

2. МНОЖЕСТВЕННЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

2.1. Множественный регрессионный анализ. Стандартизированные коэффициенты регрессии и коэффициенты эластичности

Парная регрессия дает хороший результат при моделировании, если влиянием других факторов, воздействующих на объект исследования можно пренебречь. Между тем экономические явления часто определяются большим числом одновременно действующих факторов. Поэтому возникает задача исследования одной зависимой переменной Y от нескольких объясняющих переменных X_1, X_2, …, X_p. Эта задача решается с помощью *множественного регрессионного анализа*. Множественная регрессия широко используется в решении проблем спроса, доходности акций, при изучении функции издержек производства и ряда других вопросов эконометрики. *Основная цель множественной регрессии* – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на зависимую переменную.

Модель множественной линейной регрессии можно представить в виде:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i, \quad i = \overline{1, n}. \tag{2.1}$$

где i – номер наблюдения, n – число наблюдений, ε_i – возмущения, удовлетворяющие основным предпосылкам регрессионного анализа [3].

В матричном виде модель (2.1) имеет вид:

$$Y = X\beta + \varepsilon, \tag{2.2}$$

где $Y = (y_1 \ y_2 \ \ldots \ y_n)^T$ – матрица-столбец значений зависимой переменной размера n;

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$ – матрица значений объясняющих переменных размера $n \times (p + 1)$;

$\beta = (\beta_0\ \beta_1 \dots \beta_p)^T$ – матрица-столбец параметров размера $(p + 1)$;

$\varepsilon = (\varepsilon_1\ \varepsilon_2\ \dots \varepsilon_n)^T$ – матрица-столбец возмущений размера n.

Замечание. В матрицу X дополнительно введен столбец, все элементы которого равны 1, т.е. в модели (2.1) свободный член β_0 умножается на фиктивную переменную x_{i0}, принимающую значение 1 для всех i:

$x_{i0} = 1$ ($i = 1, 2, \dots n$).

Оценкой модели (2.2) является уравнение

$$Y = Xb + e, \tag{2.3}$$

где $b = (b_0\ b_1 \dots b_p)^T$, $\quad e = (e_1\ e_n \dots e_n)^T$.

Для оценки неизвестных параметров применим метод наименьших квадратов (МНК). Для этого минимизируем остаточную сумму квадратов:

$$S = \sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2 = (Y - Xb)^T (Y - Xb) \to \mathbf{min}. \tag{2.4}$$

На основании необходимого условия экстремума функции нескольких переменных $S(b_0, b_1, \dots, b_p)$, представляющей (2.4), необходимо приравнять нулю частные производные по этим переменным или в матричном виде – вектор частных производных $\frac{\partial S}{\partial b} = \begin{pmatrix} \frac{\partial S}{\partial b_0} & \frac{\partial S}{\partial b_1} \dots & \frac{\partial S}{\partial b_p} \end{pmatrix} = 0$. Получим систему нормальных уравнений в матричном виде для определения вектора b:

$$X^T X b = X^T Y, \tag{2.5}$$

где

$$X^T X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \cdot \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} = \tag{2.6}$$

$$= \begin{pmatrix} n & \sum x_{i1} & \dots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}{}^2 & \dots & \sum x_{i1} x_{ip} \\ \dots & \dots & \dots & \dots \\ \sum x_{ip} & \sum x_{i1} x_{ip} & \dots & \sum x_{ip}{}^2 \end{pmatrix};$$

$$X^T Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \dots \\ \sum y_i x_{ip} \end{pmatrix} \tag{2.7}$$

– матрица-столбец свободных членов.

Решением уравнения (2.5) является вектор

$$b = \left(X^T X\right)^{-1} X^T Y, \tag{2.8}$$

где $\left(X^T X\right)^{-1}$ – матрица, обратная матрице коэффициентов системы (2.5).

На практике бывает необходимо сравнить влияние на зависимую переменную различных объясняющих переменных, когда последние выражаются разными единицами измерения. В этом случае используют *стандартизированные коэффициенты регрессии* b'_j и *коэффициенты эластичности* E_j (j = 1, 2, … p):

$$b'_j = b_j \frac{s_{x_j}}{s_y}; \tag{2.9}$$

где s_{x_j} – среднее квадратическое отклонение по переменной x_j:

$$s_{x_j} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_{ij}^2 - (\bar{x}_j)^2}, \tag{2.10}$$

s_y – среднее квадратическое отклонение по переменной y:

$$s_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^{\,2} - \bar{y}^2}\,. \qquad (2.11)$$

$$E_j = b_j \frac{\bar{x}_j}{\bar{y}}, \qquad (2.12)$$

где $\bar{x}_j$ – среднее значение переменной x_j; $\bar{y}$ – среднее значение переменной y.

Стандартизированный коэффициент регрессии b'_j показывает, на сколько величин s_y изменится в среднем зависимая переменная Y, при увеличении только j-й объясняющей переменной на s_{x_j}, а коэффициент эластичности E_j – на сколько процентов (от средней) изменится в среднем Y при увеличении только x_j на 1%.

2.2. Доверительные интервалы для коэффициентов регрессии

Значимость коэффициентов регрессии b_j *и доверительный интервал для параметров регрессионной модели* βj (j = 1, 2, … p).

Проверим значимость коэффициента регрессии b_j по критерию Стьюдента. Для этого проверим гипотезу $H_0 : \beta_j = 0$ (о статистической незначимости коэффициента регрессии). Альтернативная гипотеза $H_1 : \beta_j \neq 0$.

Если

$$|t| = \frac{|b_j|}{s_{b_j}} > t_{1-\alpha, n-p-1}, \qquad (2.13)$$

то гипотеза H_0 отвергается и принимается гипотеза H_1 о статистической значимости коэффициента регрессии b_j на уровне значимости α, где

$$s_{b_j} = \sqrt{s^2 \cdot \left[(X^T X)^{-1}\right]_{jj}} \qquad (2.14)$$

– среднее квадратическое отклонение (стандартная ошибка) коэффициента регрессии b_j;

$\left[(X^T X)^{-1}\right]_{jj}$ – диагональный элемент матрицы $(X^T X)^{-1}$; $t_{1-\alpha,n-p-1}$ – табличное значение критерия Стьюдента при уровне значимости α и с n-p-1 степенями свободы;

s^2 – несмещенная оценка параметра σ^2:

$$s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1}, \quad e_i = y_i - y_{i\text{оцен}}. \tag{2.15}$$

Доверительный интервал для параметра βj имеет вид:

$$b_j - t_{1-\alpha;n-p-1} \cdot s_{b_j} \le \beta_j \le b_j + t_{1-\alpha;n-p-1} \cdot s_{b_j}. \tag{2.16}$$

Доверительный интервал для параметра σ^2 имеет вид:

$$\frac{ns^2}{\chi^2_{\alpha/2;n-p-1}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{1-\alpha/2;n-p-1}}, \tag{2.17}$$

где $\chi^2_{\alpha/2;n-p-1}$, $\chi^2_{1-\alpha/2;n-p-1}$ – табличные значения χ^2 – распределения с n - p - 1 степенями свободы и уровнем значимости α.

2.3. Оценка значимости множественной регрессии.

Коэффициент детерминации.

Скорректированный коэффициент детерминации

В случае множественной регрессии коэффициент детерминации определяется по формуле:

$$R^2 = \frac{Q_R}{Q} = \frac{b^T X^T Y^T - n\bar{y}^2}{Y^T Y - n\bar{y}^2} = 1 - \frac{e^T e}{y^T y}, \tag{2.18}$$

где Q_R – сумма квадратов, обусловленная регрессией; Q_e – остаточной сумма квадратов, характеризующая влияние неучтенных факторов:

$$Q_e = e^T e = \sum_{i=1}^{n} e_i^2, \quad y^T y = \sum_{i=1}^{n} (y_i - \bar{y})^2. \qquad (\text{`}2.19)$$

Критерий значимости уравнения регрессии определяется по формуле:

$$F = \frac{R^2(n-p-1)}{(1-R^2)p} > F_{\alpha;p;n-p-1}, \qquad (2.20)$$

где $F_{\alpha;p;n-p-1}$ – табличное значение F-критерия Фишера-Снедекора, определенное на уровне значимости α при p и n-p-1 степенях свободы.

Использование только одного коэффициента детерминации R^2 для выбора лучшего уравнения регрессии может оказаться недостаточным. На практике бывают случаи, когда плохо определенная модель регрессии дает высокий коэффициент R^2. Недостатком этого коэффициента является то, что он увеличивается при добавлении новых объясняющих переменных, хотя это и не всегда означает улучшение качества регрессионной модели. В этом случае лучше использовать скорректированный коэффициент детерминации $\hat{R}^2$, определяемый по формуле:

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1}\left(1-R^2\right). \qquad (2.21)$$

Чем больше число объясняющих переменных p, тем меньше $\hat{R}^2$ по сравнению с R^2. Скорректированный коэффициент $\hat{R}^2$ может уменьшаться при введении в уравнение новых объясняющих переменных, не оказывающих заметного влияния на зависимую переменную.

Пример 1. Изучается влияние стоимости основных и оборотных средств на величину валового дохода торговых предприятий. Для этого по 12 торговым предприятиям были получены следующие данные (табл. 2.1).

Таблица 2.1

Номер предприятия	Валовый доход за год, млн. руб.	Среднегодовая стоимость, млн. руб.	
		основных фондов	оборотных средств
1	203	118	105
2	63	28	56
3	45	17	54
4	113	50	63
5	121	56	28

Номер предприятия	Валовый доход за год, млн. руб.	Среднегодовая стоимость, млн. руб. основных фондов	Среднегодовая стоимость, млн. руб. оборотных средств
6	88	102	50
7	110	116	54
8	56	124	42
9	80	114	36
10	237	154	106
11	160	115	88
12	75	98	46

Задание:

1) Построить линейное уравнение множественной регрессии и пояснить экономический смысл его параметров.

2) Сравнить раздельное влияние на зависимую переменную каждой из объясняющих переменных, используя стандартизированные коэффициенты регрессии и коэффициенты эластичности.

3) Проверить значимость коэффициентов регрессии при $\alpha = 0{,}05$ и найти 95%-е доверительные интервалы для коэффициентов регрессии.

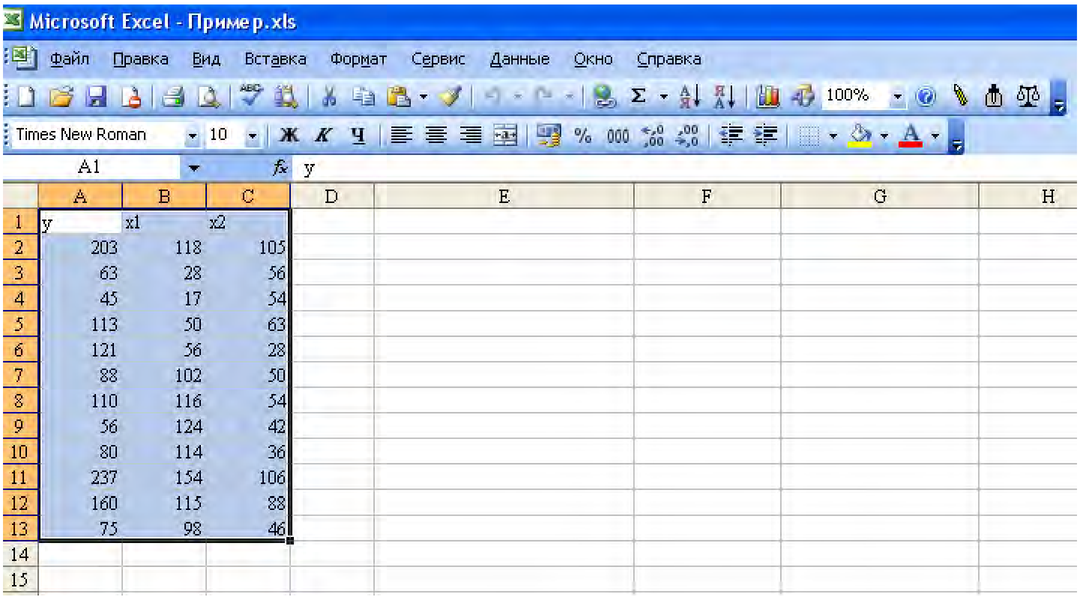
4) Найти коэффициент детерминации и скорректированный коэффициент детерминации и пояснить их смысл. Проверить значимость полученного уравнения регрессии на уровне $\alpha = 0{,}05$.

5) Построить на графике наблюдаемые и предсказанные значения уравнения множественной регрессии.

Решение. Все вычисления будем проводить в среде MS Excel.

Линейное уравнение множественной регрессии имеет вид: $y(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. Оценкой данного уравнения будет уравнение вида: $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$. Операция вычисления его параметров проводится с помощью инструмента анализа данных *Регрессия*. В результате будут вычислены коэффициенты по формулам (2.8), (2.13-14), (2.16), (2.18), (2.20). Порядок выполнения следующий:

1) Ввести исходные данные



	A	B	C
1	y	x1	x2
2	203	118	105
3	63	28	56
4	45	17	54
5	113	50	63
6	121	56	28
7	88	102	50
8	110	116	54
9	56	124	42
10	80	114	36
11	237	154	106
12	160	115	88
13	75	98	46

Рис. 2.1. Исходные данные

2) В главном меню выбрать **Сервис/ Анализ данных/Регрессия.** Щелкнуть по кнопке **ОК**;

3) заполнить диалоговое окно ввода данных и параметров вывода:

Входной интервал Y — диапазон, содержащий данные зависимой переменной;

Входной интервал X — диапазон, содержащий данные независимых переменных;

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 — указать левую верхнюю ячейку будущего диапазона, если результаты будут располагаться на текущем листе;

Новый рабочий лист — можно задать произвольное имя нового листа, если результаты необходимо разместить на отдельном листе;

Остатки — флажок, включающий остатки в выходном диапазоне (необходим для представления предсказанных значений зависимой переменной). Щелкнуть по кнопке **ОК.** (рис. 2.2).

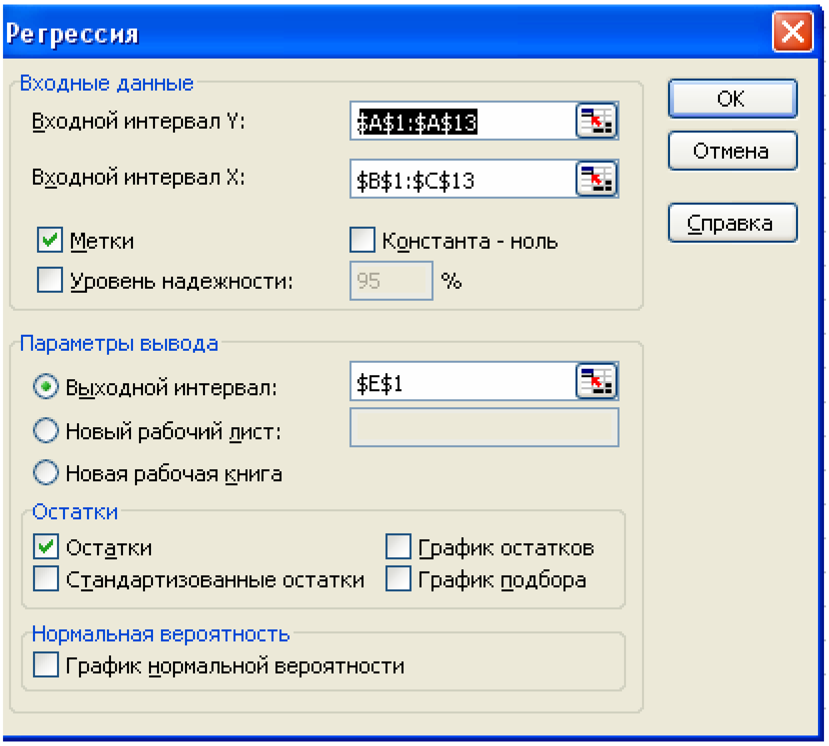
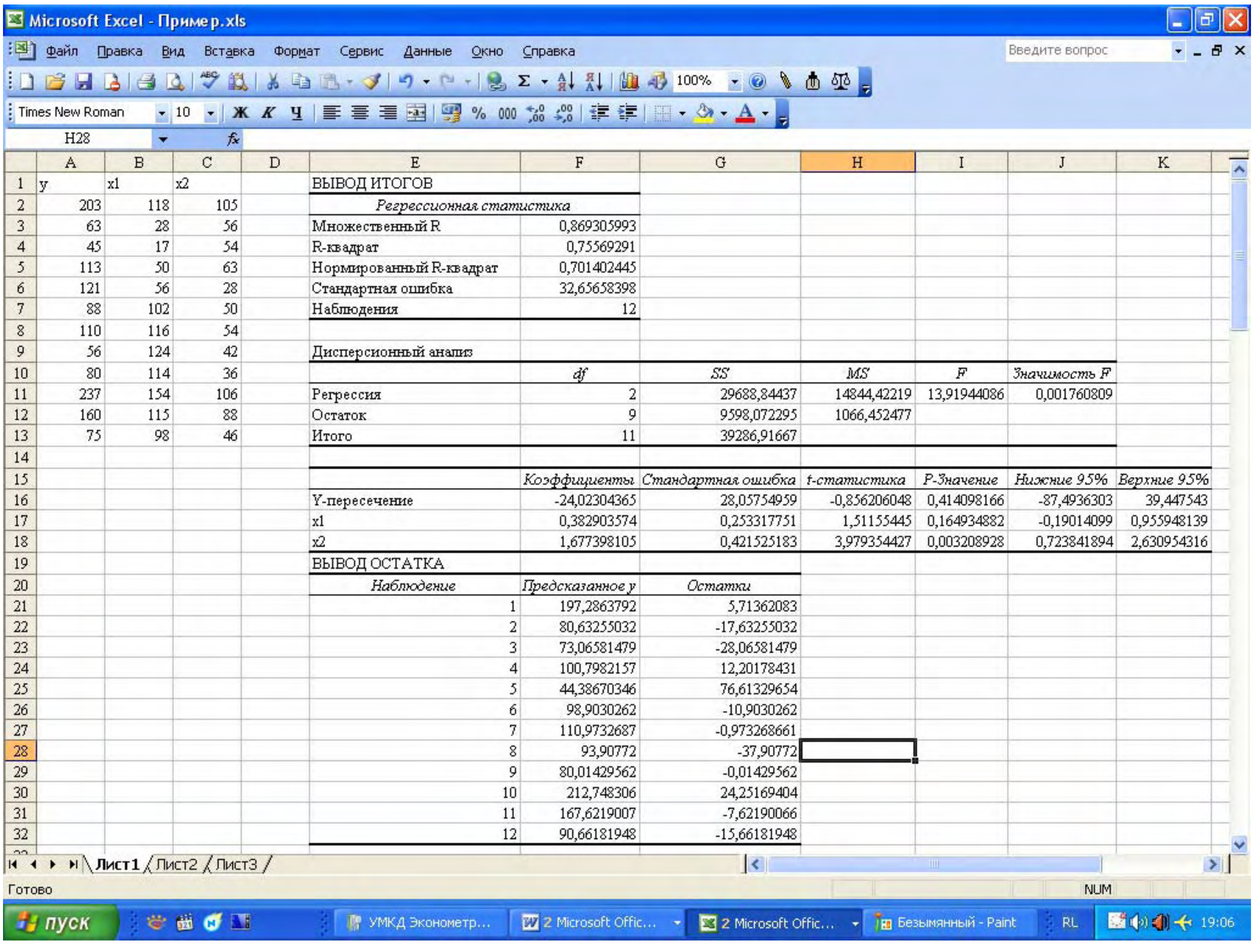


Рис.2.2 Диалоговое окно ввода параметров Регрессия

Результаты регрессионного анализа представлены на рис.2.3.



	A	B	C	D	E	F	G	H	I	J	K
1	y	x1	x2		ВЫВОД ИТОГОВ						
2	203	118	105		*Регрессионная статистика*						
3	63	28	56		Множественный R	0,869305993					
4	45	17	54		R-квадрат	0,75569291					
5	113	50	63		Нормированный R-квадрат	0,701402445					
6	121	56	28		Стандартная ошибка	32,65658398					
7	88	102	50		Наблюдения	12					
8	110	116	54								
9	56	124	42		Дисперсионный анализ						
10	80	114	36			*df*	*SS*	*MS*	*F*	*Значимость F*	
11	237	154	106		Регрессия	2	29688,84437	14844,42219	13,91944086	0,001760809	
12	160	115	88		Остаток	9	9598,072295	1066,452477			
13	75	98	46		Итого	11	39286,91667				
14											
15						*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
16					Y-пересечение	-24,02304365	28,05754959	-0,856206048	0,414098166	-87,4936303	39,447543
17					x1	0,382903574	0,253317751	1,51155445	0,164934882	-0,19014099	0,955948139
18					x2	1,677398105	0,421525183	3,979354427	0,003208928	0,723841894	2,630954316
19					ВЫВОД ОСТАТКА						
20					*Наблюдение*	*Предсказанное y*	*Остатки*				
21					1	197,2863792	5,71362083				
22					2	80,63255032	-17,63255032				
23					3	73,06581479	-28,06581479				
24					4	100,7982157	12,20178431				
25					5	44,38670346	76,61329654				
26					6	98,9030262	-10,9030262				
27					7	110,9732687	-0,973268661				
28					8	93,90772	-37,90772				
29					9	80,01429562	-0,01429562				
30					10	212,748306	24,25169404				
31					11	167,6219007	-7,62190066				
32					12	90,66181948	-15,66181948				

Рис. 2.3. Результат применения инструмента Регрессия

Таким образом, уравнение множественной регрессии имеет вид:

$$\hat{y} = -24{,}023 + 0{,}3829x_1 + 1{,}6774x_2 .$$

Оно показывает, что при увеличении только среднегодовой стоимости основных фондов x_1 (при неизменном x_2) на 1 млн. руб., величина валового дохода торговых предприятий увеличивается в среднем на 0,3829 млн. руб., а при увеличении только среднегодовой стоимости оборотных средств x_2 (при неизменном x_1) на 1 млн. руб. величина валового дохода торговых предприятий увеличивается в среднем на 1,6774 млн. руб. Коэффициент -24,023 не имеет экономического смысла.

Для вычисления стандартизированных коэффициентов регрессии и коэффициентов эластичности необходимо вычислить средние значения и стандартные отклонения по переменным y, x_1 и x_2. Для этого выполнить следующие действия:

а) выделить ячейку A16, на панели инструментов **Стандартная** щелкнуть по кнопке f_x;

б) в окне *Категория* выбрать *Статистические*, в окне *Функция* – *СРЗНАЧ*. Щелкнуть по кнопке **ОК.**

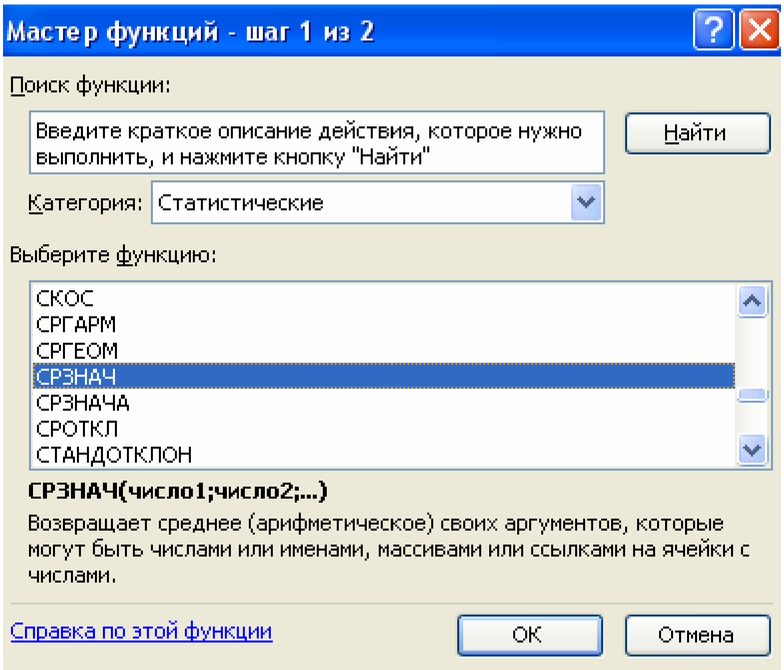


Рис.2.4. Диалоговое окно «Мастер функций»

в) заполнить аргументы функции (рис. 2.5): *Число* 1 – диапазон, содержащий данные зависимой переменной *y*. Щелкнуть по кнопке **OK.**

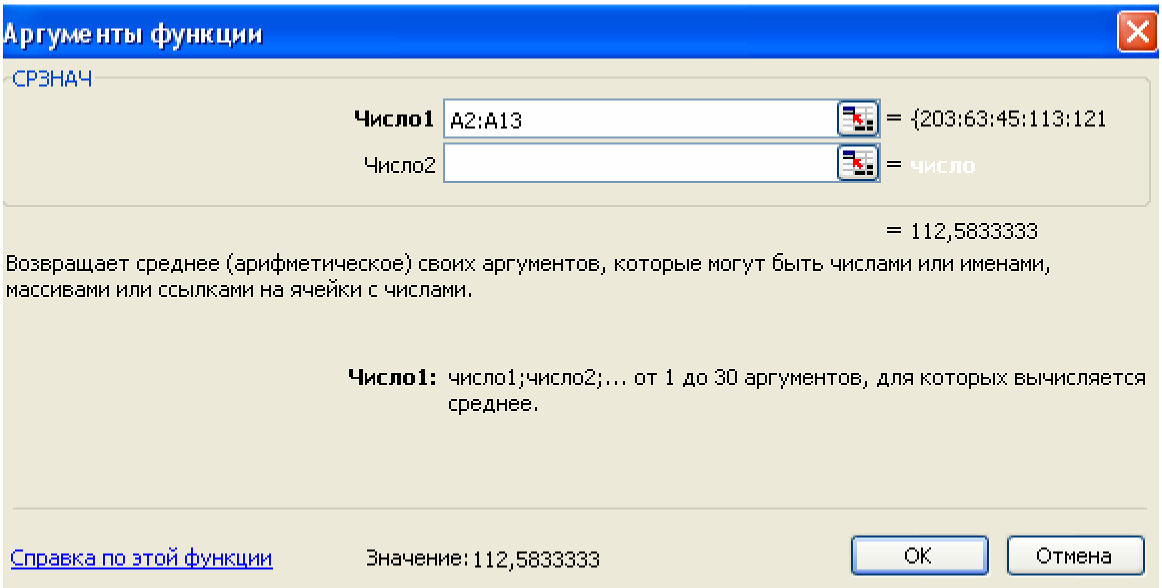


Рис. 2.6. Диалоговое окно ввода аргументов функции *СРЗНАЧ*

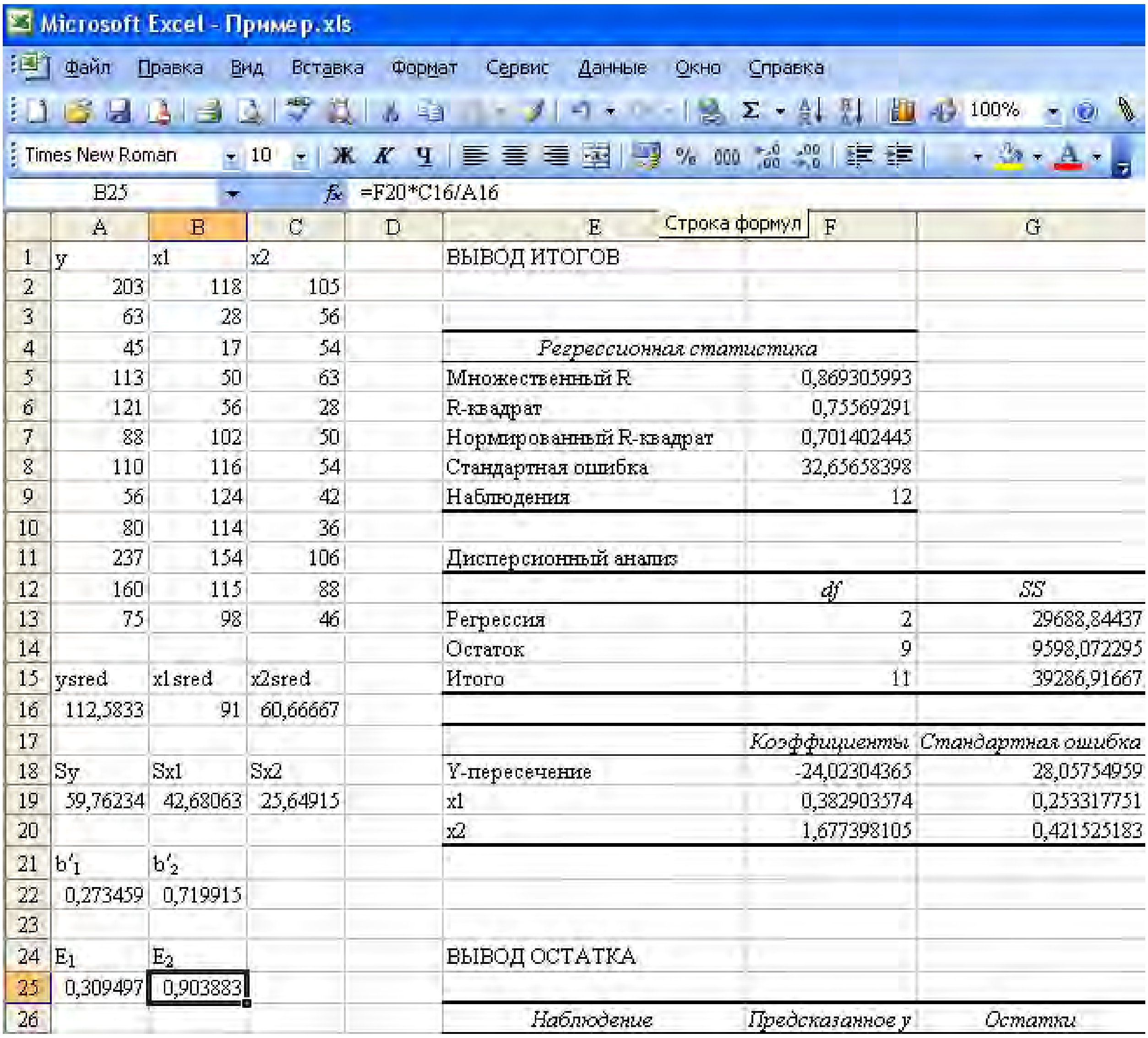
В ячейке А16 появится результат 112,5853. Выполнить те же действия для переменных x_1 и x_2.

Для вычисления среднего квадратического отклонения необходимо выбрать функцию *СТАНДОТКЛОН* и выполнить действия, аналогичные функции *СРЗНАЧ*. В итоге получим:

	A	B	C
1	y	x1	x2
2	203	118	105
3	63	28	56
4	45	17	54
5	113	50	63
6	121	56	28
7	88	102	50
8	110	116	54
9	56	124	42
10	80	114	36
11	237	154	106
12	160	115	88
13	75	98	46
14			
15	ysred	x1 sred	x2sred
16	112,5833	91	60,66667
17			
18	Sy	Sx1	Sx2
19	59,76234	42,68063	25,64915

Рис.2.7. Результат вычисления функции *СРЗНАЧ* и *СТАНДОТКЛОН*

Используя вычисленные средние значения и стандартные отклонения переменных y, x_1 и x_2 и формулы (2.9) и (2.12) вычислим стандартизированные коэффициенты регрессии и коэффициенты эластичности. Для этого введем следующие формулы — в ячейку A22: =F19*B19/A19; в ячейку B22: =F20*C20/A19; в ячейку A25: =F19*B16/A16; в ячейку B25: =F20*C16/A16:

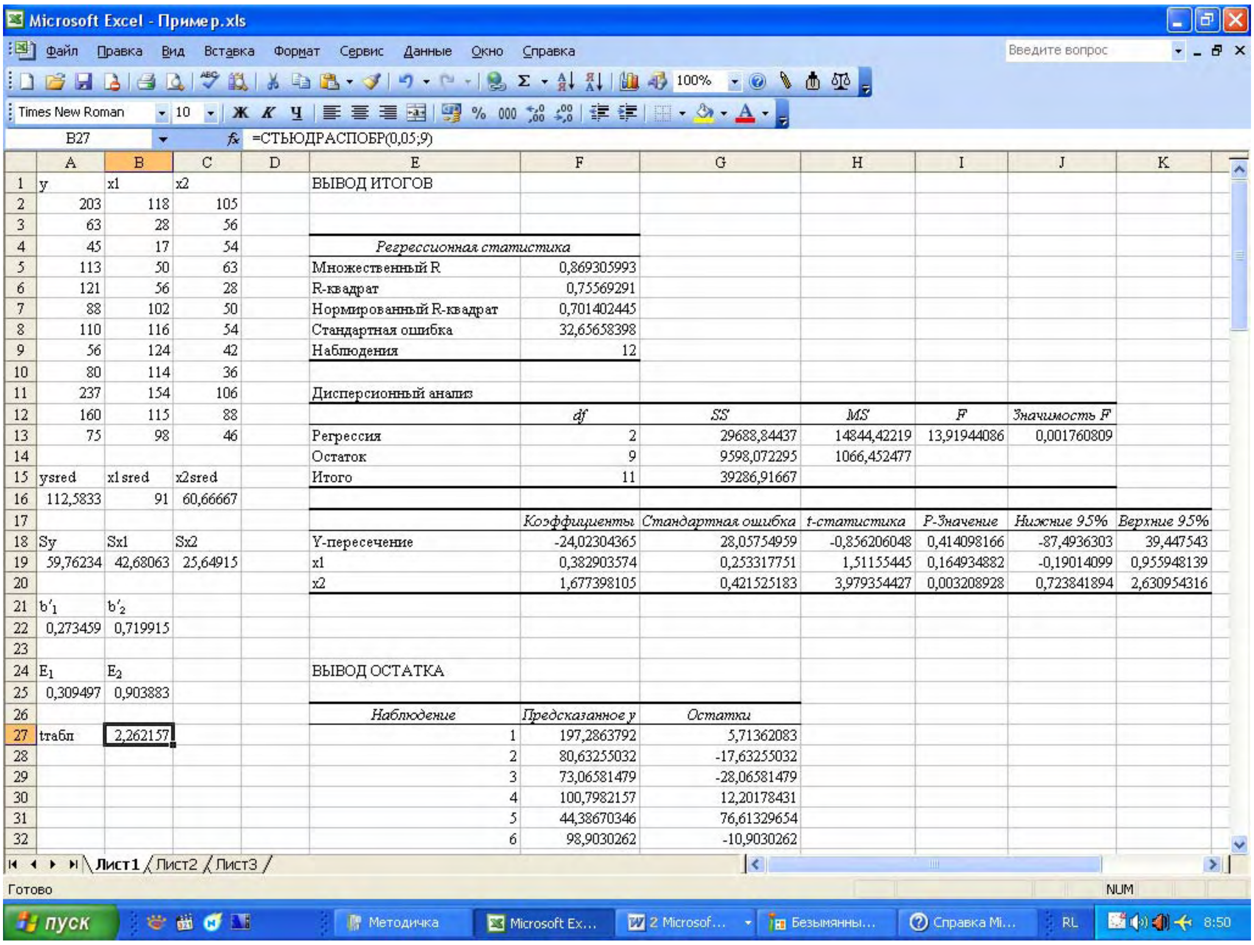


	A	B	C	D	E	F	G
1	y	x1	x2		ВЫВОД ИТОГОВ		
2	203	118	105				
3	63	28	56				
4	45	17	54		*Регрессионная статистика*		
5	113	50	63		Множественный R	0,869305993	
6	121	56	28		R-квадрат	0,75569291	
7	88	102	50		Нормированный R-квадрат	0,701402445	
8	110	116	54		Стандартная ошибка	32,65658398	
9	56	124	42		Наблюдения	12	
10	80	114	36				
11	237	154	106		Дисперсионный анализ		
12	160	115	88			*df*	*SS*
13	75	98	46		Регрессия	2	29688,84437
14					Остаток	9	9598,072295
15	ysred	x1sred	x2sred		Итого	11	39286,91667
16	112,5833	91	60,66667				
17						*Коэффициенты*	*Стандартная ошибка*
18	Sy	Sx1	Sx2		Y-пересечение	-24,02304365	28,05754959
19	59,76234	42,68063	25,64915		x1	0,382903574	0,253317751
20					x2	1,677398105	0,421525183
21	b'_1	b'_2					
22	0,273459	0,719915					
23							
24	E_1	E_2			ВЫВОД ОСТАТКА		
25	0,309497	0,903883					
26					*Наблюдение*	*Предсказанное y*	*Остатки*

Рис. 2.8. Результат вычисления коэффициентов эластичности и стандартизированных коэффициентов регрессии

Следовательно, увеличение среднегодовой стоимости основных фондов и оборотных средств только на одно s_{x1} или на одно s_{x2} увеличивает в среднем величину валового дохода торговых предприятий на $0{,}2735s_y$ или на $0{,}7199s_y$, а увеличение этих переменных на 1% приводит в среднем к росту валового дохода соответственно на 0,3095% и на 0,9039%. Таким образом, по обоим показателям, на валовый доход большее влияние оказывает фактор «среднегодовая стоимость оборотных средств», по сравнению с фактором «среднегодовая стоимость основных фондов».

Для проверки значимости коэффициентов регрессии необходимо найти табличное значение критерия Стьюдента, применяя *Статистическую* функцию – *СТЬЮДРАСПОБР*(α, *n*-*p*-1), где α = 0,05; *n* = 12; *p* = 2. Для этого в ячейку B27 ввести формулу: = СТЬЮДРАСПОБР(0,05;9).



	A	B	C	D	E	F	G	H	I	J	K
1	y	x1	x2		ВЫВОД ИТОГОВ						
2	203	118	105								
3	63	28	56								
4	45	17	54		*Регрессионная статистика*						
5	113	50	63		Множественный R	0,869305993					
6	121	56	28		R-квадрат	0,75569291					
7	88	102	50		Нормированный R-квадрат	0,701402445					
8	110	116	54		Стандартная ошибка	32,65658398					
9	56	124	42		Наблюдения	12					
10	80	114	36								
11	237	154	106		Дисперсионный анализ						
12	160	115	88			*df*	*SS*	*MS*	*F*	*Значимость F*	
13	75	98	46		Регрессия	2	29688,84437	14844,42219	13,91944086	0,001760809	
14					Остаток	9	9598,072295	1066,452477			
15	ysred	x1sred	x2sred		Итого	11	39286,91667				
16	112,5833	91	60,66667								
17						*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
18	Sy	Sx1	Sx2		Y-пересечение	-24,02304365	28,05754959	-0,856206048	0,414098166	-87,4936303	39,447543
19	59,76234	42,68063	25,64915		x1	0,382903574	0,253317751	1,51155445	0,164934882	-0,19014099	0,955948139
20					x2	1,677398105	0,421525183	3,979354427	0,003208928	0,723841894	2,630954316
21	b'1	b'2									
22	0,273459	0,719915									
23											
24	E1	E2			ВЫВОД ОСТАТКА						
25	0,309497	0,903883									
26					*Наблюдение*	*Предсказанное y*	*Остатки*				
27	tтабл	2,262157			1	197,2863792	5,71362083				
28					2	80,63255032	-17,63255032				
29					3	73,06581479	-28,06581479				
30					4	100,7982157	12,20178431				
31					5	44,38670346	76,61329654				
32					6	98,9030262	-10,9030262				

Рис. 2.9 Результат вычисления функции *СТЬЮДРАСПОБР*

Так как $t_{b_1} = 1{,}5115445 < t_{\text{табл}} = 2{,}262157$, то коэффициент b_1 незначим на 5%-ом уровне.

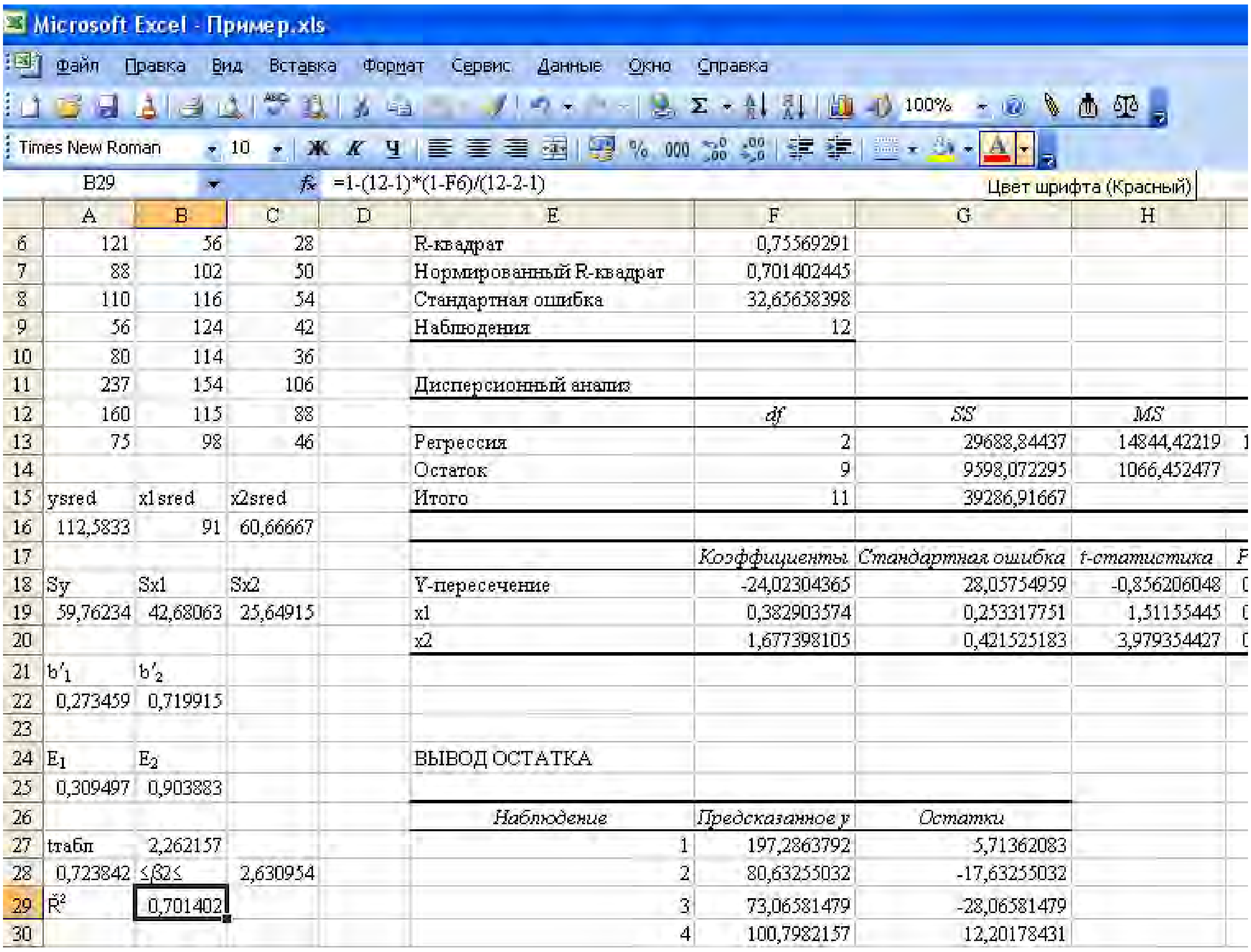
Так как $t_{b_2} = 3{,}979354427 > t_{\text{табл}} = 2{,}262157$, то коэффициент b_2 значим на 5%-ном уровне.

Вычислим доверительный интервал только для значимого коэффициента b_2 по (2.16), введя в ячейку A28 формулу: = F20 – B27*G20; а в ячейку

C28: = F20 + B27*G20. В итоге получим: $0{,}7238 \le \beta_2 \le 2{,}631$.

Таким образом, с надежностью 0,95 при изменении среднегодовой стоимости оборотных средств на 1 млн. руб. (при неизменной среднегодовой стоимости основных фондов) величина валового дохода торговых предприятий будет меняться в пределах от 0,7238 до 2,631 млн. руб.

Коэффициент детерминации равен $R^2 \approx 0{,}76$ (см. рис. 2.9). Следовательно, вариация зависимой переменной *y* — величина валового дохода — на 76% объясняется изменчивостью включенных в модель объясняющих переменных – среднегодовой стоимости основных фондов x_1 и среднегодовой стоимости оборотных средств x_2, оставшиеся 24% приходятся на неучтенные факторы и ошибки наблюдений. Скорректированный коэффициент детерминации $\hat{R}^2$ по формуле (2.21) примет вид {в ячейку B29 ввести формулу =1-(12-1)*(1-F6)/(12-2-1)}:

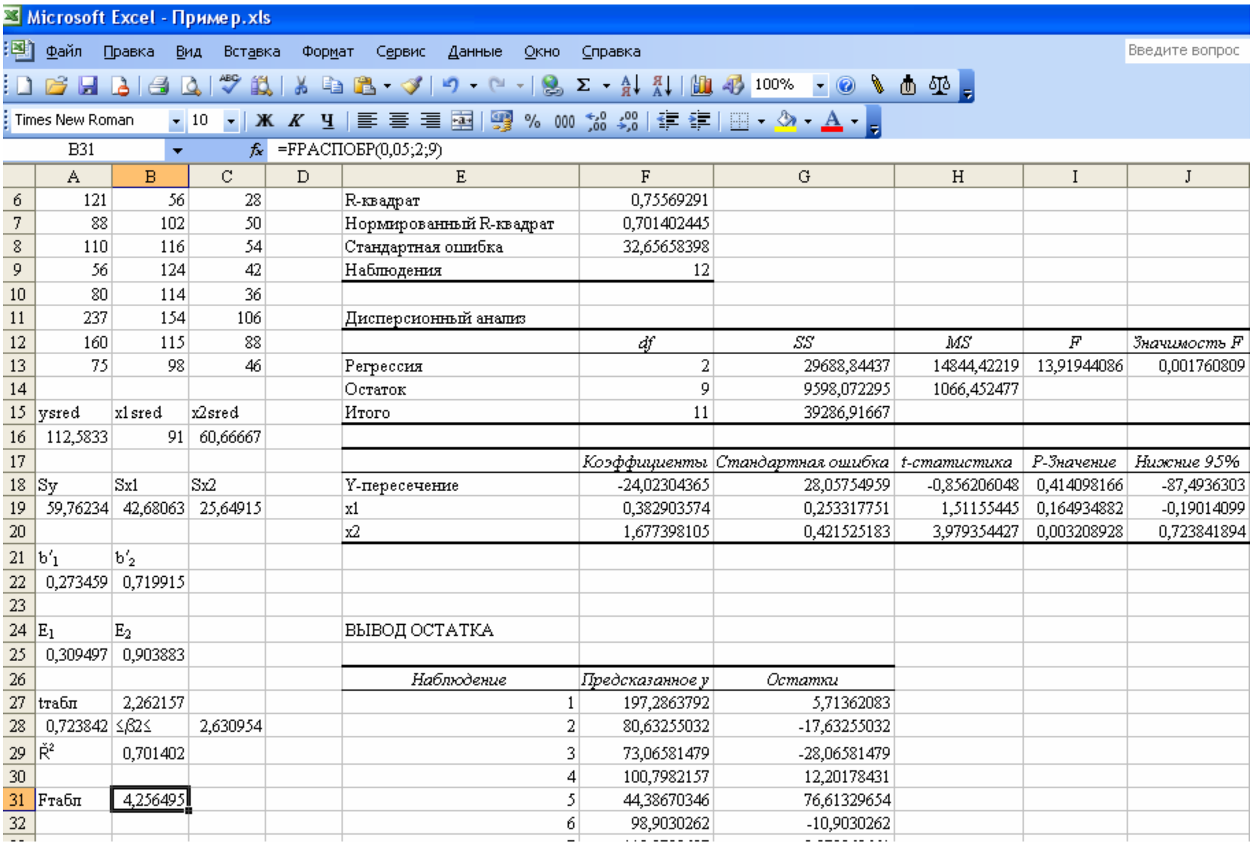


	A	B	C	D	E	F	G	H
6	121	56	28		R-квадрат	0,75569291		
7	88	102	50		Нормированный R-квадрат	0,701402445		
8	110	116	54		Стандартная ошибка	32,65658398		
9	56	124	42		Наблюдения	12		
10	80	114	36					
11	237	154	106		Дисперсионный анализ			
12	160	115	88			*df*	*SS*	*MS*
13	75	98	46		Регрессия	2	29688,84437	14844,42219
14					Остаток	9	9598,072295	1066,452477
15	ysred	x1sred	x2sred		Итого	11	39286,91667	
16	112,5833	91	60,66667					
17						*Коэффициенты*	*Стандартная ошибка*	*t-статистика*
18	Sy	Sx1	Sx2		Y-пересечение	-24,02304365	28,05754959	-0,856206048
19	59,76234	42,68063	25,64915		x1	0,382903574	0,253317751	1,51155445
20					x2	1,677398105	0,421525183	3,979354427
21	b′1	b′2						
22	0,273459	0,719915						
23								
24	E1	E2			ВЫВОД ОСТАТКА			
25	0,309497	0,903883						
26					*Наблюдение*	*Предсказанное y*	*Остатки*	
27	tтабл	2,262157			1	197,2863792	5,71362083	
28	0,723842	≤β2≤	2,630954		2	80,63255032	-17,63255032	
29	$\hat{R}^2$	0,701402			3	73,06581479	-28,06581479	
30					4	100,7982157	12,20178431	

Рис. 2.10. Результат вычисления скорректированного коэффициента детерминации

Скорректированный коэффициент детерминации $\hat{R}^2 = 0{,}701$ уменьшился по сравнению с коэффициентом детерминации R^2 = 0,76. Это говорит о том, что в уравнении присутствует незначимая переменная (в нашем случае x_1).

Для проверки значимости уравнения множественной регрессии необходимо найти табличное значение критерия Фишера, используя *Статистическую* функцию – *FРАСПРОБР*(α, *p*, *n*-*p*-1), где α = 0,05; *n* = 12; *p* = 2. Для этого в ячейку B31 ввести формулу: = FРАСПРОБР(0,05;2;9):



	A	B	C	D	E	F	G	H	I	J
6	121	56	28		R-квадрат	0,75569291				
7	88	102	50		Нормированный R-квадрат	0,701402445				
8	110	116	54		Стандартная ошибка	32,65658398				
9	56	124	42		Наблюдения	12				
10	80	114	36							
11	237	154	106		Дисперсионный анализ					
12	160	115	88			*df*	*SS*	*MS*	*F*	*Значимость F*
13	75	98	46		Регрессия	2	29688,84437	14844,42219	13,91944086	0,001760809
14					Остаток	9	9598,072295	1066,452477		
15	ysred	x1sred	x2sred		Итого	11	39286,91667			
16	112,5833	91	60,66667							
17						*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*
18	Sy	Sx1	Sx2		Y-пересечение	-24,02304365	28,05754959	-0,856206048	0,414098166	-87,4936303
19	59,76234	42,68063	25,64915		x1	0,382903574	0,253317751	1,51155445	0,164934882	-0,19014099
20					x2	1,677398105	0,421525183	3,979354427	0,003208928	0,723841894
21	b′1	b′2								
22	0,273459	0,719915								
23										
24	E1	E2			ВЫВОД ОСТАТКА					
25	0,309497	0,903883								
26					*Наблюдение*	*Предсказанное y*	*Остатки*			
27	tтабл	2,262157			1	197,2863792	5,71362083			
28	0,723842	≤β2≤	2,630954		2	80,63255032	-17,63255032			
29	Ř²	0,701402			3	73,06581479	-28,06581479			
30					4	100,7982157	12,20178431			
31	Fтабл	4,256495			5	44,38670346	76,61329654			
32					6	98,9030262	-10,9030262			

Рис. 2.11. Результат вычисления функции *FРАСПРОБР*

Так как F = 13,9194 > $F_{табл}$ = 4,256495, то уравнение множественной регрессии значимо на уровне значимости α = 0,05, следовательно, исследуемая переменная y хорошо описывается включенными в модель переменными x_1 и x_2.

Для построения наблюдаемых и предсказанных значений y выполнить следующие действия:

1) с помощью мышки выделить диапазон данных A1:A13,

2) на панели инструментов щелкнуть значок  ,

3) на вкладке *стандартные* выбрать тип *график* и нажать 2 раза кнопку *далее*,

4) на вкладке *заголовки* очистить ячейку *Название диаграммы*, на вкладке *линии сетки* убрать флажок *ось Y (основные линии)*, на вкладке *легенда* размещение сделать внизу и нажать кнопку *готово*.

После этого на листе появится диаграмма:

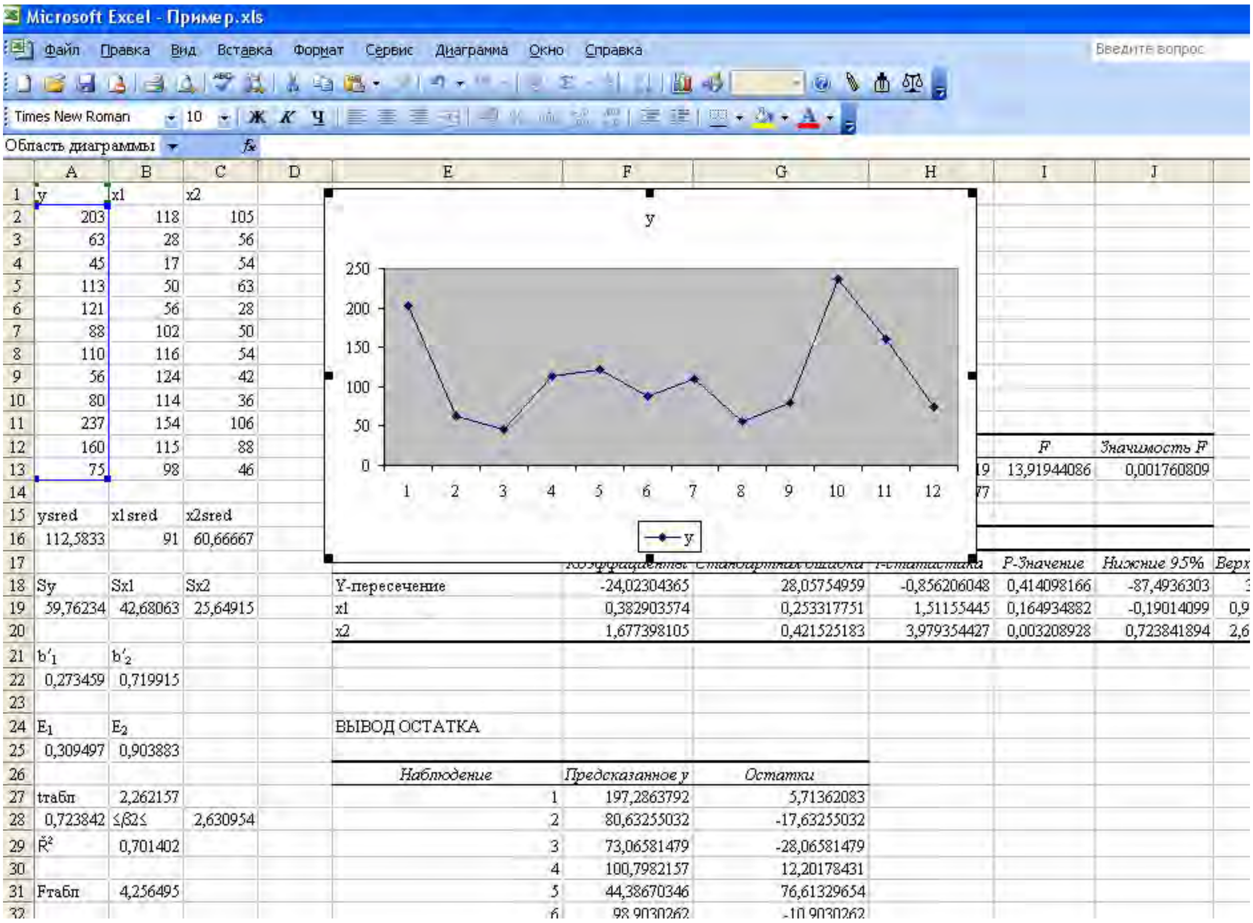


Рис. 2.12. Построение наблюдаемых значений переменной *y*

Затем выбрать в меню **Диаграмма/Добавить данные** и в ячейку *Диапазон* ввести =ЛИСТ1!$F$26:$F$38 и нажать 2 раза кнопку **ОК**. В результате получим:

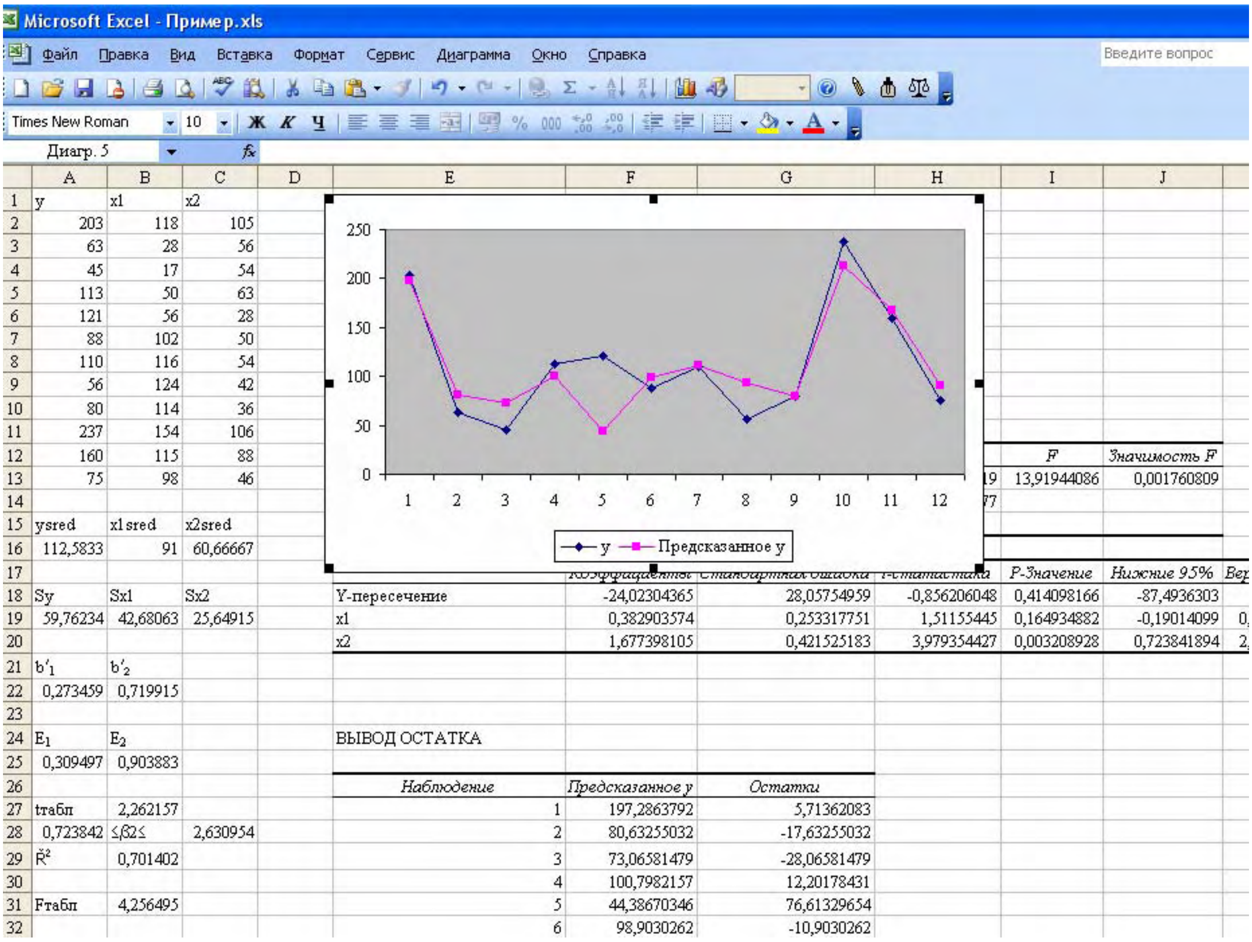


Рис. 2.13. График наблюдаемых и предсказанных значений переменной *y*

2.4. Мультиколлинеарность

Под *мультиколлинеарностью* понимается высокая взаимная коррелированность объясняющих переменных. Мультиколлинеарность может проявляться в функциональной (явной) и стохастической (скрытой) формах.

При *функциональной форме* мультиколлинеарности по крайней мере одна из парных связей между объясняющими переменными является линейной функциональной зависимостью. В этом случае матрица $X^T X$ особенная, так как содержит линейно зависимые векторы-столбцы и ее определитель равен нулю. Это приводит к невозможности решения соответствующей системы нормальных уравнений (2.5) и получения оценок параметров регрессионной модели. Чаще в экономических исследованиях мультиколлинеарность проявляется в *стохастической форме*, когда между хотя бы двумя объясняющими переменными существует тесная корреляционная связь. Матрица $X^T X$ в этом случае является неособенной, но ее определитель очень мал. В то же время вектор оценок b в соответствии с формулой (2.8) пропорционален матрице $\left(X^T X\right)^{-1}$, а значит, его элементы обратно пропорциональны величине определителя $\left|X^T X\right|$. В результате получаются значительные средние квадратические отклонения (стандартные ошибки) коэффициентов регрессии b_0, b_1, $\ldots b_p$ и оценка их значимости по t-критерию не имеет смысла, хотя в целом регрессионная модель может оказаться значимой по F-критерию.

Оценки становятся очень чувствительными к незначительному изменению результатов наблюдений и объема выборки. Уравнения регрессии, как правило, не имеют реального смысла, так как некоторые из его коэффициентов могут иметь неправильные с точки зрения экономической теории знаки и неоправданно большие значения.

Для определения мультиколлинеарности применяют подход, заключающийся в анализе корреляционной матрицы между объясняющими переменными X_1, X_2, $\ldots X_p$ и выявлении пар переменных, имеющих высокие коэффициенты корреляции (обычно не менее 0,7). Если такие переменные существуют, то говорят о мультиколлинеарности между ними.

Одним из методов устранения или уменьшения мультиколлинеарности является *использование пошаговых процедур отбора наиболее информативных переменных*.

На первом шаге рассматривается лишь одна объясняющая переменная, имеющая с зависимой переменной Y наибольший коэффициент детерминации. На втором шаге включается в регрессию новая объясняющая переменная, которая с первоначально отобранной образует пару объясняющих переменных, имеющую с Y наиболее высокий (скорректированный) коэффициент детерминации. На третьем шаге вводится в регрессию еще одна объясняющая переменная, которая вместе с двумя первоначально отобранными образует тройку объясняющих переменных, имеющую с Y наиболее высокий (скорректированный) коэффициент детерминации и т.д.

Процедура введения новых переменных продолжается до тех пор, пока будет увеличиваться соответствующий (скорректированный) коэффициент детерминации $\hat{R}^2$.

Пример 2. По данным n = 20 сельскохозяйственных районов области исследуется зависимость переменной Y — урожайности зерновых культур (в ц/га) от ряда переменных — факторов сельскохозяйственного производства:

X_1 — число тракторов (приведенной мощности на 100 га);

X_2 — число зерноуборочных комбайнов на 100 га;

X_3 — число орудий поверхностной обработки почвы на 100 га;

X_4 — количество удобрений, расходуемых на 1 га (т/га);

X_5 — количество химических средств защиты растений, расходуемых на 1 га (ц/га). Исходные данные приведены в таблице 2.2.

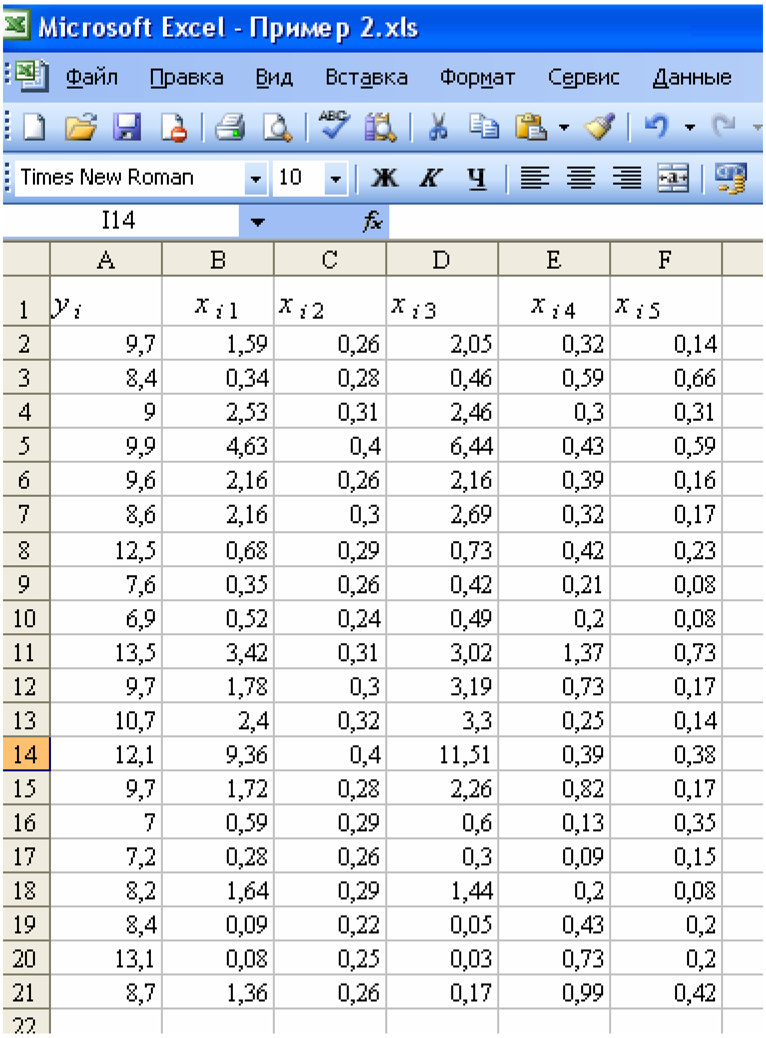
Таблица 2.2

i (номер района)	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
1	9,7	1,59	0,26	2,05	0,32	0,14
2	8,4	0,34	0,28	0,46	0,59	0,66
3	9	2,53	0,31	2,46	0,3	0,31
4	9,9	4,63	0,4	6,44	0,43	0,59
5	9,6	2,16	0,26	2,16	0,39	0,16
6	8,6	2,16	0,3	2,69	0,32	0,17
7	12,5	0,68	0,29	0,73	0,42	0,23
8	7,6	0,35	0,26	0,42	0,21	0,08
9	6,9	0,52	0,24	0,49	0,2	0,08
10	13,5	3,42	0,31	3,02	1,37	0,73
11	9,7	1,78	0,3	3,19	0,73	0,17
12	10,7	2,4	0,32	3,3	0,25	0,14
13	12,1	9,36	0,4	11,51	0,39	0,38
14	9,7	1,72	0,28	2,26	0,82	0,17
15	7	0,59	0,29	0,6	0,13	0,35
16	7,2	0,28	0,26	0,3	0,09	0,15
17	8,2	1,64	0,29	1,44	0,2	0,08

i (номер района)	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
18	8,4	0,09	0,22	0,05	0,43	0,2
19	13,1	0,08	0,25	0,03	0,73	0,2
20	8,7	1,36	0,26	0,17	0,99	0,42

Используя пошаговую процедуру отбора наиболее информативных объясняющих переменных, определить подходящую регрессионную модель, исключив при этом мультиколлинеарность. Оценить значимость коэффициентов регрессии полученной модели по t-критерию на уровне значимости α = 0,05.

Решение. Введем исходные данные:


Microsoft Excel - Пример 2.xls

Файл Правка Вид Вставка Формат Сервис Данные

Times New Roman 10

I14 fx

	A	B	C	D	E	F
1	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
2	9,7	1,59	0,26	2,05	0,32	0,14
3	8,4	0,34	0,28	0,46	0,59	0,66
4	9	2,53	0,31	2,46	0,3	0,31
5	9,9	4,63	0,4	6,44	0,43	0,59
6	9,6	2,16	0,26	2,16	0,39	0,16
7	8,6	2,16	0,3	2,69	0,32	0,17
8	12,5	0,68	0,29	0,73	0,42	0,23
9	7,6	0,35	0,26	0,42	0,21	0,08
10	6,9	0,52	0,24	0,49	0,2	0,08
11	13,5	3,42	0,31	3,02	1,37	0,73
12	9,7	1,78	0,3	3,19	0,73	0,17
13	10,7	2,4	0,32	3,3	0,25	0,14
14	12,1	9,36	0,4	11,51	0,39	0,38
15	9,7	1,72	0,28	2,26	0,82	0,17
16	7	0,59	0,29	0,6	0,13	0,35
17	7,2	0,28	0,26	0,3	0,09	0,15
18	8,2	1,64	0,29	1,44	0,2	0,08
19	8,4	0,09	0,22	0,05	0,43	0,2
20	13,1	0,08	0,25	0,03	0,73	0,2
21	8,7	1,36	0,26	0,17	0,99	0,42
22						

Рис. 2.14. Ввод исходных данных

Для вычисления матрицы парных коэффициентов корреляции воспользуемся инструментом анализа данных **Корреляция**. Для этого

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

2) заполнить диалоговое окно ввода данных и параметров вывода (рис.2.15)

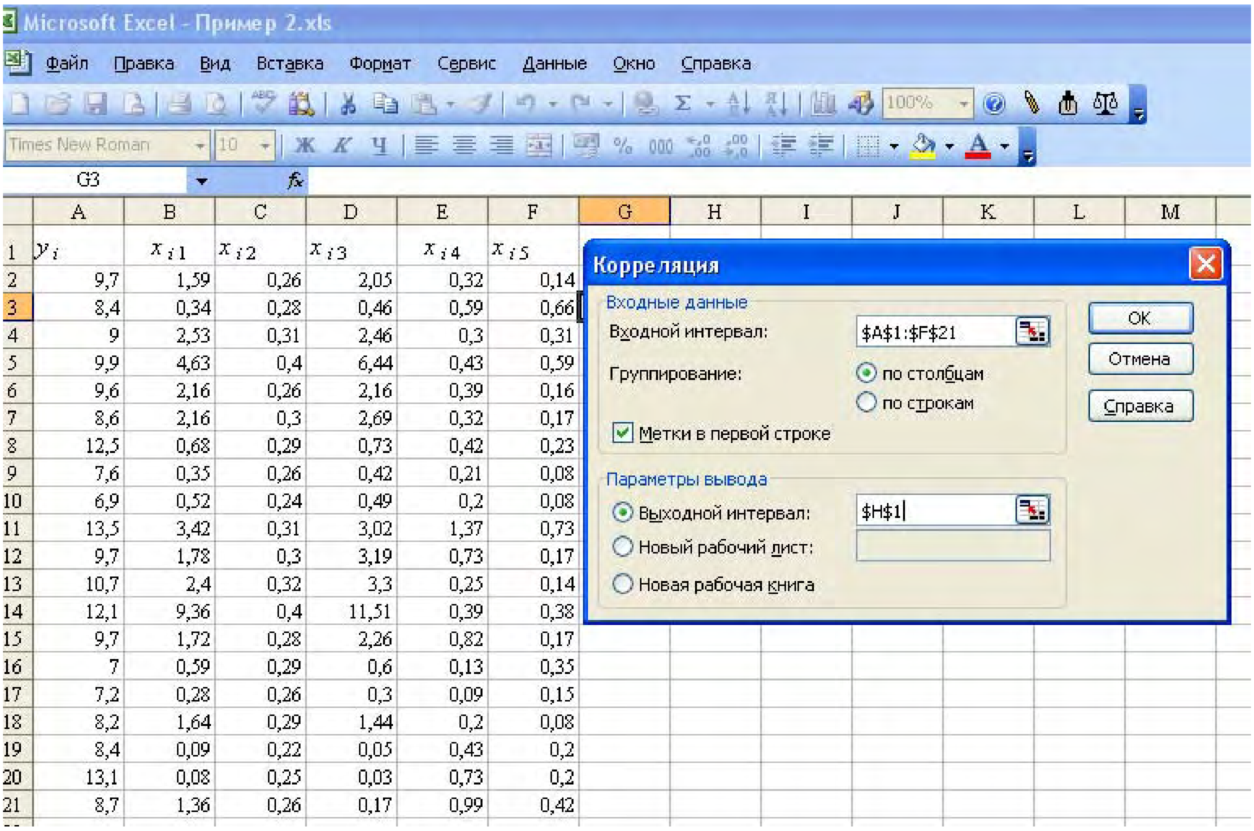
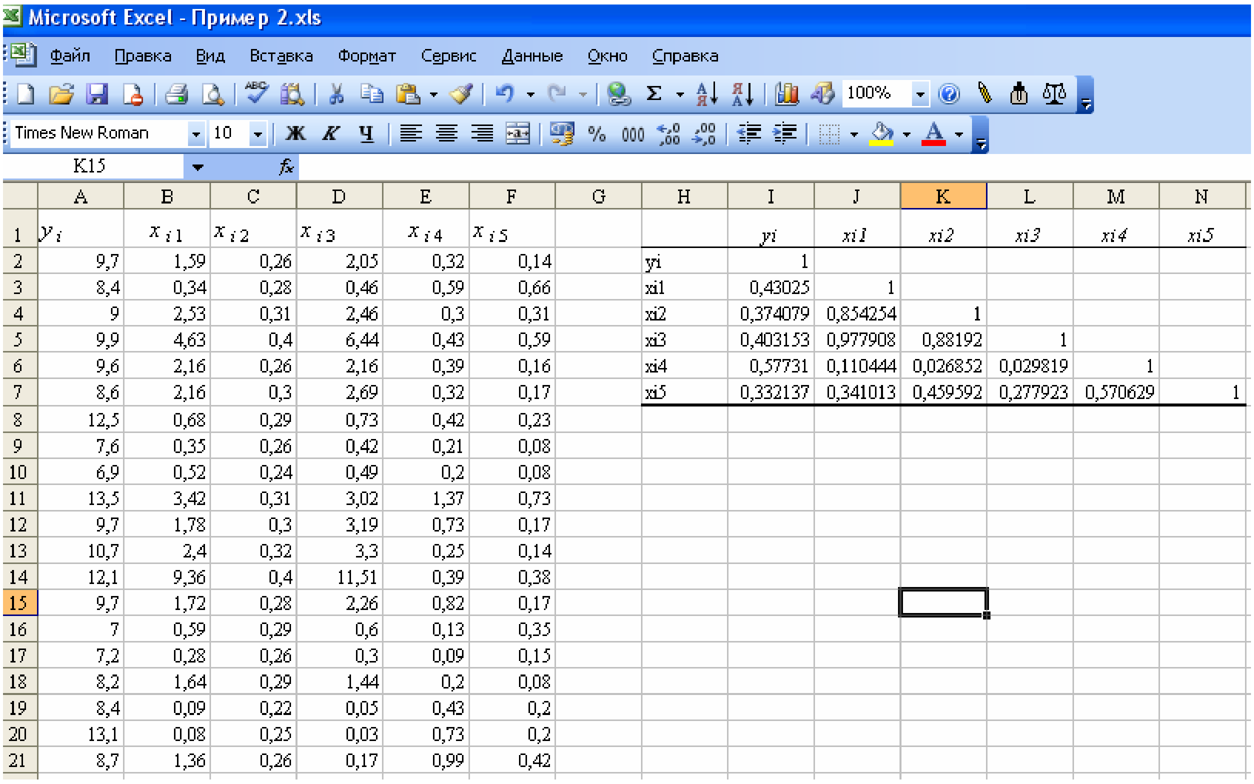


Рис. 2.15. Диалоговое окно ввода параметров инструмента Корреляция

3) результаты вычислений – матрица коэффициентов парной корреляции – представлены на рис. 2.16.



	A	B	C	D	E	F
1	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
2	9,7	1,59	0,26	2,05	0,32	0,14
3	8,4	0,34	0,28	0,46	0,59	0,66
4	9	2,53	0,31	2,46	0,3	0,31
5	9,9	4,63	0,4	6,44	0,43	0,59
6	9,6	2,16	0,26	2,16	0,39	0,16
7	8,6	2,16	0,3	2,69	0,32	0,17
8	12,5	0,68	0,29	0,73	0,42	0,23
9	7,6	0,35	0,26	0,42	0,21	0,08
10	6,9	0,52	0,24	0,49	0,2	0,08
11	13,5	3,42	0,31	3,02	1,37	0,73
12	9,7	1,78	0,3	3,19	0,73	0,17
13	10,7	2,4	0,32	3,3	0,25	0,14
14	12,1	9,36	0,4	11,51	0,39	0,38
15	9,7	1,72	0,28	2,26	0,82	0,17
16	7	0,59	0,29	0,6	0,13	0,35
17	7,2	0,28	0,26	0,3	0,09	0,15
18	8,2	1,64	0,29	1,44	0,2	0,08
19	8,4	0,09	0,22	0,05	0,43	0,2
20	13,1	0,08	0,25	0,03	0,73	0,2
21	8,7	1,36	0,26	0,17	0,99	0,42

	yi	xi1	xi2	xi3	xi4	xi5
yi	1					
xi1	0,43025	1				
xi2	0,374079	0,854254	1			
xi3	0,403153	0,977908	0,88192	1		
xi4	0,57731	0,110444	0,026852	0,029819	1	
xi5	0,332137	0,341013	0,459592	0,277923	0,570629	1

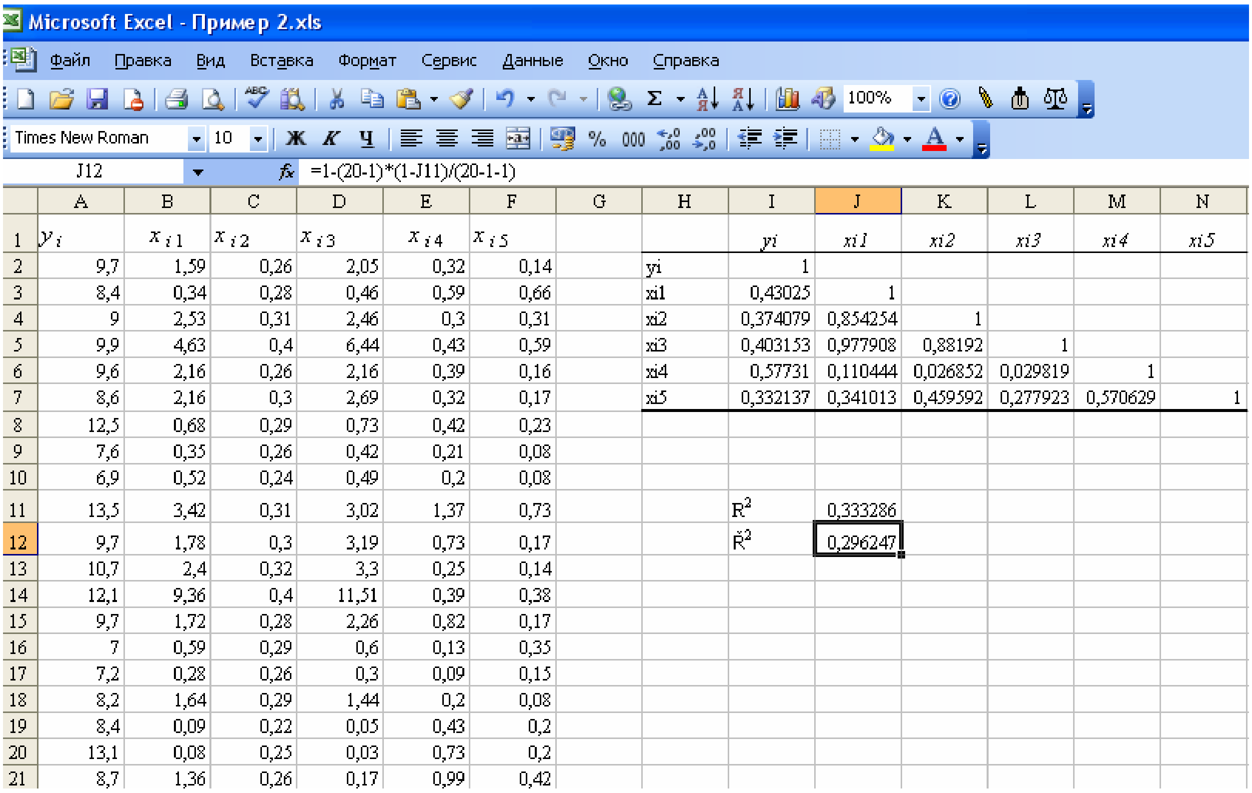
Рис. 2.16. Матрица коэффициентов парной корреляции

Анализируя матрицу парных коэффициентов корреляции, можно отметить тесную корреляционную связь между переменными X_1 и X_2

(r_{12} = 0,854254), X_1 и X_3 (r_{13} = 0,977908), X_2 и X_3 (r_{23} = 0,88192), что свидетельствует о мультиколлинеарности объясняющих переменных.

Для устранения мультиколлинеарности применим процедуру *пошагового отбора* наиболее информативных переменных.

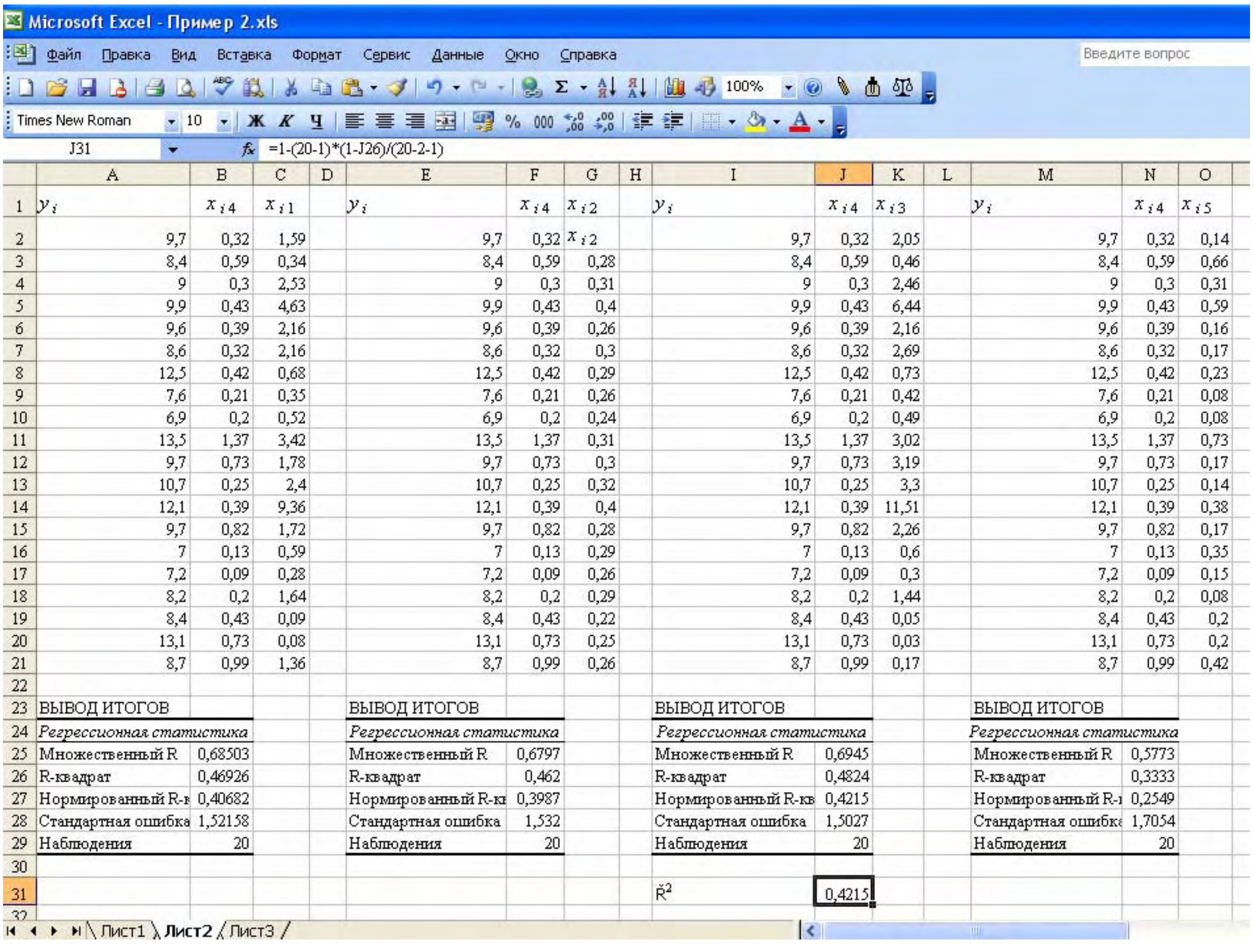
1-й шаг. Из объясняющих переменных X_1—X_5 выберем переменную X_4, имеющую с зависимой Y наибольший коэффициент детерминации $R^2_{y \cdot 4} = r^2_{y4} = 0{,}333$. С учетом поправки на несмещенность по формуле (2.21) при n = 20, p = 1 скорректированный коэффициент детерминации равен $\hat{R}^2_{y \cdot 4} = 0{,}296$ (см. рис. 2.17):



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}			yi	xi1	xi2	xi3	xi4	xi5
2	9,7	1,59	0,26	2,05	0,32	0,14		yi	1					
3	8,4	0,34	0,28	0,46	0,59	0,66		xi1	0,43025	1				
4	9	2,53	0,31	2,46	0,3	0,31		xi2	0,374079	0,854254	1			
5	9,9	4,63	0,4	6,44	0,43	0,59		xi3	0,403153	0,977908	0,88192	1		
6	9,6	2,16	0,26	2,16	0,39	0,16		xi4	0,57731	0,110444	0,026852	0,029819	1	
7	8,6	2,16	0,3	2,69	0,32	0,17		xi5	0,332137	0,341013	0,459592	0,277923	0,570629	1
8	12,5	0,68	0,29	0,73	0,42	0,23								
9	7,6	0,35	0,26	0,42	0,21	0,08								
10	6,9	0,52	0,24	0,49	0,2	0,08								
11	13,5	3,42	0,31	3,02	1,37	0,73			R^2	0,333286				
12	9,7	1,78	0,3	3,19	0,73	0,17			$\tilde{R}^2$	0,296247				
13	10,7	2,4	0,32	3,3	0,25	0,14								
14	12,1	9,36	0,4	11,51	0,39	0,38								
15	9,7	1,72	0,28	2,26	0,82	0,17								
16	7	0,59	0,29	0,6	0,13	0,35								
17	7,2	0,28	0,26	0,3	0,09	0,15								
18	8,2	1,64	0,29	1,44	0,2	0,08								
19	8,4	0,09	0,22	0,05	0,43	0,2								
20	13,1	0,08	0,25	0,03	0,73	0,2								
21	8,7	1,36	0,26	0,17	0,99	0,42								

Рис. 2.17. Обыкновенный и скорректированный коэффициенты детерминации

2-й шаг. Среди всевозможных пар объясняющих переменных (X_4, X_j), j = 1, 2, 3, 5, выберем ту, которая имеет с зависимой переменной Y наиболее высокий коэффициент детерминации. Для этого воспользуемся инструментом анализа данных **Регрессия** (см. пример 1) и найдем коэффициенты детерминации для уравнений $Y = f(X_4, X_1)$, $Y = f(X_4, X_2)$, $Y = f(X_4, X_3)$ и $Y = f(X_4, X_5)$:

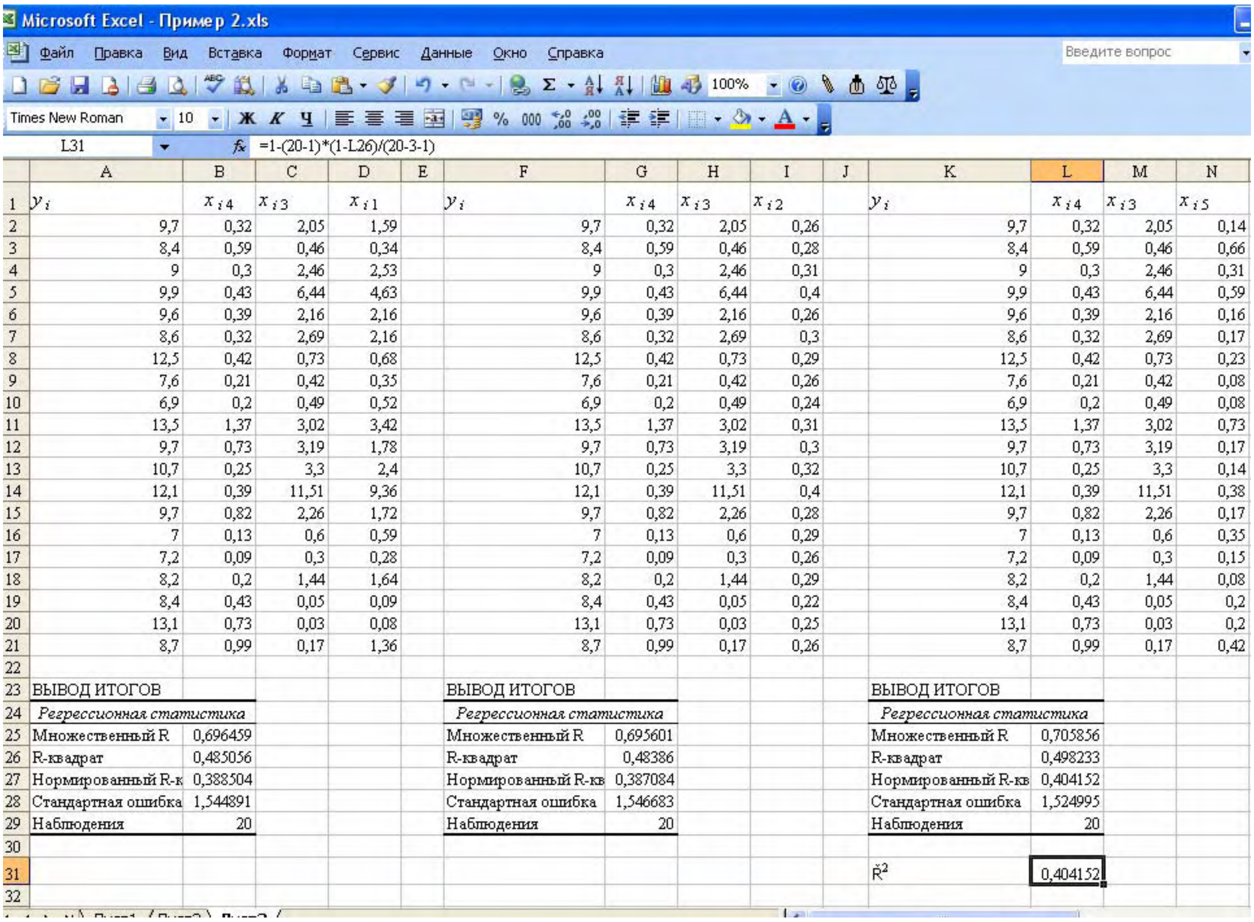


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	y_i	x_{i4}	x_{i1}		y_i	x_{i4}	x_{i2}		y_i	x_{i4}	x_{i3}		y_i	x_{i4}	x_{i5}
2	9,7	0,32	1,59		9,7	0,32	x_{i2}		9,7	0,32	2,05		9,7	0,32	0,14
3	8,4	0,59	0,34		8,4	0,59	0,28		8,4	0,59	0,46		8,4	0,59	0,66
4	9	0,3	2,53		9	0,3	0,31		9	0,3	2,46		9	0,3	0,31
5	9,9	0,43	4,63		9,9	0,43	0,4		9,9	0,43	6,44		9,9	0,43	0,59
6	9,6	0,39	2,16		9,6	0,39	0,26		9,6	0,39	2,16		9,6	0,39	0,16
7	8,6	0,32	2,16		8,6	0,32	0,3		8,6	0,32	2,69		8,6	0,32	0,17
8	12,5	0,42	0,68		12,5	0,42	0,29		12,5	0,42	0,73		12,5	0,42	0,23
9	7,6	0,21	0,35		7,6	0,21	0,26		7,6	0,21	0,42		7,6	0,21	0,08
10	6,9	0,2	0,52		6,9	0,2	0,24		6,9	0,2	0,49		6,9	0,2	0,08
11	13,5	1,37	3,42		13,5	1,37	0,31		13,5	1,37	3,02		13,5	1,37	0,73
12	9,7	0,73	1,78		9,7	0,73	0,3		9,7	0,73	3,19		9,7	0,73	0,17
13	10,7	0,25	2,4		10,7	0,25	0,32		10,7	0,25	3,3		10,7	0,25	0,14
14	12,1	0,39	9,36		12,1	0,39	0,4		12,1	0,39	11,51		12,1	0,39	0,38
15	9,7	0,82	1,72		9,7	0,82	0,28		9,7	0,82	2,26		9,7	0,82	0,17
16	7	0,13	0,59		7	0,13	0,29		7	0,13	0,6		7	0,13	0,35
17	7,2	0,09	0,28		7,2	0,09	0,26		7,2	0,09	0,3		7,2	0,09	0,15
18	8,2	0,2	1,64		8,2	0,2	0,29		8,2	0,2	1,44		8,2	0,2	0,08
19	8,4	0,43	0,09		8,4	0,43	0,22		8,4	0,43	0,05		8,4	0,43	0,2
20	13,1	0,73	0,08		13,1	0,73	0,25		13,1	0,73	0,03		13,1	0,73	0,2
21	8,7	0,99	1,36		8,7	0,99	0,26		8,7	0,99	0,17		8,7	0,99	0,42
22															
23	ВЫВОД ИТОГОВ				ВЫВОД ИТОГОВ				ВЫВОД ИТОГОВ				ВЫВОД ИТОГОВ		
24	*Регрессионная статистика*				*Регрессионная статистика*				*Регрессионная статистика*				*Регрессионная статистика*		
25	Множественный R	0,68503			Множественный R	0,6797			Множественный R	0,6945			Множественный R	0,5773	
26	R-квадрат	0,46926			R-квадрат	0,462			R-квадрат	0,4824			R-квадрат	0,3333	
27	Нормированный R-к	0,40682			Нормированный R-кв	0,3987			Нормированный R-кв	0,4215			Нормированный R-к	0,2549	
28	Стандартная ошибка	1,52158			Стандартная ошибка	1,532			Стандартная ошибка	1,5027			Стандартная ошибка	1,7054	
29	Наблюдения	20			Наблюдения	20			Наблюдения	20			Наблюдения	20	
30															
31									$\hat{R}^2$	0,4215					

Рис. 2.18. Результат применения инструмента Регрессия

Наибольший коэффициент детерминации с переменной Y имеет пара (X_4, X_3): R^2_{y43} = 0,4824 и с учетом поправки по (2.21) (при n = 20, p = 2) $\hat{R}^2_{y43} = 0{,}4215$ (см. рис. 2.18).

3-й шаг. Среди всевозможных троек объясняющих переменных (X_4, X_3, X_j), j =1, 2, 5 наиболее информативной оказалась тройка (X_4, X_3, X_5), имеющая максимальный коэффициент детерминации R^2_{y435} = 0,4982 и соответственно скорректированный коэффициент (при n = 20, p = 3) $\hat{R}^2_{y43} = 0{,}4042$ (см. рис. 2.19):

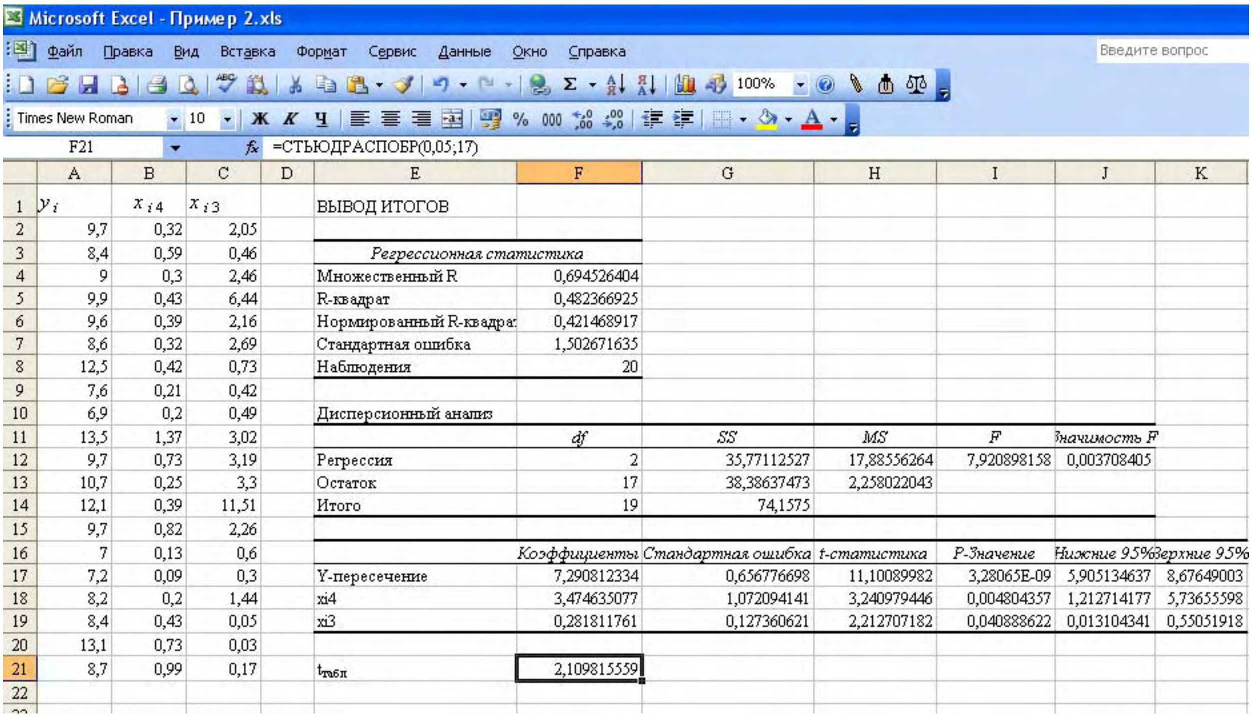


	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	y_i	x_{i4}	x_{i3}	x_{i1}		y_i	x_{i4}	x_{i3}	x_{i2}		y_i	x_{i4}	x_{i3}	x_{i5}
2	9,7	0,32	2,05	1,59		9,7	0,32	2,05	0,26		9,7	0,32	2,05	0,14
3	8,4	0,59	0,46	0,34		8,4	0,59	0,46	0,28		8,4	0,59	0,46	0,66
4	9	0,3	2,46	2,53		9	0,3	2,46	0,31		9	0,3	2,46	0,31
5	9,9	0,43	6,44	4,63		9,9	0,43	6,44	0,4		9,9	0,43	6,44	0,59
6	9,6	0,39	2,16	2,16		9,6	0,39	2,16	0,26		9,6	0,39	2,16	0,16
7	8,6	0,32	2,69	2,16		8,6	0,32	2,69	0,3		8,6	0,32	2,69	0,17
8	12,5	0,42	0,73	0,68		12,5	0,42	0,73	0,29		12,5	0,42	0,73	0,23
9	7,6	0,21	0,42	0,35		7,6	0,21	0,42	0,26		7,6	0,21	0,42	0,08
10	6,9	0,2	0,49	0,52		6,9	0,2	0,49	0,24		6,9	0,2	0,49	0,08
11	13,5	1,37	3,02	3,42		13,5	1,37	3,02	0,31		13,5	1,37	3,02	0,73
12	9,7	0,73	3,19	1,78		9,7	0,73	3,19	0,3		9,7	0,73	3,19	0,17
13	10,7	0,25	3,3	2,4		10,7	0,25	3,3	0,32		10,7	0,25	3,3	0,14
14	12,1	0,39	11,51	9,36		12,1	0,39	11,51	0,4		12,1	0,39	11,51	0,38
15	9,7	0,82	2,26	1,72		9,7	0,82	2,26	0,28		9,7	0,82	2,26	0,17
16	7	0,13	0,6	0,59		7	0,13	0,6	0,29		7	0,13	0,6	0,35
17	7,2	0,09	0,3	0,28		7,2	0,09	0,3	0,26		7,2	0,09	0,3	0,15
18	8,2	0,2	1,44	1,64		8,2	0,2	1,44	0,29		8,2	0,2	1,44	0,08
19	8,4	0,43	0,05	0,09		8,4	0,43	0,05	0,22		8,4	0,43	0,05	0,2
20	13,1	0,73	0,03	0,08		13,1	0,73	0,03	0,25		13,1	0,73	0,03	0,2
21	8,7	0,99	0,17	1,36		8,7	0,99	0,17	0,26		8,7	0,99	0,17	0,42
22														
23	ВЫВОД ИТОГОВ					ВЫВОД ИТОГОВ					ВЫВОД ИТОГОВ			
24	*Регрессионная статистика*					*Регрессионная статистика*					*Регрессионная статистика*			
25	Множественный R	0,696459				Множественный R	0,695601				Множественный R	0,705856		
26	R-квадрат	0,485056				R-квадрат	0,48386				R-квадрат	0,498233		
27	Нормированный R-к	0,388504				Нормированный R-кв	0,387084				Нормированный R-кв	0,404152		
28	Стандартная ошибка	1,544891				Стандартная ошибка	1,546683				Стандартная ошибка	1,524995		
29	Наблюдения	20				Наблюдения	20				Наблюдения	20		
30														
31											$\check{R}^2$	0,404152		
32														

Рис. 2.19. Результат применения инструмента Регрессия

Так как скорректированный коэффициент детерминации на 3-м шаге уменьшился (0,4042) по сравнению со 2-м шагом (0,4215), то в регрессионной модели достаточно ограничиться двумя объясняющими переменными X_4 и X_3. В итоге уравнение множественной регрессии примет вид: $\hat{y} = 7{,}291 + 0{,}282X_3 + 3{,}475X_4$ (см. рис. 2.20).

Таким образом, при увеличении только числа орудий поверхностной обработки почвы на 100 га — X_3 на 1 единицу (при неизменном X_4), урожайность зерновых культур — Y увеличится на 0,282 ц/га, а при увеличении только количества удобрений, расходуемых на 1 га — X_4 на 1 единицу (при неизменном X_3), урожайность зерновых культур увеличится на 3,475 ц/га.



Microsoft Excel - Пример 2.xls

F21 =СТЬЮДРАСПОБР(0,05;17)

	A	B	C	D	E	F	G	H	I	J	K
1	y_i	x_{i4}	x_{i3}		ВЫВОД ИТОГОВ						
2	9,7	0,32	2,05								
3	8,4	0,59	0,46		*Регрессионная статистика*						
4	9	0,3	2,46		Множественный R	0,694526404					
5	9,9	0,43	6,44		R-квадрат	0,482366925					
6	9,6	0,39	2,16		Нормированный R-квадрат	0,421468917					
7	8,6	0,32	2,69		Стандартная ошибка	1,502671635					
8	12,5	0,42	0,73		Наблюдения	20					
9	7,6	0,21	0,42								
10	6,9	0,2	0,49		Дисперсионный анализ						
11	13,5	1,37	3,02			*df*	*SS*	*MS*	*F*	*Значимость F*	
12	9,7	0,73	3,19		Регрессия	2	35,77112527	17,88556264	7,920898158	0,003708405	
13	10,7	0,25	3,3		Остаток	17	38,38637473	2,258022043			
14	12,1	0,39	11,51		Итого	19	74,1575				
15	9,7	0,82	2,26								
16	7	0,13	0,6			*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
17	7,2	0,09	0,3		Y-пересечение	7,290812334	0,656776698	11,10089982	3,28065E-09	5,905134637	8,67649003
18	8,2	0,2	1,44		xi4	3,474635077	1,072094141	3,240979446	0,004804357	1,212714177	5,73655598
19	8,4	0,43	0,05		xi3	0,281811761	0,127360621	2,212707182	0,040888622	0,013104341	0,55051918
20	13,1	0,73	0,03								
21	8,7	0,99	0,17		$t_{табл}$	2,109815559					
22											

Рис. 2.20. Результат применения инструмента Регрессия для уравнения $Y = f(X_4, X_3)$

Все коэффициенты регрессии оказались значимыми на уровне α = 0,05 (рис. 2.20), так как каждое из значений t– статистики t_{b0} = 11,1; t_{b3} = 1,072; t_{b4} = 2,213 больше соответствующего табличного значения $t_{табл}$ = 2,11. $t_{табл}$ вычислено с помощью функции *СТЬЮДРАСПОБР*(α, n-p-1), при α = 0,05; n = 20; p = 2.

2.5. Фиктивные переменные

До сих пор мы рассматривали регрессионную модель, в которой в качестве объясняющих переменных выступали количественные переменные (доход, себестоимость продукции, стоимость основных фондов и т.д.). Однако на практике часто возникает необходимость исследования влияния качественных признаков, имеющих два или несколько уровней (градаций). К числу таких признаков можно отнести: пол (мужской, женский), образование (начальное, среднее, высшее), фактор сезонности (зима, весна, лето, осень) и т.п.

Для оценивания влияния значений количественных переменных и уровней качественных признаков уравнения регрессии вводятся так называемые *фиктивные переменные* или *манекены*.

В качестве фиктивных переменных обычно используются *дихотомические* (*бинарные, булевы*) переменные, которые принимают всего два значения «0» или «1» (например, значение такой переменной Z_1 по фактору «пол»: $Z_1 = 0$ для работников – женщин и $Z_1 = 1$ – для мужчин).

Если рассматриваемый качественный признак имеет несколько (k) уровней (градаций), то вводят (k–1) бинарных переменных.

Например, надо изучить зависимость размера заработной платы Y работников не только от количественных факторов $X_1, X_2, \ldots, X_p$, но и от качественного признака Z (уровень образования). Так как $k = 3$ (начальное, среднее и высшее образования), то нужно ввести $k - 1= 3 - 1 = 2$ бинарные переменные Z_1 и Z_2:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \alpha_1 z_{i1} + \alpha_2 z_{i2} + \varepsilon_i,$$

$$\text{где } z_{i1} = \begin{cases} 1, & \text{если } i \text{ - й работник имеет высшее образование;} \\ 0 & \text{во всех остальных случаях;} \end{cases}$$

$$z_{i2} = \begin{cases} 1, & \text{если } i \text{ - й работник имеет среднее образование;} \\ 0 & \text{во всех остальных случаях.} \end{cases}$$

Третьей бинарной переменной Z_3 не требуется: если i-й работник имеет начальное образование, это будет отражено парой значений $z_{i1} = 0$, $z_{i2} = 0$.

Пример 3. Необходимо исследовать зависимость между результатами письменных вступительных и курсовых (на 1-м курсе) экзаменов по математике. Получены следующие данные о числе решенных задач на вступительных экзаменах X (задание — 10 задач) и курсовых экзаменах Y (задание — 7 задач) 12 студентов, а также распределение этих студентов по фактору «пол» (табл. 2.3):

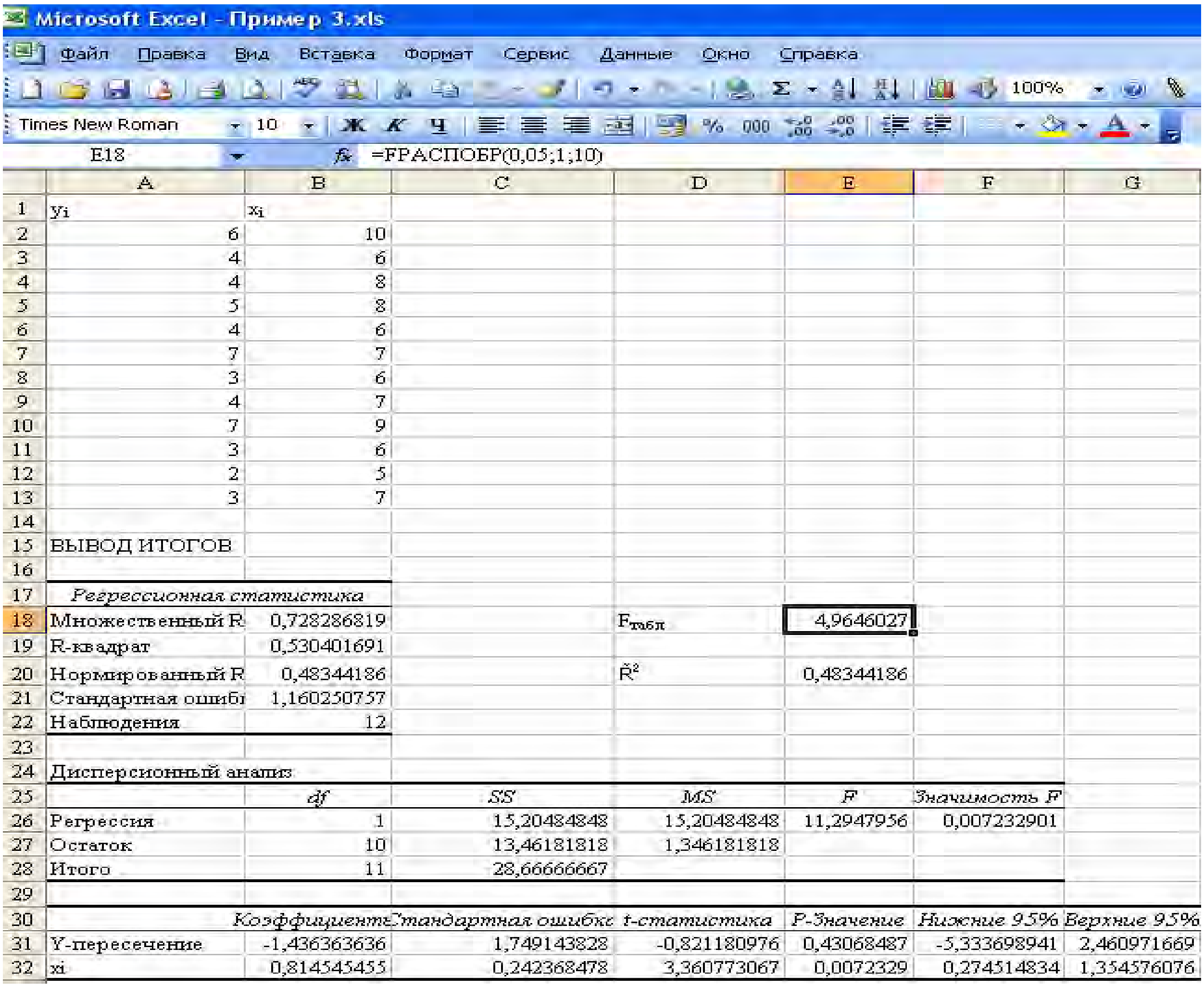
Таблица 2.3

№ студента	Число решенных задач		Пол студента	№ студента	Число решенных задач		Пол студента
i	x_i	y_i		i	x_i	y_i	
1	10	6	муж.	7	6	3	жен.

№ студента	Число решенных задач		Пол студента	№ студента	Число решенных задач		Пол студента
i	x_i	y_i		i	x_i	y_i	
2	6	4	жен.	8	7	4	муж.
3	8	4	муж.	9	9	7	муж.
4	8	5	жен.	10	6	3	жен.
5	6	4	жен.	11	5	2	муж.
6	7	7	муж.	12	7	3	жен.

Построить линейную регрессионную модель Y по X с использованием фиктивной переменной по фактору «пол». Можно ли считать, что эта модель одна и та же для юношей и девушек?

Решение. Построим уравнение парной регрессии с помощью анализа данных **Регрессия**. В итоге получим:


Microsoft Excel - Пример 3.xls

E18 =FРАСПОБР(0,05;1;10)

	A	B	C	D	E	F	G
1	yi	xi					
2	6	10					
3	4	6					
4	4	8					
5	5	8					
6	4	6					
7	7	7					
8	3	6					
9	4	7					
10	7	9					
11	3	6					
12	2	5					
13	3	7					
14							
15	ВЫВОД ИТОГОВ						
16							
17	*Регрессионная статистика*						
18	Множественный R	0,728286819		$F_{табл}$	4,9646027		
19	R-квадрат	0,530401691					
20	Нормированный R	0,48344186		$\tilde{R}^2$	0,48344186		
21	Стандартная ошибі	1,160250757					
22	Наблюдения	12					
23							
24	Дисперсионный анализ						
25		*df*	*SS*	*MS*	*F*	*Значимость F*	
26	Регрессия	1	15,20484848	15,20484848	11,2947956	0,007232901	
27	Остаток	10	13,46181818	1,346181818			
28	Итого	11	28,66666667				
29							
30		*Коэффициенть*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
31	Y-пересечение	-1,436363636	1,749143828	-0,821180976	0,43068487	-5,333698941	2,460971669
32	xi	0,814545455	0,242368478	3,360773067	0,0072329	0,274514834	1,354576076

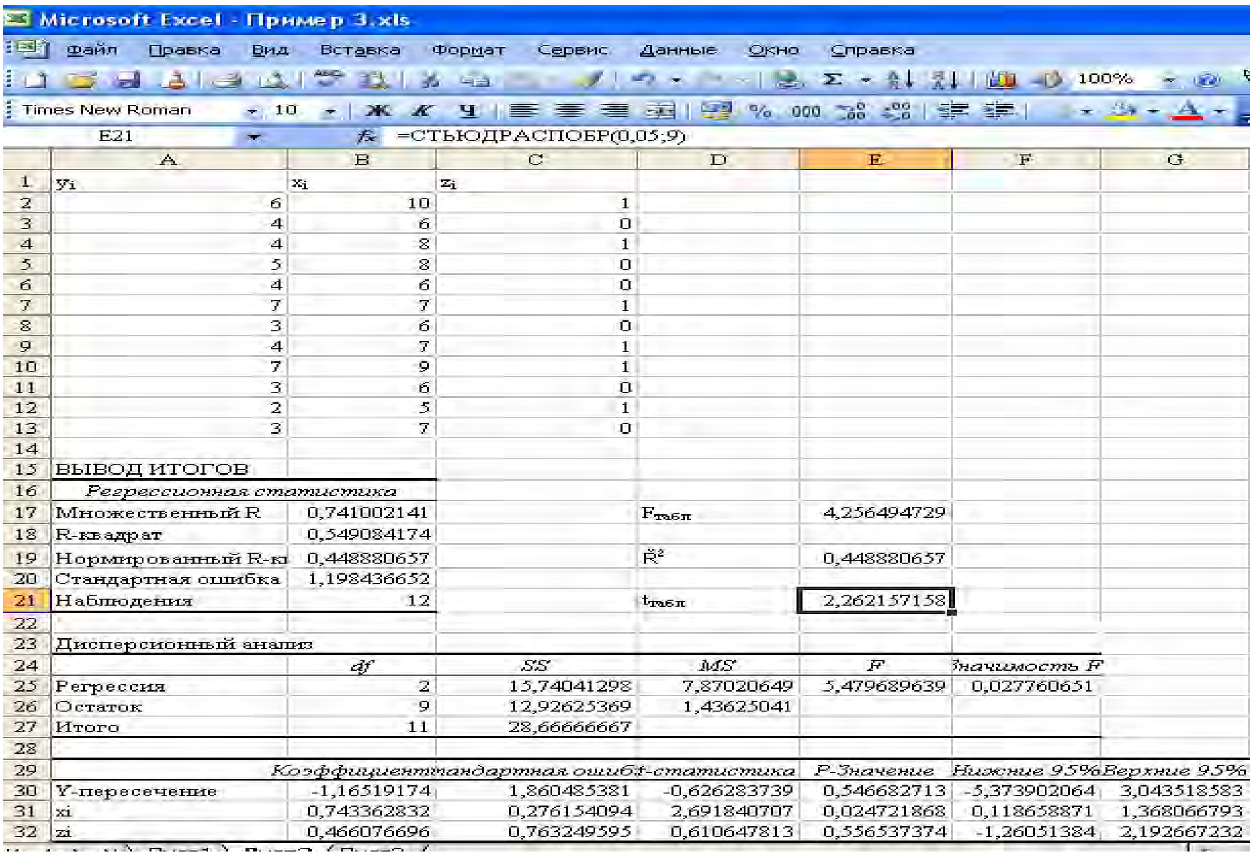
Рис. 2.21. Результат выполнения инструмента Регрессия

Таким образом, уравнение парной регрессии имеет вид: $\hat{y} = -1{,}437 + 0{,}815x$. Коэффициент детерминации R^2 = 0,53, т.е. 53% вариации зависимой переменной Y обусловлено регрессией. Скорректированный коэффициент детерминации $\hat{R}^2$ = 0,483. Уравнение регрессии значимо по F – критерию, так как F = 11,295 > $F_{табл}$ = 4,965 (см. рис. 2.21). Тем не менее, данное уравнение не учитывает влияние качественного признака – фактора «пол». Для ее учета введем в регрессионную модель фиктивную переменную Z:

$$z_i = \begin{cases} 1, \text{если } i\text{ - й студент мужского пола} \\ 0 \quad \text{если } i\text{ - й студент женского пола} \end{cases}.$$

Тогда уравнение множественной регрессии примет вид (рис. 2.22):

$\hat{y} = -1{,}165 + 0{,}743x + 0{,}466z$. Из уравнения следует, что при том же числе решенных задач на вступительных экзаменах X, на курсовых экзаменах юноши решают в среднем на 0,466 ≈ 0,5 задачи больше.



	A	B	C	D	E	F	G
1	yi	xi	zi				
2	6	10	1				
3	4	6	0				
4	4	8	1				
5	5	8	0				
6	4	6	0				
7	7	7	1				
8	3	6	0				
9	4	7	1				
10	7	9	1				
11	3	6	0				
12	2	5	1				
13	3	7	0				
14							
15	ВЫВОД ИТОГОВ						
16	*Регрессионная статистика*						
17	Множественный R	0,741002141		$F_{табл}$	4,256494729		
18	R-квадрат	0,549084174					
19	Нормированный R-кв	0,448880657		$\hat{R}^2$	0,448880657		
20	Стандартная ошибка	1,198436652					
21	Наблюдения	12		$t_{табл}$	2,262157158		
22							
23	Дисперсионный анализ						
24		*df*	*SS*	*MS*	*F*	*Значимость F*	
25	Регрессия	2	15,74041298	7,87020649	5,479689639	0,027760651	
26	Остаток	9	12,92625369	1,43625041			
27	Итого	11	28,66666667				
28							
29		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
30	Y-пересечение	-1,16519174	1,860485381	-0,626283739	0,546682713	-5,373902064	3,043518583
31	xi	0,743362832	0,276154094	2,691840707	0,024721868	0,118658871	1,368066793
32	zi	0,466076696	0,763249595	0,610647813	0,556537374	-1,26051384	2,192667232

Рис. 2.22. Результат выполнения инструмента Регрессия (для множественной регрессии)

Коэффициент детерминации R^2 = 0,549. Уравнение регрессии значимо по F – критерию, так как F = 5,48 > $F_{табл}$ = 4,26 (см. рис. 2.22). Однако коэффициент регрессии α при фиктивной переменной Z незначим по t–критерию: t = 0,611 < $t_{табл}$ = 2,26 (см. рис. 2.22). К тому же скорректированный коэффициент детерминации для множественной регрессии $\hat{R}^2$ = 0,449 уменьшился по сравнению с коэффициентом парной регрессии $\hat{R}^2$ = 0,483.

Таким образом, по имеющимся данным влияние фактора «пол» оказалось несущественным и можно считать, что регрессионная модель вступительных экзаменов по математике одна и та же для юношей и девушек.

2.6. Нелинейные модели регрессии

Мы рассматривали линейные модели регрессии, в которых переменные имели первую степень (модели, линейные по переменным), а параметры выступали в виде коэффициентов при этих переменных (модели, линейные по параметрам). Однако соотношения между социально-экономическими явлениями часто носят нелинейный характер: производственные функции (функция Кобба-Дугласа), функции спроса и т.д.

Для оценивания параметров нелинейной модели используют подход, основанный на линеаризации модели и заключающийся в том, что, с помощью подходящих алгебраических преобразований исходных переменных, исследуемую зависимость представляют в виде линейного соотношения между преобразованными переменными.

Если модель не линейна по переменным, то введением новых переменных, ее можно привести к линейному виду:

$$y_i = \beta_0 + \beta_1 x_{i1}^2 + \beta_2 \sqrt{x_{i2}} + \varepsilon_i, \quad i = \overline{1,n}. \qquad (2.22)$$

Введя переменные $z_1 = x_1^2, \quad z_2 = \sqrt{x_2}$, модель (2.22) приводится к линейному виду:

$$y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \varepsilon_i, \quad i = \overline{1,n}. \qquad (2.23)$$

Более сложной проблемой является нелинейность модели по параметрам. К числу таких моделей относятся мультипликативная модель

$$y_i = \beta_0 \cdot x_{i1}^{\beta_1} \cdot x_{i2}^{\beta_2} \cdot \varepsilon_i, \quad i = \overline{1,n} \tag{2.24}$$

и экспоненциальная модель

$$y_i = e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}} \cdot \varepsilon_i, \quad i = \overline{1,n}. \tag{2.25}$$

Эти модели можно привести к линейному виду с помощью логарифмирования обеих частей уравнения. Тогда (2.24) примет вид:

$$\ln y_i = \ln \beta_0 + \beta_1 \cdot \ln x_{i1} + \beta_2 \cdot \ln x_{i2} + \ln \varepsilon_i, \quad i = \overline{1,n}, \tag{2.26}$$

а модель (2.25):

$$\ln y_i = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \ln \varepsilon_i, \quad i = \overline{1,n}. \tag{2.27}$$

Пример 4. На основании данных между личным располагаемым доходом x_1 в млрд.$, x_2 — индекс относительной цены на отдых в млрд.$ и y — услуги на отдых жителей США в млрд.$ за 10 лет (табл. 2.4):

Таблица 2.4

Год	Личный располагаемый доход x_1	Индекс относительной цены на отдых x_2	Отдых y
1	479,7	86,5	9,6
2	489,7	88,6	10
3	503,8	90,6	10,4
4	524,9	91,6	10,9
5	542,3	92,8	11,3
6	580,8	94,6	11,6
7	616,3	95,7	11,9
8	646,8	96,5	12,4
9	673,5	97,8	12,7
10	701,3	99,8	13,4

и предположения, что уравнение регрессии имеет вид:

$$y(x) = \beta_0 \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \varepsilon, \tag{2.28}$$

требуется:

1) Найти оценку коэффициентов регрессии.

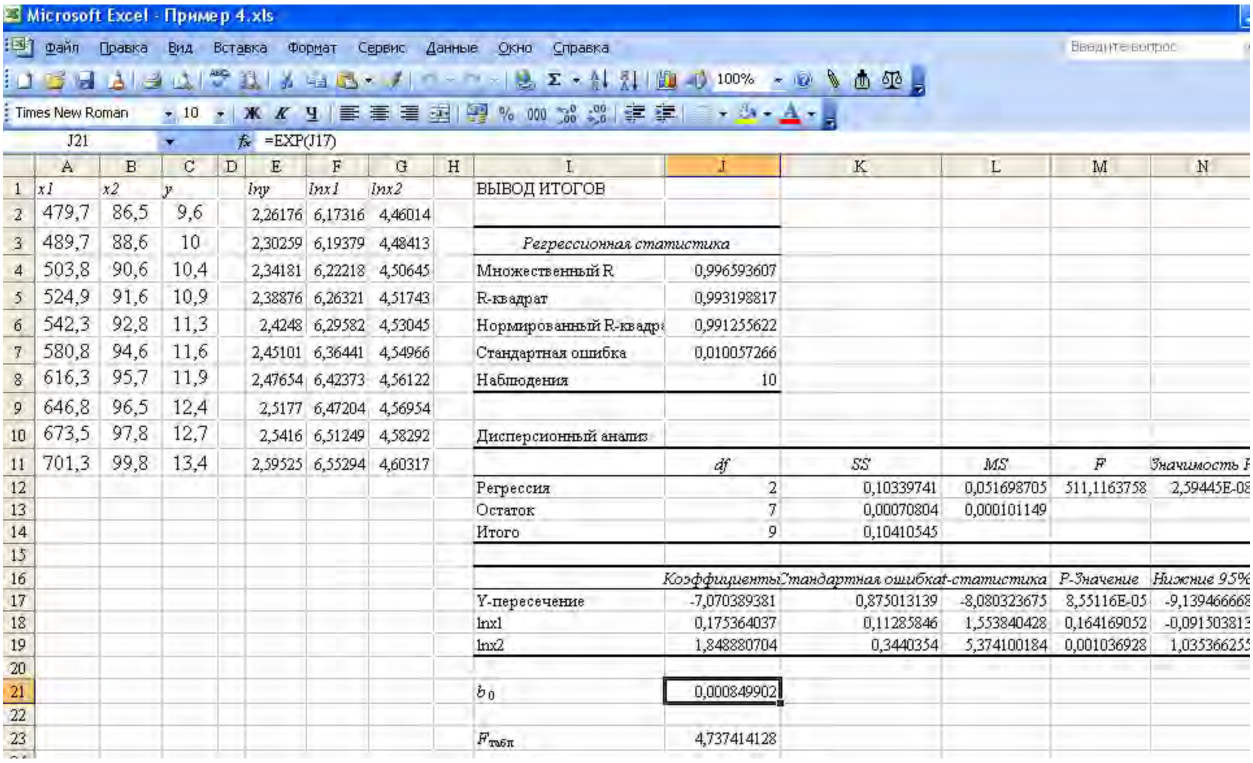
2) Оценить на уровне 5% значимость уравнения регрессии с помощью F-критерия Фишера-Снедекора.

Решение. Оценкой уравнения (2.28) является уравнение

$$\hat{y} = b_0 \cdot x_1^{b_1} \cdot x_2^{b2}. \qquad (2.29)$$

Приведем его к линейному виду с помощью алгебраического преобразования: $\mathbf{ln}\,\hat{y} = \mathbf{ln}\,b_0 + b_1 \cdot \mathbf{ln}\,x_1 + b_2 \cdot \mathbf{ln}\,x_2$.

Для вычисления коэффициентов b_j (j = 0, 1, 2) нелинейной множественной регрессии воспользуемся инструментом анализа данных **Регрессия**, применив его не к исходным данным y, x_1, x_2, а к lny, lnx_1, lnx_2. В итоге получим:



	A	B	C	D	E	F	G
1	x1	x2	y		lny	lnx1	lnx2
2	479,7	86,5	9,6		2,26176	6,17316	4,46014
3	489,7	88,6	10		2,30259	6,19379	4,48413
4	503,8	90,6	10,4		2,34181	6,22218	4,50645
5	524,9	91,6	10,9		2,38876	6,26321	4,51743
6	542,3	92,8	11,3		2,4248	6,29582	4,53045
7	580,8	94,6	11,6		2,45101	6,36441	4,54966
8	616,3	95,7	11,9		2,47654	6,42373	4,56122
9	646,8	96,5	12,4		2,5177	6,47204	4,56954
10	673,5	97,8	12,7		2,5416	6,51249	4,58292
11	701,3	99,8	13,4		2,59525	6,55294	4,60317

	I	J	K	L	M	N
1	ВЫВОД ИТОГОВ					
2						
3	*Регрессионная статистика*					
4	Множественный R	0,996593607				
5	R-квадрат	0,993198817				
6	Нормированный R-квадр	0,991255622				
7	Стандартная ошибка	0,010057266				
8	Наблюдения	10				
9						
10	Дисперсионный анализ					
11		*df*	*SS*	*MS*	*F*	*Значимость F*
12	Регрессия	2	0,10339741	0,051698705	511,1163758	2,59445E-08
13	Остаток	7	0,00070804	0,000101149		
14	Итого	9	0,10410545			
15						
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*
17	Y-пересечение	-7,070389381	0,875013139	-8,080323675	8,55116E-05	-9,13946666
18	lnx1	0,175364037	0,11285846	1,553840428	0,164169052	-0,09150381
19	lnx2	1,848880704	0,3440354	5,374100184	0,001036928	1,03536625
20						
21	b_0	0,000849902				
22						
23	$F_{табл}$	4,737414128				

Рис. 2.23. Результат выполнения инструмента Регрессия (для нелинейной множественной регрессии)

Т.о. $\mathbf{ln}\,b_0 = -7{,}07039 \Rightarrow b_0 = e^{-7{,}07039} = 0{,}00085;\quad b_1 = 0{,}175364;\quad b_2 = 1{,}84888$.

Уравнение нелинейной множественной регрессии примет вид:

$$\hat{y} = 0{,}00085 \cdot x_1^{0{,}175364} \cdot x_2^{1{,}84888}.$$

Коэффициент детерминации $R^2 = 0,99$ (рис. 2.23). Таким образом, вариация зависимой переменной на 99% объясняется изменчивостью объясняющих переменных, входящих в модель. Уравнение регрессии значимо по F – критерию, так как $F = 511,1164 > F_{табл} = 4,737$ (рис. 2.23).

3. Временные ряды

При построении эконометрической модели используется два типа данных: данные, характеризующие совокупность различных объектов в определенный момент времени; данные, характеризующие один объект в различные моменты времени.

Модели, построенные по данным первого типа, называются *пространственными моделями*. Модели, построенные на основе второго типа данных, называются моделями *временных рядов*.

Временной ряд (ряд динамики) — это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени.

При изучении временного ряда в нем обычно выделяют неслучайную и случайную составляющие. В свою очередь, неслучайная составляющая разбивается на два вида в зависимости от образующих их факторов.

Долговременные факторы формируют общую длительную тенденцию изменения анализируемого признака. Эта тенденция задается монотонной (возрастающей или убывающей) функцией и носит название *тренда*. На рис. 3.1 показан временной ряд, имеющий возрастающую тенденцию.

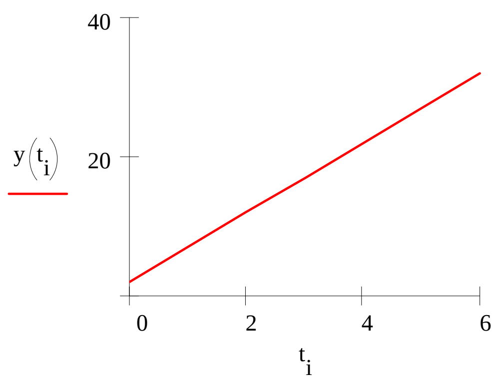


Рис. 3.1. Тренд, имеющий возрастающую тенденцию

Изучаемый признак может быть подвержен циклическим колебаниям. Эти колебания могут носить сезонный характер, поскольку экономическая деятельность ряда отраслей зависит от времени года (сезонов) (например, цены на сельхозпродукцию, уровень безработицы в курортных городах и т.д.). Такие факторы носят название *сезонных*. При наличии больших массивов данных за длительные промежутки времени можно выявить циклические колебания, связанные с общей динамикой конъюнктуры рынка. На рис. 3.2. представлен временной ряд, имеющий только сезонную компоненту.

Таким образом, каждый уровень временного ряда формируется из трендовой (T), циклической (S) и случайной (E) компонент:

T – тренд, плавно меняющаяся компонента, описывающая чистое влияние долговременных факторов, т.е.длительную тенденцию изменения признака;

S – циклическая компонента, отражающая повторяемость экономических процессов в течение длительных периодов. Эта повторяемость может носить сезонный характер (повторяемость процесса в течение не очень длительного периода: года, квартала, месяца, недели);

E – случайная компонента, отражающая влияние неподдающихся учету и регистрации случайных факторов.

В большинстве случаев уровень временного ряда можно представить как сумму или произведение этих компонент.

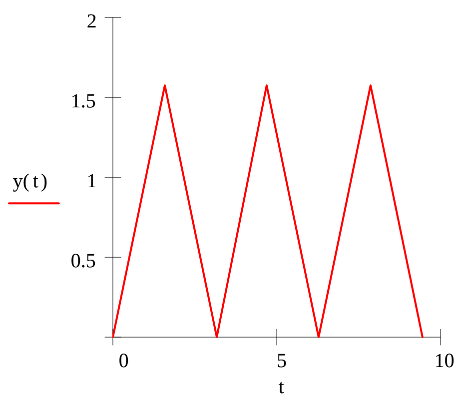


Рис. 3.2. Временной ряд, имеющий только сезонную компоненту

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

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

$Y = T + S + E$ — аддитивная модель;

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

Построение аддитивной и мультипликативной моделей отдельного временного ряда — это выявление и придание количественного значения каждой компоненте *T*, *S*, *E*. Полученную информацию используют для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

3.1. Автокорреляция уровней временного ряда

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих уровней.

Корреляционную зависимость между последовательными уровнями ряда называют *автокорреляцией* уровней ряда.

Количественно её можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Формула для расчета коэффициента автокорреляции имеет вид:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}}, \qquad (3.1)$$

где

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

Эту величину называют *коэффициентом автокорреляции* уровней ряда *первого порядка*, так как он измеряет зависимость между соседними уровнями ряда y_t и y_{t-1}.

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями y_t и y_{t-2} и определяется по формуле:

$$r_2 = \frac{\sum_{t=3}^{n} (y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n} (y_t - \bar{y}_3)^2 \sum_{t=3}^{n} (y_{t-2} - \bar{y}_4)^2}}, \tag{3.2}$$

где

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

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют *лагом*. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Считается целесообразным для обеспечения статистической достоверности коэффициентов автокорреляции использовать правило — максимальный лаг должен быть не больше $n/4$.

По коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. По его знаку нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержат положительную автокорреляцию уровней, однако при этом могут иметь убывающую тенденцию.

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

Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором автокорреляция наиболее высокая, а, следовательно, и лаг, при котором связь между текущими и предыдущими уровнями ряда наиболее тесная, т.е. выявить структуру ряда.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ, то ряд содержит циклические колебания с периодичностью τ моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать одно из двух предположений относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний; либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

Поэтому коэффициент автокорреляции уровней и автокорреляционную функцию целесообразно использовать для выявления во временном ряде наличия или отсутствия трендовой и сезонной компоненты.

Пример 5. Имеются условные данные об общем количестве правонарушений на таможне одного из субъектов РФ (табл. 3.1). Определить наличие или отсутствие во временном ряде трендовой и сезонной компонент.

Таблица 3.1

Год	Квартал	t	Количество возбужденных дел, y_i
1999	I	1	375
	II	2	371
	III	3	869
	IV	4	1015
2000	I	5	357
	II	6	471
	III	7	992
	IV	8	1020
2001	I	9	390
	II	10	355
	III	11	992
	IV	12	905
2002	I	13	461
	II	14	454

Год	Квартал	t	Количество возбужденных дел, y_i
	III	15	920
	IV	16	927

Решение. Построим поле корреляции (рис. 3.3).

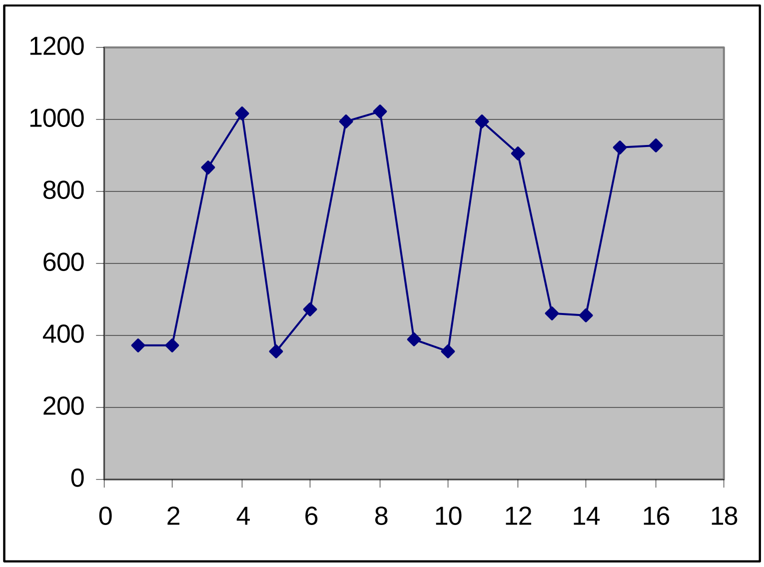


Рис. 3.3 Поле корреляции

Из графика видно, что значения y имеют колебательный характер. Рассчитаем несколько последовательных коэффициентов автокорреляции. Для этого составим вспомогательную таблицу 3.2. Все вычисления проводились в среде MS Excel.

Таблица 3.2

t	y_t	y_{t-1}	$y_t - \bar{y}_1$	$y_{t-1} - \bar{y}_2$	$(y_t - \bar{y}_1)\cdot(y_{t-1} - \bar{y}_2)$	$(y_t - \bar{y}_1)^2$	$(y_{t-1} - \bar{y}_2)^2$
1	375	—	—	—	—	—	—
2	371	375	-328,93	-288,13	94776,66	108197,14	83020,82
3	869	371	169,07	-292,13	-49390,01	28583,54	85341,88
4	1015	869	315,07	205,87	64861,72	99267,00	42381,08
5	357	1015	-342,93	351,87	-120666,81	117603,27	123810,15
6	471	357	-228,93	-306,13	70084,12	52410,47	93717,62
7	992	471	292,07	-192,13	-56115,74	85302,94	36915,22
8	1020	992	320,07	328,87	105259,26	102442,67	108153,28
9	390	1020	-309,93	356,87	-110604,88	96058,67	127353,82
10	355	390	-344,93	-273,13	94212,79	118979,00	74601,82
11	992	355	292,07	-308,13	-89995,48	85302,94	94946,15
12	905	992	205,07	328,87	67439,59	42052,34	108153,28
13	461	905	-238,93	241,87	-57790,01	57089,14	58499,48
14	454	461	-245,93	-202,13	49711,32	60483,20	40857,88

t	y_t	y_{t-1}	$y_t-\bar{y}_1$	$y_{t-1}-\bar{y}_2$	$(y_t-\bar{y}_1)\cdot(y_{t-1}-\bar{y}_2)$	$(y_t-\bar{y}_1)^2$	$(y_{t-1}-\bar{y}_2)^2$
15	920	454	220,07	-209,13	-46023,28	48429,34	43736,75
16	927	920	227,07	256,87	58325,86	51559,27	65980,48
Сумма	10499	9947	—	—	74085,13	1153760,93	1187469,73
Среднее значение	699,33	663,13	—	—	—	—	—

$$\bar{y}_1=\frac{1}{n-1}\sum_{t=2}^{n}y_t=\frac{1}{15}\cdot 10499=699{,}33;\quad \bar{y}_2=\frac{1}{n-1}\sum_{t=2}^{n}y_{t-1}=\frac{1}{15}\cdot 9947=663{,}13.$$

Коэффициент автокорреляции первого порядка:

$$r_1=\frac{\sum_{t=2}^{n}(y_t-\bar{y}_1)(y_{t-1}-\bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t-\bar{y}_1)^2\sum_{t=2}^{n}(y_{t-1}-\bar{y}_2)^2}}=\frac{74085{,}13}{\sqrt{1153760{,}93\cdot 1187469{,}73}}=0{,}063.$$

Для расчета коэффициента автокорреляции второго порядка составим таблицу 3.3.

Таблица 3.3

t	y_t	y_{t-2}	$y_t-\bar{y}_3$	$y_{t-2}-\bar{y}_4$	$(y_t-\bar{y}_3)\cdot(y_{t-2}-\bar{y}_4)$	$(y_t-\bar{y}_3)^2$	$(y_{t-2}-\bar{y}_4)^2$
1	375	—	—	—	—	—	—
2	371	—	—	—	—	—	—
3	869	375	145,57	-269,79	-39273,09	21191,04	72784,33
4	1015	371	291,57	-273,79	-79828,09	85013,90	74958,62
5	357	869	-366,43	224,21	-82158,52	134269,90	50272,05
6	471	1015	-252,43	370,21	-93452,66	63720,18	137058,62
7	992	357	268,57	-287,79	-77291,02	72130,61	82820,62
8	1020	471	296,57	-173,79	-51539,88	87954,61	30201,47
9	390	992	-333,43	347,21	-115771,16	111174,61	120557,76
10	355	1020	-368,43	375,21	-138239,66	135739,61	140785,76
11	992	390	268,57	-254,79	-68428,16	72130,61	64915,76
12	905	355	181,57	-289,79	-52616,81	32968,18	83975,76
13	461	992	-262,43	347,21	-91118,95	68868,76	120557,76
14	454	905	-269,43	260,21	-70109,16	72591,76	67711,47
15	920	461	196,57	-183,79	-36127,02	38640,33	33777,19
16	927	454	203,57	-190,79	-38838,52	41441,33	36399,19
Сумма	10128	9027	—	—	-1034792,71	1037835,43	1116776,36
Среднее значение	723,43	644,79	—	—	—	—	—

Тогда

$$r_2 = \frac{-1034792{,}71}{\sqrt{1037835{,}43 \cdot 1116776{,}36}} = -0{,}961.$$

Аналогично находим коэффициенты автокорреляции более высоких порядков, а все полученные значения заносим в таблицу 3.4.

Таблица 3.4

Лаг	Коэффициент автокорреляции
1	0,063
2	–0,961
3	–0,036
4	0,964
5	0,051
6	–0,977
7	–0,069
8	0,965
9	0,162
10	–0,973
11	–0,065
12	0,986

По результатам табл. 3.4 построим коррелограмму (рис. 3.4). Анализ коррелограммы и графика исходных уровней временного ряда позволяет сделать вывод о наличии в изучаемом временном ряде сезонных колебаний периодичностью в четыре квартала, т.к. коэффициенты автокорреляции 4-го, 8-го и 12-го порядков оказались самыми большими.

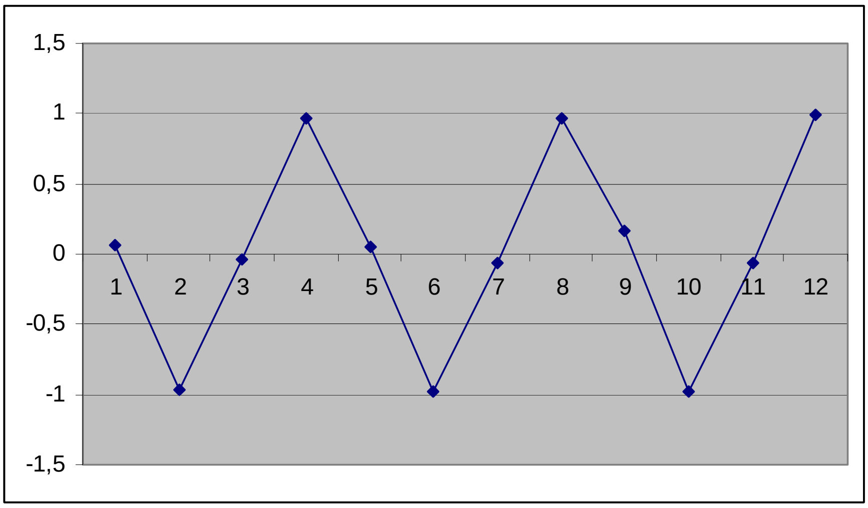


Рис. 3.4. Коррелограмма

3.2. Моделирование тенденции временного ряда

Распространенным способом моделирования тенденции временного ряда является построение аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда. Этот способ называется *аналитическим выравниваем* временного ряда.

Поскольку зависимость от времени может принимать разные формы, то для построения тренда используют различные виды функций. Чаще всего применяются следующие функции:

линейный тренд $y_t = a + bt$;

гиперболический тренд $y_t = a + \frac{b}{t}$;

экспоненциальный тренд $y_t = e^{a+bt}$ или $y_t = ab^t$;

степенной тренд $y_t = at^b$;

полиномы различных степеней $y_t = b_0 + b_1 t + b_2 t^2 + ... + b_k t^k$.

Параметры каждого из перечисленных трендов можно определить обычным методом наименьших квадратов, используя в качестве независимой переменной время $t = 1, 2, ..., n$, а в качестве зависимой переменной — фактические уровни временного ряда. Существует несколько способов определения типа тенденции. К числу наиболее распространенных способов относится качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени. В этих же целях можно использовать и коэффициенты автокорреляции уровней ряда.

Выбор наилучшего уравнения в случае, когда ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации и средней ошибки аппроксимации. Этот метод легко реализуется при компьютерной обработке данных.

Примечание. В Excel трендовые модели строятся на основе диаграмм, определяющих уровни динамики ряда. Для построения линии тренда необходимо выделить временной ряд и выбрать команду *Добавить линию тренда в* меню *Диаграмма*. Появится диалоговое окно *Линия тренда*, содержащее вкладку *Тип*, на которой задается тип тренда. Вкладка *Параметры* предназначена для задания параметров линии тренда:

1. имя тренда,
2. прогноз вперед на количество периодов, на которое линия тренда проектируется в будущее,
3. прогноз назад на количество периодов, на которое линия тренда проектируется в прошлое,
4. пересечение кривой с осью Y,
5. показывать уравнение на диаграмме,
6. поместить на диаграмме величину достоверности аппроксимации R^2.

3.3. Моделирование сезонных колебаний

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

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которые ставит уровни ряда в зависимость от значений сезонной компоненты. Построение обеих моделей сводится к расчету значений T, S и E для каждого уровня ряда. Процесс построения модели включает в себя 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 ($T + E$) в аддитивной или ($T \cdot E$) в мультипликативной модели;

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

5) расчет полученных по модели значений ($T + E$) или ($T \cdot E$) и оценка качества полученной модели;

6) прогнозирование с помощью полученной модели.

Методику построения каждой из моделей рассмотрим на примерах.

Метод скользящей средней. Это один из самых старых и широко известных способов сглаживания временного ряда. Сглаживание представляет собой некоторый способ локального усреднения данных, при котором несистематические компоненты взаимно погашают друг друга. Метод скользящей средней основан на переходе от начальных значений ряда к их средним значениям на интервале времени, длина которого выбрана заранее. Данный интервал времени принято называть «окном». Полученный таким образом ряд скользящих средних ведет себя более гладко, чем исходный ряд, за счет усреднения отклонений исходного ряда. Эта процедура дает представление об общей тенденции ряда.

Предположим, что длина окна сглаживания *p* выражается нечетным числом: $p = 2m + 1$. Даны дискретные во времени наблюдения над некоторым изучаемым процессом: $y_1, y_{2.},..., y_n$ (*n* — объем выборки).

Тогда метод скользящей средней состоит в том, что исходный эмпирический временной ряд $y_1, y_{2.},..., y_n$ преобразуется в ряд сглаженных значений (оценок) по формуле (3.3):

$$\bar{y}_t = \frac{1}{p} \sum_{j=t-m}^{t+m} y_j \, , \tag{3.3}$$

где p — размер окна, j — порядковый номер уровня в окне сглаживания, m — величина, определяемая по формуле $m = (p-1)/2$.

Для определения сглаженных уровней при $p = 2m$ применяется метод центрирования, который заключается в нахождении средней из двух смежных скользящих средних для отнесения полученного уровня к определенной дате.

Примечание. В среде MS Excel имеется режим работы «*Скользящее среднее*» в *Анализе данных* в меню *Сервис* для сглаживания эмпирического временного ряда на основе метода простой скользящей средней.

Пример 6. Построим аддитивную и мультипликативную модели временного ряда по данным табл. 3.1.

Решение. Мы показали (см. пример 5), что данный временной ряд содержит сезонные колебания периодичностью 4, т.е. количество правонарушений в первый и второй квартал ниже, чем в третий и четвертый. Рассчитаем компоненты аддитивной модели временного ряда.

Шаг 1. Проведем выравнивание исходных уровней ряда методом скользящей средней, используя режим «Скользящее среднее» в Excel, учитывая, что p = 4. Результаты получим в виде таблицы 3.5.

Таблица 3.5

t	y_t	Скользящее среднее	Стандартные погрешности
1	375	#Н/Д	#Н/Д
2	371	657,50	#Н/Д
3	869	653,00	#Н/Д
4	1015	678,00	#Н/Д
5	357	708,75	290,90
6	471	710,00	276,95
7	992	718,25	285,89
8	1020	689,25	314,56
9	390	689,25	319,07
10	355	660,50	304,51
11	992	678,25	278,55
12	905	703,00	255,26
13	461	685,00	236,75
14	454	690,50	234,71
15	920		
16	927		

Так как $p = 4$, то процедура центрирования обязательна. Для этого найдем средние значения из двух последовательных скользящих средних, используя функцию СРЗНАЧ. Результат поместим в табл. 3.6.

Таблица 3.6

Квартал	t	y_t	Скользящая средняя	Центрированная скользящая средняя	Оценка сезонной компоненты
1	**2**	**3**	**4**	**5**	**6**
I	1	375	—	—	—
II	2	371	657,50	—	—
III	3	869	653,00	655,25	213,75
IV	4	1015	678,00	665,50	349,50
I	5	357	708,75	693,38	-336,38
II	6	471	710,00	709,38	-238,38
III	7	992	718,25	714,13	277,88
IV	8	1020	689,25	703,75	316,25
I	9	390	689,25	689,25	-299,25
II	10	355	660,50	674,88	-319,88
III	11	992	678,25	669,38	322,63
IV	12	905	703,00	690,63	214,38
I	13	461	685,00	694,00	-233,00
II	14	454	690,50	687,75	-233,75
III	15	920	—	—	—
IV	16	927	—	—	—

Шаг 2. Найдем оценку сезонной компоненты, как разность между фактическими уровнями ряда и центрированными скользящими средними (табл. 3.6, гр. 6). Используем эти оценки для расчета значений сезонной компоненты S (табл. 3.7). Для этого найдем средние за каждый квартал (по всем годам) оценки сезонной компоненты. В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимно погашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по все кварталам должна быть равна нулю.

Таблица 3.7

Показатели	Год	№ квартала			
		I	II	III	IV
	1999	—	—	213,75	349,5
	2000	-336,38	-238,38	277,88	316,25
	2001	-299,25	-319,88	322,63	214,38
	2002	-233,00	-233,75	—	—

Всего за i-й квартал		-868,63	-792,01	814,26	880,13
Средняя оценка сезонной компоненты $\overline{S}_i$		-289,54	-264,00	271,42	293,38
Скорректированная сезонная компонента S_i		-292,36	-266,82	268,61	290,56

Для данной модели имеем:

$$\sum_{i=1}^{4} \overline{S}_i = -289{,}54 - 264 + 271{,}42 + 293{,}38 = 11{,}25 .$$

Вычислим корректирующий коэффициент: $k = 1/4 \sum_{i=1}^{4} \overline{S}_i = 2{,}8125.$

Рассчитываем скорректированные значения сезонной компоненты по формуле $S_i = \overline{S}_i - k$ и заносим полученные данные в последнюю строку табл. 3.7. Проверим равенство нулю суммы значений сезонной компоненты:

–292,36 – 266,82 + 268,61 +290,56 = 0.

Шаг 3. Исключим влияние сезонной компоненты, вычитая её значение из каждого уровня временного ряда, так как аддитивная модель имеет вид $Y = T + S + E$, откуда $T + E = Y - S$ (табл. 3.8, гр. 4).

Таблица 3.8

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	-292,36	667,36	672,69	380,33	-5,33	28,36
2	371	-266,82	637,82	673,61	406,79	-35,79	1280,98
3	869	268,61	600,39	674,54	943,15	-74,15	5497,66
4	1015	290,56	724,44	675,46	966,02	48,98	2398,88
5	357	-292,36	649,36	676,39	384,03	-27,03	730,46
6	471	-266,82	737,82	677,31	410,49	60,51	3661,17
7	992	268,61	723,39	678,24	946,85	45,15	2038,72
8	1020	290,56	729,44	679,16	969,72	50,28	2527,76
9	390	-292,36	682,36	680,09	387,73	2,27	5,16
10	355	-266,82	621,82	681,01	414,19	-59,19	3503,93
11	992	268,61	723,39	681,94	950,55	41,45	1718,15
12	905	290,56	614,44	682,86	973,42	-68,42	4681,95
13	461	-292,36	753,36	683,79	391,43	69,57	4839,96
14	454	-266,82	720,82	684,72	417,90	36,10	1303,53
15	920	268,61	651,39	685,64	954,25	-34,25	1173,13
16	927	290,56	636,44	686,57	977,13	-50,13	2512,66
						Сумма	37902,46

Шаг 4. Определим компоненту T данной модели. Для этого проведем аналитическое выравнивание ряда $(T + E)$ с помощью линейного тренда (рис. 3.5).

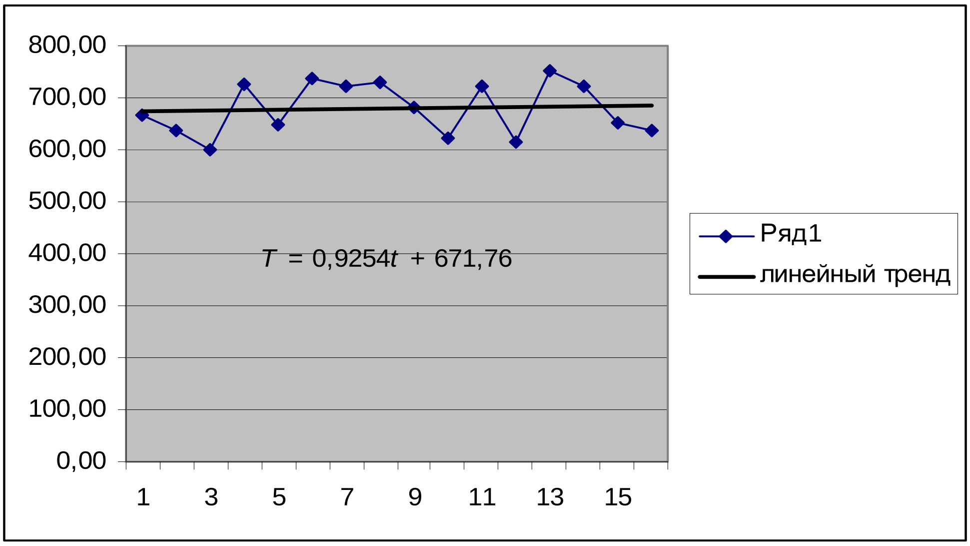


Рис. 3.5. Выравнивание линейным трендом

Подставляя в уравнение тренда t =1, 2,…, 16, найдем уровни T для каждого момента времени (табл. 3.8, гр. 5).

Шаг 5. Найдем значения уровней ряда, полученные по аддитивной модели. Для этого прибавим к уровням T значения сезонной компоненты для соответствующих кварталов (табл. 3.8, гр. 6). На одном графике отложим фактические значения уровней временного ряда и теоретические, полученные по аддитивной модели (рис. 3.6).

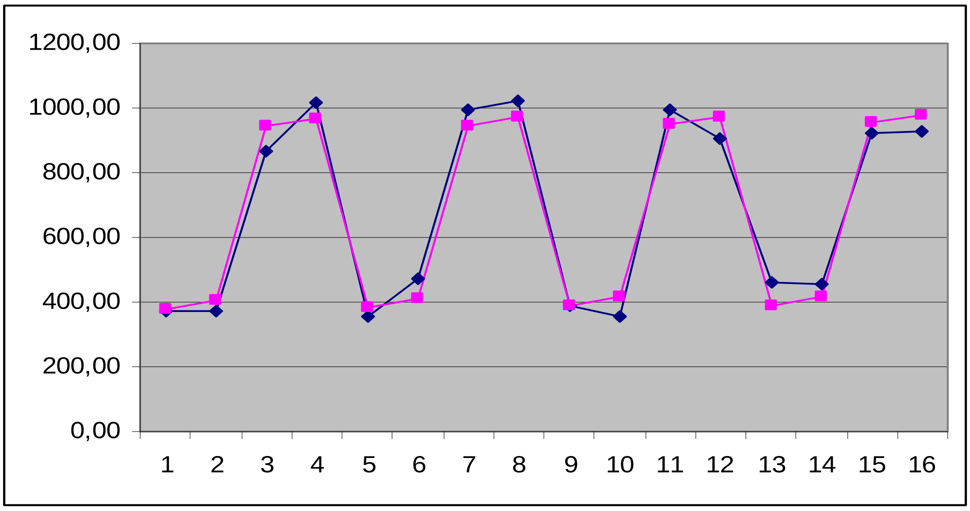


Рис. 3.6. Фактические и теоретические значения уровней временного ряда по аддитивной модели

Для оценки качества построенной модели используем сумму квадратов полученных абсолютных ошибок:

$$R^2 = 1 - \frac{E^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} = 1 - \frac{37902,46}{1252743,75} = 0,97\,.$$

Следовательно, аддитивная модель объясняет 97% общей вариации уровней временного ряда количества правонарушений по кварталам за 4 года. *Шаг* 6. Прогнозирование по аддитивной модели. Предположим, что необходимо дать прогноз об общем объеме правонарушений на I и II квартал 2003 года. Прогнозное значение Y_t уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда (рис. 3.5):

$$T = 0,9254t + 671,76\,.$$

Получим

$$T_{17} = 0,9254 \cdot 17 + 671,76 = 687,49\,;$$

$$T_{18} = 0,9254 \cdot 18 + 671,76 = 688,42\,.$$

Значения сезонных компонент за соответствующие кварталы равны $S_1 = -292,36$ и $S_2 = -266,82$.

Тогда

$$Y_{17} = T_{17} + S_1 = 687,49 - 292,36 \approx 395\,;$$

$$Y_{18} = T_{18} + S_2 = 688,42 - 266,82 \approx 422.$$

Таким образом, за первые два квартала 2003 года следовало ожидать порядка 395 и 422 правонарушений соответственно.

Перейдем к расчету компонент мультипликативной модели временного ряда.

Шаг 1. Методика, применяемая на этом шаге, полностью совпадает с методикой построения аддитивной модели (табл. 3.5).

Шаг 2. Найдем оценки сезонной компоненты как частное от деления фактических уровней на центрированные скользящие средние (табл. 3.9, гр. 6).

Таблица 3.9

Квартал	t	y_t	Скользящая средняя	Центрированная скользящая средняя	Оценка сезонной компоненты
1	**2**	**3**	**4**	**5**	**6**
I	1	375	—	—	—
II	2	371	657,50	—	—
III	3	869	653,00	655,25	1,33
IV	4	1015	678,00	665,50	1,53
I	5	357	708,75	693,38	0,51
II	6	471	710,00	709,38	0,66
III	7	992	718,25	714,13	1,39
IV	8	1020	689,25	703,75	1,45
I	9	390	689,25	689,25	0,57
II	10	355	660,50	674,88	0,53
III	11	992	678,25	669,38	1,48
IV	12	905	703,00	690,63	1,31
I	13	461	685,00	694,00	0,66
II	14	454	690,50	687,75	0,66
III	15	920	—	—	—
IV	16	927	—	—	—

Эти оценки используем для расчета сезонной компоненты S (табл. 3.10). Так же как и в аддитивной модели считается, что сезонные воздействия за период взаимно погашаются. В мультипликативной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов в цикле, т.е. 4 в нашем случае.

Таблица 3.10

Показатели	Год	№ квартала			
		I	II	III	IV
	1999	—	—	1,33	1,53
	2000	0,51	0,66	1,39	1,45
	2001	0,57	0,53	1,48	1,31
	2002	0,66	0,66	—	—
Всего за i-ый квартал		1,74	1,85	4,20	4,28
Средняя оценка сезонной компоненты $\bar{S}_i$	—	0,58	0,62	1,40	1,43
Скорректированная сезонная компонента S_i	—	0,58	0,61	1,39	1,42

Для определения корректирующего коэффициента, найдем сумму средних оценок сезонной компоненты и разделим число периодов в цикле на эту сумму:

$$\sum_{i=1}^{4} \overline{S}_i = 0{,}58 + 0{,}62 + 1{,}40 + 1{,}43 = 4{,}03.$$

Тогда корректирующий коэффициент: $k = 4 \Big/ \sum_{i=1}^{4} \overline{S}_i = 0{,}99.$

Скорректированные значения сезонной компоненты получаются при умножении её средней оценки на корректирующий коэффициент. Найдем сумму значений сезонной компоненты и убедимся, что она равна 4:

0,58 + 0,61 + 1,39 + 1,42 = 4.

Шаг 3. Устраним сезонную компоненту из исходных уровней ряда: $Y = T \cdot S \cdot E$, тогда $T \cdot E = Y/S$. Разделим каждый уровень ряда на соответствующие значения сезонной компоненты. В результате получим величины, которые содержат только тенденцию и случайную компоненту (табл. 3.11, гр. 4).

Таблица 3.11

t	y_t	S_i	y_t / S_i	T	$T \cdot S$	$E = (y_t - T \cdot S)^2$
1	**2**	**3**	**4**	**5**	**6**	**7**
1	375	0,58	648,87	654,91	378,49	12,21
2	371	0,61	605,46	658,19	403,31	1043,97
3	869	1,39	625,12	661,47	919,54	2554,54
4	1015	1,42	715,21	664,76	943,40	5126,31
5	357	0,58	617,72	668,04	386,08	845,63
6	471	0,61	768,66	671,32	411,35	3557,75
7	992	1,39	713,60	674,60	937,79	2938,91
8	1020	1,42	718,73	677,88	962,03	3360,68
9	390	0,58	674,82	681,16	393,67	13,43
10	355	0,61	579,35	684,44	419,40	4146,80
11	992	1,39	713,60	687,72	956,03	1293,54
12	905	1,42	637,70	691,01	980,66	5723,76
13	461	0,58	797,67	694,29	401,25	3569,99
14	454	0,61	740,92	697,57	427,44	705,53
15	920	1,39	661,80	700,85	974,28	2946,33
16	927	1,42	653,20	704,13	999,28	5224,75
					Сумма	43064,13

Шаг 4. Определим компоненту T в мультипликативной модели. Для этого рассчитаем параметры линейного тренда, используя уровни ($T \cdot E$).

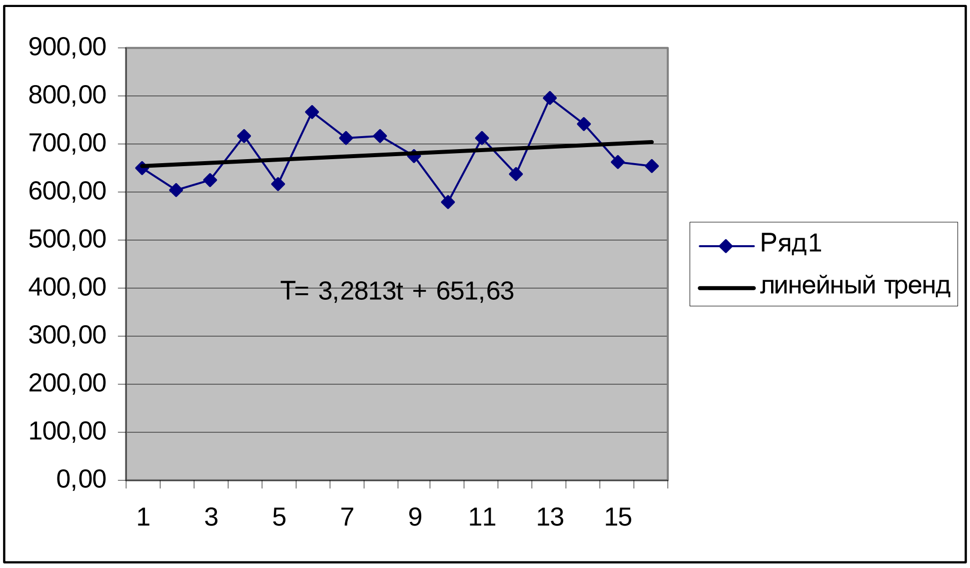


Рис. 3.7. Параметры линейного тренда

Подставляя в уравнение тренда $T = 3{,}2813t + 651{,}63$ значения t =1, 2,…, 16, найдем уровни T для каждого момента времени (табл. 3.11, гр. 5).

Шаг 5. Найдем уровни ряда, умножив значения T на соответствующие им значения сезонной компоненты (табл. 3.11, гр. 6). Построим график фактических уровней временного ряда и теоретических, полученных по мультипликативной модели (рис. 3.8).

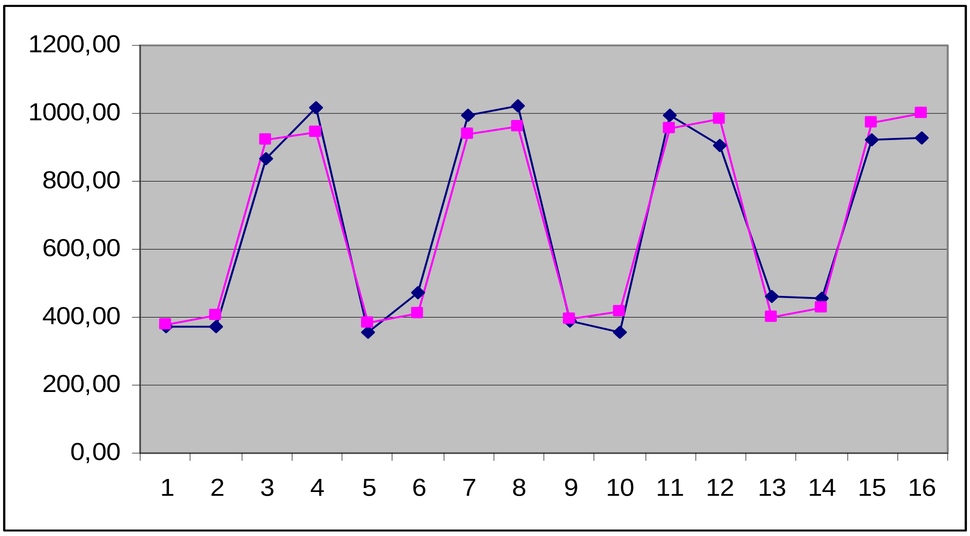


Рис. 3.8. Фактические и теоретические уровни временного ряда по мультипликативной модели

Для оценки качества построенной модели вычислим ошибку:

$$R^2 = 1 - \frac{E^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} = 1 - \frac{43064{,}13}{1252743{,}75} = 0{,}966\,.$$

Для расчета ошибки в мультипликативной модели можно использовать формулу $E = Y/(T \cdot S)$.

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

Шаг 6. Осуществим прогноз об общем объеме правонарушений по мультипликативной модели на I и II кварталы 2003.

Для определения трендовой компоненты используем уравнение тренда:

$$T = 3{,}2813t + 651{,}63\,.$$

Тогда

$$T_{17} = 3{,}2813 \cdot 17 + 651{,}63 = 707{,}41;$$

$$T_{18} = 3{,}2813 \cdot 18 + 651{,}63 = 710{,}69.$$

Значения сезонных компонент за соответствующие кварталы равны:

$S_1 = 0{,}58$ и $S_2 = 0{,}61$.

Тогда

$$Y_{17} = T_{17} \cdot S_1 = 707{,}41 \cdot 0{,}58 \approx 409;$$

$$Y_{18} = T_{18} \cdot S_2 = 710{,}69 \cdot 0{,}61 \approx 436.$$

Т.е. следовало ожидать порядка 409 и 436 правонарушений соответственно в первом и втором кварталах 2003 г. В данном примере аддитивная и мультипликативная модели дают примерно одинаковый результат.

3.4. Автокорреляция в остатках. Критерий Дарбина-Уотсона

Рассмотрим регрессионную модель временного ряда. Упорядоченность наблюдений оказывается существенной в тех случаях, если прослеживается механизм влияния результатов предыдущих наблюдений на результаты последующих. Математически это выражается в том, что случайные величины ε_i в регрессионной модели не оказываются независимыми, т.е. условие $r(\varepsilon_i, \varepsilon_j) = 0$ не выполняется. Такие модели называются моделями с *наличием автокорреляции*. На практике именно ими оказываются временные ряды.

Автокорреляция остатков регрессионной модели ε_i (или случайных ошибок уравнения регрессии) это *корреляционная зависимость* между настоящими и прошлыми значениями остатков.

Рассмотрим в качестве примера временной ряд — ряд последовательных значений курса ценной бумаги, наблюдаемых в определенные моменты времени. Очевидно, что результаты предыдущих торгов оказывают влияние на результаты последующих: если в какой-то момент курс окажется завышенным по сравнению с реальным, тогда скорее всего он будет завышен и на следующих торгах. В этом случае имеет место *положительная автокорреляция*.

Графически положительная автокорреляция выражается в чередовании зон, где наблюдаемые значения оказываются выше предсказанных, и зон, где наблюдаемые значения ниже предсказанных (рис. 3.7).

Отрицательная автокорреляция встречается в тех случаях, когда наблюдения действуют друг на друга по принципу «маятника»: завышенные значения в предыдущих наблюдениях приводят к занижению их в последующих наблюдениях.

Графически это выражается в том, что результаты наблюдений слишком часто пересекают линию тренда.

Автокорреляция в остатках может быть вызвана несколькими причинами, имеющими различную природу:

– она может быть связана с исходными данными и вызвана наличием ошибок измерения в значениях результативного признака;

– в ряде случаев она может быть следствием неправильной спецификации модели. Модель может не включать фактор, который оказывает существенное воздействие на результат и влияние которого отражается в остатках. Вследствие чего, последние могут оказаться автокоррелированными. Очень часто этим фактором является фактор времени.

От истинной автокорреляции остатков следует отличать ситуации, когда причина автокорреляции заключается в неправильной спецификации функциональной формы модели. В этом случае следует изменить форму модели, а не использовать специальные методы расчета параметров уравнения регрессии при наличии автокорреляции в остатках.

Одним из наиболее распространенных методов определения автокорреляции в остатках является тест Дарбина-Уотсона, который основан на простой идее: если корреляция ошибок регрессии не равна нулю, то она присутствует и в остатках регрессии ε_i. В данном тесте для оценки корреляции используется статистика:

$$d = \frac{\sum_{t=2}^{n}\left(\varepsilon_t - \varepsilon_{t-1}\right)^2}{\sum_{t=1}^{n}\varepsilon_t^2}. \qquad (3.4)$$

Величина d есть отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по выбранной модели регрессии.

Можно доказать, что при больших значениях n справедлива формула:

$$d = 2 \cdot \left(1 - r_1\right), \qquad (3.5)$$

где r_1 — коэффициент автокорреляции остатков первого порядка.

Из формулы (3.5) следует что, если в остатках существует полная положительная автокорреляция и $r_1 = 1$, то $d = 0$. Если в остатках полная отрицательная автокорреляция и $r_1 = -1$, то $d = 4$. Если автокорреляция отсутствует, то $r_1 = 0$ и $d = 2$. Таким образом, $0 \le d \le 4$.

Для выявления автокорреляции остатков на основе критерия Дарбина-Уотсона используют следующий алгоритм.

1. Выдвигается гипотеза H_0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H_1 и H_1^* состоят соответственно в наличии положительной или отрицательной автокорреляции в остатках.

2. Задается уровень значимости α.

3. По специальным таблицам определяются критические значения критерия Дарбина-Уотсона d_L и d_U для заданного числа наблюдений *n*, числа независимых переменных m и уровня значимости α.

4. По этим значениям числовой промежуток [0; 4] разбивают на пять отрезков. Принятие или отклонение каждой из гипотез с вероятностью $p = 1 - \alpha$ осуществляется следующим образом:

$0 < d < d_L$ — есть положительная автокорреляция остатков, гипотеза H_0 отклоняется и принимается гипотеза H_1 с вероятностью *p*;

$d_L < d < d_U$ — зона неопределенности;

$d_U < d < 4 - d_U$ — нет оснований отклонить H_0, т.е. автокорреляция остатков отсутствует;

$4 - d_U < d < 4 - d_L$ — зона неопределенности;

$4 - d_L < d < 4$ есть отрицательная автокорреляция остатков, гипотеза H_0 отклоняется и принимается гипотеза H_1^* с вероятностью *p*.

Алгоритм проверки гипотезы о наличии автокорреляции в остатках изображен на рис. 3.9.

Рис. 3.9. Алгоритм проверки о наличии автокорреляции в остатках

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

Пример 7. Проверим гипотезу о наличии автокорреляции в остатках для аддитивной модели временного ряда, рассмотренного в примере 6.

Решение. Исходные данные и промежуточные расчеты заносим в табл. 3.12. Третий столбец копируем из табл. 3.8.

Таблица 3.12

t	y_t	$\varepsilon_t = E$	ε_{t-1}	$(\varepsilon_t - \varepsilon_{t-1})^2$	ε_t^2
1	375	-5,33	—	0,00	28,41
2	371	-35,79	-5,33	927,81	1280,92
3	869	-74,15	-35,79	1471,49	5498,22
4	1015	48,98	-74,15	15161,00	2399,04
5	357	-27,03	48,98	5777,52	730,62
6	471	60,51	-27,03	7663,25	3661,46
7	992	45,15	60,51	235,93	2038,52
8	1020	50,28	45,15	26,32	2528,08
9	390	2,27	50,28	2304,96	5,15
10	355	-59,19	2,27	3777,33	3503,46
11	992	41,45	-59,19	10128,41	1718,10
12	905	-68,42	41,45	12071,42	4681,30
13	461	69,57	-68,42	19041,24	4839,98
14	454	36,10	69,57	1120,24	1303,21
15	920	-34,25	36,10	4949,12	1173,06
16	927	-50,13	-34,25	252,17	2513,02
	Сумма	0,02	50,15	84908,21	37902,56

Фактическое значение критерия Дарбина-Уотсона для данной модели составляет:

$$d = \frac{84908,21}{37902,56} = 2,24\,.$$

Зададим уровень значимости $\alpha = 0,05$. По таблице значений критерия Дарбина-Уотсона по числу наблюдений n = 16 и числа независимых параметров модели m = 1 (мы рассматриваем зависимость от времени) критические значения $d_L = 1,1$ и $d_U = 1,37$. Фактическое значение d попадает в интервал $d_U < d < 4 - d_U$ (1,37 < d < 2,63), следовательно, H_0 принимается, т.е. отсутствует автокорреляция в остатках.

Существует несколько ограничений на применение критерия Дарбина-Уотсона.

1. Он не применим к моделям, включающим в качестве независимых переменных лаговые значения результативного признака.

2. Методика расчета и использование критерия Дарбина-Уотсона направлена на выявление автокорреляции остатков первого порядка.

3. Критерий Дарбина-Уотсона дает достоверные результаты только для больших выборок.

Существуют также и другие тесты на наличие автокорреляции, которые можно использовать на практике [1, 3].

4. Системы эконометрических уравнений

При использовании отдельных уравнений регрессии, например для экономических расчетов, в большинстве случаев предполагается, что факторы можно изменять независимо друг от друга. Однако это предположение является ошибочным, так как изменение одной переменной, не может происходить при абсолютной неизменности других. Её изменение повлечет за собой изменения во всей системе взаимосвязанных признаков. Следовательно, отдельно взятое уравнение множественной регрессии не может характеризовать истинные влияния отдельных признаков на вариацию результативной переменной. Именно поэтому в последние десятилетия в экономических исследованиях важное место заняла проблема описания структуры связей между переменными системой, называемой *системой одновременных уравнений или структурными уравнениями*.

Система уравнений в эконометрических исследованиях может быть построена по-разному.

Возможна система *независимых уравнений*, когда каждая зависимая переменная y рассматривается как функция одного и того же набора факторов x:

$$
\begin{cases}
y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\
y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
y_m = a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m.
\end{cases}
\tag{4.1}
$$

Набор факторов x_j в каждом уравнении может варьироваться. Каждое уравнение системы независимых уравнений может рассматриваться самостоятельно. По существу, каждое уравнение этой системы является уравнением регрессии. Для определения параметров системы (4.1) используют метод наименьших квадратов.

Если зависимая переменная y одного уравнения выступает в виде фактора в другом уравнении, то исследователь может построить модель в виде системы *рекурсивных уравнений*:

$$
\begin{cases}
y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\
y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\
y_3 = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 + \ldots + a_{3n}x_n + \varepsilon_3, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
y_m = b_{m1}y_1 + \ldots. + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m.
\end{cases}
\tag{4.2}
$$

В данной системе в каждое последующее уравнение включаются в качестве факторов зависимые переменные предшествующих уравнений наряду с набором исходных факторов. Параметры системы (4.2) определяются методом наименьших квадратов.

Наибольшее распространение получили системы *взаимозависимых уравнений*, в которых одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других уравнениях — в правую часть системы:

$$
\begin{cases}
y_1 = b_{12}y_2 + b_{13}y_3 + \ldots. + b_{1m}y_m + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\
y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2m}y_m + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\
y_3 = b_{31}y_1 + b_{32}y_2 + \ldots + b_{3m}y_m + a_{31}x_1 + a_{32}x_2 + \ldots + a_{3n}x_n + \varepsilon_3, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
y_m = b_{m1}y_1 + b_{m2}y_2 + \ldots. + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m.
\end{cases}
\tag{4.3}
$$

Система (4.3) взаимозависимых уравнений получила название системы *совместных*, *одновременных уравнений*. В эконометрике эта система уравнений называется также *структурной формой модели*. В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим. Для этой цели используются специальные приемы оценивания.

4.1. Структурная и приведенная формы модели

Структурная форма модели содержит эндогенные и экзогенные переменные. Эндогенные переменные — это зависимые переменные, число которых равно числу уравнений в системе и которые обозначены через y. Экзогенные переменные — это предопределенные переменные, влияющие на эндогенные переменные, но не зависящие от них, обозначены x.

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. Внеэкономические переменные (например, климатические условия, социальное положение, пол, возраст и т.д.) входят в систему только как экзогенные переменные. В качестве экзогенных переменных могут рассматриваться значения эндогенных переменных за предшествующий период времени (лаговые переменные). Структурная форма модели позволяет увидеть влияние изменений любой экзогенной переменной на значения эндогенной переменной. Целесообразно в качестве экзогенных переменных выбирать такие переменные, которые могут быть объектом регулирования. Изменяя их и управляя ими, можно заранее иметь целевые значения эндогенных переменных.

Структурная форма модели (4.3) в правой части содержит при эндогенных переменных коэффициенты b_{ij} и экзогенных переменных — коэффициенты a_{ij}, которые называются *структурными коэффициентами модели*. Все переменные в модели должны быть выражены в отклонениях от среднего уровня, т.е. под x подразумевается $x-\bar{x}$, а под y соответственно $y-\bar{y}$. Поэтому свободный член в каждом уравнении системы (4.3) отсутствует.

Использование МНК для оценивания структурных коэффициентов модели приводит к получению смещенных и несостоятельных оценок, поэтому структурная форма модели преобразуется в *приведенную форму модели*. Так как препятствием к применению МНК является коррелированность эндогенных переменных со случайными членами, следует разрешить систему уравнений (4.3) относительно y так, чтобы в правых частях уравнений оставались только экзогенные переменные x.

Приведенная форма модели представляет собой систему линейных функций эндогенных переменных от экзогенных:

$$\begin{cases} y_1=\delta_{11}x_1+\delta_{12}x_2+\ldots+\delta_{1n}x_n+u_1, \\ y_2=\delta_{21}x_1+\delta_{22}x_2+\ldots+\delta_{2n}x_n+u_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m=\delta_{m1}x_1+\delta_{m2}x_2+\ldots+\delta_{mn}x_n+u_m, \end{cases} \tag{4.4}$$

где δ_{ij} — коэффициенты приведенной формы модели, u_i — остаточная величина для приведенной формы.

Коэффициенты приведенной формы модели представляют собой нелинейные функции коэффициентов структурной формы модели. Рассмотрим это положение на примере простейшей структурной модели.

Для структурной модели вида

$$\begin{cases} y_1=b_{12}y_2+a_{11}x_1+\varepsilon_1, \\ y_2=b_{21}y_1+a_{22}x_2+\varepsilon_2 \end{cases} \tag{4.5}$$

приведенная форма модели имеет вид

$$\begin{cases} y_1=\delta_{11}x_1+\delta_{12}x_2+u_1, \\ y_2=\delta_{21}x_1+\delta_{22}x_2+u_2. \end{cases} \tag{4.6}$$

Из первого уравнения (4.5) выразим y_2:

$$y_2=\frac{y_1-a_{11}x_1-\varepsilon_1}{b_{12}}=\frac{y_1}{b_{12}}-\frac{a_{11}}{b_{12}}x_1-\frac{\varepsilon_1}{b_{12}},$$

подставляя во второе уравнение (4.5), имеем

$$\frac{y_1}{b_{12}} - \frac{a_{11}}{b_{12}}x_1 - \frac{\varepsilon_1}{b_{12}} = b_{21}y_1 + a_{22}x_2 + \varepsilon_2,$$

откуда

$$y_1 = \frac{a_{11}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}b_{12}}{1 - b_{12}b_{21}}x_2 + \frac{\varepsilon_1 + b_{12}\varepsilon_2}{1 - b_{12}b_{21}}.$$

Если из второго уравнения (4.5) выразить y_1 и подставить в первое уравнение, то получим

$$y_2 = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}}{1 - b_{12}b_{21}}x_2 + \frac{b_{21}\varepsilon_1 + \varepsilon_2}{1 - b_{12}b_{21}}.$$

Таким образом, коэффициенты приведенной формы модели будут выражаться через коэффициенты структурной формы следующим образом:

$$\delta_{11} = \frac{a_{11}}{1 - b_{12}b_{21}}; \qquad \delta_{12} = \frac{a_{22}b_{12}}{1 - b_{12}b_{21}}; \qquad u_1 = \frac{\varepsilon_1 + b_{12}\varepsilon_2}{1 - b_{12}b_{21}};$$

$$\delta_{21} = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}; \qquad \delta_{22} = \frac{a_{22}}{1 - b_{12}b_{21}}; \qquad u_2 = \frac{b_{21}\varepsilon_1 + \varepsilon_2}{1 - b_{12}b_{21}}.$$

К системе уравнений (4.4) применяют обычный метод наименьших квадратов и получают оценки некоторых выражений от исходных параметров, из которых потом находят оценки и самих параметров. Такая процедура называется *косвенным методом наименьших квадратов*.

Следует отметить, что приведенная форма модели хотя и позволяет получить значения эндогенной переменной через значения экзогенных переменных, но аналитически она уступает структурной форме модели, так как в ней отсутствуют оценки взаимосвязи между эндогенными переменными.

4.2. Проблема индентификации

Экономический смысл и интерес для анализа представляют параметры структурной формы, так как именно структурная форма раскрывает экономический механизм формирования значений эндогенных переменных.

При переходе от приведенной формы модели к структурной приходиться сталкиваться с проблемой идентификации.

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

Структурная модель (4.3) в полном варианте содержит $m(m + n - 1)$ параметров, а соответствующая приведенная ей модель содержит mn параметров, т.е. структурная модель содержит большее число параметров, чем приведенная форма модели. Соответственно $m(m + n - 1)$ параметров структурной модели не могут быть однозначно определены через mn параметров приведенной формы модели.

Чтобы получить единственно возможное решение для структурной модели, необходимо предположить, что некоторые из структурных коэффициентов модели ввиду слабой взаимосвязи признаков с эндогенной переменной из левой части системы равны нулю. Тем самым уменьшается число структурных коэффициентов модели. Уменьшение числа структурных коэффициентов модели возможно и другим способом, например, приравниванием некоторых коэффициентов друг другу, т.е. при этом предполагается, что их взаимодействие на формируемую эндогенную переменную одинаково. На структурные коэффициенты могут накладываться ограничения. Например, ограничения вида: $b_{ij} + a_{ij} = 0$.

С позиции идентифицируемости структурные модели можно подразделить на три вида: идентифицируемые, неидентифицируемые, сверхидентифицируемые.

Модель *идентифицируема*, если все структурные её коэффициенты определяются однозначно, т.е. число параметров структурной модели равно числу параметров приведенной формы модели.

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

Модель *сверхидентифицируема*, если число приведенных коэффициентов больше числа структурных коэффициентов, т.е. можно получить несколько оценок одного структурного коэффициента. Сверхидентифицируемая модель в отличие от неидентифицируемой модели практически решаема, но требует для этого специальных методов вычисления параметров.

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых требуется проверить на идентификацию. Модель считается идентифицируемой, если каждое уравнение в системе идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение.

Чтобы уравнение системы было идентифицируемо, необходимо, чтобы число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного.

Если обозначить число эндогенных переменных в i-м уравнении системы через L, а число экзогенных (предопределенных) переменных, которые содержатся в системе, но не входят в данное уравнение, через H, то условие идентифицируемости модели может быть записано в виде следующего счетного правила:

1) при $H + 1 = L$ уравнение идентифицируемо;
2) при $H + 1 < L$ уравнение неидентифицируемо;
3) при $H + 1 > L$ уравнение сверхидентифицируемо.

Для оценки параметров структурной модели система должна быть идентифицируема или сверхидентифицируема.

Рассмотренное счетное правило отражает необходимое, но недостаточное условие идентификации. Более точно условия идентификации определяются, если накладывать ограничения на коэффициенты матриц параметров структурной модели. Уравнение идентифицируемо, если по отсутствующим в нем переменным (эндогенным и экзогенным) можно из коэффициентов при них в других уравнениях системы получить матрицу, определитель которой отличен от нуля, а ранг матрицы не меньше, чем число эндогенных переменных в системе без одного.

Целесообразность проверки условия идентификации модели через определитель матрицы коэффициентов, отсутствующих в данном уравнении. Но присутствующих в других, объясняется тем, что возможна ситуация, когда для каждого уравнения выполнено счетное правило, а названный определитель равен нулю. В этом случае выполняется только необходимое, но недостаточное условие идентификации.

В эконометрических моделях часто наряду с уравнениями, параметры которых должны быть статистически оценены, используются балансовые тождества переменных, коэффициенты при которых равны ±1. В этом случае, хотя само тождество и не требует проверки на идентификацию, так как коэффициенты при переменных в тождестве известны, но в проверке на идентификацию структурных уравнений системы тождества участвуют.

Пример 8. Пусть изучается модель вида:

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21}r_t + b_{22}I_{t-1} + \varepsilon_2, \\ r_t = a_3 + b_{31}Y_t + b_{32}M_t + \varepsilon_3, \\ Y_t = C_t + I_t + G_t, \end{cases} \tag{4.7}$$

где C — расходы на потребление, Y — совокупный доход, I — инвестиции, r — процентная ставка, M — денежная масса, G — государственные расходы; t и $t-1$ обозначают текущий и предыдущий периоды времени; ε_1, ε_2, ε_3 — случайные ошибки.

Первое уравнение системы (4.7) есть функция потребления, второе уравнение — функция инвестиций, третье уравнение — функция денежного рынка, четвертое уравнение — тождество дохода. Модель представляет собой систему одновременных уравнений. Проверим каждое уравнение системы на идентификацию.

Решение. Модель включает четыре эндогенные переменные C_t, Y_t, I_t, r_t и четыре предопределенные переменные (две экзогенные переменные M_t, G_t и две лаговые переменные $C_{t\text{-}1}$ и $I_{t\text{-}1}$). Проверим необходимое условие идентификации для каждого из уравнений системы.

Первое уравнение содержит две эндогенные переменные C_t, Y_t и одну предопределенную переменную $C_{t\text{-}1}$, т.е. $L = 2$, $H = 4 - 1 = 3$, так как

$H + 1 > L$, то уравнение сверхидентифицируемо.

Второе уравнение содержит две эндогенные переменные I_t, и r_t и одну лаговую $I_{t\text{-}1}$, $L = 2$, $H = 4 - 1 = 3$, так как $H + 1 > L$, то уравнение сверхидентифицируемо.

Третье уравнение включает две эндогенные переменные I_t,и r_t и одну экзогенную M_t, следовательно, сверхидентифицируемо.

Четвертое уравнение представляет собой тождество, параметры которого известны, поэтому необходимости в идентификации нет.

Проверим дл каждого уравнения достаточное условие идентификации. Для этого составим матрицу коэффициентов при переменных модели.

	C_t	I_t	r_t	Y_t	$C_{t\text{-}1}$	$I_{t\text{-}1}$	M_t	G_t
Первое уравнение	–1	0	0	b_{11}	b_{12}	0	0	0
Второе уравнение	0	–1	b_{21}	0	0	b_{22}	0	0
Третье уравнение	0	0	–1	b_{31}	0	0	b_{32}	0
Тождество	1	1	0	–1	0	0	0	1

В соответствии с достаточным условием идентификации модели ранг матрицы при переменных, не входящих в исследуемое уравнение, должен быть равен числу эндогенных переменных без одного.

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

	I_t	r_t	I_{t-1}	M_t	G_t
Второе уравнение	–1	b_{21}	b_{22}	0	0
Третье уравнение	0	–1	0	b_{32}	0
Тождество	1	0	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной матрицы третьего порядка не равен нулю:

$$\begin{vmatrix} b_{22} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{22} b_{32} \neq 0 .$$

Достаточное условие идентификации для данного уравнения выполняется.

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

	C_t	Y_t	C_{t-1}	M_t	G_t
Первое уравнение	–1	b_{11}	b_{12}	0	0
Третье уравнение	0	b_{31}	0	b_{32}	0
Тождество	1	–1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной матрицы третьего порядка не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12} b_{32} \neq 0 .$$

Достаточное условие идентификации для данного уравнения выполняется

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

	C_t	I_t	C_{t-1}	I_{t-1}	G_t
Первое уравнение	–1	0	b_{12}	0	0
Второе уравнение	0	–1	0	b_{22}	0
Тождество	1	1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной матрицы третьего порядка не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12} b_{22} \neq 0 .$$

Достаточное условие идентификации для данного уравнения выполняется.

Таким образом, все уравнения модели сверхидентифицируемы. Приведенная форма модели в общем виде будет выглядеть следующим образом:

$$\begin{cases} C_t = A_1 + \delta_{11}C_{t-1} + \delta_{12}I_{t-1} + \delta_{13}M_t + \delta_{14}G_t + u_1, \\ I_t = A_2 + \delta_{21}C_{t-1} + \delta_{22}I_{t-1} + \delta_{23}M_t + \delta_{24}G_t + u_2, \\ r_t = A_3 + \delta_{31}C_{t-1} + \delta_{32}I_{t-1} + \delta_{33}M_t + \delta_{34}G_t + u_3, \\ Y_t = A_4 + \delta_{41}C_{t-1} + \delta_{42}I_{t-1} + \delta_{43}M_t + \delta_{44}G_t + u_4. \end{cases}$$

4.3. Методы оценки параметров структурной формы модели

Коэффициенты структурной модели могут быть оценены разными способами в зависимости от вида системы одновременных уравнений. Наибольшее распространение получили следующие методы оценивания коэффициентов структурной модели:

1) косвенный метод наименьших квадратов (КМНК);

2) двухшаговый метод наименьших квадратов (ДМНК);

3) трехшаговый метод наименьших квадратов;

4) метод максимального правдоподобия с полной информацией;

5) метод максимального правдоподобия при ограниченной информации.

Косвенный метод наименьших квадратов применяется в случае точно *идентифицируемой* структурной модели. КМНК состоит в следующем:

1. составляют приведенную форму модели и определяют численные значения параметров для каждого её уравнения в отдельности с помощью обычного 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 простым является ДМНК.

Основная идея ДМНК состоит в том, чтобы на основе приведенной формы модели получить для сверхидентифицируемого уравнения теоретические значения эндогенных переменных, содержащихся в правой части уравнения. Далее, подставив их вместо фактических значений, можно применять обычный МНК к структурной форме сверхидентифицируемого уравнения. Метод получил название двухшагового МНК, так как дважды используется МНК: на первом шаге при определении приведенной формы модели и нахождении на её основе оценок теоретических значений эндогенной переменной $y_i = \delta_{i1}x_1 + \delta_{i2}x_2 + ... + \delta_{in}x_n$

и на втором шаге применительно к структурному сверхидентифицируемому уравнению при определении структурных коэффициентов модели по данным теоретических (расчетных) значений эндогенных переменных.

Сверхидентифицируемая структурная модель может быть двух типов: все уравнения системы сверхидентифицируемы; система содержит наряду со сверхидентифицируемыми точно идентифицируемые уравнения.

В первом случае для каждого уравнения используется ДМНК. Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

Для примера 8 необходимо применить именно ДМНК, но можно поступить иначе. Если из модели исключить тождество дохода, число эндогенных переменных модели снизится на единицу и переменная Y_t станет экзогенной, а число предопределенных переменных не изменится, так как из модели будет исключена эндогенная переменная G_t.. В правых частях функции потребления и функции денежного рынка будут находиться только предопределенные переменные. Функция инвестиций выражает зависимость эндогенной переменной I_t от эндогенной переменной r_t (которая зависит только от предопределенных переменных) и предопределенной переменной I_{t-1}. Таким образом, получилась рекурсивная система, параметры которой можно оценивать обычным МНК, и нет необходимости исследования системы на идентификацию.

Метод максимального правдоподобия рассматривается как наиболее общий метод оценивания, результаты которого при нормальном распределении признаков совпадают с МНК. Однако при большом числе уравнений системы этот метод приводит к достаточно сложным вычислительным процедурам. Поэтому используют модифицированный метод максимального правдоподобия при ограниченной информации (метод наименьшего дисперсионного отношения). В данном методе сняты ограничения на параметры, связанные с функционированием системы в целом. Это делает решение системы более простым, но трудоемкость вычислений остается достаточно высокой. Несмотря на его значительную популярность, к середине 60-х годов он был практически вытеснен ДМНК в связи с гораздо большей простотой последнего.

Дальнейшим развитием ДМНК является трехшаговый МНК (ТМНК). Этот метод оценивания пригоден для всех видов уравнений структурной модели. Однако при некоторых ограничениях на параметры более эффективным оказывается ДМНК. Подробнее обо всех методах можно прочитать в литературе [1-3].

Рассмотрим в качестве примера применение ДМНК к макроэкономической модели Кейнса.

Пример 9. Дана модифицированная модель Кейнса:

$$\begin{cases} C_t = a_1 + b_{11}Y_t + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + b_{22}Y_{t-1} + \varepsilon_2, \\ Y_t = C_t + I_t + G_t, \end{cases} \tag{4.8}$$

где Y — валовой национальный доход; C — личное потребление; I — инвестиции; G — государственные расходы; t и $t-1$ обозначают текущий и предыдущий периоды времени; ε_1, ε_2 — случайные ошибки.

Информация об уровнях всех показателей за двенадцать лет дана в табл. 4.1.

Таблица 4.1

Год наблюдения	C_t	I_t	Y_t	Y_{t-1}	G_t
1	1016,6	267	1412,7	—	486,1
2	1435,9	376	1978,9	1412,7	652,7
3	1776,1	408,8	2292	1978,9	839
4	2003,8	407,1	2514,4	2292	842,1
5	3265,7	670,4	4632	2514,4	1258
6	4476,9	1165,2	7116,6	4632	1960,1
7	5886,9	1504,7	8819,9	7116,6	2419,4
8	7443,2	1762,4	10627,5	8819,9	3422,3
9	9024,8	2186,4	12886,1	10627,5	3964,9
10	11401,4	2865	16679,9	12886,1	4669,7
11	14363,5	3611,1	21079,5	16679,9	6820,6
12	17742,6	4580,7	26009,7	21079,5	8375,2

Необходимо получить структурную форму модели.

Решение. В модели содержится три эндогенных переменных Y_t, C_t, I_t и две предопределенные переменные Y_{t-1} и G_t. Первое уравнение сверхидентифицируемо, а второе уравнение идентифицируемо (убедитесь в этом самостоятельно). Применим к каждому уравнению исходной системы обычный МНК и получим систему приведенных уравнений. Будем решать задачу в Excel и использовать режим «Регрессия». Результаты вычислений даны в табл. 4.2.

Таблица 4.2

	Первое уравнение	Второе уравнение	Третье уравнение
Входной интервал Y	Ссылка на столбец C_t	Ссылка на столбец I_t	Ссылка на столбец Y_t
Входной интервал X	Ссылка на столбцы Y_{t-1} и G_t	Ссылка на столбцы Y_{t-1} и G_t	Ссылка на столбцы Y_{t-1} и G_t
Полученные коэффициенты			
Y – пересечение	377,5173769	19,22987478	412,5083336
Y_{t-1}	0,581716738	0,15389729	0,817044334
G_t	0,63282491	0,155275844	1,037152085

Таким образом, система приведенных уравнений будет иметь вид:

$$\begin{cases} C_t = 377{,}5 + 0{,}582 Y_{t-1} + 0{,}632 G_t, \\ I_t = 19{,}3 + 0{,}154 Y_{t-1} + 0{,}155 G_t, \\ Y_t = 412{,}5 + 0{,}817 Y_{t-1} + 1{,}037 G_t. \end{cases} \quad (4.9)$$

Подставим данные наблюдений из табл. 4.1 в третье уравнение приведенной системы (4.9) и определим расчетные значения $\hat{Y}_t$, соответствующие эндогенной переменной Y_t. (табл. 4.3).

Таблица 4.3

Год наблюдения	Y_t	Y_{t-1}	$\hat{Y}_t$
1	1412,7	—	—
2	1978,9	1412,7	2243,70
3	2292	1978,9	2899,53
4	2514,4	2292	3158,56
5	4632	2514,4	3771,62
6	7116,6	4632	6229,98
7	8819,9	7116,6	8736,37
8	10627,5	8819,9	11168,20
9	12886,1	10627,5	13207,85
10	16679,9	12886,1	15784,21
11	21079,5	16679,9	21114,73
12	26009,7	21079,5	26321,75

Структурная форма модели будет иметь вид:

$$\begin{cases} C_t = a_1 + b_{11}\hat{Y}_t + \varepsilon_1, \\ I_t = a_2 + b_{12}\hat{Y}_t + b_{22}Y_{t-1} + \varepsilon_2. \end{cases} \quad (4.10)$$

Применим еще раз МНК к каждому уравнению системы 4.10, результаты использования режима «Регрессия» даны в табл. 4.4.

Таблица 4.4

	Первое уравнение	Второе уравнение
Входной интервал Y	Ссылка на столбец C_t	Ссылка на столбец I_t
Входной интервал X	Ссылка на столбец $\hat{Y}_t$	Ссылка на столбцы $\hat{Y}_t$ и Y_{t-1}
Полученные коэффициенты		
Y – пересечение	97,66181762	-42,52826142
Y_t	0,678200399	0,149713669
Y_{t-1}	—	0,031574585

Получим окончательный вид структурной модели:

$$\begin{cases} C_t = 97{,}66 + 0{,}678Y_t + \varepsilon_1, \\ I_t = -42{,}53 + 0{,}15Y_t + 0{,}031Y_{t-1} + \varepsilon_2. \end{cases} \quad (4.11)$$

Из уравнения (4.11) следует, что 67,8% прироста национального дохода идет на увеличение потребления. На увеличение инвестиций направляется соответственно 15% и 3,1% прироста национального дохода текущего года и предыдущего года.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Доугерти К. Введение в эконометрику: Пер. с англ. — М: ИНФРА-М, 2001—XIV, 402 с.

2. Елисеева И.И., Курышева С.В., Гордиенко Н.М. и др. Практикум по эконометрике. — М: Финансы и статистика, 2002, 192 с.

3. Кремер Н.Ш., Путко Б.А. Эконометрика. — М: ЮНИТИ-ДАНА, 2003, 311 с.

План выпуска учеб.-метод. документ. 2016 г., поз. 20

Публикуется в авторской редакции

Минимальные систем. требования:
PC 486 DX-33; Microsoft Windows XP; Internet Explorer 6.0; Adobe Reader 6.0

Подписано в свет 04.08.2016.
Гарнитура «Таймс». Уч.-изд. л. 2,3. Объем данных 1,6 Мбайт

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Волгоградский государственный архитектурно-строительный университет»
Редакционно-издательский отдел
400074, Волгоград, ул. Академическая, 1
http://www.vgasu.ru, info@vgasu.ru