ВЫСШЕЕ ОБРАЗОВАНИЕ

соответствует ФГОС

А.Н. ГЕРАСИМОВ,
Е.И. ГРОМОВ, Ю.С. СКРИПНИЧЕНКО

ЭКОНОМЕТРИКА

ФЕНИКС

Серия
«Высшее образование»

А. Н. ГЕРАСИМОВ, Е. И. ГРОМОВ,
Ю. С. СКРИПНИЧЕНКО

ЭКОНОМЕТРИКА

УЧЕБНОЕ ПОСОБИЕ

*Рекомендовано УМО РАЕ
по классическому университетскому
и техническому образованию в качестве учебного пособия
для студентов высших учебных заведений,
обучающихся по направлению подготовки
38.03.01 Экономика
(протокол № 528 от «10» августа 2015 г.)*

Ростов-на-Дону
«ФЕНИКС»
2017

УДК 330(075.8)
ББК656я73
КТК 0906
Г37

Рецензенты:
доктор экономических наук, профессор кафедры экономической теории и мировой экономики Северо-Кавказского федерального университета, заслуженный работник высшей школы РФ
А.В. Гладилин;
доктор экономических наук, профессор кафедры экономики и управления Адыгейского государственного университета
Е.Н. Захарова;
доктор технических наук, профессор кафедры математического моделирования и информатики Волгоградского государственного университета *А.Ф. Рогачев*

Герасимов А. Н.
Г37 Эконометрика : учебное пособие / А. Н. Герасимов, Е. И. Громов, Ю. С. Скрипниченко. — Ростов н/Д : Феникс, 2017. — 540, [1] с. — (Высшее образование).

ISBN 978-5-222-27576-4

Рассмотрены примеры решения задач с использованием основных приемов и методов эконометрического анализа социально-экономических явлений и процессов, особенности и порядок спецификации, параметризации, идентификации и верификации моделей парной и множественной регрессии. Отдельные разделы посвящены изучению взаимосвязей, анализу и моделированию временных рядов данных и системам одновременных уравнений. Содержит задачи для самостоятельного решения, вопросы для самоконтроля и тестовые задания.

Для бакалавров, магистров и аспирантов экономических специальностей, преподавателей, научных работников и специалистов аналитических служб.

УДК 330(075.8)
ББК656я73

ISBN 978-5-222-27576-4

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Эконометрика – относительно недавно появившаяся, но быстро развивающаяся научная дисциплина. Постоянно усложняющиеся экономические процессы ведут к необходимости создания и совершенствования методов их изучения и анализа. Современная наука все шире использует математический и статистический аппараты для своих исследований. На практике распространение получили количественный анализ и моделирование.

Сегодня эконометрика получила всемирное признание. Если в период плановой экономики в нашей стране упор делался на балансовые и оптимизационные методы и модели, то с переходом к рыночной системе эти модели выглядят архаичными, все более возрастает роль эконометрических методов. Только на их основе возможны исследование и теоретическое обобщение эмпирических зависимостей экономических переменных и построение качественного прогноза в бизнесе.

Основная задача эконометрики заключается в наполнении эмпирическим содержанием априорных экономических рассуждений. Следовательно, предметом эконометрики являются социально-экономические явления и процессы. Изучение экономических явлений в эконометрике осуществляется посредством специальных методов и моделей, позволяющих на базе экономической теории, социально-экономической статистики и математико-статистического инструментария придавать количественные выражения качественным зависимостям.

В научных исследованиях концептуальные и прикладные аспекты эконометрики зачастую отделяются друг от друга. Безусловно, в изучении социально-экономических процессов первичной является собственно экономическая составляющая эконометрики, поскольку экономика определяет постановку задачи и ключевые предпосылки. Математически выраженный результат интересен только с позиции его экономической интерпретации.

Во второй половине XX века широкому распространению и развитию эконометрических методов способствовало появление электронных вычислительных машин. Эконометрические пакеты для персональных компьютеров сделали методы более доступными, упростили расчеты всевозможных статистических характеристик, параметров, построение таблиц и графиков. Это позволило снизить трудоемкость работы эконометриста и сконцентрировать усилия на постановке задачи, выборе соответствующей модели, метода ее решения, интерпретации полученных результатов.

ГЛАВА 1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ЭКОНОМЕТРИЧЕСКОГО ИССЛЕДОВАНИЯ

1.1. Эконометрика как наука

Эконометрика — быстроразвивающаяся отрасль науки, целью которой является количественное измерение экономических отношений. Свидетельством всемирного признания эконометрики является присуждение Нобелевской премии Р. Фришу и Я. Тинбергу (1969 г.), Л. Клейну (1980 г.), Т. Хаавельмо (1989 г.), Дж. Хекману и Д. Макфаддену (2000 г.).

По мнению директора ЦЭМИ РАН академика В.Л. Макарова, современное экономическое образование держится на трех китах: макро-, микроэкономике и эконометрике. В условиях рыночной экономики возрастает роль эконометрических методов и теряет свою актуальность построение оптимизационных моделей отраслей и предприятий. Без знания эконометрических методов невозможны исследование и теоретическое обобщение эмпирических зависимостей переменных, а также построение надежного прогноза в банковском деле, финансах и бизнесе.

Слово «эконометрика» представляет собой комбинацию двух слов — «экономика» и «метрика». Сам термин был введен, по одним источникам, в 1926 году норвежским ученым Р. Фришем и подчеркивает специфику, содержание эконометрики как науки: количественное выражение связей и соотношений, раскрытых и обоснованных экономической теорией. По другим данным, термин «эконометрика» введен бухгалтером П. Цьемпой в 1910 году (Австро-Венгрия), причем у Цьемпы — «эконометрия».

Один из первых сторонников выделения этой новой дисциплины И. Шумпетер полагал, что в соответствии со своим назначением эта наука должна называться «эконометрика». Советский ученый А.Л. Вайнштейн считал, что название науки основывается на греческом слове «метрия» (геометрия, планиметрия и т. д.), соответственно по аналогии — эконометрия. Однако в мировой науке общепринятым стал термин «эконометрика».

Зарождение эконометрики является следствием междисциплинарного подхода к изучению экономики. Эта наука возникла в результате взаимодействия и объединения в особый «сплав» трех дисциплин: экономической теории, математической экономики и статистики (экономической и математической). Впоследствии к ним присоединилось развитие вычислительной техники как особое условие дисциплины «эконометрика».

В журнале «Эконометрика», основанном в 1933 году Р. Фришем, он дал следующее определение науке: «Эконометрика — это не то же самое, что экономическая статистика. Она не идентична и тому, что мы называем экономической теорией, хотя значительная часть этой теории носит количественный характер. Эконометрика не является синонимом приложений математики к экономике. Как показывает опыт, каждая из трех отправных точек — статистика, экономическая теория и математика — необходимое, но недостаточное условие для понимания количественных отношений в современной экономической жизни. Это — единство всех трех составляющих. И это единство образует эконометрику».

Таким образом, *эконометрика* — это наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов. Нельзя утверждать, что достигнуто однозначное определение эконометрики. Кроме того, есть мнение, что единое общепринятое определение эконометрики в настоящее время отсутствует.

Так, С. Фишер отмечал, что эконометрика — это раздел экономики, занимающийся разработкой и применением статистических методов для измерения взаимосвязей между экономическими переменными.

Э. Маленво придерживался широкого понимания, интерпретируя эконометрику как «любое приложение математики и статистических методов к изучению экономических явлений».

О. Ланге писал, что эконометрика занимается определением наблюдаемых в экономической жизни конкретных количественных закономерностей, применяя для этой цели статистические методы.

С.А. Айвазян полагает, что эконометрика объединяет совокупность методов и моделей, позволяющих на базе экономической теории, экономической статистики и математического инструментария придавать количественные выражения качественным зависимостям.

Основные результаты экономической теории носят качественный характер, а эконометрика вносит в них эмпирическое содержание. Математическая экономика выражает экономические законы в виде математических соотношений, а эконометрика осуществляет опытную проверку этих законов. Экономическая статистика дает информационное обеспечение исследуемого процесса в виде исходных статистических данных и экономических показателей, а эконометрика, используя традиционные математико-статистические и специально разработанные методы, проводит анализ количественных взаимосвязей между этими показателями.

Базовые понятия эконометрики имеют два определения — «экономическое» и «математическое». Подобная двойственность имеет место и в формулировках результатов от экономических работ, в которых почти не используется математический аппарат, до математических трудов, использующих аппарат современной математики.

Экономическая составляющая эконометрики является первичной. Именно экономика определяет постановку задачи и исходные предпосылки, а результат, формируемый на математическом языке, представляет интерес лишь в том случае, если удается его экономическая интерпретация.

Широкому внедрению экономических методов способствует развитие электронных вычислительных машин. Компьютерные эконометрические пакеты сделали более доступными и наглядными эти методы, так как наиболее трудоемкую работу по расчету различных статистик, параметров, характеристик, по построению таблиц и графиков в основном стал выполнять компьютер, а эконометристу осталась творческая работа: постановка задачи, выбор соответствующей модели и метода ее решения, интерпретация результатов.

1.2. Предмет эконометрики

Рыночная экономика требует улучшения использования статистической и экономической информации, характеризующей результаты хозяйственной деятельности. Создание информационной базы невозможно без учета действия различных факторов, формирующих финансовые результаты. Одновременно целесообразно выделить влияние факторов, которые зависят непосредственно от принятия управленческих решений данным субъектом хозяйствования, и влияние факторов, не зависящих от менеджмента (изменение цен, тарифов, налогов, инфляции и т. д.).

Элиминирование (устранение) влияния таких факторов в эконометрических моделях и характеристика их влияния путем соответствующих вычислений позволяют более правильно прогнозировать результаты исследуемых явлений в будущем периоде.

Эконометрические расчеты позволяют лучше понять хозяйственные процессы и явления, что помогает достоверно формулировать результаты и делать прогнозы. Более эффективное функционирование хозяйственной системы требует умелой и реальной экономической политики, а эффективная политика требует понимания различных взаимосвязей между факторами и результатами. При принятии управленческих решений необходимо знать систему этих взаимосвязей и соответствующим образом влиять на них.

Предметом исследования эконометрики можно назвать (так же как и для экономической теории) экономические явления. Но в отличие от экономической теории эконометрика акцентирует внимание на количественных, а не на качественных аспектах этих явлений. Например, экономическая теория утверждает, что спрос на товар с ростом его цены убывает, но при этом не отвечает на вопрос — как быстро, согласно какому условию и т. д. Эконометрика отвечает на эти вопросы применительно к каждому конкретному случаю.

Таким образом, *предметом эконометрики* являются факторы, формирующие развитие экономических явлений и процессов, а эконометрика — это искусство предвидения, разработки экономических прогнозов и гипотез.

Предпосылки, на которых основываются оценки факторов развития экономики, связаны с риском. Для уменьшения ошибок эконометристы должны решить, как включить в эконометрические расчеты все без исключения факторы и выбрать наиболее эффективные методы их оценки, обеспечивающие их достоверность.

1.3. Цели и задачи эконометрики

Устойчивое экономическое развитие (государства, отрасли или объекта хозяйствования) зависит от надежности и обоснованности управленческих решений. Обоснование процесса принятия решений является наиболее важной задачей эконометрики. Ход принятия управленческих решений должен учитывать их многовариантность, наличие неопределенности, оценку влияния факторов на отдельно взятый вариант и т. д. Выбор наилучшего варианта производится путем эконометрических расчетов. Для каждого из вариантов предпочтение отдается обеспечивающему наивысшую эффективность управленческих решений.

Эконометрика обеспечивает непрерывный процесс принятия управленческих решений, дающих возможность достигать намеченных целей. В этом контексте задачей являются выявление возможной цели менеджмента и определение ее фактического достижения при различных вариантах осуществления экономического процесса. Следовательно, задачей является оценка направленных действий на достижение и повышение экономической эффективности.

Кроме того, задача эконометрики — прогнозирование путей развития макро- и микроэкономических факторов. Прогнозная информация должна давать возможность принятия решений в зависимости от экономической конъюнктуры. Такие решения разрабатываются на основе статистических данных, обработанных и обобщенных эконометрическими методами.

Таким образом, ключевая задача эконометрики — наполнение эмпирическим содержанием априорных экономических рассуждений (количественная характеристика известных качественных экономических явлений).

Подобная трактовка задачи отвечает ключевой цели эконометрики — эмпирическому обобщению экономических законов.

С практической точки зрения к основным задачам эконометрики можно отнести:

- построение эконометрических моделей — представление эконометрических моделей в математической форме, удобной для проведения эмпирического анализа. Данную проблему называют проблемой *спецификации*, которую можно решить несколькими способами;
- оценку параметров построенной модели, позволяющую характеризовать адекватность модели реальными данными. Указанная задача решается на этапе *параметризации*;
- проверку качества полученных параметров модели в целом. Данная задача реализуется на этапе *верификации* модели;
- использование построенных моделей для объяснения поведения исследуемых экономических показателей, прогнозирования, осмысления экономических решений.

1.4. Критерии и принципы эконометрики

Успешное выполнение поставленных задач эконометрики зависит от следующих критериев и принципов эконометрических расчетов (рис. 1.1).

Выявление цели позволяет менеджменту выбрать возможные варианты действий. Выбор способов достижения (альтернатив) поставленной цели может быть осуществлен по ранжированию их пользы. Последовательное взвешивание затрат по отношению к их эффективности делает процесс выдвижения альтернатив наиболее важным критерием эконометрических расчетов, требующим формулирования исходных предпосылок, определения сферы и элементов эконометрических расчетов, сбора информации и разработки эконометрических гипотез, их оценки.

Рис. 1.1. Классификация критериев и принципов эконометрики

Ресурсы, необходимые для осуществления производственного процесса, имеют стоимостную оценку, а их истинная мера выражается возможностями реализации альтернативных действий. Эконометрические расчеты позволяют выявить последствия и результаты каждой из альтернатив, дать характеристику как уровню затрат и их эффективности, так и степени достижения целей.

Эконометрические расчеты нужно проводить постоянно, систематически проверяя их критерии от постановки проблемы, отбора цели, составления альтернативных действий, сбора данных, выбора метода их оценки и построения эконометрических прогнозов и моделей, взвешивания затрат по отношению к результатам, дополнительной проверки предпосылок, исходных данных, перепроверки целей, выявления новых альтернатив до построения улучшенных моделей.

Принцип правильной формулировки проблемы требует определения, насколько широко она поставлена, выяснения целей ее достижения и поиска методов и способов оценки альтернативных действий.

Принцип систематической направленности требует определения наиболее важных взаимосвязей между факторами и результатами, привлечения к выяснению проблем соответствующих специалистов (технологов, инженеров, социологов и т. д.).

Принцип учета неопределенности и его применение в эконометрических исследованиях необходимы для выявления неопределенных ситуаций и оценки их влияния на конечную эффективность. Наиболее сложным аспектом этого принципа является прогнозирование изменения управленческих решений относительно затрат и эффективности в зависимости от изменения экономических предпосылок и факторов развития экономики.

1.5. Основные этапы эконометрического моделирования

Реализация поставленных задач дисциплины требует разделения процедуры построения эконометрической модели на несколько взаимосвязанных этапов. Остановимся подробнее на каждом из них.

1-й этап — ***постановочный*** — формируются цель исследования, набор участвующих в модели экономических переменных. В качестве цели эконометрического моделирования обычно рассматриваются анализ изучаемого экономического процесса (объекта), прогноз его экономических показателей, предложение о возможном развитии явления при различных значениях экзогенных (независимых) переменных, выработка управленческих решений. При выборе переменных необходимо теоретически обосновать каждую (рекомендуется, чтобы их число было как минимум в пять раз меньше числа наблюдений). Объясняющие переменные не должны быть связаны функциональной или тесной корреляционной зависимостью, так как это сопряжено с невозможностью оценки параметров модели (явление *мультиколлинеарности*). Для отбора переменных может использоваться процедура пошагового отбора переменных, а для оценки влияния качественных признаков (пол, образование) — ***фиктивные*** переменные. Определяющим при включении в модель тех или иных переменных является экономический анализ исследуемого объекта.

2-й этап — ***априорный*** — проводятся анализ сущности изучаемого объекта, формирование и формализация априорной (известной заранее, до начала моделирования) информации.

3-й этап — ***информационный*** — осуществляется сбор необходимой статистической информации, значений экономических переменных. Здесь используются данные наблюдения, полученные в условиях активного (с участием исследователя) и пассивного (без участия эконометриста) эксперимента.

4-й этап — ***спецификация*** модели — в математической форме выражаются обнаруженные связи и соотношения, устанавливается состав экзогенных и эндогенных (соответственно внутренних, результативных) переменных; формируются исходные предпосылки и ограничения модели. От того, насколько точно выполнена задача спецификации, зависит успех эконометрического моделирования.

5-й этап — ***параметризация*** — оцениваются параметры (коэффициенты) выбранной зависимости. Эта оценка осуществляется на основе имеющихся статистических данных (обычно для получения количественных оценок используются методы регрессионного анализа).

6-й этап — ***идентификация*** — осуществляются статистический анализ модели и оценка ее параметров. При этом не следует путать проблему идентификации модели с ее *идентифицируемостью* (возможностью получения определенных параметров модели, заданной системой одновременных уравнений).

7-й этап — ***верификация*** — проводится проверка адекватности модели, выясняется, насколько удачно решены проблемы спецификации, идентификации, идентифицируемости модели, какова точность прогнозных расчетов по данной модели, насколько соответствует построенная модель реальному экономическому явлению. На этом этапе может совершенствоваться форма модели, уточняться состав переменных.

Выделенные этапы построения модели носят условный характер. Состав используемых процедур, приемов и методов, их очередность зависят от типа разрабатываемой эконометрической модели, особенностей исследуемых процессов. На наш взгляд, центральное место в эконометрическом исследовании занимает этап оценки параметров модели. По его результатам конкретизируется уравнение, полученные оценки параметров которого играют ведущую роль при проверке качества модели, обосновании направлений ее дальнейшей модификации.

Если модель удовлетворяет всем требованиям качества, то она может быть использована для прогнозирования или объяснения внутренних механизмов исследуемых процессов. Такая модель позволяет с высокой долей надежности предсказать среднее значение исследуемого экономического показателя на основе прогнозируемых или фиксируемых значений факторов, предвидеть вероятности отклонения конкретных значений изучаемой величины от предсказываемого по модели. Модель поможет определить, на какие факторы, в каком направлении и объеме следует воздействовать, чтобы значение исследуемого показателя лежало в определенных числовых границах.

Предлагаемая схема демонстрирует суть и последовательность эконометрических исследований (рис. 1.2).

Данная схема отражает циклическую природу эконометрического исследования: от экономической теории к моделированию; от моделирования к совершенствованию теории и более глубокому пониманию сути происходящих процессов; от понимания сути к осуществлению продуманной и целенаправленной экономической политики.

1.6. Общее представление о стохастических и детерминированных процессах

Изучение механизма рыночных связей, взаимодействие спроса и предложения, влияние объема и состава предложения на объем и структуру товарооборота, формирование товарных запасов, издержек обращения, прибыли и других качественных

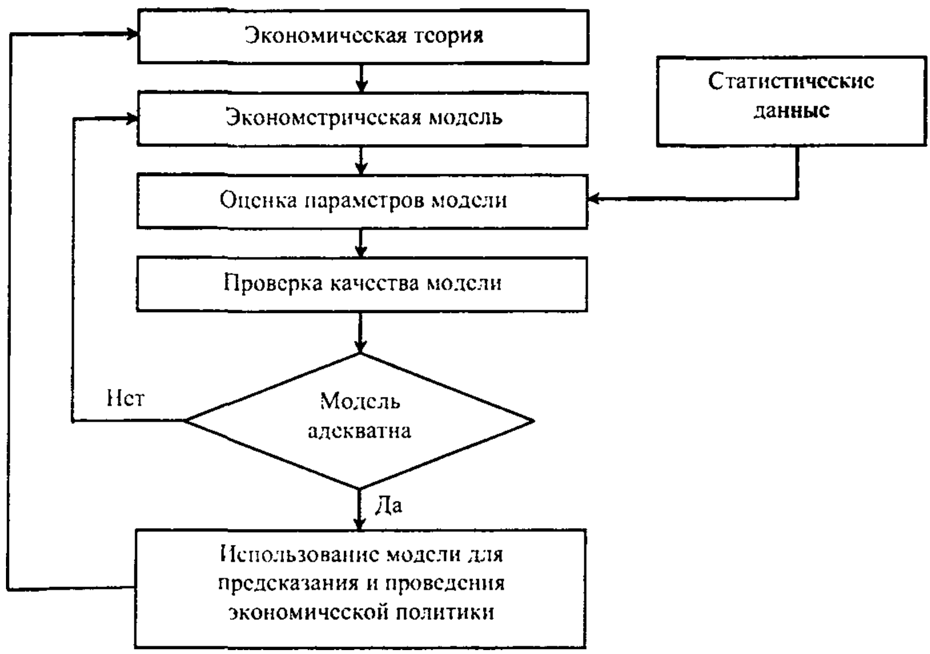


Рис. 1.2. Циклический характер эконометрического моделирования

показателей имеют первостепенное значение для прогнозирования конъюнктуры рынка. В целях научно обоснованного прогнозирования и рационального управления механизмом рыночных отношений важно выявленным связям придать математическую определенность. Без количественной оценки закономерности связи невозможно доводить результаты экономических разработок до такого уровня, чтобы они могли использоваться для практических целей.

В решении этих задач важная роль принадлежит эконометрике, в частности эконометрическим моделям, которые представляют собой совокупность уравнений, описывающих связи между некоторыми экономическими показателями. Соотношения между переменными могут носить *стохастический* и *детерминированный* характер.

Стохастика происходит от греческого слова *stochastikos* и означает «умеющий угадывать, проницательный», прогнозирование стохастических (случайных) процессов определяется как предсказание значения случайного процесса в некоторый будущий момент времени по наблюденным значениям этого процесса в прошлом и настоящем с некоторой степенью вероятности и описывается регрессионными уравнениями.

Термин *детерминированное* в переводе с латыни означает «предсказуемое». Детерминированные соотношения выражаются тождествами и не содержат случайных величин. Примером детерминированной связи может быть соотношение между валовым национальным доходом и валовыми доходами отдельных отраслей.

Для получения долгосрочных и качественных прогнозов возникает необходимость синтеза комбинированных *детерминированно-стохастических* моделей. Оценка объекта прогнозирования в связи с этим должна базироваться на сочетании аспектов детерминированности и неопределенности, где под детерминизмом будет пониматься философская концепция, признающая объективную закономерность и причинную обусловленность всех явлений природы и общества.

Детерминированные модели предполагают жесткие функциональные связи между переменными величинами модели.

При *функциональной связи* изменение результативного признака y всецело обусловлено действием факторного признака x:

$$Y = f(x). \tag{1.1}$$

(Примером функциональной связи является зависимость длины окружности l от радиуса r: $l = 2\pi r$.)

Стохастические модели допускают наличие случайных воздействий на исследуемые показатели. Такой вид зависимости называется *корреляционной*. При корреляционной связи изменение результативного признака y обусловлено влиянием факторного признака x не всецело, а лишь частично, так как возможно влияние прочих факторов Δ:

Наблюдаемое значение зависимой переменой	=	Объясненная часть, зависящая от значений объясняющих переменных	+	Случайная составляющая
Y	$=$	$f(x)$	$+$	Δ,

где Δ — случайная величина, называемая ***возмущением***, или *ошибкой*.

Сформулировать задачу моделирования самым общим образом можно так: на основании экспериментальных данных определить объясненную часть и, рассматривая случайную составляющую как случайную величину, получить оценки параметров ее распределения.

Если в стохастической системе удается строить прогнозы, то это только означает, что нащупана детерминированная структура, суть или стержень системы или процесса, и эта суть прогнозируется (экстраполируется «назад» или «вперед») в море случайностей (временных, пространственных или структурных). Таким образом, прогнозирование базируется на неявном детерминизме существа объекта или процесса, «прорывающегося» сквозь случайно проявляющиеся и измеряемые параметры.

1.7. Методы прогнозирования: интуитивный, формализованный

Под *интуитивными методами* прогнозирования понимают методы экспертных оценок с научно обоснованной организацией проведения всех этапов экспертизы и применением количественных методов при оценке суждений экспертов и при формальной групповой обработке результатов.

К *формализованным методам* прогнозирования, относятся методы экстраполяции и методы математического моделирования.

Интуитивные методы базируются на интуитивно-логическом мышлении человека в сочетании с количественными методами оценки и обработки полученных результатов. Различают: а) методы индивидуальных экспертных оценок: метод интервью, аналитический метод, метод написания сценария; б) методы коллективных экспертных оценок: метод «комиссий», метод «Дельфи», метод «коллективной генерации идей»; в) методы морфологического анализа.

Формализованные методы базируются на математической теории, которая обеспечивает повышение достоверности и точности прогнозов, значительно сокращает сроки их выполнения, позволяет автоматизировать обработку информации и оценку результатов. Различают: а) методы экстраполяции; б) методы математического моделирования.

При проведении эконометрических исследований используются формализованные методы прогнозирования (в частности методы экстраполяции и математического моделирования). Под *экстраполяцией* понимается метод научного исследования, заключающийся в распространении выводов, полученных из наблюдений над одной частью явления, на другую его часть. Или в более узком смысле — это нахождение по ряду данных временного ряда других его значений, находящихся вне этого ряда.

Экстраполяция заключается в изучении сложившихся в прошлом и настоящем устойчивых тенденций экономического развития и перенесении их на будущее. В прогнозировании экстраполяция (экстраполирование) реализуется при работе с временными рядами и представляет собой нахождение значений функции за пределами ее области определения с использованием информации о поведении данной функции в некоторых точках, принадлежащих области ее определения. Формальная экстраполяция базируется на продолжении и сохранении в будущем прошлых и настоящих тенденций развития экономического объекта.

Различают перспективную и ретроспективную экстраполяцию. *Перспективная* экстраполяция предполагает продолжение уровней ряда динамики (временных рядов) на будущее на основе выявленной закономерности изменений уровней в изучаемом отрезке времени. *Ретроспективная* экстраполяция характеризуется продолжением уровней ряда динамики (временных рядов) в прошлое.

Понятие экстраполяции неразрывно связано с *интерполяцией* (интерполирование), применяющейся, как правило, на этапе предварительной обработки данных при отсутствии информации для отдельных интервалов времени. Методы интерполяции позволяют определить промежуточные значения функции внутри области ее определения (либо интерполяция – нахождение таблично заданной функции внутри заданного интервала, где она не задана).

Для повышения точности экстраполяции используются различные приемы. Например, экстраполируемая часть общей кривой развития (тренд) корректируется с учетом реального опыта функционирования отраслей — аналогов исследований или объектов, отражающих в своем развитии прогнозируемый объект.

Метод экстраполяции применяется при стабильности экономической системы, устойчивости явлений, когда динамика процессов, показателей в перспективе определяется тенденциями их изменения в прошедшем периоде. Прогноз становится проекцией прошлого в будущем. В экстраполяции требуется анализировать временные ряды; обрабатывать ретроспективный ряд, применяя сглаживание и выравнивание; искать коэффициенты, минимизирующие отклонения; использовать методы скользящего и экспоненциального сглаживания.

В эконометрическом исследовании особое внимание уделяется тренду («тенденции развития»). При разработке модели тренд оказывается основой прогнозируемого временного ряда, на который накладываются другие составляющие. *Тренд* — это длительная тенденция изменения стохастического процесса, определяющая общее направление развития, основную тенденцию изменения экономических показателей, их временных рядов, характеристика основной закономерности движения во времени, в некоторой мере свободной от случайных воздействий.

Модели тренда могут различаться по виду. Их выбор в каждом конкретном случае осуществляется в соответствии с рядом статистических критериев. Наибольшее распространение в практических исследованиях получили следующие функции: линейная, квадратическая, степенная (частные случаи степенного полинома), показательная, экспоненциальная, логистическая (частные случаи экспоненциального многочлена).

Контрольные вопросы

1. Эконометрика как наука.
2. Предмет эконометрики.
3. Цели эконометрики.
4. Задачи эконометрики.
5. Критерии и принципы эконометрики.
6. Этапы эконометрического моделирования.
7. Общее представление о детерминированных и стохастических процессах.
8. Общее понятие эконометрических моделей, их типы.
9. Модели временных рядов.
10. Регрессия модели с одним уравнением.

ТЕСТОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ ПРОВЕРКИ ЗНАНИЙ

1. Эконометрика — это:

а) наука, которая дает количественное выражение взаимосвязей в экономике;

б) учение о системе показателей, дающих представление об экономике;

в) различного рода цифровые данные.

2. Предметом эконометрики является:

а) сбор цифровых данных;

б) определение наблюдаемых в экономике количественных закономерностей;

в) изучение экономических законов.

3. Эконометрическая модель описывает:
а) состав переменных;
б) функциональные связи между переменными;
в) набор цифровых данных;
г) стохастические связи между переменными.

4. Переменные, определяемые из уравнений модели, называются:
а) независимые;
б) зависимые;
в) предопределенные.

5. Переменные, задаваемые «извне», в определенной степени управляемые (планируемые), называются:
а) предопределенные;
б) эндогенные;
в) экзогенные.

6. Пространственные данные фиксируются:
а) по одному объекту за период времени;
б) в один и тот же момент времени по нескольким объектам;
в) по нескольким объектам за период времени.

7. Идентификация модели — это:
а) статистическое оценивание неизвестных параметров модели;
б) сбор необходимой статистической информации;
в) статистическая оценка параметров и модели в целом;
г) проверка точности модельных данных.

8. Нулевой называется гипотеза:
а) которая отклоняется;
б) подвергающаяся проверке;
в) которая содержит одно конкретное предположение.

9. Альтернативной называется гипотеза:
а) необходимая для проверки нулевой гипотезы;
б) которая отклоняется;
в) которая содержит несколько конкретных предположений.

10. Уровнем значимости называется:
а) совокупность значений критерия проверки, при которых нулевую гипотезу не отклоняют;
б) совокупность значений критерия проверки, при которых нулевую гипотезу отклоняют;
в) вероятность отвергнуть правильную нулевую гипотезу.

11. Задача спецификации модели не включает определение:
а) вида зависимости;
б) набора экзогенных и эндогенных переменных;
в) значений параметров модели;
г) состава системы уравнений и их структуру.

12. К проблеме построения эконометрической модели не относится:
а) отбор существенных факторов, включаемых в уравнение регрессии;
б) разработка программного обеспечения для реализации эконометрических методов;
в) выбор метода оценки параметров модели;
г) определение спецификации модели.

13. Проверка адекватности эконометрической модели эмпирическим данным проводится на этапе:
а) линеаризации;
б) параметризации;
в) спецификации;
г) верификации.

14. Экономическая статистика используется эконометрикой в качестве:

а) информационного обеспечения необходимыми исходными данными;
б) экономического обоснования полученных результатов эконометрического моделирования;
в) прикладной дисциплины для обеспечения проведения автоматизированных эконометрических расчетов;
г) математического инструментария.

15. Эконометрические модели являются:

а) структурными;
б) оптимизационными;
в) стохастическими;
г) нормативными.

ПРАКТИКУМ

Задача 1.1. По данным таблицы 1.1 предположите наличие связи между признаками и определите ее форму, используя графический метод.

Таблица 1.1

Данные о стоимости основных производственных фондов и продукции сельского хозяйства в разрезе районов Ставропольского края, млн руб.

№ п/п	Район	Стоимость основных фондов	Продукция сельского хозяйства
		X	*Y*
1	Александровский	1195,1	331,9
2	Андроповский	439,3	1919
3	Апанасенковский	1857,8	2195,9
4	Арзгирский	1612	1296
5	Благодарненский	786,3	2425,3

Окончание табл. 1.1

№ п/п	Район	Стоимость основных фондов	Продукция сельского хозяйства
		X	*Y*
6	Буденновский	1415	3657
7	Георгиевский	3551	5768
8	Грачевский	1050,1	880
9	Изобильненский	1732	5018
10	Ипатовский	4669,1	4678,8
11	Кировский	1903	2814
12	Кочубеевский	6021,6	1253,1
13	Красногвардейский	4600.3	5874
14	Курский	1278,5	1902,1
15	Левокумский	863,3	1730
16	Минераловодский	818,8	2566,4
17	Нефтекумский	492,05	2323
18	Новоалександровский	3082,1	1873,5
19	Новоселицкий	1717,4	2323
20	Петровский	1157	760,9
21	Предгорный	1441	1247,5
22	Советский	2114,7	3955,7
23	Степновский	192,5	2685
24	Труновский	2048,4	3021,7
25	Туркменский	1151,7	2232
26	Шпаковский	3063.9	9259

Решение

а) *Инструменты графического анализа SPSSStatistics 20.*

Для реализации графического метода определения зависимости между исследуемыми переменными с помощью средств графического отображения программного продукта *SPSSStatistics 20* первоначально необходимо создать таблицу исходных данных, в которой укажем два ряда значений переменных x и y. Затем в меню «Графика» найдем инструмент «Конструктор диаграмм». В появившемся диалоговом окне необходимо выполнить следующие операции:

- в качестве типа диаграммы выбрать «Простая диаграмма рассеяния»;
- соотнести переменные x и y соответственно осям абсцисс и ординат (рис. 1.3), нажать «OK».

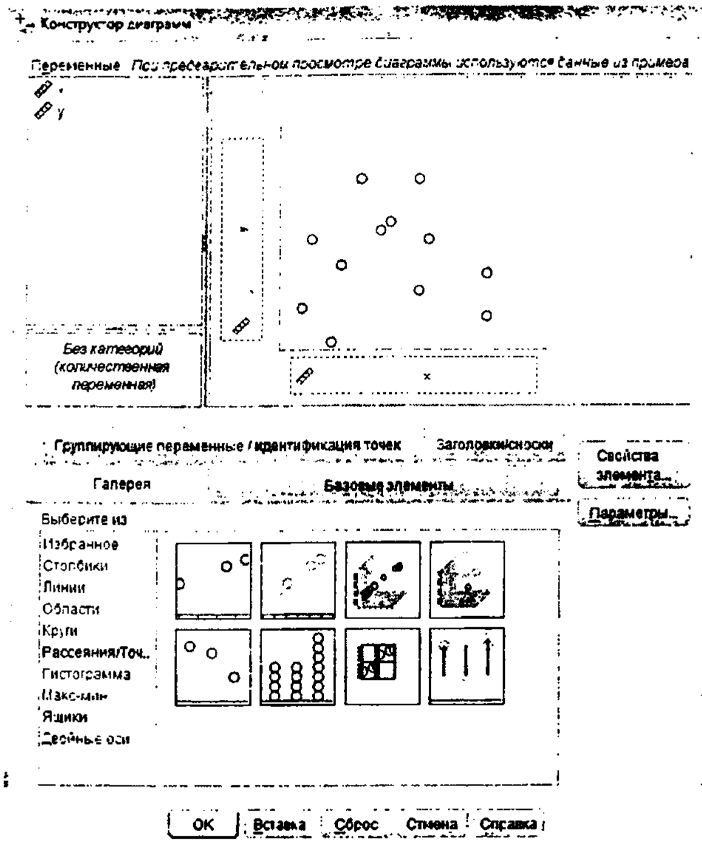


Рис. 1.3. Инструмент *SPSSStatistics 20* «Конструктор диаграмм»

В результате получим диаграмму рассеивания фактических значений исследуемых переменных, которая называется полем корреляции (рис. 1.4).

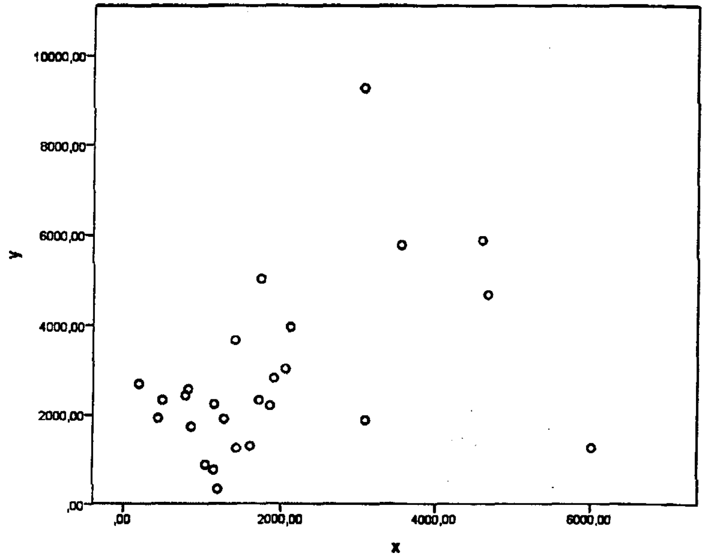


Рис. 1.4. Диаграмма рассеивания множества точек, выполненная средствами *SPSSStatistics 20*

б) *Инструменты графического анализа Statistica 6.0.*

Для построения диаграммы в *Statistica 6.0* необходимо создать таблицу исходных данных со значениями переменных x и y. Далее в меню «Графика» откроем инструмент «Диаграммы рассеяния» (рис. 1.5).

Выполним необходимые настройки:

- во вкладке «Переменные» соотнесем переменные x и y соответственно осям абсцисс и ординат;
- в поле «Подгонка» выключим линейную подгонку;
- нажмем «ОК».

В результате получим поле корреляции, аналогичное полю корреляции, построенному в *SPSSStatistics 20* (рис. 1.6).

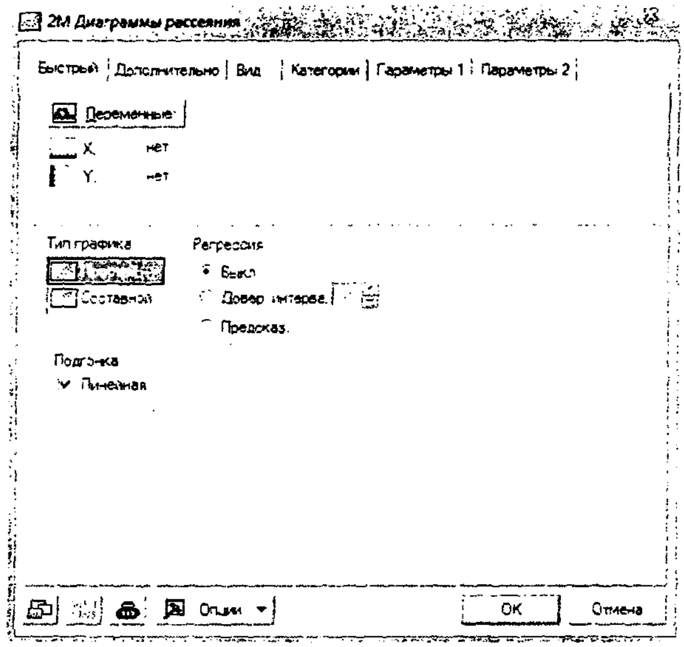


Рис. 1.5. Окно «Диаграммы рассеяния» в *Statistica* 6.0

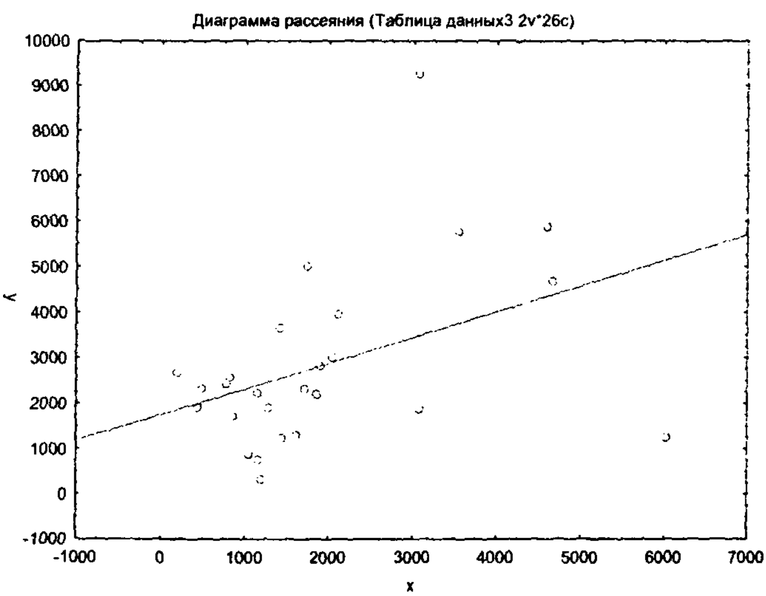


Рис. 1.6. Диаграмма рассеивания множества точек в *Statistica 6.0*

Таким образом, по расположению точек можно предположить наличие прямолинейной корреляционной связи. Однако для проверки данного предположения целесообразно было бы провести более углубленный анализ исследуемой связи, например при помощи эмпирической регрессии.

Задача 1.2. Имеются данные об урожайности зерновых культур (Y) в регионе и обеспеченности сельскохозяйственных организаций (СХО) зерноуборочными комбайнами (X). Необходимо определить вид корреляционной зависимости между рассматриваемыми признаками на основании эмпирической регрессии.

Таблица 1.2

Данные об урожайности зерновых культур (Y) в регионе и обеспеченности сельхозтоваропроизводителей (СХО) зерноуборочными комбайнами (X) в 2012 г.

№ п/п	Район	Урожайность зерновых культур в сельхозпредприятиях, ц/га	Приходится зерноуборочных комбайнов на 1000 га посевов зерновых, шт.
1	2	3	4
1	Александровский	23,9	2,8
2	Андроповский	13,5	3,9
3	Апанасенковский	19,6	3,4
4	Арзгирский	14,8	2,1
5	Благодарненский	19,1	4,4
6	Буденновский	17,3	2,9
7	Георгиевский	28,9	2,6
8	Грачевский	18,2	1,8
9	Изобильненский	28,7	4,2
10	Ипатовский	22,4	2,2
11	Кировский	26,5	7,2
12	Кочубеевский	44,4	2,7
13	Красногвардейский	31,8	4,0

Окончание табл. 1.2

1	2	3	4
14	Курский	23,5	5,3
15	Левокумский	11,0	6,0
16	Минераловодский	20,7	2,9
17	Нефтекумский	16,8	1,9
18	Новоалександровский	37,1	2,6
19	Новоселицкий	21,4	3,7
20	Петровский	20,5	3,8
21	Предгорный	25,6	2,1
22	Советский	27,2	2,9
23	Степновский	20,7	6,7
24	Труновский	29,4	3,4
25	Туркменский	11,7	2,5
26	Шпаковский	22,3	2,5

Решение

Эмпирическая регрессия строится по данным аналитической или комбинационной группировок и представляет собой зависимость групповых средних значений признака-результата от групповых средних значений признака-фактора. Поэтому на первоначальном этапе осуществим аналитическую группировку представленных данных в разрезе пяти групп. Для этого сначала упорядочим исходную совокупность по признаку-фактору:

Таблица 1.3

Ранжирование по признаку-фактору

№ п/п	Район	Приходится зерноуборочных комбайнов на 1000 га посевов зерновых, шт.	Урожайность зерновых культур в сельхозпредприятиях, ц/га	Номер интервала по признаку-результату
1	2	3	4	5
1	Грачевский	1,8	18,2	2
2	Нефтекумский	1,9	16,8	1

Окончание табл. 1.3

1	2	3	4	5
3	Арзгирский	2,1	14,8	1
4	Предгорный	2,1	25,6	3
5	Ипатовский	2,2	22,4	2
6	Туркменский	2,5	11,7	1
7	Шпаковский	2,5	22,3	2
8	Георгиевский	2,6	28,9	3
9	Новоалександровский	2,6	37,1	4
10	Кочубеевский	2,7	44,4	5
11	Александровский	2,8	23,9	2
12	Буденновский	2,9	17,3	1
13	Минераловодский	2,9	20,7	2
14	Советский	2,9	27,2	3
15	Апанасенковский	3,4	19,6	2
16	Труновский	3,4	29,4	3
17	Новоселицкий	3,7	21,4	2
18	Петровский	3,8	20,5	2
19	Андроповский	3,9	13,5	1
20	Красногвардейский	4	31,8	4
21	Изобильненский	4,2	28,7	3
22	Благодарненский	4,4	19,1	2
23	Курский	5,3	23,5	2
24	Левокумский	6	11	1
25	Степновский	6,7	20,7	2
26	Кировский	7,2	26,5	3

При использовании алгоритма равноинтервальных группировок были получены следующие интервалы:

– для признака-фактора 1,8–2,9,
2,9–4,0,
4,0–5,0,

5,0–6,1,
6,1–7,2;
– для признака-результата 11,0–17,7,
17,7–24,4,
24,4–31,0,
31,0–37,7,
37,7–44,4.

Результаты аналитической группировки представлены в таблице 1.4.

Таблица 1.4

Аналитическая группировка районов края по урожайности зерновых и обеспеченности зерноуборочными комбайнами

Группы	Приходится зерноуборочных комбайнов на 1000 га посевов зерновых, шт.	Урожайность зерновых культур в сельхозпредприятиях, ц/га					Всего районов
		11,0–17,7	17,7–24,4	24,4–31,0	31,0–37,7	37,7–44,4	
1	1,8–2,9	3	4	2	1	1	11
2	2,9–4,0	2	4	2			8
3	4,0–5,0		1	1	1		3
4	5,0–6,1	1	1				2
5	6,1–7,2		1	1			2
Итого		6	11	6	2	1	26

Групповые средние по признаку-фактору определяются как серединные значения интервалов, а по признаку-результату – по формуле средней арифметической взвешенной из серединных значений интервалов.

Таким образом, были получены следующие групповые средние.

Таблица 1.5

Групповые средние по признаку-фактору и признаку-результату

$\bar{X}$	2,35	3,45	4,5	5,55	6,65
$\bar{Y}$	23,5	21,04	27,7	17,7	24,4

В заключение отобразим в одной координатной плоскости поле корреляции по исходным данным и эмпирическую линию регрессии по групповым средним признака-фактора и признака-результата, как показано на рисунке 1.7.

Как видно на рисунке 1.7, количество точек линии эмпирической регрессии соответствует количеству групп, в разрезе которых произведена группировка исходной совокупности районов территории по признаку-фактору.

В соответствии с полученными результатами можно предположить наличие тесной, прямолинейной зависимости в регионе вариации урожайности от обеспеченности сельскохозяйственных предприятий уборочной техникой.

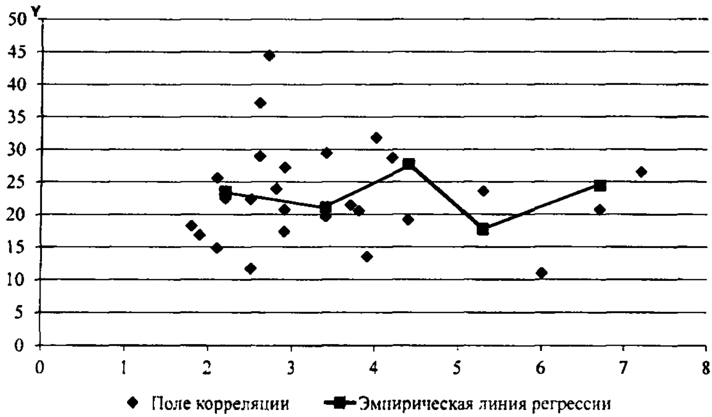


Рис. 1.7. Исходное поле корреляции и линия эмпирической регрессии зависимости урожайности зерновых в регионе от обеспеченности СХО техникой

Задача 1.3. Произведите механическое выравнивание динамического ряда, характеризующего изменение стоимости продукции сельского хозяйства за период 1993–2012 годов (в сопоставимой оценке), методом трехлетних укрупненных интервалов и трехлетней скользящей средней. Отобразите графически результаты выравнивания, сделайте выводы.

Таблица 1.6

Динамика стоимости продукции сельского хозяйства (*Y*) в регионе, млн руб.

Год	*Y*	Год	*Y*
1993	331,9	2003	953,8
1994	323,7	2004	1253,1
1995	859,2	2005	533,8
1996	614,9	2006	655,1
1997	587,8	2007	515,3
1998	931	2008	500,7
1999	201,6	2009	666,1
2000	384,6	2010	984,4
2001	814,1	2011	1030,0
2002	741,5	2012	982,9

Решение

Метод укрупненных интервалов основан на укрупнении периодов времени, к которым относятся уровни ряда (в нашем примере за три года), при этом одновременно уменьшается количество интервалов. Для каждого образованного таким образом периода рассчитывается свой показатель уровня ряда, обычно средняя величина.

Для этого в таблице 1.7 сначала была определена общая сумма продукции сельского хозяйства по каждому из трехлетних периодов (гр. 3), а затем среднее значение за три года (гр. 5).

Таблица 1.7

Расчетная таблица трехлетней скользящей средней, млн руб.

Год	Продукция сельского хозяйства	Трехлетние суммы	Трехлетние скользящие суммы	Трехлетняя средняя	Трехлетняя скользящая средняя
1	2	3	4	5	6
1993	331,9	–	–	–	–
1994	323,7	1514,80	1514,80	504,93	504,93
1995	859,2	–	1797,80	–	599,27

Окончание табл. 1.7

1	2	3	4	5	6
1996	614,9	–	2061,90	–	687,30
1997	587,8	2133,70	2133,70	711,23	711,23
1998	931	–	1720,40	–	573,47
1999	201,6	–	1517,20	–	505,73
2000	384,6	1400,30	1400,30	466,77	466,77
2001	814,1	–	1940,20	–	646,73
2002	741,5	–	2509,40	–	836,47
2003	953,8	2948,40	2948,40	982,80	982,80
2004	1253,1	–	2740,70	–	913,57
2005	533,8	–	2442,00	–	814,00
2006	655,1	1704,20	1704,20	568,07	568,07
2007	515,3	–	1671,10	–	557,03
2008	500,7	–	1682,10	–	560,70
2009	666,1	2151,2	2151,20	717,07	717,07
2010	984,4	–	2680,50	–	893,50
2011	1030,0	–	2997,30	–	999,10
2012	982,9	2012,9	–	1006,45	–

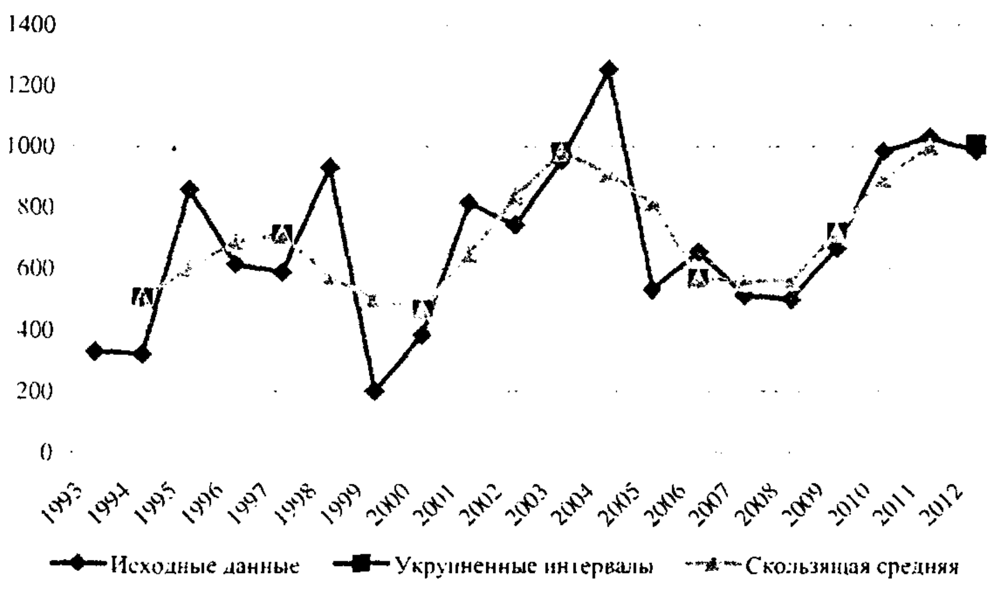


Рис. 1.8. Динамика стоимости продукции сельского хозяйства

Метод скользящей средней предполагает формирование укрупненных интервалов, состоящих из одинакового числа уровней. Каждый последующий интервал получаем, сдвигаясь на один уровень вниз. Трехлетние скользящие суммы приведены в графе 4, на основании их определены скользящие трехлетние средние (гр. 6).

Фактические значения анализируемого показателя и результаты выравнивания графически представлены на рисунке 1.8.

Задача 1.4. По данным таблицы 1.8 с использованием программного продукта *Statistica* 6.0 определите основные статистические характеристики представленных переменных. Проанализируйте полученные результаты. Сделайте выводы о характере распределения генеральной совокупности.

Таблица 1.8

Вариационный ряд значений трех переменных x_1, x_2, x_3, содержащий 45 наблюдений

x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
81,8	35,8	34	64,2	44,5	25,9	76,5	36	7,19
77,2	57,7	25,7	75,7	25,4	34,2	63,3	23,2	1,19
75,3	31,8	17,9	81,1	38,8	1,92	69,8	33,7	8,69
75,2	27,3	30,9	80,7	25,9	17,9	80,5	50,1	15,2
74,5	34,7	22,9	64,1	48	22,9	79,4	42,3	49,8
73,4	28,3	23,7	82,8	51,8	18,4	71,6	52,2	10,5
91,8	46,4	30,4	65,5	41,8	17,3	71,2	38	22,2
79,5	39,6	24,9	82,7	46,2	9,61	78,1	25,5	22,2
71,4	24,5	28,5	76,3	49	9,87	74,5	26,1	5,74
73,5	52	16,2	69	30,1	19,7	75,1	38,3	17,6
75,7	42	23,1	75	57,7	10,4	68,8	52,6	21
90,1	55,9	18,4	87,2	47,4	13,9	90,6	58,5	21,9
88,4	35,4	39,1	72,5	49,5	16,8	92,1	57,3	26
82,1	31,7	11	80,7	23,5	19,6	62,8	39	6,82
63,9	50,5	4,35	83,5	40,8	33,8	69,5	49,4	33,3

Решение

Откройте рабочее окно *Statistica 6.0.* В меню «Файл» выберите пункт «Создать». В появившемся диалоговом окне выберите подменю «Рабочая книга». В «Рабочей книге » создайте папку с названием «Задача 1.4».

Для создания таблицы данных в меню «Файл» выберите пункт «Создать» → «Таблица». В появившемся окне сделайте следующие отметки:

- число переменных: 3;
- число наблюдений: 45;
- формат отображения: числовой, с одним десятичным разрядом;
- префикс переменных: *x* (рис. 1.9).

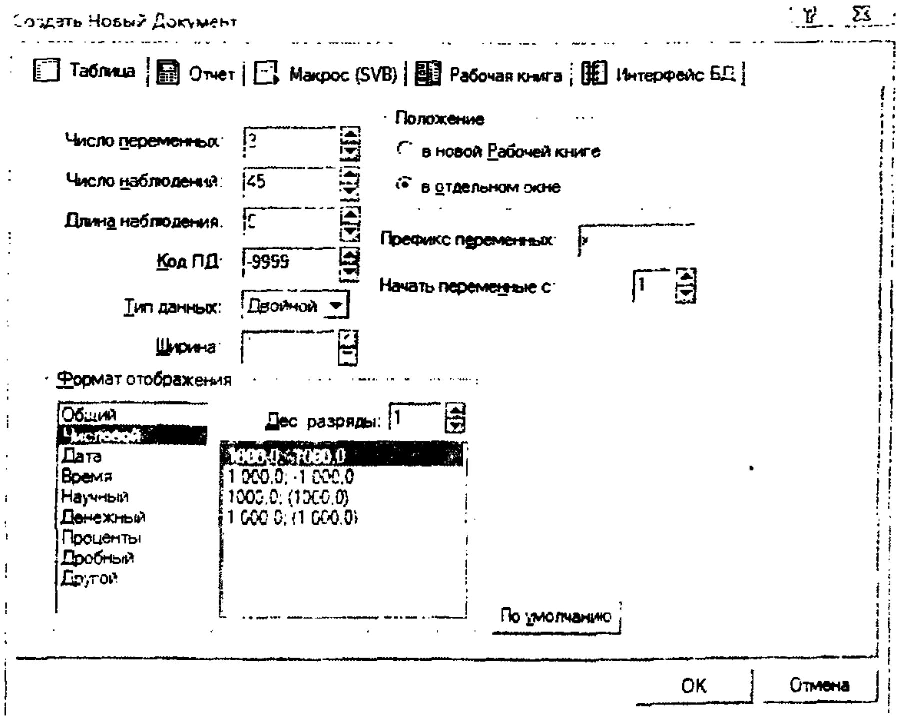


Рис. 1.9. Создание новой таблицы данных в *Statistica 6.0*

Созданную таблицу необходимо заполнить значениями из таблицы 1.8, а затем переместить ее в папку «Задача 1.4»: «Вставить» → «Документ STATISTICA» → «Из окна» → «Таблица 1». Рабочая книга с папкой «Задача 1.4», содержащей таблицу 1, представлена на рисунке 1.10:

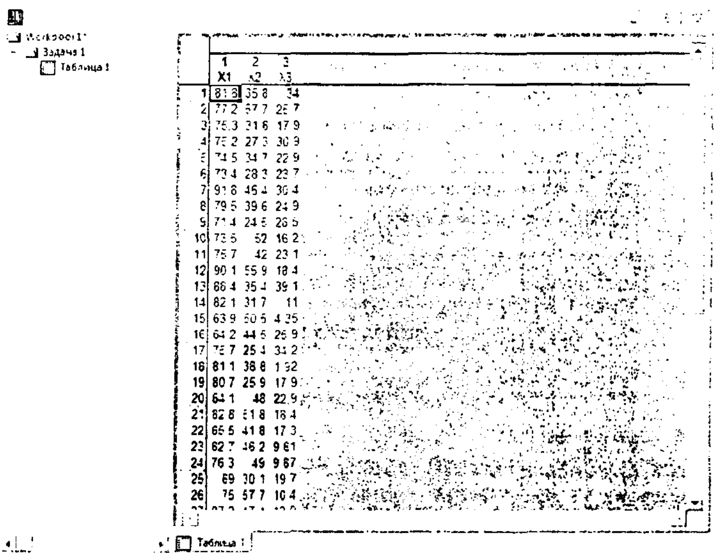


Рис. 1.10. Структура рабочей книги пакета *Statistica 6.0*

Основные статистические характеристики для исходной совокупности данных можно получить, используя модуль «Основные статистики и таблицы», находящийся в меню «Анализ» (рис. 1.11).

К основным статистикам, использующимся по умолчанию, относятся: число наблюдений N, среднее значение переменной, стандартное отклонение, минимум и максимум. Кроме того, значительный интерес для исследователя составляют такие по-

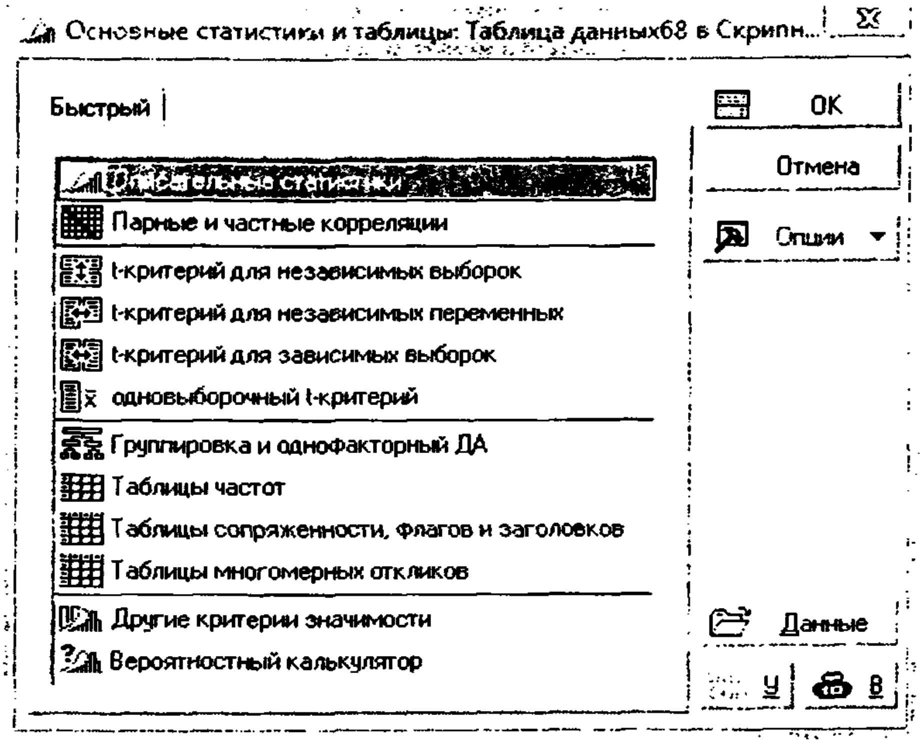


Рис. 1.11. Модуль «Основные статистики и таблицы» в *Statistica 6.0*

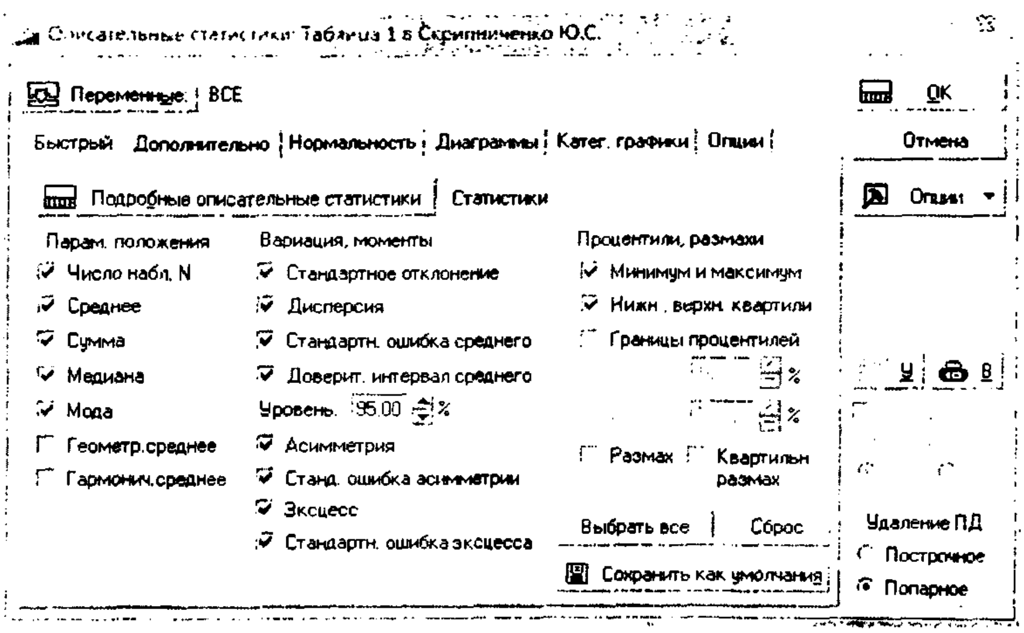


Рис. 1.12. Опции модуля «Описательные статистики» программы *Statistica 6.0*

казатели, как сумма, медиана, мода, дисперсия, стандартная ошибка среднего, асимметрия, эксцесс, нижняя и верхняя квартили. Отметим указанные статистики во вкладке «Дополнительно» (рис. 1.12).

Далее появится окно выбора переменных. В этом окне следует выбрать все переменные. В результате получим таблицу, содержащую выбранные статистические показатели исходной совокупности (рис. 1.13).

Описательные статистики [illegible]

Переменная	N набл.	Среднее	Медиана	Мода	Частота моды	Сумма	Минимум	Максимум	Нижняя квартиль	Верхняя квартиль	Дисперсия	Ст. откл.	Станд. ошибка	Асимметрия	Эксцесс
X1	45	[illegible]	[illegible]	Множест	2	3439	[illegible]	92.1	[illegible]	[illegible]	[illegible]	[illegible]	1.13	[illegible]	[illegible]
X2	45	[illegible]	40.6	[illegible]	2	[illegible]	23.2	56.5	31.6	49.5	[illegible]	[illegible]	[illegible]	[illegible]	-1.16
X3	45	[illegible]	[illegible]	Множест	2	[illegible]	[illegible]	45.8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Рис. 1.13. Таблица «Описательные статистики» в *Statistica 6.0*

Полученную таблицу результатов можно переместить в папку «Задача 1.4».

Остановимся подробнее на каждом показателе:

- *число наблюдений N* отражает численность единиц совокупности по каждому из признаков;
- *среднее значение* переменной исчисляется по формуле простой арифметической средней из всех значений переменных. Чем больше размер выборки, тем более надежна оценка среднего. Чем больше изменчивость данных (больше разброс), тем оценка менее надежна;
- *стандартное отклонение* — широко используемая мера вариабельности (изменчивости) данных, отражающая среднюю величину отклонения всех наблюдений от арифметической средней;
- *минимум и максимум* — минимальное и максимальное значения уровней исследуемого ряда данных;
- *сумма* — арифметическое значение суммы всех уровней ряда данных;
- *медиана* — значение, которое при ранжировании (сортировке значений ряда по порядку возрастания или убывания) разбивает выборку на две равные части. Половина

наблюдений лежит ниже медианы и половина наблюдений лежит выше медианы. Если число наблюдений в выборке нечетное, то медиана вычисляется как среднее двух средних значений;

- *мода* — это значение, наиболее часто встречающееся в выборке;
- *дисперсия* — средний квадрат отклонений индивидуальных значений признака от средней арифметической;
- *стандартная ошибка среднего* — это теоретическое стандартное отклонение всех средних выборки размера n, извлекаемое из совокупности и зависящее от совокупной дисперсии (сигма) и размера выборки (n);
- *асимметрия* является мерой несимметричности распределения. Если асимметрия отчетливо отличается от нуля, распределение асимметричное (если разделить распределение на две части по его среднему значению (или медиане), то распределение значений по двум сторонам от этой центральной точки будет неодинаковым), плотность нормального распределения симметрична (если разбить распределение пополам в точке среднего (или медианы), то распределение значений с двух сторон от этой центральной точки будет «зеркальным отображением» друг друга; нормальное распределение — пример симметричного распределения) относительно среднего;
- *эксцесс* — измеряет «пикообразность» распределения. Если эксцесс значимо отличим от нуля, то функция плотности либо имеет более закругленный пик, либо имеет более острый пик, чем пик плотности нормального распределения. Функция плотности нормального распределения имеет эксцесс, равный 0;
- ***нижняя и верхняя квартили*** равны 25-й и 75-й процентилям распределения (соответственно). 25-я процентиль переменной — это такое значение, ниже которого расположены 25% значений переменной; 75-я процентиль — это такое значение, ниже которого находятся 75% значений переменной.

Проанализируем полученные результаты для переменной x_1.

На основании полученных результатов можно заключить, что значения переменной сгруппированы вокруг средней 76,4, при этом оценка медианы равна 75,7. Поскольку оценка средней незначительно отличается от оценки медианы, можно предположить, что распределение случайной величины x_1 близко к симметричному. Для проверки этого предположения необходимо проанализировать коэффициент асимметрии.

Коэффициент асимметрии переменной равен 0,19. Коэффициент близок к нулю, следовательно, можно предположить, что распределение является симметричным.

Для характеристики вариации признака используем размах вариации, оценки дисперсии, среднего квадратического отклонения и квартильные отклонения. В нашем случае указан диапазон значений выборочной совокупности: минимальное зна-

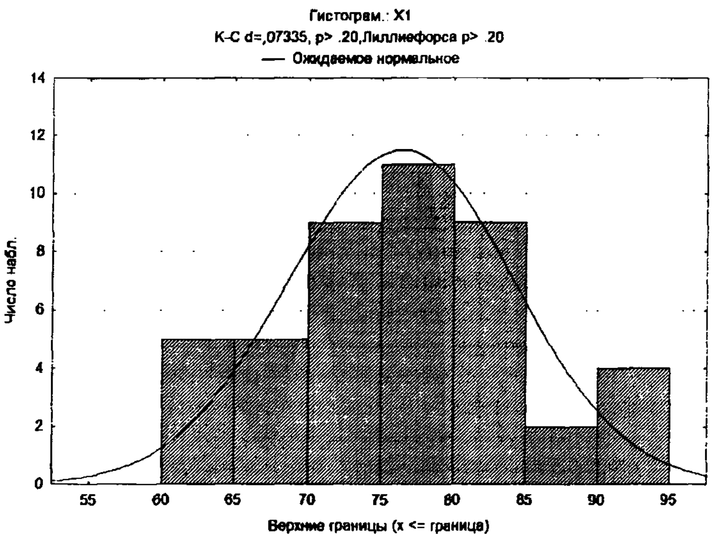


Рис. 1.14. Гистограмма распределения переменной x_1 в *Statistica 6.0*

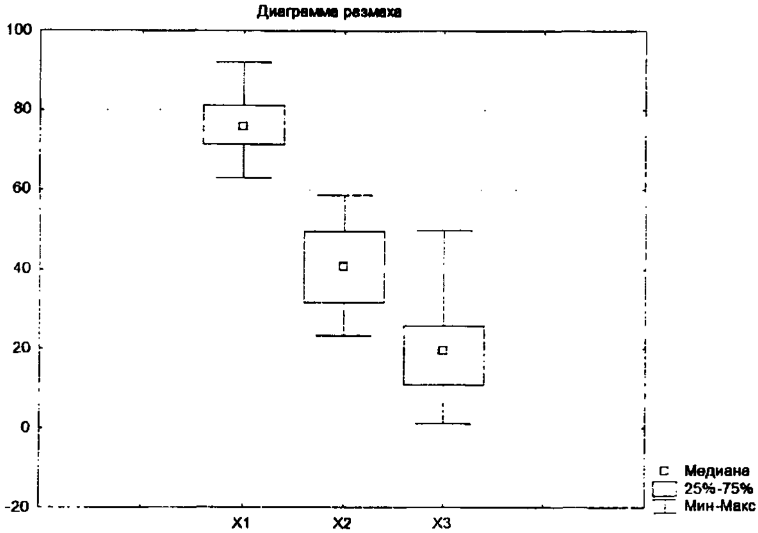


Рис. 1.15. Диаграмма размаха переменных *Statistica 6.0*

чение — 62,8, максимальное — 92,1, таким образом, размах вариации равен 92,1 – 62,8 = 29,3. Но оценка размаха вариации приемлема только для однородных совокупностей, поэтому целесообразнее использовать другие показатели, которые характеризуют отклонение значений признака отдельных единиц совокупности от средней величины. Для переменной x_1 оценка дисперсии равна 60,95, соответственно оценка среднего квадратического отклонения 7,8.

Аналогично определим основные статистические характеристики переменных x_2 и x_3.

Для графического изображения результатов анализа можно воспользоваться гистограммами распределения по переменным x_1, x_2 и x_3 (рис. 1.14, 1.15).

Задача 1.5. По данным таблицы 1.9 с использованием программного продукта *SPSSStatistics 20* определите основные статистические характеристики представленных переменных. Проанализируйте полученные результаты. Сделайте выводы о характере распределения генеральной совокупности.

Таблица 1.9

Вариационный ряд значений трех переменных x_1, x_2, x_3

x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
56,5	53,1	48,9	65,4	55,5	40,9	61,8	47,1	50,6
62,1	49.8	46,2	62.4	56,2	39,5	62,3	51.7	42
75,4	49,6	43,2	52,7	57,5	53,2	69,7	45,8	51
63,7	53.6	37,1	64,5	57,5	33,3	61,2	52,1	46,4
61,2	43,8	46.4	69.9	59	54	59,4	58.4	45,2
49,5	51,5	55,2	71,8	45,5	43,9	59.6	49,8	49,1
65,9	44,7	49,6	58,8	49,6	55,8	48,3	57,4	49,7
74,3	55	49,6	64,1	59,5	40,5	77.5	44,7	45,5
58,2	58,9	55,2	44,5	62,6	27,7	51,2	49,4	46,2
65,2	54,5	49,7	57,6	50,6	41,3	71,9	53.9	51
73,1	54,3	35,9	65,2	58,2	51,7	69	51,4	34,4
71,8	51,3	53.9	71,8	55,2	44,6	50,6	40,2	50,4
57,3	43,1	50.4	80,3	46,2	42,4	68.1	54,8	45,9
68,6	55,5	52,1	47,4	62,6	43,9	55,1	54,5	46,7
72,3	45,7	45,7	68,1	53,5	54,6	60,2	56,5	41

Решение

В окне ввода данных *SPSSStatistics 20*, используя информацию из таблицы 1.9, создайте таблицу исходных данных, содержащую 3 переменные и 45 наблюдений.

Откройте инструмент «Анализ» → «Описательные статистики» → «Описательные». В открывшемся диалоговом окне поместите имена переменных x_1, x_2, x_3 в поле «Переменные» (рис. 1.16).

Во вкладке «Параметры» необходимо указать следующие статистические характеристики:

- среднее;
- сумма;

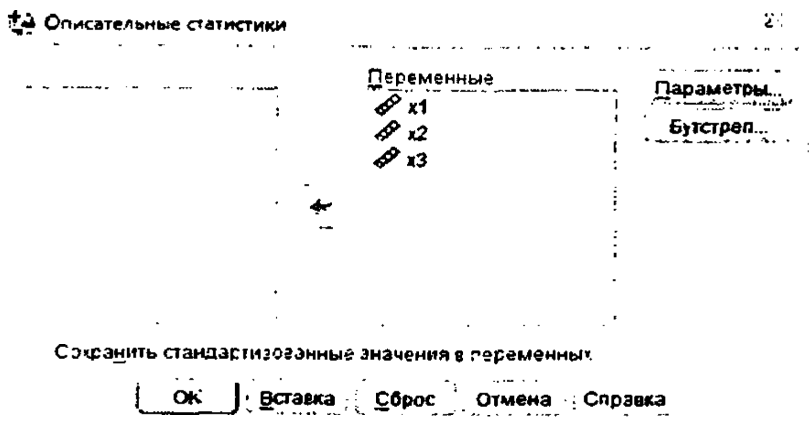


Рис. 1.16. Окно «Описательные статистики» в *SPSSStatistics 20*

- стандартное отклонение;
- дисперсия;
- размах;
- минимум;
- максимум;
- стандартная ошибка среднего;
- эксцесс;
- асимметрия (рис. 1.17).

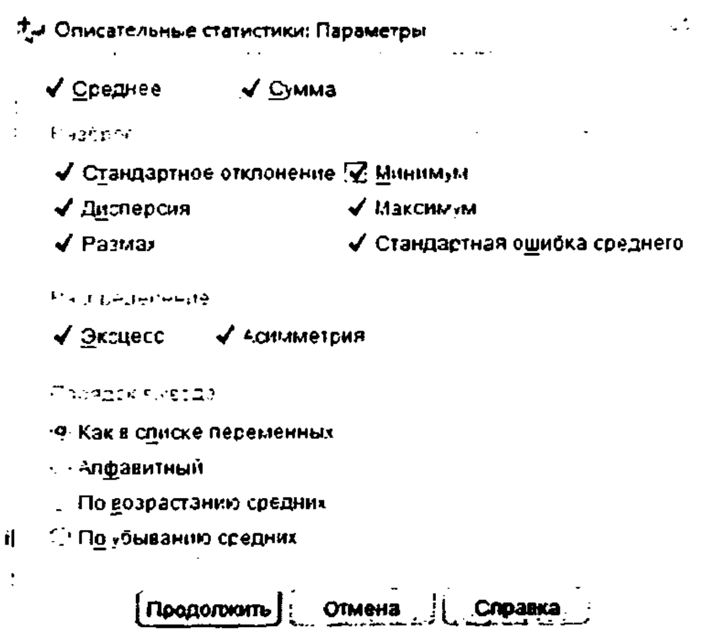


Рис. 1.17. Параметры окна «Описательные статистики» в *SPSS Statistics 20*

В окне вывода результатов найдем таблицу «Описательные статистики» (рис. 1.18).

Описательные статистики

Размах	Мин.	Макс.	Сумма	Среднее	Станд. отклонение	Дисперсия	Асимметрия	Эксцесс
35,80	44,50	80,30	2845,5	63,2333	8,47486	71,823	-,255	-,447
22,40	40,20	62,60	2361,3	52,4733	5,34039	28,520	-,274	-,509
28,10	27,70	55,80	2081,5	46,2556	6,34699	40,284	-,753	,492

Рис. 1.18. Описательные статистики в *SPSSStatistics 20*

Проанализируем полученные результаты для переменной x_1.

На основании полученных результатов можно сделать вывод, что значения переменной сгруппированы вокруг средней величины 63,23.

Коэффициент асимметрии переменной равен –0,26. Коэффициент отличается от нуля, следовательно, распределение не является симметричным.

Для характеристики вариации признака используем размах вариации, оценки дисперсии, среднего квадратического отклонения и квартильные отклонения. В нашем случае указан диапазон значений выборочной совокупности: минимальное значение – 44,5, максимальное значение – 80,3, размах вариации R = 35,8. Но оценка размаха вариации приемлема только для однородных совокупностей, поэтому целесообразнее использовать другие показатели, которые характеризуют отклонение значений признака отдельных единиц совокупности от средней величины. Для переменной x_1 оценка дисперсии равна 71,8, соответственно, оценка среднего квадратического отклонения 8,47.

Аналогично определим основные статистические характеристики переменных x_2 и x_3.

Для графического изображения результатов анализа можно воспользоваться гистограммами распределения по переменным x_1, x_2 и x_3. Откроем инструмент «Графика» → «Панель выбора диаграмм». В появившемся диалоговом окне выберем переменную x_1; укажем тип гистограммы: «Гистограмма с нормальной кривой». В результате получим график, изображенный на рисунке 1.19.

На этом рисунке изображена линия, характеризующая нормальное распределение значений исследуемой переменной. Так как гистограмма выходит за рамки данной линии, то можно считать, что в нашем случае значения переменной x_1 не подчиняются закону нормального распределения.

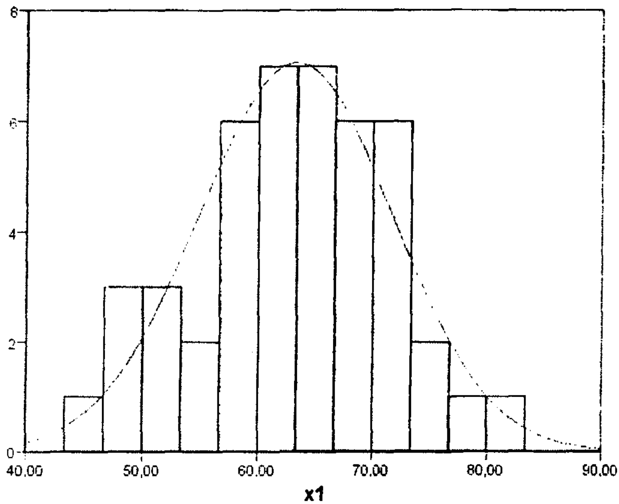


Рис. 1.19. Гистограмма распределения переменной x_1 в *SPSSStatistics 20*

Для построения диаграммы «Ящики» воспользуемся инструментом «Графика» → «Устаревшие диалоговые окна» → «Ящики». В настройках отметим пункты «Простые» и «Итоги по отдельным переменным» (рис. 1.20).

В диалоговом окне «Простые ящики» в поле «Ящики представляют» переместим набор исследуемых переменных, нажмем «ОК». В результате получим диаграммы «Ящики» для переменных x_1, x_2, x_3 (рис. 1.21).

Полученная диаграмма позволяет визуально определить степень размаха значений исследуемых переменных, смещение

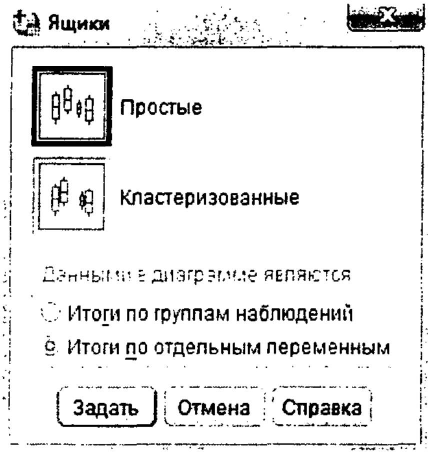


Рис. 1.20. Настройки диалогового окна «Ящики» в *SPSSStatistics 20*

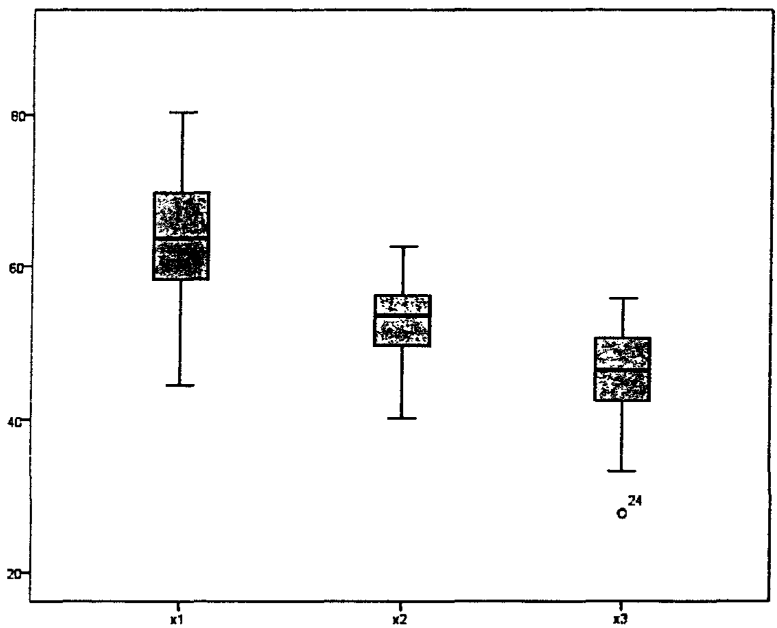


Рис. 1.21. Диаграмма размаха переменных «Ящики» в *SPSSStatistics 20*

средних значений относительно максимальных и минимальных значений, область сосредоточения основной совокупности значений переменных и степень разброса аномальных значений признаков в изучаемых совокупностях.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 1.6. Используя графический метод, по данным таблицы 1.10 предположите наличие связи между признаками и определите ее форму.

Таблица 1.10

Данные об индексе физического объема инвестиций и стоимости сельскохозяйственной продукции в 2012 г. в разрезе районов Ставропольского края

№ п/п	Район	Индекс физического объема инвестиций, в % к предыдущему году	Объем сельскохозяйственной продукции СХО, млн руб.
		X	Y
1	2	3	4
1	Александровский	200,0	1568,8
2	Андроповский	47,6	645
3	Апанасенковский	68,9	1271,2
4	Арзгирский	96,8	800
5	Благодарненский	26,4	1454,1
6	Буденновский	143,5	1678
7	Георгиевский	177,8	2912
8	Грачевский	49,6	830
9	Изобильненский	127,2	2298
10	Ипатовский	84,6	3101,8
11	Кировский	269,7	1363
12	Кочубеевский	68,5	5790,6

Окончание табл. 1.10

1	2	3	4
13	Красногвардейский	161,4	4721
14	Курский	98,5	1219,5
15	Левокумский	50,9	640
16	Минераловодский	130,7	1113,7
17	Нефтекумский	92,0	934
18	Новоалександровский	116,8	2432,1
19	Новоселицкий	156,7	1635
20	Петровский	208,1	2153
21	Предгорный	90,7	2168,3
22	Советский	192,2	2643
23	Степновский	85,9	599
24	Труновский	98,5	2166,1
25	Туркменский	84,3	980
26	Шпаковский	116,1	7142

Задача 1.7. Используя графический метод, по данным таблицы 1.11 предположите наличие связи между признаками и определите ее форму.

Таблица 1.11

Данные об уровне рентабельности и урожайности зерновых культур СХО в 2012 г. в разрезе районов Ставропольского края

№ п/п	Район	Уровень рентабельности деятельности сельхозпредприятий, %	Урожайность зерновых культур в сельхозпредприятиях, ц/га
		X	Y
1	2	3	4
1	Александровский	39,79	23,9
2	Андроповский	1,45	13,5

Окончание табл. 1.11

1	2	3	4
3	Апанасенковский	25,54	19,6
4	Арзгирский	17,59	14,8
5	Благодарненский	1,20	19,1
6	Буденновский	11,16	17,3
7	Георгиевский	31,78	28,9
8	Грачевский	5,24	18,2
9	Изобильненский	18,70	28,7
10	Ипатовский	14,45	22,4
11	Кировский	18,71	26,5
12	Кочубеевский	23,77	44,4
13	Красногвардейский	24,99	31,8
14	Курский	14,03	23,5
15	Левокумский	15,82	11
16	Минераловодский	10,67	20,7
17	Нефтекумский	29,73	16,8
18	Новоалександровский	22,03	37,1
19	Новоселицкий	29,76	21,4
20	Петровский	14,47	20,5
21	Предгорный	11,27	25,6
22	Советский	35,03	27,2
23	Степновский	24,63	20,7
24	Труновский	29,43	29,4
25	Туркменский	0,91	11,7
26	Шпаковский	11,10	22,3

Задача 1.8. Имеются данные об объеме продукции животноводства (X) в регионе и о фактическом потреблении мясных продуктов в расчете на 1 человека (Y). Необходимо определить вид корреляционной зависимости между рассматриваемыми признаками на основании эмпирической регрессии.

Таблица 1.12

Данные об объеме продукции животноводства (*X*) в регионе и о фактическом потреблении мясных продуктов в расчете на 1 человека (*Y*)

№ п/п	Район	Объем продукции животноводства, произведенной хозяйствами всех категорий всего, в фактических ценах, млн руб.	Фактическое потребление мясных продуктов на 1 человека в год, кг
1	Александровский	396,1	51,0
2	Андроповский	1112	51,5
3	Апанасенковский	1364,4	51,6
4	Арзгирский	336,9	51,3
5	Благодарненский	716,7	51,7
6	Буденновский	1243	51,8
7	Георгиевский	3043	55,8
8	Грачевский	76	51,8
9	Изобильненский	1641	51,9
10	Ипатовский	1584,2	51,6
11	Кировский	661	52,2
12	Кочубеевский	739,3	52,1
13	Красногвардейский	1565	51,9
14	Курский	1074,8	52,5
15	Левокумский	1089	51,9
16	Минераловодский	798,2	51,8
17	Нефтекумский	863	52,1
18	Новоалександровский	619,2	51,4
19	Новоселицкий	845	52,2
20	Петровский	1031,9	51,7
21	Предгорный	780,4	52,4
22	Советский	1039,6	51,7
23	Степновский	1197	51,9
24	Труновский	953,2	51,3
25	Туркменский	877	51,6
26	Шпаковский	8073	53,6

Задача 1.9. Имеются данные об уровне рентабельности сельхозпредприятий (X) и о размерах среднемесячной заработной платы работников (Y). Необходимо определить вид корреляционной зависимости между рассматриваемыми признаками на основании эмпирической регрессии.

Таблица 1.13

Данные об уровне рентабельности сельхозпредприятий (X) и о размерах среднемесячной заработной платы работников (Y)

№ п/п	Район	Среднемесячная начисленная заработная плата работающих, руб.	Уровень рентабельности деятельности сельхозпредприятий, %
1	2	3	4
1	Александровский	8843	39,79
2	Андроповский	8866	1,45
3	Апанасенковский	7938,3	25,54
4	Арзгирский	11089	17,59
5	Благодарненский	10413	1,20
6	Буденновский	11573	11,16
7	Георгиевский	13599	31,78
8	Грачевский	11636	5,24
9	Изобильненский	15168	18,70
10	Ипатовский	13712	14,45
11	Кировский	11398	18,71
12	Кочубеевский	17028,8	23,77
13	Красногвардейский	20416	24,99
14	Курский	3395	14,03
15	Левокумский	6903	15,82
16	Минераловодский	10343	10,67
17	Нефтекумский	9847	29,73
18	Новоалександровский	16581,64	22,03

Окончание табл. 1.13

1	2	3	4
19	Новоселицкий	12612	29,76
20	Петровский	10549	14,47
21	Предгорный	14274	11,27
22	Советский	18749	35,03
23	Степновский	4275	24,63
24	Труновский	16077	29,43
25	Туркменский	8945	0,91
26	Шпаковский	18360	11,10

Задача 1.10. Проведите механическое выравнивание динамического ряда для результативной переменной методом укрупнения интервалов и методом простой скользящей средней, основываясь на данных таблицы 1.14.

Таблица 1.14

Стоимость основных производственных фондов СХО региона за период 2001–2012 гг., млн руб.

Период	Стоимость основных производственных фондов
2001	40182
2002	37977
2003	34512
2004	25128
2005	24581
2006	21767
2007	20763
2008	19291
2009	18134
2010	18377
2011	18525
2012	20142

Задача 1.11. Проведите механическое выравнивание динамического ряда для результативной переменной методом укрупнения интервалов и методом простой скользящей средней, основываясь на данных таблицы 1.15.

Таблица 1.15

Валовая продукция СХО региона за период 2001–2012 гг., млн руб.

Период	Валовая продукция
2001	15605,0
2002	19636,0
2003	24828,0
2004	27068,0
2005	30174,8
2006	46280,1
2007	45393,2
2008	52159,3
2009	44580,2
2010	48174,5
2011	66846,8
2012	68524,2

Задача 1.12. Проведите анализ основных статистических характеристик следующей совокупности.

Таблица 1.16

Вариационный ряд значений переменных x_1, x_2, x_3

x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
44	83,7	45,9	57,5	98,5	45,3	50	94,4	44,6
67,5	83,8	61,4	61,7	98,2	43,9	55,1	86,1	49,2
60,4	88,7	66,6	64,5	97	55,2	52,4	84,1	50,4
57,3	87,6	50,3	34,6	100	45,9	48,9	85,1	53,5
63,5	86,7	59,6	51,7	89,6	54,2	45,3	88,2	47,8
61	86,3	57,3	67	85,1	46,1	42,5	80,6	60,8
53,9	100	44,3	65,9	83	45,6	71,1	83	36,2
42,8	87,9	36,1	55,2	83,5	53	55	92,4	48,7
51,2	77,9	41,6	67,4	88,6	37	50	97	56,8
59,9	82,5	50,9	57	90,8	55,9	63	94,5	61,6
68,6	83,2	39,4	51,2	98,5	36,9	52,4	93,5	50,6

Окончание табл. 1.16

x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
65.8	85,5	58.4	57,3	91,9	54.8	45,6	93,4	42.6
57,1	97	47,3	56	89,9	48,8	58,5	94.5	38
59.9	80,7	50.6	54,5	84,8	57	42,4	84,8	50,4
45,3	93,3	37	53,1	93,6	43,3	63,2	92,5	49.9

Задача 1.13. Осуществите анализ основных статистических характеристик вариационного ряда.

Таблица 1.17

Вариационный ряд значений переменных x_1, x_2, x_3

x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
59,1	92,4	60,5	63	98,6	61,3	54,7	99,4	64,4
60,6	100	67,9	58,1	99,3	68,1	63,2	98	59.9
48,1	102	62,4	55,7	94,7	64,6	56,2	100	65.8
64,7	98,1	69,5	57,5	101	65,4	55,1	95,4	60,6
59,4	96,5	66,8	68,9	100	67,2	71,1	91,4	61,1
61,7	91,5	65,9	59,1	91,7	63,1	60,5	94,5	70,1
53,4	90,7	64,3	56,4	95,1	62,5	64,1	96	68,5
47,1	90,3	69,7	61,1	98,1	66	49,4	89,2	67,9
53,4	101	63,2	63,1	102	65,3	62	95,2	65,8
56,5	100	65,8	56,3	94,4	60	54,1	95,8	64,2
66,1	94	70,4	52,5	92,1	60,6	56,3	91,1	64,6
59,2	90,6	64,1	53,5	95,2	62,6	65	90,3	70,2
54	93	61,1	62	91,5	69,2	71,7	102	68,7
61,8	94,4	59,9	63,3	94	65,5	65,2	89,2	61,3
54,4	93,3	63	57,6	100	66,6	58,8	93,2	64,4

Задача 1.14. Проведите анализ основных статистических характеристик следующей совокупности.

Таблица 1.18

Вариационный ряд значений переменных x_1, x_2, x_3

x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
102	68,5	75,1	110	73,1	72,4	76,7	59,5	69,7
103	76,9	71,6	115	66,9	67	87,3	59	80,1
114	71,1	56,9	102	62,8	65,2	114	61,7	67,6

Окончание табл. 1.18

x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
117	70,6	77,9	116	59,1	62,8	122	66,9	66,7
114	61	65,3	108	63,1	68,9	113	59,6	72,7
88,8	70,6	65,3	108	60,2	74,7	81,8	68	81,3
92,4	54,3	72,1	89,3	63,9	71,7	106	60,1	69,1
87,7	67,4	71,4	94	57,7	80,2	108	59,6	65,7
77,2	70,2	70,6	101	66,7	75,3	85,2	60,4	65,2
89,7	56,2	64,3	100	65,8	67,6	96,1	63	63,6
89,1	62,6	69	97,6	68,9	72,5	90,2	61,3	67,2
84,9	60,7	70,3	89,9	66,4	76,2	97,5	71,5	79,7
111	74,3	78,3	97,7	51,2	72	91,7	55,8	67,4
109	67,3	68,6	86,7	71,4	63,9	61,5	62,4	72
72,3	70,4	77,3	96,5	76,3	64,5	115	54,8	79,9

Задача 1.15. Осуществите анализ основных статистических характеристик вариационного ряда.

Таблица 1.19

Вариационный ряд значений переменных x_1, x_2, x_3

x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
56,5	29,2	79,6	60,9	33,3	86,9	62,6	18,2	85,1
62,3	9,4	92,7	71,1	40,4	79,4	71,3	5,63	88,4
63,9	13	90	52,2	17	89,3	68	23	91,2
62,3	19,9	89,2	67,9	30	82	61	4.52	90,2
61	9,89	84,5	54,8	10,5	78,7	65,3	16,8	91,5
55,9	17,6	81,4	73	37,6	91,7	65	25,5	79,3
53,6	1,8	78,1	62,6	8,92	90,1	67,7	1,43	88,9
65,8	13,3	78,5	66,7	0,725	81,8	68,9	11,5	86,2
70,9	11,7	85,2	61	18,8	82,8	66,9	8,37	91,9
64,1	33,5	89,2	67,7	4,46	88,7	62,7	14,4	80,6
59,8	35,5	80,1	67,8	0,898	92,3	59,1	57,8	83,2
67,5	19	84,9	65,7	12	90,9	62,3	7,95	84
59,4	0,957	79,1	62,1	4,64	80,8	68,9	19,2	80,4
57,8	9,35	89,3	62	1,55	87,1	73,9	6,17	92,4
67,5	14,1	80,9	61,5	3,08	90,8	65,6	35,7	81,6

ГЛАВА 2. ПОДГОТОВКА ЭМПИРИЧЕСКОЙ БАЗЫ ЭКОНОМЕТРИЧЕСКОГО ИССЛЕДОВАНИЯ

2.1. Формирование эмпирической базы исследования

Основой эконометрических моделей являются статистические данные. Содержание собираемой информации зависит от вида анализа и назначения модели, а объективную характеристику развития явления обеспечивают правильно подобранные статистические и математические методы.

Статистическая база для эконометрической модели может состоять из временных и пространственных (структурных) рядов данных.

Временным рядом называется ряд значений статистического показателя, упорядоченного по времени. В зависимости от длительности охватываемого периода временные ряды делятся на кратко- (до 1 года), средне- (1—3 года) и долгосрочные (свыше 5 лет).

Наиболее приемлемы следующие единицы измерения времени: месяцы, кварталы, годы.

Месячные и квартальные ряды данных можно разложить на три составляющие: тренд, сезонное движение и остаток. Трендовая составляющая выражает тенденцию развития показателя. Сезонное движение представляет собой строго чередующиеся отклонения от тренда, связанные с сезонными изменениями в экономике. Остаток содержит все другие колебания показателя, объясняемые действием множества факторов или возникающие совершенно случайно. Сезонное движение часто затрудняет анализ остатков, поэтому в эконометрических моделях обычно используются ряды данных, очищенных от сезонности.

Годовые временные ряды состоят из двух составляющих: тренда и остатка. Остаток в таких рядах может содержать циклическую компоненту (не обязательно периодическую).

Временные ряды при оценке эконометрической модели должны соответствовать требованию сравнимости, т. е. составляются в хронологической последовательности и являются соизмеримыми. Эконометрические модели базируются на моментных и интервальных рядах.

Пространственные (структурные) ряды представляют собой совокупность значений одного признака в разных экономических группах (структура производства по секторам, отраслям; распределение уровня потребления по семьям и др.). Пространственные модели позволяют изучить структурные соотношения.

Структурные ряды рассматриваются на макро- и микроуровне для исследования спроса и потребления, импорта и экспорта, а также в задачах, где необходимо исключить влияние изменений в ценах.

В случае, когда экономические явления рассматриваются как бы с двух сторон (исследование уровня жизни по группам семей во временных интервалах), динамический анализ соединяют со структурным, образуя так называемый перекрестный анализ данных.

В эконометрических моделях чаще всего используются временные ряды, а к структурному анализу данных обращаются как к источнику дополнительной информации.

Как отмечалось выше, эконометрические модели характеризуют явления макро- и микроэкономики. Микроэкономические ряды количественно описывают процессы на отдельных предприятиях. Макроэкономические ряды составляются на основе микроэкономических данных и характеризуют отдельные отрасли и хозяйство в целом. Следовательно, макроэкономические ряды — это результат агрегирования микроэкономических показателей. Однако *агрегирование* — не простое суммирование микроэкономических данных, а многоступенчатый процесс отбора, классификации и обработки статистической информации. Особое значение в этой связи приобретает создание интегрированной статистической информационной системы.

Методы агрегирования можно разделить на:

1. Иерархическое агрегирование — формирование головной организацией совокупного показателя, учитывающего данные

по каждому филиалу; пространственное укрупнение данных: район — субъект федерации — федеральный округ — федерация;

2. Вещественное агрегирование — задача состоит в объединении рядов, относящихся к определенной группе (урожайность зернобобовых культур включает в себя урожайность озимой пшеницы, ярового ячменя, гороха и т. д.). Цены при вещественном агрегировании используются в качестве весов;

3. Агрегирование во времени — формирование квартальных, годовых рядов на основе месячных данных показателя.

2.2. Предварительная обработка статистических данных

При подготовке информационной базы для эконометрической модели необходимо, чтобы временные ряды адекватно отражали сущность экономических связей и были взаимно согласованы. Эконометристу нужно обеспечить:

- соответствие временных рядов содержанию экономических показателей модели;
- согласованность единиц измерения;
- использование единой системы цен;
- форму выражения (абсолютные значения, индексы, темпы роста и прироста);
- общность интервалов времени (год, квартал и т. д.);
- соответствие степеней укрупнения (отрасль, вид товара и т. д.).

Выполнение этих условий может потребовать предварительной обработки данных:

- агрегирование и дезагрегирование имеющихся в распоряжении эконометриста статистических данных;
- экстраполяция и интерполяция рядов при отсутствии информации для отдельных интервалов времени;
- вычисление индексов, статистических показателей рядов динамики;
- формирование лаговых рядов.

Экспериментальная разработка регрессионных уравнений показывает, что абсолютные величины часто не могут быть использованы при оценивании коэффициентов в эконометрических моделях. Более точные оценки получаются при переходе от рядов абсолютных значений к темпам роста и прироста, индексам, средним и относительным величинам.

Абсолютные значения представляют собой разностное значение показателя в текущем и предыдущем (базисном) временных интервалах:

$$\Delta K_t = K_t - K_{t-1}, \qquad (2.1)$$

где K — изучаемый показатель за соответствующий период времени t.

Темпы роста представляют собой отношение абсолютных значений изучаемого показателя текущего периода времени к предшествующему или базисному:

$$T_p = K_t / K_{t-1} \cdot 100. \qquad (2.2)$$

Темпы прироста представляют собой отношение абсолютных приростов к абсолютному значению показателя в предыдущем временном интервале (базисном):

$$T_n = \Delta K_t / K_{t-1} \cdot 100. \qquad (2.3)$$

Использование абсолютных и относительных приростов позволяет исключить мультиколлинеарность независимых переменных и автокорреляцию остатков, тем самым облегчает задачу оценивания коэффициентов в регрессионных уравнениях. Однако это преобразование не является необходимым во всех случаях, поскольку затрудняет интерпретацию эконометрических моделей и часто характеризуется незначительной теснотой связи.

Относительные переменные представляют собой соотношение двух показателей. Обычно это отношение выражается в процентах. В качестве относительных переменных могут выступать показатели интенсивности, структуры, координации и т. д.

Средние значения характеризуют изучаемый показатель за определенный интервал времени, вычисляются на основе его моментных значений и экономически интерпретируют явление одной числовой величиной. Выделяют следующие виды средних: арифметическая, гармоническая, геометрическая и т. д.

Комбинированные переменные представляют собой данные, полученные в результате различных комбинаций нескольких переменных. Типичным примером комбинированной переменной является индексное значение.

На движение некоторых показателей могут оказывать влияние факторы, которые трудно выразить в форме статистического ряда. Например, в распоряжении исследователя имеется информация о наступлении или ненаступлении какого-либо события в определенный момент.

Иногда известно, что хотя некоторый фактор и может быть выражен статистическим рядом, однако влияние его на какой-либо процесс определяется не непосредственным изменением его числовых значений, а изменением его качественных характеристик. Во всех этих случаях используются искусственные переменные, которые описывают влияние исследуемых факторов и принимают одно из двух значений — ноль или единицу. Эти переменные называют также *качественными*, или *фиктивными*.

Потребность введения фиктивных переменных может возникнуть в процессе оценки регрессионных уравнений. В этом случае их экономическая интерпретация может быть неоднозначной.

Имеет смысл введение искусственной переменной в том случае, если для зависимой переменной характерны единичные скачки, не объясняемые действием факторов правой части уравнения регрессии. Эта переменная принимается равной нулю вне интервала, где наблюдался скачок, и, следовательно, она не оказывает влияния на прогноз показателя, а ее роль состоит в очищении параметров остальных переменных от искажений, вызванных единичными скачками.

Некоторые непосредственно не измеряемые факторы оказывают такое влияние на результативный показатель, в силу которого он растет, снижается или циклически изменяется. В качестве примера такого рода факторов можно привести научно-технический прогресс, демографические и социальные процессы. Влияние этих переменных описывается функцией времени.

Если показатель изменяется линейно, то в качестве тренда используют линейную функцию, а в случае прогрессивного роста — квадратичный или экспоненциальный тренд. Могут применяться такие функции времени, как квадратичный корень и различные комбинации трендов.

Циклические функции служат для описания правильных систематических движений показателя, вызванных прямо не измеряемыми экономическими факторами. Чаще всего периодическое движение аппроксимируют модифицированной синусоидой. Ее параметры определяют на основе анализа циклического движения остатков регрессионных уравнений.

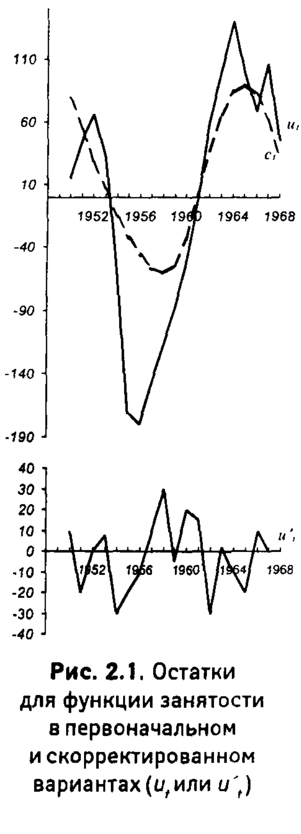


Рис. 2.1. Остатки для функции занятости в первоначальном и скорректированном вариантах (u_t или u'_t)

Необходимость введения циклической переменной проиллюстрирована на рисунке 2.1, из которого видно, что траектория движения остатка описывает волну, а циклическая переменная позволяет избежать волнообразного движения в новом остатке.

Комбинирование динамических факторов — распространенный прием в эконометрическом моделировании. Динамика вводится чаще всего с помощью лагов, разностных преобразований, посредством трендовых переменных. Рассмотрим несколько примеров такого комбинирования.

1. Линейная комбинация переменных, представляющая собой ряды одного показателя, сдвинутые назад на один, два или большие временные интервалы. В качестве примера можно привести разложе-

ние показателей во времени с помощью системы весовых коэффициентов. Это разложение впервые было использовано в инвестиционных функциях Уортонской модели, оно имело вид:

$$X_t = \sum_{i=0}^{7} A_i X_{t-i}, \tag{2.4}$$

где A_i — весовые коэффициенты.

2. Комбинирование переменных, представляющих собой ряды одного показателя, сдвинутые на целое или дробное число временных интервалов. Сдвиг на 1,5 года можно выразить с помощью средней из двух показателей, сдвинутой на 1 и 2 года:

$$X_{t-\frac{3}{2}} = \frac{1}{2}(X_{t-1} + X_{t-2}). \tag{2.5}$$

Сдвиг на один месяц учитывается следующим образом:

$$X_{t-\frac{1}{12}} = \frac{11}{12}X_t + \frac{1}{12}X_{t-1}.$$

3. Применение квазиразностей — комбинирование абсолютных значений ряда с его разностями:

$$X_t - q \cdot X_{t-1} = (1-q)\cdot X_t + q \cdot \Delta X_t,\ 0 < q < 1, \tag{2.6}$$

где X_t — ряд абсолютных значений показателя;

ΔX_t — ряд его первых разностей.

4. Авторегрессионные преобразования. Этот метод заключается во введении в уравнение переменной u_{t-1}:

$$u_{t-1} = y_{t-1} - a_0 - a_1 x_{1,t-1} - a_2 x_{2,t-1} - \ldots - a_n x_{n,t-1}. \tag{2.7}$$

2.3. Интерполирование статистических данных

Решение задач интерполяции и экстраполяции обеспечивается построением интерполяционной функции $L(x)$, приближенно заменяющей исходную $f(x)$,заданную таблично, и проходящей через все заданные точки — *узлы интерполяции*. С помощью этой функции можно рассчитать искомое значение исходной функции в любой точке.

В связи с интерполяцией рассматриваются три основные проблемы:

1) выбор интерполяционной функции $L(x)$;

2) оценка погрешности интерполяции $R(x)$;

3) размещение узлов интерполяции для обеспечения наивысшей возможной точности восстановления функции.

Специальные методы интерполяции позволяют определить искомое значение функции без непосредственного прямого построения интерполяционной функции. В принципе все интерполяционные методы, базирующиеся на использовании в качестве интерполяционной функции полиномов, дают одни и те же результаты, но с разными затратами. Это объясняется тем, что полином n-й степени, содержащий $n + 1$ параметр и проходящий через все заданные $n + 1$ точки, — единственный. Поэтому чаще первая проблема интерполяции решается выбором в качестве интерполяционной функции именно полинома, хотя могут применяться и другие функции (например тригонометрические полиномы).

Одним из наиболее распространенных методов данной группы является *интерполяционный метод Лагранжа*. Содержание этого метода заключается в следующем.

Пусть известны значения некоторой функции $f(x)$ в $n + 1$ различных произвольных точках $y_i = f(x_i)$, где $i = 0, 1, ..., n$. Для интерполирования (восстановления) функции в какой-либо точке x, принадлежащей отрезку $[x_0; x_n]$, необходимо построить интерполяционный полином n-го порядка, который в методе Лагранжа представляется следующим образом:

$$L_n(x) = y_0 \cdot Q_0(x) + y_1 \cdot Q_1(x) + ... + y_j \cdot Q_j(x) + ... + y_n \cdot Q_n(x) = \sum_{i=0}^{n} y_i \cdot Q_i(x), \qquad (2.8)$$

где

$$Q_j(x) = \frac{(x - x_0) \cdot (x - x_1) \cdot ... \cdot (x - x_{j-1}) \cdot (x - x_{j+1}) \cdot ... \cdot (x - x_n)}{(x_j - x_0) \cdot (x_j - x_1) \cdot ... \cdot (x_j - x_{j-1}) \cdot (x_j - x_{j+1}) \cdot ... \cdot (x_j - x_n)}. \qquad (2.9)$$

Оценить погрешность интерполяции в точке x, принадлежащей отрезку $[x_0; x_n]$, т. е. решить вторую проблему интерполяции, можно по формуле:

$$R(x) = |f(x) - L_n(x)| \le \frac{M_{n+1}}{(n+1)!} \cdot \prod_{i=0}^{n} (x - x_i), \qquad (2.10)$$

где $M_{n+1} = \max\left|f^{(n+1)}(x)\right|$ — максимальное значение производной $(n + 1)$ исходной функции $f(x)$ на отрезке $[x_0; x_n]$.

2.4. Методы многомерных сравнений

Кластерный анализ — это совокупность методов, позволяющих классифицировать многомерные наблюдения, каждое из которых описывается набором исходных переменных $x_1, x_2, \ldots, x_n$.

Целью кластерного анализа является образование групп схожих между собой объектов, которые принято называть кластерами.

В отличие от комбинационных группировок кластерный анализ приводит к разбиению на группы с учетом всех группировочных признаков одновременно. Например, если каждый наблюдаемый объект характеризуется двумя признаками x_1 и x_2, то при выполнении комбинационной группировки вся совокупность объектов будет разбита на группы по x_1, а затем внутри каждой выделенной группы будут образованы подгруппы x_2. Такой подход получил название *монотетического.*

В кластерном анализе используется иной принцип образования групп, который называется *политетический* подход. В данном подходе все группировочные признаки одновременно участвуют в группировке, то есть они учитываются все сразу при отнесении наблюдения в ту или иную группу. При этом не указываются четкие границы каждой группы, а также неизвестно заранее, сколько групп целесообразно выделить в исследуемой совокупности.

Методы кластерного анализа помогают построить научно обоснованные классификации, выявить внутренние связи между единицами наблюдаемой совокупности.

Методы кластерного анализа можно разделить на две большие группы: *агломеративные* (объединяющие) и *дивизимные* (разделяющие). *Агломеративные* методы последовательно объединяют отдельные объекты в группы (кластеры), а *дивизимные* методы расчленяют группы на отдельные объекты. Каждый метод агломеративного и дивизимного кластерного анализа может быть реализован при помощи различных алгоритмов, наиболее доступным из которых является иерархический кластерный анализ.

В иерархическом кластерном анализе каждое наблюдение образовывает сначала свой отдельный кластер. Процесс объединения кластеров происходит последовательно: на основании матрицы расстояний или матрицы сходства объединяются наиболее близкие объекты. Если матрица сходства первоначально имеет размерность $n \times n$, то полностью процесс кластеризации завершается за $n - 1$ шагов, в итоге все объекты будут объединены в один кластер. Последовательность объединения легко поддается графической интерпретации и может быть представлена в виде *дендрограммы*. На дендрограмме указывают номера объединяемых объектов и расстояния, при которых произошли объединения.

Множество методов иерархического кластерного анализа различается не только используемыми мерами сходства, но и алгоритмами классификации. Из них наиболее распространены метод одиночной связи, метод полных связей, метод средней связи, метод Уорда.

Алгоритм образования в *методе одиночной связи* следующий: на основании матрицы сходства (различия) определяется два наиболее схожих или близких объекта, они образуют первый кластер. На втором этапе выбирается объект, который имеет наибольшее сходство хотя бы с одним из объектов, уже включенных в кластер. При совпадении данных будет идти образование сразу нескольких кластеров.

Недостатками этого метода являются необходимость постоянного хранения матрицы сходства и невозможность определения по результатам кластеризации, сколько же кластеров можно образовать в исследуемой совокупности объектов.

Метод полных связей — включение нового объекта в кластер происходит только в том случае, если расстояние между объектами не меньше некоторого заданного уровня.

Метод средней связи — для решения вопроса о включении нового объекта в уже существующий кластер вычисляется среднее значение меры сходства, которое затем сравнивается с заданным пороговым уровнем. Если необходимо объединить два кластера, то вычисляют расстояния между их центрами и сравнивают его с заданными пороговыми значениями.

Метод Уорда предполагает, что на первом шаге каждый кластер состоит из одного объекта. Первоначально объединяются два ближайших кластера. Для них определяются средние значение каждого признака и рассчитывается сумма квадратов отклонений V_k:

$$V_k = \sum_{i=1}^{n_k} \sum_{j=1}^{p} \left(x_{ij} - \overline{x_{ik}}\right)^2, \tag{2.11}$$

где k — номер кластера;

i — номер объекта;

j — номер признака;

p — количество признаков, характеризующих каждый объект;

n_k — количество объектов в k-м кластере.

В дальнейшем на каждом шаге работы алгоритма объединяются те объекты или кластеры, которые дают наименьшее приращение величины V_k. Метод Уорда приводит к образованию кластеров приблизительно равных размеров с минимальной внутрикластерной вариацией. В итоге все объекты оказываются объединенными в один кластер.

Иерархические методы объединения трудоемки. На каждом шаге необходимо выстраивать дистанционную матрицу для всех кластеров, поэтому при наличии большого количества наблюдений применяют другие методы. К быстродействующим относятся итеративные методы, что позволяет использовать их для обработки больших массивов статистической информации. Характерной особенностью итеративных методов кластер-анализа является то, что кластеры формируются исходя из задаваемых условий разбиения, которые могут быть изменены в процессе исследования. Итеративные методы группировки данных могут привести к образованию пересекающихся кластеров, когда один объект может одновременно принадлежать к нескольким кластерам. К итеративным методам кластерного анализа относят метод поиска сгущений и метод k-средних.

Метод k-средних принадлежит к группе итеративных методов эталонного типа. Само название метода было предложено Дж. Мак-Куином в 1967 году. В отличие от иерархических процедур метод k-средних не требует вычисления и хранения матрицы расстояний или сходств между объектами. Алгоритм этого метода предполагает использование только исходных значений переменных.

Алгоритм иерархического кластерного анализа представим в виде последовательных процедур:

1. Формируем матрицу исходных данных X размером $n\times m$, где n — число объектов наблюдения; m — число признаков, по которым происходит группировка.

2. Переходим от матрицы исходных данных X к матрице нормированных данных Z, где разнородные по своей физической природе признаки приведем к основанию (условным единицам). Такой переход осуществляется по одному из следующих признаков:

а) $z_{ij} = \dfrac{x_{ij} - \bar{x}_j}{\sigma_j}$;

б) $z_{ij} = \dfrac{x_{ij}}{\bar{x}_j}$;

в) $z_{ij} = \dfrac{x_{ij}}{x_j}$, где x_j — эталонное значение;

г) $z_{ij} = x_{ij}$;

д) $z_{ij} = \dfrac{x_{ij} - \bar{x}_j}{x_{j(\max)} - x_{j(\min)}}$.

3. Определим расстояния между всеми парами объектов d_{ij} и построим исходную матрицу расстояний D_0. Для расчета используют одну из следующих метрик:

- метрика l-норма: $d_l = \sum\limits_{l}^{m} |x_{ik} - x_{jk}|$;
- метрика Минковского: $d_p = (\sum (x_{ik} - x_{jk})^p)^{\frac{1}{p}}$;
- метрика Евклидово расстояние: $d_e = (\sum (x_{ik} - x_{jk})^2)^{\frac{1}{2}}$;
- метрика взвешенное Евклидово расстояние:

$$d_e = (\sum w_k (x_{ik} - x_{jk})^2)^{\frac{1}{2}}.$$

4. Используя процедуру иерархического кластерного анализа, по данным матрицы D_0 выбираем группы однородных объектов.

5. При помощи специальных показателей оцениваем результаты анализа, в случае необходимости производится перегруппировка данных.

6. Результаты кластерного анализа обобщаем при помощи графиков или таблиц, даем им экономическую интерпретацию.

Контрольные вопросы

1. Временные структурные ряды данных.
2. Методы агрегирования статистических данных.
3. Адекватность и согласованность исходной информации.
4. Применение относительных показателей и средних величин.
5. Комбинирование динамических факторов.
6. Интерполяционный метод Лагранжа.
7. Оценка погрешности интерполяции.
8. Методы кластерного анализа.
9. Алгоритмы классификации данных.
10. Алгоритм иерархического кластерного анализа.

☑ ТЕСТОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ ПРОВЕРКИ ЗНАНИЙ

1. Кластерный анализ — это:
а) совокупность методов, позволяющих классифицировать многомерные наблюдения, каждое из которых описывается набором исходных переменных;
б) разновидность дисперсионного анализа;
в) совокупность методов статистического изучения интерполированных данных, позволяющих выявить погрешность интерполяции;
г) совокупность методов, позволяющих оценить устойчивость тенденции временного ряда.

2. Цель кластерного анализа состоит в:
а) в образовании групп схожих между собой объектов;
б) изучении тенденции временного ряда;
в) оценке погрешности интерполированных данных;
г) оценке вариации результативного признака в модели.

3. Образование кластеров в анализе базируется на основе:
а) политетического подхода;
б) монотетического подхода;
в) гемотетического подхода;
г) мегатетического подхода.

4. Известны две группы методов кластерного анализа:
а) агломеративные и дивизимные;
б) дивизимные и иерархические;
в) иерархические и итерационные;
г) итерационные и агломеративные.

5. Методы кластерного анализа реализуются при помощи следующих алгоритмов:
а) агломеративного;

б) итерационного;
в) комбинационного;
г) логического;
д) аналитического.

6. Агломеративный кластерный анализ предполагает:
а) последовательное объединение отдельных объектов в группы;
б) последовательное разделение группы на отдельные объекты;
в) последовательное выделение компонент временного ряда;
г) устранение тенденции временного ряда.

7. Дивизимный кластерный анализ предполагает:
а) последовательное объединение отдельных объектов в группы;
б) последовательное разделение группы на отдельные объекты;
в) последовательное выделение компонент временного ряда;
г) изучение тенденции временного ряда.

8. Графическое изображение результатов кластерного анализа называют:
а) дендрограммой;
б) мезограммой;
в) тентограммой;
г) изограммой.

9. К методам иерархического кластерного анализа относят:
а) метод Уорда;
б) метод k-средних;
в) метод поиска сгущений;
г) метод Мак-Куина;
д) метод средней связи;
е) метод Махалонобиуса;
ж) метод одиночной связи.

10. К итеративным методам кластерного анализа относят:
а) метод k-средних;
б) метод Уорда;
в) метод Махалонобиуса;
г) метод средней связи;
д) метод поиска сгущений.

11. Переход от матрицы исходных данных к матрице нормированных данных осуществляется по признаку:

а) $z_{ij} = \dfrac{x_{ij} - \overline{x}_j}{\sigma_j}$;

б) $z_{ij} = \dfrac{x_{ij} - \overline{x}_j}{\sigma_j^2}$;

в) $z_{ij} = \dfrac{x_{ij} - \overline{x}_j}{V_j}$;

г) $z_{ij} = \dfrac{x_{ij} - \overline{y}_j}{\sigma_j}$.

12. Преобразование матрицы исходных данных в матрицу нормированных данных осуществляется по признаку:

а) $z_{ij} = \dfrac{x_{ij} - \overline{x}_j}{x_{j(\max)} - x_{j(\min)}}$;

б) $z_{ij} = \dfrac{x_{ij} - \overline{x}_j}{x_{j(\max)} - x_{j(\max)}}$;

в) $z_{ij} = \dfrac{x_{ij} - \overline{x}_j}{x_{j(\min)} - x_{j(\min)}}$;

г) $z_{ij} = \dfrac{y_{ij} - \overline{x}_j}{x_{j(\max)} - x_{j(\min)}}$.

13. Приведение матрицы исходных данных к матрице нормированных данных осуществляется по признаку:

а) $z_{ij} = \dfrac{x_{ij}}{\overline{x}_j}$;

б) $z_{ij} = \dfrac{x_{ij}}{\overline{y}_j}$;

в) $z_{ij} = \dfrac{y_{ij}}{\overline{x}_j}$;

г) $z_{ij} = \dfrac{x_{ij} \cdot x_{ji}}{\overline{x}_j}$.

14. Расстояния между всеми парами объектов d_{ij} определяют по метрике:

а) l-норма: $d_l = \sum_{1}^{m} \left| x_{ik} - x_{jk} \right|$;

б) n-норма: $d_l = \sum_{1}^{m} \left| x_{ik} - x_{jk} \right|$;

в) m-норма: $d_l = \sum_{1}^{m} \left| x_{ik} - x_{jk} \right|$;

г) l-норма: $d_l = \sum_{1}^{m} \left| x_{ik} - \overline{x}_{jk} \right|$.

15. Методику расчета между всеми парами объектов d_{ij} осуществляют по метрике:

а) Минковского: $d_p = (\sum \left(x_{ik} - x_{jk} \right)^p)^{\frac{1}{2}}$; ;

б) Пинкорского: $d_p = (\sum \left(x_{ik} - x_{jk} \right)^p)^{\frac{1}{p}}$;

в) Койковского: $d_p = (\sum (x_{ik} - x_{jk})^p)^{\frac{1}{p}}$;

г) Минковского: $d_p = (\sum (x_{ik} - \bar{x}_{jk})^p)^{\frac{1}{p}}$.

ПРАКТИКУМ

Задача 2.1. По данным временного ряда об уровне рентабельности продукции в сельском хозяйстве в регионе необходимо рассчитать индивидуальные показатели динамики, сделать выводы.

Таблица 2.1

Уровень рентабельности производства продукции отрасли (*У*) за период 1990–2008 гг., млн руб.

Период	Уровень рентабельности производства продукции отрасли (*У*)
1990	158,6
1991	107,3
1992	292,1
1993	242,7
1994	113,9
1995	92,5
1996	47,4
1997	27,5
1998	14,6
1999	60,4
2000	58,8
2001	52,8
2002	34,3
2003	41,6
2004	47,1
2005	24,1
2006	40,1
2007	59,8
2008	37,4

Решение

Согласно формулам (2.1), (2.2), (2,3), представленным в теоретической части раздела, рассчитаем индивидуальные показатели динамики временного ряда с помощью программы *Statistica 6.0.*

Для этого вначале необходимо создать таблицу исходных данных в окне *Statistica 6.0*, обозначив исследуемый динамический ряд переменной *Y*. На следующем этапе необходимо создать ряд значений исследуемой переменной, смещенный на 1 период времени вперед, обозначив новую переменную как «Y_1» (рис. 2.2).

	1 Год	2 Y	3 Y 1
1	1990	158,6	
2	1991	107,3	158,6
3	1992	292,1	107,3
4	1993	242,7	292,1
5	1994	113,9	242,7
6	1995	92,5	113,9
7	1996	47,4	92,5
8	1997	27,5	47,4
9	1998	14,6	27,5
10	1999	60,4	14,6
11	2000	58,8	60,4
12	2001	52,8	58,8
13	2002	34,3	52,8
14	2003	41,6	34,3
15	2004	47,1	41,6
16	2005	24,1	47,1
17	2006	40,1	24,1
18	2007	59,8	40,1
19	2008	37,4	59,8

Рис. 2.2. Исходные данные для расчета показателей ряда динамики средствами *Statistica 6.0*

Для вычисления базисных абсолютных приростов в меню «Вставка» выберем инструмент «Добавить переменную». В появившемся диалоговом окне выполним последовательность действий:

- в поле «Имя» введем имя новой переменной, например «$АПУ_б$»;
- в поле «Длинная метка или формула» введем выражение «=*Y-158,6*»;
- в поле «Формат отображения» укажем числовой тип переменной с одним десятичным разрядом, нажмем «ОК» (рис. 2.3).

В результате получим значения абсолютных приростов исследуемого динамического ряда.

Для вычисления цепных абсолютных приростов в диалоговом окне «Добавить переменную» выполним последовательность действий:

- в поле «Имя» введем имя новой переменной, например «$АПУ_ц$»;
- в поле «Длинная метка или формула» введем выражение «=*Y-Y_1*»;

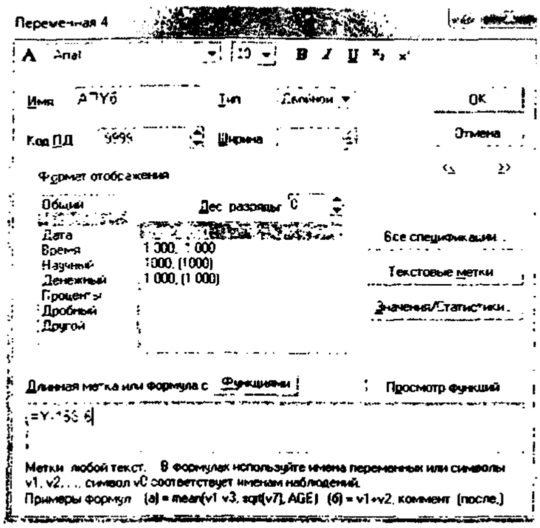


Рис. 2.3. Вычисление новой переменной средствами *Statistica 6.0*

- в поле «Формат отображения» укажем числовой тип переменной с одним десятичным разрядом, нажмем «OK».

Для вычисления базисных темпов роста в диалоговом окне «Добавить переменную» выполним последовательность действий:

- в поле «Имя» введем имя новой переменной, например «$Тр_б$»;
- в поле «Длинная метка или формула» введем выражение «*=Y/158,6*100*»;
- в поле «Формат отображения» укажем числовой тип переменной с одним десятичным разрядом, нажмем «OK».

Для вычисления цепных темпов роста в диалоговом окне «Добавить переменную» выполним последовательность действий:

- в поле «Имя» введем имя новой переменной, например «$Тр_ц$»;
- в поле «Длинная метка или формула» введем выражение «*=Y/Y_1*100*»;
- в поле «Формат отображения» укажем числовой тип переменной с одним десятичным разрядом, нажмем «OK».

Для вычисления базисных темпов прироста в диалоговом окне «Добавить переменную» выполним последовательность действий:

- в поле «Имя» введем имя новой переменной, например «$Tпр_б$»;
- в поле «Длинная метка или формула» введем выражение «$=АПУ_б/158{,}6*100$»;
- в поле «Формат отображения» укажем числовой тип переменной с одним десятичным разрядом, нажмем «OK».

Для вычисления цепных темпов прироста в диалоговом окне «Добавить переменную» выполним последовательность действий:

- в поле «Имя» введем имя новой переменной, например «$Tпр_ц$»;
- в поле «Длинная метка или формула» введем выражение «$=АПУ_ц/Y\_1*100$»;
- в поле «Формат отображения» укажем числовой тип переменной с одним десятичным разрядом, нажмем «OK».

В результате получим значения индивидуальных показателей динамики (рис. 2.4).

	1 Год	2 Y	3 Y 1	4 АПУб	5 АПУц	6 Трб	7 Трц	8 Тпрб	9 Тпрц
1	1990	158,6							
2	1991	107,3	158,6	-51,3	-51,3	67,7	67,7	-32,3	-32,3
3	1992	292,1	107,3	133,5	184,8	184,2	272,2	84,2	172,2
4	1993	242,7	292,1	84,1	-49,4	153,0	83,1	53,0	-16,9
5	1994	113,9	242,7	-44,7	-128,8	71,8	46,9	-28,2	-53,1
6	1995	92,5	113,9	-66,1	-21,4	58,3	81,2	-41,7	-18,8
7	1996	47,4	92,5	-111,2	-45,1	29,9	51,2	-70,1	-48,8
8	1997	27,5	47,4	-131,1	-19,9	17,3	58,0	-82,7	-42,0
9	1998	14,6	27,5	-144,0	-12,9	9,2	53,1	-90,8	-46,9
10	1999	60,4	14,6	-98,2	45,8	38,1	413,7	-61,9	313,7
11	2000	58,8	60,4	-99,8	-1,6	37,1	97,4	-62,9	-2,6
12	2001	52,8	58,8	-105,8	-6,0	33,3	89,8	-66,7	-10,2
13	2002	34,3	52,8	-124,3	-18,5	21,6	65,0	-78,4	-35,0
14	2003	41,6	34,3	-117,0	7,3	26,2	121,3	-73,8	21,3
15	2004	47,1	41,6	-111,5	5,5	29,7	113,2	-70,3	13,2
16	2005	24,1	47,1	-134,5	-23,0	15,2	51,2	-84,8	-48,8
17	2006	40,1	24,1	-118,5	16,0	25,3	166,4	-74,7	66,4
18	2007	59,8	40,1	-98,8	19,7	37,7	149,1	-62,3	49,1
19	2008	37,4	59,8	-121,2	-22,4	23,6	62,5	-76,4	-37,5

Рис. 2.4. Расчет показателей ряда динамики в *Statistica 6.0*

Задача 2.2. Постройте интерполяционную функцию на основе решетчато заданной функции, принимающую в данных точках $x_0 = 1$, $x_1 = 3$, $x_2 = 6$ значения функции $y_0 = 10$, $y_1 = 16$, $y_2 = 4$.

Решение

Построим интерполяционную функцию $L_2(x)$:

$$L_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} \cdot y_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} \cdot y_1 +$$

$$+\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} \cdot y_2 = \frac{(x-3)(x-6)}{(1-3)(1-6)} \cdot 10 +$$

$$+\frac{(x-1)(x-6)}{(3-1)(3-6)} \cdot 16 + \frac{(x-1)(x-3)}{(6-1)(6-3)} \cdot 4 =$$

$$= \frac{x^2-9x+18}{10} \cdot 10 + \frac{x^2-7x+6}{-6} \cdot 16 + \frac{x^2-4x+3}{15} \cdot 4 =$$

$$= x^2 - 9x + 18 - 2{,}67x^2 + 18{,}7x - 16 + 0{,}27x^2 - 1{,}1x + 0{,}8 =$$

$$= -1{,}4x^2 + 8{,}6x + 2{,}8$$

Найдем значения функции в точках $x = 2{,}5; 4{,}2; 5{,}7$:

$$L_2(2{,}5) = -1{,}4 \cdot 2{,}5^2 + 8{,}6 \cdot 2{,}5 + 2{,}8 = 15{,}55;$$
$$L_2(4{,}2) = -1{,}4 \cdot 4{,}2^2 + 8{,}6 \cdot 4{,}2 + 2{,}8 = 14{,}22;$$
$$L_2(5{,}7) = -1{,}4 \cdot 5{,}7^2 + 8{,}6 \cdot 5{,}7 + 2{,}8 = 6{,}33.$$

Задача 2.3. Дана функция $y = \sqrt{x}$, известны три узла интерполяции:

X	100	121	144
Y	10	11	12

Необходимо оценить погрешность интерполяции функции при $x = 115$; 127; 134.

Решение

Для определения погрешности интерполяции в точке x из $[x_0;x_n]$ в соответствии с формулой (2.6) необходимо сначала найти производную ($n + 1$)-порядка (в нашем случае третьего) исходной функции $f(x)$:

$$y''' = \left(x^{\frac{1}{2}}\right)''' = \left(\frac{1}{2}\cdot x^{-\frac{1}{2}}\right)'' = \left(-\frac{1}{4}\cdot x^{-\frac{3}{2}}\right)' = \frac{3}{8}\cdot x^{-\frac{5}{2}}.$$

Полученная производная представляет собой обратную зависимость, поэтому максимальное значение функции на отрезке [100; 144] будет приниматься при минимальном значении аргумента, т. е. при $x = 100$:

$$M_3 = \frac{3}{8}\cdot 100^{-\frac{5}{2}}.$$

На завершающем этапе оценим погрешность интерполяции функции:

- при $x = 115$:

$$R(115) \le \frac{\frac{3}{8}\cdot 100^{-\frac{5}{2}}}{3!}\cdot|115-100|\cdot|115-121|\cdot|115-144| =$$
$$= 163{,}13\cdot 10^{-5} = 1{,}6313\cdot 10^{-3},$$

- при $x = 127$:

$$R(127) \le \frac{\frac{3}{8}\cdot 100^{-\frac{5}{2}}}{3!}\cdot|127-100|\cdot|127-121|\cdot|127-144| =$$
$$= 172{,}13\cdot 10^{-5} = 1{,}7213\cdot 10^{-3},$$

- при $x = 134$:

$$R(134) \le \frac{\frac{3}{8}\cdot 100^{-\frac{5}{2}}}{3!}\cdot|134-100|\cdot|134-121|\cdot|134-144| =$$
$$= 276{,}25\cdot 10^{-5} = 2{,}7625\cdot 10^{-3}.$$

Задача 2.4. Используя данные таблицы 2.2, рассмотрим порядок применения иерархического агломеративного кластерного анализа.

Таблица 2.2

Данные о потреблении продуктов питания в расчете на одного жителя территории

Виды продуктов питания	Регион *A*	Регион *B*	Регион *C*	Регион *D*	Регион *E*
Мясо и мясопродукты на 1 человека в год, кг	59	113	70	94	37
Молоко и молочные продукты на 1 человека в год, кг	363	263	313	412	81
Яйца на 1 человека в год, шт.	268	229	193	257	298
Хлеб и хлебобулочные изделия на 1 человека в год, кг	129	100	90	84	119

Решение

Согласно алгоритму иерархического агломеративного кластерного анализа последовательно проведем указанные процедуры:

1. Сформируем исходную матрицу данных размерностью $n \times m$:

$$X = \begin{bmatrix} 59 & 363 & 268 & 129 \\ 113 & 263 & 229 & 100 \\ 70 & 313 & 193 & 90 \\ 94 & 412 & 257 & 84 \\ 37 & 81 & 298 & 119 \end{bmatrix}.$$

2. Проведем нормирование исходных данных по варианту $z_{ij} = \frac{x_{ij}}{\bar{x}_j}$

$$\bar{x}_1 = \frac{59+113+70+94+37}{5} = 74,6;$$

$$\bar{x}_2 = 286,4;$$

$$\bar{x}_3 = 249,0;;$$

$$\bar{x}_4 = 104,4.$$

Тогда $z_{11} = \frac{59}{74,6} = 0,79;$

$$z_{12} = \frac{363}{286,4} = 1,27;$$

$$z_{21} = \frac{113}{74,6} = 1,51;$$

$$z_{22} = \frac{263}{286,4} = 0,92.$$

В результате матрица нормированных данных Z имеет вид:

$$Z = \begin{Bmatrix} 0,79 & 1,27 & 1,08 & 1,24 \\ 1,51 & 0,92 & 0,92 & 0,96 \\ 0,94 & 1,09 & 0,78 & 0,86 \\ 1,26 & 1,44 & 1,03 & 0,80 \\ 0,50 & 0,28 & 1,20 & 1,14 \end{Bmatrix}.$$

3. Используя метрику l-норма: $d_l = \sum_1^m \left| x_{ik} - x_{jk} \right|$, перейдем к матрице расстояний D_0:

$d_{12} = |0,79 - 1,51| + |1,27 - 0,92| + |1,08 - 0,92| + |1,24 - 0,96| = 1,51;$

$d_{13} = |0,79 - 0,94| + |1,27 - 1,09| + |1,08 - 0,78| + |1,24 - 0,86| = 1,01;$
и т. д.

В результате матрица расстояний D_0 имеет вид:

$$D_0 = \begin{Bmatrix} 0 & 1,51 & 1,01 & 1,13 & 1,5 \\ 1,51 & 0 & 0,98 & 1,04 & 2,11 \\ 1,01 & 0,98 & 0 & 0,98 & 1,96 \\ 1,13 & 1,04 & 0,98 & 0 & 2,43 \\ 1,50 & 2,11 & 1,96 & 2,43 & 0 \end{Bmatrix}.$$

В агломеративном кластерном анализе могут применяться три алгоритма кластеризации данных: «ближайшего соседа», «дальнего соседа» и «средней связи». Выберем алгоритм «дальнего соседа», тогда на первом шаге объединим ближайшие объекты $d_{23} = 0,98$ и $d_{34} = 0,98$ и получим следующие кластеры:

Кластер	1	2	3
Объект	1	2, 3, 4	5

Теперь необходимо определить расстояния до второго кластера от всех других объектов, кроме 2-го, 3-го, 4-го, и сформируем матрицу расстояний D_1:

$$D_1 = \begin{Bmatrix} 0 & 1,51 & 1,50 \\ 1,51 & 0 & 2,43 \\ 1,5 & 2,43 & 0 \end{Bmatrix}.$$

Минимальное расстояние $d_{13} = 1,50$, следовательно, имеем кластеры:

Кластер	1	2
Объект	1, 5	2, 3, 4

Расстояние между двумя кластерами $d_{12} = 2,43$.

Задача 2.5. По данным таблицы 2.3 проведите иерархический агломеративный кластерный анализ.

Таблица 2.3

Уровень обеспеченности населения основными продуктами питания за 2012 г. в разрезе районов Ставропольского края

№ п/п	Район	Коэффициент обеспеченности					
		хлебными продуктами	мясными продуктами	молочными продуктами	картофелем	сахаром	яйцами
1	Александровский	1,092	0,654	0,465	1,176	1,330	0,901
2	Андроповский	1,102	0,660	0,469	1,187	1,342	0,909
3	Апанасенковский	1,105	0,662	0,470	1,190	1,346	0,911
4	Арзгирский	1,098	0,658	0,467	1,183	1,338	0,906
5	Благодарненский	1,107	0,663	0,471	1,192	1,349	0,913
6	Буденновский	1,109	0,664	0,472	1,194	1,351	0,914
7	Георгиевский	1,194	0,715	0,508	1,286	1,454	0,985
8	Грачевский	1,108	0,664	0,471	1,193	1,350	0,914
9	Изобильненский	1,109	0,665	0,472	1,195	1,352	0,915
10	Ипатовский	1,103	0,661	0,470	1,188	1,344	0,910
11	Кировский	1,117	0,670	0,476	1,204	1,361	0,922
12	Кочубеевский	1,114	0,667	0,474	1,200	1,357	0,918
13	Красногвардейский	1,110	0,665	0,472	1,195	1,352	0,915
14	Курский	1,124	0,674	0,478	1,211	1,370	0,927
15	Левокумский	1,110	0,665	0,473	1,196	1,353	0,916
16	Минераловодский	1,109	0,664	0,472	1,194	1,350	0,914
17	Нефтекумский	1,114	0,668	0,474	1,200	1,357	0,919
18	Новоалександровский	1,101	0,660	0,468	1,186	1,341	0,908
19	Новоселицкий	1,118	0,670	0,476	1,204	1,361	0,922
20	Петровский	1,106	0,663	0,471	1,191	1,347	0,912
21	Предгорный	1,122	0,672	0,478	1,209	1,367	0,925
22	Советский	1,107	0,663	0,471	1,192	1,349	0,913
23	Степновский	1,111	0,666	0,473	1,197	1,354	0,916
24	Труновский	1,098	0,658	0,467	1,183	1,338	0,907
25	Туркменский	1,105	0,662	0,470	1,190	1,346	0,911
26	Шпаковский	1,146	0,687	0,488	1,235	1,396	0,945

Решение

Осуществим решение с помощью программного продукта *SPSSStatistics* 20. Первый шаг алгоритма иерархического агломеративного кластерного анализа предусматривает формирование матрицы исходных данных *X*. В данном случае размерность матрицы составляет 26×6. Классификационные признаки обозначим именами *K1*, *K2*, *K3*, *K4*, *K5*, *K6*.

Далее найдем инструмент «Анализ» → «Классификация» → «Иерархическая кластеризация». В появившемся диалоговом окне «Иерархический кластерный анализ» перенесем все классификационные признаки в окно «Переменные» (рис. 2.5).

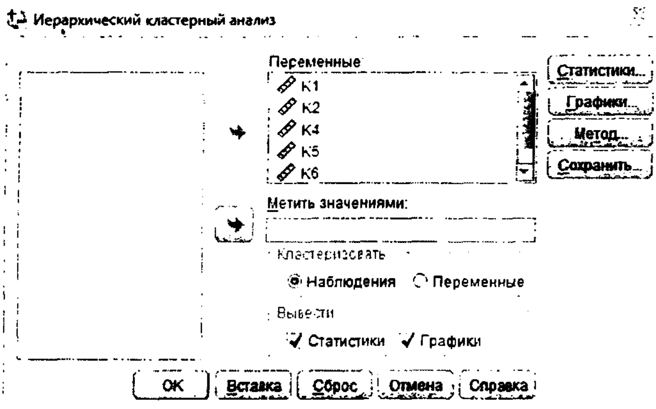


Рис. 2.5. Окно «Иерархический кластерный анализ» *SPSSStatistics 20*

На следующем этапе укажем необходимые настройки.

Во вкладке «Статистики» отметим таблицы «Порядок агломерации» и «Матрица близостей», нажмем «Продолжить».

Во вкладке «Графики» отметим пункт «Дендрограмма», нажмем «Продолжить».

Во вкладке «Метод» необходимо указать метод, с помощью которого будет производиться объединение объектов совокупности в кластеры. Укажем метод дальнего соседа. В качестве меры близости можно выбрать одну из нескольких метрик расчета матрицы расстояний. Отметим метрику «Расстояние Евклида». Кроме того, в данной вкладке можно указать способ расчета матрицы стандартизированных коэффициентов *Z*. В нашем случае необходимости расчета такой матрицы нет, поскольку все классификационные признаки представляют собой коэффициенты, выражены одинаковыми единицами измерения и являются сопоставимыми между собой.

Шаги агломерации

Этап	Кластер объединен с		Коэффициенты	Этап первого появления кластера		Следующий этап
	Кластер 1	Кластер 2		Кластер 1	Кластер 2	
1	3	25	,000	0	0	11
2	5	22	,000	0	0	13
3	6	16	,001	0	0	8
4	4	24	,001	0	0	18
5	11	19	,001	0	0	17
6	9	13	,001	0	0	12
7	12	17	,001	0	0	17
8	6	8	,002	3	0	13
9	15	23	,002	0	0	12
10	2	18	,002	0	0	15
11	3	20	,002	1	0	16
12	9	15	,004	6	9	19
13	5	6	,004	2	8	16
14	14	21	,005	0	0	20
15	2	10	,005	10	0	18
16	3	5	,009	11	13	19
17	11	12	,009	5	7	20
18	2	4	,011	15	4	21
19	3	9	,014	16	12	22
20	11	14	,023	17	14	23
21	1	2	,025	0	18	22
22	1	3	,043	21	19	23
23	1	11	,071	22	20	25
24	7	26	,105	0	0	25
25	1	7	,225	23	24	0

Рис. 2.6. Этапы объединения объектов в кластеры

Указав все необходимые настройки, нажмем «ОК».

В окне вывода результатов анализа найдем матрицу исходных расстояний, на основе которой осуществлялось объединение объектов в кластеры.

Кроме того, окно вывода результатов содержит таблицу «Шаги агломерации» (рис. 2.6). Из таблицы видно, что на первом этапе на расстоянии 0,000 были объединены 3-й и 25-й объекты совокупности (Апанасенковский и Туркменский муниципальные

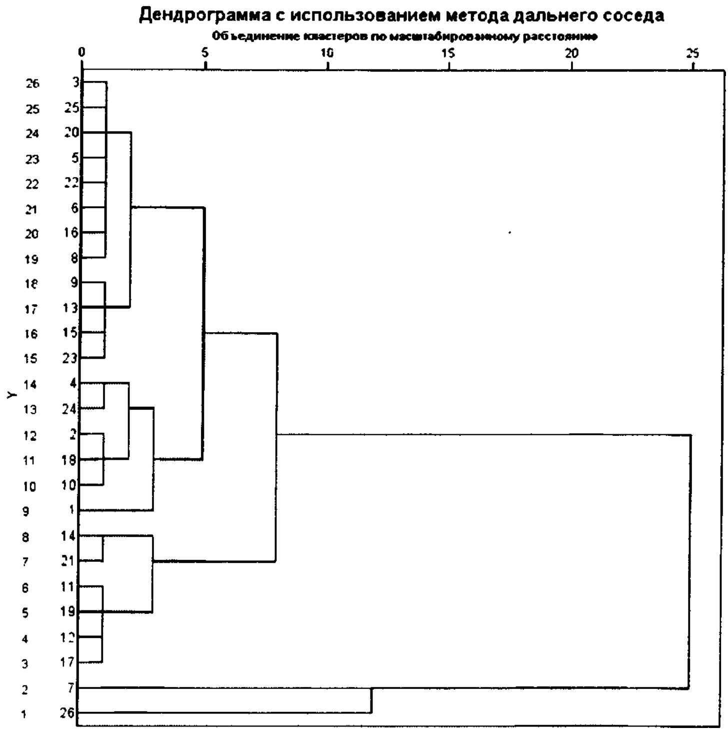


Рис. 2.7. Графическое изображение результатов кластерного анализа (дендрограмма) в *SPSSStatistics* 20

районы), затем 5-й и 22-й объекты (Благодарненский и Советский районы) – на таком расстоянии.

Аналогично было осуществлено в общем итоге 25 итераций, которые позволили в итоге получить многомерную классификацию муниципальных районов Ставропольского края по шести классификационным признакам. В завершение анализа была построена дендрограмма (рис. 2.7).

Анализируя итог кластеризации, можно сделать вывод, что все муниципальные районы Ставропольского края можно разделить на 3 группы по степени обеспеченности основными продуктами питания (в разрезе шести групп продуктов). В первую группу входит 18 районов, во вторую – 6, в третью – 2.

Задача 2.6. По данным таблицы 2.4 проведите иерархический агломеративный кластерный анализ.

Таблица 2.4

Показатели численности и занятости населения за 2012 г. в разрезе районов Ставропольского края

№ п/п	Район	Плотность населения, чел/км²	Численность безработных, чел.	Численность детей в ДОУ, тыс. чел.	Численность работающих в сельском хозяйстве, тыс. чел.	Численность работающих в промышленном производстве, тыс. чел.
	2	3	4	5	6	7
1	Александровский	24,5	1207	1,3	2,0	1,1
2	Андроповский	14,7	738	0,9	0,5	0,9
3	Апанасенковский	9,8	455	1,4	4,7	0,3
4	Арзгирский	8,1	873	1,0	2,5	0,3
5	Благодарненский	26,1	487	2,1	2,2	1,3
6	Буденновский	17,1	230	1,5	3,3	0,3
7	Георгиевский	47,2	1827	2,4	3,4	1,1
8	Грачевский	20,0	1827	1,0	1,2	0,2
9	Изобильненский	51,7	312	3,4	2,7	5,0
10	Ипатовский	16,5	750	1,9	4,3	1,8
11	Кировский	48,0	1255	2,2	1,6	1,2

Окончание табл. 2.4

1	2	3	4	5	6	7
12	Кочубеевский	33,1	1464	1,7	4,5	2,2
13	Красногвардейский	18,2	1066	1,3	3,2	0,5
14	Курский	14,0	2826	1,2	1,9	0,4
15	Левокумский	9,1	676	1,3	2,5	0,3
16	Минераловодский	32,1	371	4,2	0,9	4,9
17	Нефтекумский	18,3	937	2,1	1,8	4,8
18	Новоалександровский	32,2	512	2,0	4,2	2,2
19	Новоселицкий	14,9	468	0,8	2,5	0
20	Петровский	29,3	1154	2,4	3,3	2,2
21	Предгорный	51,6	678	2,1	2,7	1,3
22	Советский	34,0	1007	2,2	4,1	2,8
23	Степновский	12,1	1464	0,7	2,0	0
24	Труновский	20,6	1190	1,2	3,7	0
25	Туркменский	10,3	406	1,0	1,8	0
26	Шпаковский	46,9	1094	2,9	4,2	2,9

Решение

Осуществим решение с помощью программного продукта *Statistica 6.0.*

На первом этапе составим матрицу исходных данных – таблицу размерностью 26 × 5 (рис. 2.8).

Далее откроем модуль «Анализ» → «Многомерный разведочный анализ» → «Кластерный анализ». Выберем пункт «Иерархическая классификация». Во вкладке «Дополнительно» сделаем следующие настройки:

- выберем все переменные;
- объектами кластеризации являются наблюдения (строки), то есть 26 муниципальных районов Ставропольского края;
- выберем подходящее правило объединения и меру близости (в нашем примере целесообразно использовать метод полной связи как правило объединения и Евклидово расстояние в качестве меры близости) (рис. 2.9).

Из полученных результатов анализа значительный интерес представляет в первую очередь матрица расстояний, а также схема объединения (рис. 2.10, 2.11).

	1 X1	2 X2	3 X3	4 X4	5 X5
Александровский	24,5	1207	1,3	2	1.1
Андроповский	14,7	738	0,9	0.5	0.9
Апанасенковский	9,8	455	1,4	4.7	0,3
Арзгирский	8,1	873	1	2.5	0.3
Благодарненский	26,1	487	2,1	2,2	1.3
Буденновский	17.1	230	1,5	3,3	0.3
Георгиевский	47.2	1827	2,4	3.4	1 1
Грачевский	20	1827	1	1.2	0.2
Изобильненский	51.7	312	3,4	2,7	5
Ипатовский	16.5	750	1,9	4.3	1 8
Кировский	48	1255	2,2	1.6	1,2
Кочубеевский	33,1	1464	1.7	4 5	2.2
Красногвардейский	18.2	1066	1,3	3.2	0.5
Курский	14	2826	1,2	1.9	0,4
Левокумский	9.1	676	1.3	2.5	0.3
Минераловодский	32,1	371	4,2	0,9	4.9
Нефтекумский	18.3	937	2,1	1.8	4.8
Новоалександровский	32.2	512	2	4,2	2.2
Новоселицкий	14,9	468	0,8	2.5	0
Петровский	29.3	1154	2,4	3 3	2,2
Предгорный	51,6	678	2,1	2.7	1.3
Советский	34	1007	2.2	4.1	2,8
Степновский	12,1	1464	0.7	2	0
Труновский	20.6	1190	1,2	3.7	0
Туркменский	10,3	406	1	1.8	0
Шпаковский	46,9	1094	2.9	4,2	2.9

Рис. 2.8. Таблица исходных данных в *Statistica 6.0*

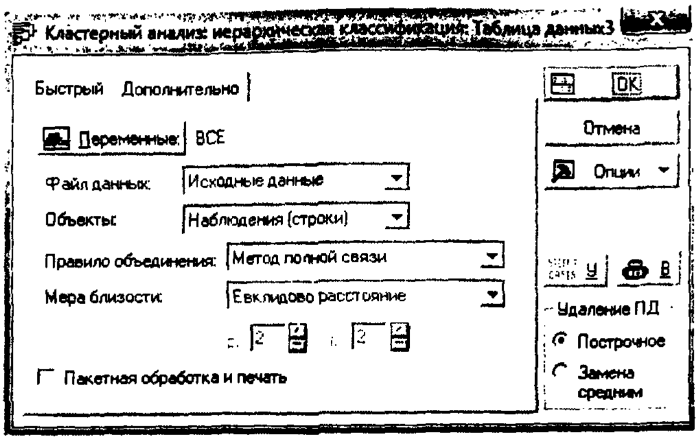


Рис. 2.9. Опции инструмента «Кластерный анализ» в *Statistica 6.0*

	Евклидово расстояние (Таблица данных3)																									
наблюд.	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10	C_11	C_12	C_13	C_14	C_15	C_16	C_17	C_18	C_19	C_20	C_21	C_22	C_23	C_24	C_25	C_26
C_1	0	469	752	334	720	977	620	620	895	457	53	257	141	1619	531	836	270	695	746	53	530	200	276	101	807	115
C_2	469	0	283	135	251	508	1089	1089	428	13	518	726	328	2088	62	367	199	227	288	416	70	270	733	463	347	357
C_3	752	283	0	418	36	225	1373	1372	149	295	801	1009	611	2371	221	87	482	61	102	699	227	553	1014	742	112	640
C_4	334	135	418	0	386	643	955	954	563	123	384	592	193	1953	197	503	65	362	417	282	200	137	600	333	478	224
C_5	720	251	36	386	0	257	1340	1340	177	263	768	977	579	2339	190	116	450	26	102	667	193	520	982	710	129	607
C_6	977	508	225	643	257	0	1597	1597	89	520	1025	1234	836	2596	446	142	707	282	258	924	449	777	1238	965	203	865
C_7	620	1089	1373	955	1340	1597	0	27	1515	1077	572	363	762	1000	1152	1456	890	1315	1363	673	1149	820	378	645	1425	733
C_8	620	1089	1372	954	1340	1597	27	0	1515	1077	573	363	761	999	1151	1456	890	1315	1363	673	1149	820	377	645	1425	734
C_9	895	428	149	563	177	89	1515	1515	0	439	943	1152	755	2514	367	62	626	201	187	842	366	695	1157	884	141	782
C_10	457	13	295	123	263	520	1077	1077	439	0	506	714	316	2076	74	379	187	239	299	404	80	258	721	451	358	345
C_11	53	518	801	384	768	1025	572	573	943	506	0	210	191	1571	580	884	319	743	794	103	577	248	234	122	856	161
C_12	257	726	1009	592	977	1234	363	363	1152	714	210	0	398	1362	788	1093	527	952	1001	310	786	457	101	292	1063	370
C_13	141	328	611	193	579	836	762	761	755	316	191	398	0	1760	390	695	129	554	606	89	389	61	411	160	668	40
C_14	1619	2088	2371	1953	2339	2596	1000	999	2514	2076	1571	1362	1760	0	2150	2455	1889	2314	2360	1672	2148	1819	1366	1639	2422	1732
C_15	531	62	221	197	190	446	1152	1151	367	74	580	788	390	2150	0	306	261	166	231	478	43	332	794	524	288	420
C_16	836	367	87	503	116	142	1456	1456	62	379	884	1093	695	2455	306	0	566	141	138	783	308	636	1097	825	105	723
C_17	270	199	482	65	450	707	890	890	626	187	319	527	129	1889	261	566	0	425	479	217	261	72	536	271	540	160
C_18	695	227	61	362	26	282	1315	1315	201	239	743	952	554	2314	166	141	425	0	110	642	167	495	957	685	147	582
C_19	746	288	102	417	102	258	1363	1363	187	299	794	1001	606	2360	231	138	479	110	0	693	235	548	996	722	62	634
C_20	53	416	699	282	667	924	673	673	842	404	103	310	89	1672	478	783	217	642	693	0	477	147	326	106	755	63
C_21	530	70	227	200	193	449	1149	1149	366	80	577	786	389	2148	43	308	261	167	235	477	0	329	793	523	293	416
C_22	200	270	553	137	520	777	820	820	695	258	248	457	61	1819	332	636	72	495	548	147	329	0	468	208	609	88
C_23	276	733	1014	600	982	1238	378	377	1157	721	234	101	411	1366	794	1097	536	957	996	326	793	468	0	274	1058	384
C_24	101	463	742	333	710	965	645	645	884	451	122	292	160	1639	524	825	271	685	722	106	523	208	274	0	784	140
C_25	807	347	112	478	129	203	1425	1425	141	358	856	1063	668	2422	288	105	540	147	62	755	293	609	1058	784	0	696
C_26	115	357	640	224	607	865	733	734	782	345	161	370	40	1732	420	723	160	582	634	63	416	88	384	140	696	0

Рис. 2.10. Матрица расстояний

	Схема объединения (Таблица данных3) Метод полной связи Евклидово расстояние																			
расст. объедин.	**Объект 1**	Объект 2	Объект 3	Объект 4	Объект 5	Объект 6	Объект 7	Объект 8	Объект 9	Объект 10	Объект 11	Объект 12	Объект 13	Объект 14	Объект 15	Объект 16	Объект 17	Объект 18	Объект 19	Объект 20
12,78632	C_2	C_10																		
25,82692	C_5	C_18																		
27,33953	C_7	C_8																		
40,21206	C_13	C_26																		
42,56677	C_15	C_21																		
53,25552	C_1	C_20																		
61,27789	C_3	C_5	C_18																	
62,17467	C_19	C_25																		
62,20169	C_9	C_16																		
64,97684	C_4	C_17																		
80,11779	C_2	C_10	C_15	C_21																
87,95408	C_13	C_26	C_22																	
101,0430	C_12	C_23																		
102,7357	C_1	C_20	C_11																	
122,2342	C_1	C_20	C_11	C_24																
141,9162	C_6	C_9	C_16																	
146,5736	C_3	C_5	C_18	C_19	C_25															
224,4097	C_4	C_17	C_13	C_26	C_22															
282,4123	C_3	C_5	C_18	C_19	C_25	C_6	C_9	C_16												
325,8249	C_1	C_20	C_11	C_24	C_12	C_23														
419,7202	C_2	C_10	C_15	C_21	C_4	C_17	C_13	C_26	C_22											
673,2389	C_1	C_20	C_11	C_24	C_12	C_23	C_7	C_8												
864,5193	C_2	C_10	C_15	C_21	C_4	C_17	C_13	C_26	C_22	C_3	C_5	C_18	C_19	C_25	C_6	C_9	C_16			
1597,284	C_1	C_20	C_11	C_24	C_12	C_23	C_7	C_8	C_2	C_10	C_15	C_21	C_4	C_17	C_13	C_26	C_22	C_3	C_5	C_18
2596,002	C_1	C_20	C_11	C_24	C_12	C_23	C_7	C_8	C_2	C_10	C_15	C_21	C_4	C_17	C_13	C_26	C_22	C_3	C_5	C_18

Рис. 2.11. Схема объединения

Результатом анализа является объединение 26 муниципальных районов в группы, причем основанием группировки будут считаться одновременно пять классификационных признаков.

В данном примере из схемы объединения видим, что в первый кластер были объединены регионы С2 и С10 (Андроповский и Ипатовский районы) на расстоянии 12,79. Во второй кластер – районы С5 и С18 на расстоянии 25,8 и т.д. Графически результаты анализа представлены на рисунке 2.12 в виде дендрограммы.

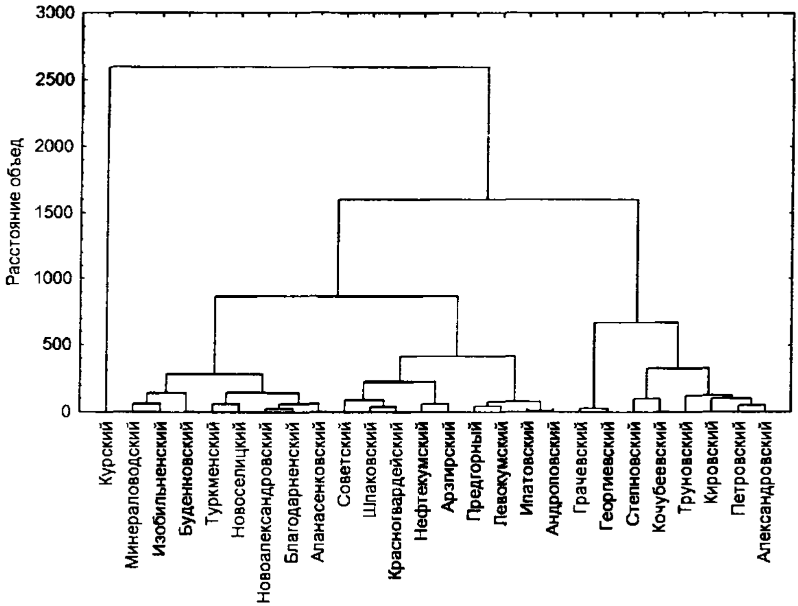


Рис. 2.12. Дендрограмма в *Statistica* 6.0

Анализируя рисунок 2.12, можно сделать вывод, что муниципальные районы Ставропольского края можно разделить на 3 группы по показателям численности и занятости населения в 2012 г. Следует отметить, что Курский район не относится ни к одной из этих групп.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 2.7. По данным временного ряда об инвестициях в сельское хозяйство региона необходимо рассчитать индивидуальные и средние показатели динамики, сделать выводы.

Таблица 2.5

Инвестиции в сельское хозяйство региона (*У*) за период 1998–2008 гг., млн руб.

Период	Инвестиции в сельское хозяйство региона (*У*)
1998	397,7
1999	584,3
2000	808,0
2001	2745,1
2002	2842,0
2003	3440,1
2004	3662,5
2005	3461,9
2006	3957,6
2007	4132,1
2008	4067,2

Задача 2.8. Имея данные о фондовооруженности труда в сельхозпредприятиях за последние 11 лет в регионе, необходимо рассчитать индивидуальные и средние показатели динамики, сделать выводы.

Таблица 2.6

Фондовооруженность труда в СХО региона (*У*) за период 1998–2008 гг., тыс. руб.

Период	Фондовооруженность труда (*У*)
1998	201,4
1999	185,0
2000	147,4

Окончание табл. 2.6

Период	Фондовооруженность труда ($У$)
2001	154,4
2002	152,8
2003	166,1
2004	193,0
2005	209,8
2006	226,1
2007	237,4
2008	260,9

Задача 2.9. Постройте интерполяционную функцию, принимающую в точках $x_0 = 1$, $x_1 = 3$, $x_2 = 6$ заданные значения функции $y_0 = 11$, $y_1 = 15{,}8$, $y_2 = -4$.

Задача 2.10. Найдите многочлен наименьшей степени, принимающий в данных точках $x_0 = 1$, $x_1 = 4$, $x_2 = 7$ значения функции $y_0 = -4{,}6$, $y_1 = -6{,}4$, $y_2 = 24{,}2$.

Задача 2.11. Используя интерполяционный метод Лагранжа, найдите функцию, принимающую в точках $x_0 = 1$, $x_1 = 4$, $x_2 = 7$ заданные значения функции $y_0 = 11{,}6$, $y_1 = 22{,}4$, $y_2 = 11{,}6$.

Задача 2.12. Постройте интерполяционную функцию по значениям решетчато заданной функции:

X	1	3	6
Y	11,6	21,2	17,6

Задача 2.13. Оцените интерполяционную функцию по ее таблично заданным значениям:

X	1	3	6
Y	11,6	21,2	17,6

Задача 2.14. Синтезируйте интерполяционную функцию, принимающую в точках $x_0 = 0, x_1 = 1, x_2 = 2, x_3 = 6$ заданные значения функции $y_0 = -1, y_1 = -3, y_2 = 3, y_3 = 1187$. Оцените погрешность интерполяции в указанных точках: $x = 4; 12$.

Задача 2.15. Постройте многочлен наименьшей степени, принимающий в точках $x_0 = 0, x_1 = 1, x_2 = 2, x_3 = 6$ значения функции $y_0 = 4, y_1 = 9{,}9, y_2 = 18{,}8, y_3 = 240{,}4$. Оцените погрешность интерполяции в указанных точках: $x = 4; 31$.

Задача 2.16. Дана функция $y = \sqrt{x}$, известны три узла интерполяции:

X	49	121	169
Y	7	10	13

Оцените погрешность интерполяции функции при $x = 15; 77; 139$.

Задача 2.17. Дана функция $y = \sqrt[3]{x}$, известны три узла интерполяции:

X	8	125	729
Y	2	5	9

Оцените погрешность интерполяции функции при $x = 13; 237$.

Задача 2.18. Дана функция $y = \sqrt{x^5}$, известны три узла интерполяции:

X	1	4	9
Y	1	32	243

Оцените погрешность интерполяции функции при $x = 2; 5; 7$.

Задача 2.19. По данным таблицы 2.7 проведите иерархический агломеративный кластерный анализ.

Таблица 2.7

Показатели деятельности сельского хозяйства за 2012 г. в разрезе районов Ставропольского края

№ п/п	Район	Прибыль (убыток) сельхозпредприятий, тыс. руб.	Объем с.-х. продукции, произведенной хозяйствами всех категорий – всего, в фактических ценах, млн руб.	Доля растениеводства в общем объеме произведенной с.-х. продукции, %	Производство картофеля во всех категориях хозяйств, тыс. т
1	2	3	4	5	6
1	Александровский	–12587,0	1424,8	63,92	10,4
2	Андроповский	–9867,0	634,8	37,61	10,0
3	Апанасенковский	42305,6	1908,1	51,62	6,0
4	Арзгирский	69602,2	1333,5	69,68	1,0
5	Благодарненский	51125,9	1699,0	66,57	10,1
6	Буденовский	132622,5	2033,2	71,78	4,8
7	Георгиевский	116920,5	2416,2	64,18	11,9
8	Грачевский	–14762,0	1262,1	60,32	7,4
9	Изобильненский	85743,0	2451,6	65,76	41,3
10	Ипатовский	95553,2	2878,3	62,01	0,1
11	Кировский	14578,0	1782,3	62,16	24,6
12	Кочубеевский	141326,5	2902,2	59,01	31,4
13	Красногвардейский	154726,0	3158,0	76,59	15,7
14	Курский	111524,0	1478,8	67,98	1,8
15	Левокумский	116381,0	1969,7	57,98	6,8
16	Минераловодский	15477,2	1219,5	67,55	20,7
17	Нефтекумский	32157,0	1796,9	66,29	9,9
18	Новоалександровски	196126,4	3787,3	76,21	27,1
19	Новоселицкий	201437,0	1878,7	67,04	8,6
20	Петровский	13471,0	2432,1	69,07	16,2
21	Предгорный	66419,4	2248,8	54,11	61,9
22	Советский	289872,2	2496,9	68,75	14,9

Окончание табл. 2.7

1	2	3	4	5	6
23	Степновский	8838,4	1208,6	65,69	1,9
24	Труновский	119266,8	2441,6	62,38	12,3
25	Туркменский	32157,0	1432,1	66,39	1,9
26	Шпаковский	172547,0	4046,8	32,21	30,4

Задача 2.20. По данным таблицы 2.8 проведите иерархический агломеративный кластерный анализ.

Таблица 2.8

Потребление основных продуктов растениеводства за 2012 г. в разрезе районов Ставропольского края

№ п/п	Район	Реализовано зерна сельскохозяйственными предприятиями, тыс. т	Реализовано подсолнечника сельскохозяйственными предприятиями, тыс. т	Реализовано картофеля сельскохозяйственными предприятиями, т	Реализовано овощей сельскохозяйственными предприятиями, т
1	2	3	4	5	6
1	Александровский	64,7	4,8	1	162
2	Андроповский	15,7	1	0	0
3	Апанасенковский	132,4	0,7	60	940
4	Арзгирский	84,5	0,5	1	0
5	Благодарненский	110,5	4	35	133
6	Буденновский	136,7	1,4	140	301
7	Георгиевский	92,3	6	8	123
8	Грачевский	74	3,2	0	5
9	Изобильненский	78,5	9,4	66	3555
10	Ипатовский	121,5	4,8	70	734
11	Кировский	50,2	5,2	0	205
12	Кочубеевский	127,5	10,5	372	564
13	Красногвардейский	173,6	17,1	4	22
14	Курский	78,2	8,1	40	465
15	Левокумский	47,4	0	45	678

Окончание табл. 2.8

1	2	3	4	5	6
16	Минераловодский	21,7	7,4	0	23
17	Нефтекумский	46,1	0	15	90
18	Новоалександровский	291,8	33,9	34	686
19	Новоселицкий	119,6	7	0	14
20	Петровский	126	4,6	69	270
21	Предгорный	47,4	5,4	983	7387
22	Советский	176,4	6,9	278	329
23	Степновский	64,9	0,2	14	23
24	Труновский	166,3	10,6	210	2689
25	Туркменский	65,6	1,1	1	0
26	Шпаковский	50,3	2,3	8	20

Задача 2.21. По данным таблицы 2.9 проведите иерархический агломеративный кластерный анализ.

Таблица 2.9

Показатели производства в животноводстве за 2012 г. в разрезе районов Ставропольского края

№ п/п	Район	Средний надой молока на одну корову в сельхозпредприятиях, кг	Средний сбор яиц от одной курицы-несушки, шт.	Средний настриг шерсти с одной овцы, кг (в физическом весе)	Уровень рентабельности (убыточности) деятельности сельхозпредприятий, %
1	2	3	4	5	6
1	Александровский	2773	0,0	2,3	–5,4
2	Андроповский	1990	0,0	2,5	–11,3
3	Апанасенковский	2268	0,0	4,3	6,7
4	Арзгирский	2041	0,0	5,1	13,1
5	Благодарненский	3801	104,0	3,8	16,7
6	Буденновский	2867	33,0	3,1	14,8
7	Георгиевский	3957	309,0	0,0	10,9
8	Грачевский	1753	307,0	1,5	–10,4
9	Изобильненский	3981	0,0	3,5	15,0

Окончание табл. 2.9

1	2	3	4	5	6
10	Ипатовский	3646	67,0	4,9	15,3
11	Кировский	3862	288,0	4,8	15,3
12	Кочубеевский	5208	148.0	3,1	12,4
13	Красногвардейский	4787	0,0	2,7	17,6
14	Курский	2057	0.0	2,6	43,4
15	Левокумский	2638	0.0	3,1	39,8
16	Минераловодский	1382	0,0	0,0	6,1
17	Нефтекумский	2288	174.0	3,0	16,4
18	Новоалександровский	5314	0,0	2,4	9,4
19	Новоселицкий	3656	0.0	3,7	40,0
20	Петровский	3283	53,0	2,6	7,1
21	Предгорный	4681	307,0	1,2	6,1
22	Советский	4349	0,0	5.4	27,6
23	Степновский	2410	0,0	4,3	2,4
24	Труновский	4940	0,0	3,0	9,9
25	Туркменский	0	0,0	3,9	15,2
26	Шпаковский	2680	245.0	3,6	10,0

Задача 2.22. По данным таблицы 2.10 проведите иерархический агломеративный кластерный анализ.

Таблица 2.10

Финансовые показатели деятельности СХО за 2012 г. в разрезе районов Ставропольского края

№ п/п	Район	Кредиторская задолженность, млн руб.	Дебиторская задолженность, млн руб.	Доля просроченной кредиторской задолженности в общей, %	Доля просроченной дебиторской задолженности в общей, %
1	2	3	4	5	6
1	Александровский	210,6	131,9	11,0	6.4
2	Андроповский	290,3	225,8	13,8	7,4
3	Апанасенковский	266,7	153,7	8,3	10,7
4	Арзгирский	284,5	96,7	3,3	8,9

Окончание табл. 2.10

1	2	3	4	5	6
5	Благодарненский	239,7	223,9	38,2	24,7
6	Буденновский	418,9	224,1	10,8	12,6
7	Георгиевский	813,8	312,4	19,2	5,8
8	Грачевский	485	212,2	23,8	22,9
9	Изобильненский	1724,8	2603,3	17,8	37,5
10	Ипатовский	449,6	261,6	11,0	7,1
11	Кировский	397,2	260,3	12,9	3,5
12	Кочубеевский	525,9	207,9	21,5	17,0
13	Красногвардейский	420,3	410,3	0,4	0,4
14	Курский	139,2	66,1	50,9	15,0
15	Левокумский	361	233,8	16,9	7,7
16	Минераловодский	1957	1677,8	16,0	9,7
17	Нефтекумский	1132,3	292,6	4,7	14,5
18	Новоалександровский	760,3	674,2	18,8	9,2
19	Новоселицкий	86,5	117,2	22,1	11,8
20	Петровский	534,2	466,0	21,8	6,2
21	Предгорный	530,5	304,0	16,4	9,6
22	Советский	446,8	410,3	15,9	9,0
23	Степновский	187,9	26,2	23,3	17,9
24	Труновский	393,7	361,9	14,2	7,2
25	Туркменский	180	66,3	17,2	30,0
26	Шпаковский	808,7	936,8	16,9	11,0

Задача 2.23. По данным таблицы 2.11 проведите иерархический агломеративный кластерный анализ.

Таблица 2.11

Состояние основных фондов в 2012 г. в разрезе районов Ставропольского края

№ п/п	Район	Основные фонды на конец года, млн руб.	Степень износа, %	Тракторы – всего, шт.	Наличие зерноуборочных комбайнов – всего, шт.
1	2	3	4	5	6
1	Александровский	646,3	54,8	427	130

Окончание табл. 2.11

1	2	3	4	5	6
2	Андроповский	718,8	48,9	126	17
3	Апанасенковский	1177,6	48,3	918	311
4	Арзгирский	776,1	39,1	568	181
5	Благодарненский	1216,8	41,4	547	196
6	Буденновский	864,2	41,8	829	286
7	Георгиевский	580,4	28,4	607	139
8	Грачевский	549,8	40,4	203	34
9	Изобильненский	10761,8	79,3	751	171
10	Ипатовский	1250,0	49,6	939	245
11	Кировский	1579,3	33,8	443	95
12	Кочубеевский	2052,6	41,5	846	146
13	Красногвардейский	944,4	32,9	677	207
14	Курский	894,0	37,2	562	166
15	Левокумский	886,7	30,4	621	176
16	Минераловодский	13770,5	27,3	394	80
17	Нефтекумский	9624,9	53,4	351	89
18	Новоалександровский	3589,9	39,0	878	239
19	Новоселицкий	495,6	37,7	483	207
20	Петровский	3151,2	34,5	932	285
21	Предгорный	1398,3	38,2	513	78
22	Советский	1867,9	54,7	769	242
23	Степновский	321,0	38,2	421	122
24	Труновский	1377,7	48,7	953	201
25	Туркменский	520,4	51,0	458	151
26	Шпаковский	1854,1	45,6	331	71

Задача 2.24. По данным таблицы 2.12 проведите иерархический агломеративный кластерный анализ.

Таблица 2.12

Показатели, характеризующие уровень загрязнения окружающей среды в 2012 г. в разрезе районов Ставропольского края

№ п/п	Район	Количество предприятий, имеющих выбросы вредных веществ в атмосферу	Количество источников выбросов — всего	Выброшено в атмосферу всего загрязняющих веществ от стационарных источников, т	Доля твердых веществ в общем объеме выброшенных в атмосферу загрязняющих веществ от стационарных источников, %
1	Александровский	6	370	84	60,7
2	Андроповский	5	166	203	2,0
3	Апанасенковский	4	195	48	77.1
4	Арзгирский	4	287	56	10.7
5	Благодарненский	11	999	459	24.4
6	Буденновский	1	23	114	0,0
7	Георгиевский	10	491	1228	7,6
8	Грачевский	6	204	81	65,4
9	Изобильненский	15	2314	25832	0,8
10	Ипатовский	9	1355	703	8.5
11	Кировский	7	324	289	38.4
12	Кочубеевский	16	755	1693	20,6
13	Красногвардейский	7	416	654	17,0
14	Курский	3	129	228	32,9
15	Левокумский	4	103	78	5,1
16	Минераловодский	6	182	165	0.0
17	Нефтекумский	6	460	7492	0,6
18	Новоалександровский	12	1312	2530	3.5
19	Новоселицкий	4	228	71	64,8
20	Петровский	10	830	203	9.9
21	Предгорный	9	331	254	3,1
22	Советский	12	574	491	30,5
23	Степновский	2	26	14	0,0
24	Труновский	5	371	457	10,3
25	Туркменский	2	111	29	0,1
26	Шпаковский	18	1386	1307	25,0

ГЛАВА

3 СПЕЦИФИКАЦИЯ ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ

3.1. Организация процесса построения эконометрических моделей

Целью рассмотрения этой главы является обоснование применения соответствующих методологических положений для каждого из указанных в первой главе этапов построения эконометрических моделей.

Применительно к постановочным и априорным этапам, на наш взгляд, предложить однозначные методологические концепции представляется затруднительным, поскольку формулировка цели, определение сущности объекта моделирования, формализация заранее известных данных определяются в большей степени экономическими потребностями, наличием времени для решения задачи, мировоззрением эконометриста и т. д.

Вместе с тем следует отметить, что рассматриваемые методологические подходы логически увязывают между собой отдельные этапы, позволяя комплексно подойти к реализации задач, стоящих перед исследователем. Так, проблема отбора наиболее существенных переменных регрессионной модели, решаемая на этапе спецификации, предполагает выделение двух основных подходов:

1) априорное (до построения модели) исследование характера и силы взаимосвязей между рассматриваемыми переменными, по результатам которого в модель включаются факторы, наиболее значимые по своему «непосредственному» влиянию на зависимую переменную. И наоборот, из модели исключаются малозначимые (с точки зрения силы влияния) факторы;

2) статистический анализ факторов, осуществляемый на основе качественных характеристик.

Таким образом, рассматриваемая в этой главе методология позволяет эконометристу достичь цели исследования и решить стоящие перед ним задачи.

Построение эконометрической модели представляет собой многоступенчатый процесс последовательного усовершенствования и расширения модели на основе статистического анализа и экспериментальной верификации результатов. Процесс построения модели представлен на рисунке 3.1.

Очевидно, что определению численных оценок модели должны предшествовать теоретический эконометрический анализ, отбор переменных, предварительная обработка статистических рядов, характеристика связей. Уравнения и переменные могут

Рис. 3.1. Процесс построения и применения эконометрической модели

быть представлены в нескольких вариантах. Экспериментальная проверка вариантов может привести к рассмотрению новых переменных, новых уравнений и новой их верификации. Следующий затем экспериментальный анализ модели в целом может внести коррективы и вызвать необходимость дополнительных проверок.

Опыт работы с моделями показал, что построение эконометрической модели лучше начинать с анализа простых — малоразмерных моделей. Простые укрупненные модели помогают глубже понять экономические процессы и связи.

3.2. Спецификация эконометрических моделей

Следуя цели количественного описания взаимосвязей между экономическими переменными, эконометрика прежде всего связана с методами регрессии и корреляции. В зависимости от числа параметров, включенных в регрессионную модель, различают *парную* и *множественную регрессии.*

Парная регрессия представляет собой регрессию между двумя переменными — x и y и имеет вид:

$$y=f(x), \tag{3.1}$$

где y – зависимая (эндогенная) переменная, результативный признак; x — независимая (экзогенная), или объясняющая, переменная, факторный признак.

Множественная регрессия соответственно представляет собой регрессию результативного признака с двумя и более числом факторов и имеет вид:

$$y=f(x_1,x_2,x_3,\ldots,x_n). \tag{3.2}$$

В целом в регрессионном анализе рассматривается односторонняя зависимость случайной переменной (y) от одной (или нескольких) неслучайной независимой переменной (x).

Из всего круга факторов, оказывающих влияние на результат, необходимо выделить наиболее существенные. Парная (простая) регрессия достаточна, если имеется доминирующий экзогенный признак, который используется в качестве объясняющей переменной.

Уравнение простой регрессии характеризует связь между двумя переменными, которая проявляется как закономерность в среднем и целом по совокупности наблюдений. Практически в каждом отдельном случае фактическая величина результативного признака (y_i) складывается из теоретического значения результативного признака, найденного по уравнению регрессии (y_{x_i}), и остатка (ошибки упрощения, случайной величины), характеризующего отклонение фактического значения результативного признака от расчетного (Δ_i):

$$y_i = y_{x_i} + \Delta_i. \tag{3.3}$$

Случайная величина (возмущение) включает влияние неучтенных в модели факторов, ее присутствие вызвано тремя источниками: спецификацией модели, выборочным характером данных, особенностями измерения переменных. Поэтому величина случайных ошибок зависит от правильного осуществления этапа спецификации модели.

К проблемам спецификации относят два типа задач:

1) определение набора объясняющих переменных;
2) выбор формы уравнения модели.

При спецификации уравнения применяются разные стратегии: сначала в правую часть уравнения вводят одну переменную, которая является наиболее важной с точки зрения экономической теории (парная регрессия). Затем начинают вводить другие переменные и оценивают их экономическую и статистическую значимость (множественная регрессия). Иногда поступают по-другому, вводят в правую часть уравнения все возможные переменные и постепенно исключают из него те факторы, которые являются статистически незначимыми.

При наличии различных теоретических гипотез необходимо исследовать их параллельно.

Оцененные параметры необходимо проверить с точки зрения их экономической интерпретируемости, т. е. выяснить, согласуются ли их абсолютные значения и знаки с основными предпосылками экономической теории. Особенно важно при этом соответствие знаков параметров. Существуют также логические ограничения и на абсолютные значения параметров.

Если оценки параметров находятся в противоречии с экономической теорией, то следует прибегнуть к другой комбинации независимых факторов или преобразовать переменные первого набора. Нереальные значения параметров часто получаются вследствие мультиколлинеарности переменных в правой части. Это происходит из-за того, что зависимость между этими переменными превышает их влияние на результативный признак. В случае мультиколлинеарности необходимо либо преобразовать взаимозависимые переменные, либо исключить некоторые из них. Если присутствие взаимозависимых переменных необходимо, то для оценки параметров следует применить условный метод наименьших квадратов. В соответствии с этим методом заранее фиксируются абсолютные и относительные значения одного или нескольких параметров.

Правильность спецификации модели проверяется также с помощью коэффициента детерминации, корреляции и стандартного отклонения (подробнее об этом — в следующих главах). При сравнении альтернативных регрессионных уравнений предпочтение отдается варианту с наименьшим стандартным отклонением и наибольшим коэффициентом детерминации. Обратная ситуация может означать, что набор переменных недостаточен для описания реальных зависимостей и, возможно, требуется введение новых факторов.

Спецификация уравнения может быть изменена и при исследовании остатка (ошибки) уравнения на автокорреляцию. Автокорреляция показывает, что в уравнении не учтены некоторые важные факторы и, следовательно, модель специфицирована неверно. Если дополнительные переменные не могут быть определены, то для исключения влияния автокорреляции используют квазиразностные преобразования или вводят трендовые переменные.

Наряду с ошибками спецификации могут иметь место ошибки выборки, возникающие в силу неоднородности данных в исходной статистической совокупности. Если совокупность неоднородна, то уравнение регрессии не имеет практического смысла. Решение этой проблемы описано нами выше.

В практическом использовании методов регрессии опасность представляют и ошибки измерения. Если ошибки спецификации можно уменьшить, например, изменяя форму модели, а ошибки выборки – увеличивая объем исходных данных, то ошибки измерения сводят на нет все усилия по количественной оценке связи между признаками.

Предполагая, что все ошибки выборки и измерения сведены к минимуму (используя предложенные методики), основное внимание сконцентрируем на ошибках спецификации модели.

В уравнениях регрессии выбор вида аппроксимирующей функции может быть осуществлен тремя методами:

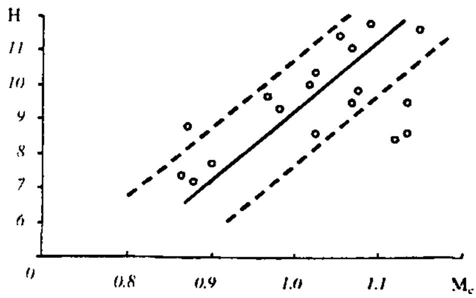
- графическим;
- аналитическим;
- экспериментальным.

В уравнении регрессии необходимо оценить тесноту связи эндогенной переменной с каждой из независимых переменных, что предполагает использование диаграмм рассеивания (корреляционных диаграмм, поля корреляции). В случае парной регрессии выбор осуществляется по графическому изображению реальных данных в виде точек в декартовой системе координат (рис. 3.2). Диаграмма показывает существование зависимости, вид и тесноту связи в исследуемом соотношении.

Поле корреляции представляет собой важный и простой инструмент исследования эконометрических взаимосвязей, однако при окончательной спецификации модели необходимы более точные критерии.

Значительный интерес представляет аналитический метод выбора типа уравнения регрессии. Он основан на изучении материальной природы связи исследуемых признаков.

При автоматизированной обработке информации выбор вида уравнения регрессии обычно осуществляется экспериментальа



б

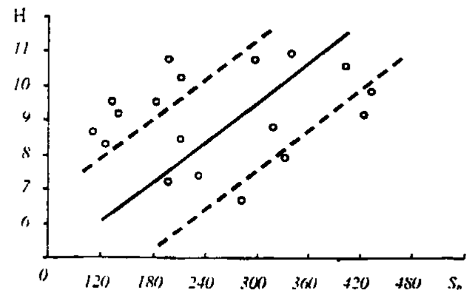


Рис. 3.2. Связь сельскохозяйственной продукции (H) с индексом погодных условий (M_e) и использованием химических удобрений (S_h)

ным методом, путем сравнения величины остаточной дисперсии ($S^2_{\text{ост}}$), рассчитанной при разных моделях:

$$S^2_{\text{ост}} = \frac{1}{n}\sum(y - y_x)^2. \tag{3.4}$$

Чем меньше величина остаточной дисперсии, тем меньше влияние не учитываемых в уравнении факторов и тем лучше уравнение подходит к исходным данным.

При количественной оценке связи между двумя переменными используются следующие классы математических функций:

$y_x = a_0 + a_1 x,$ $\qquad$ $y_x = a_0 + a_1 x + a_2 x^2,$

$y_x = a_0 + \frac{a_1}{x},$ $\qquad$ $y_x = a_0 + a_1 x + a_2 x^2 + a_3 x^3,$

$y_x = a_0 + x^{a_1},$ $\qquad$ $y_x = a_0 + a_1^x.$

Кроме того, используются другие типы кривых:

$$y_x = \frac{1}{a_0 + a_1 x}; \quad y_x = a_0 + a_1 x + a_2 \cdot \frac{1}{x}; \quad y_x = a_0 + a_1 \lg x \quad \text{и т. д.}$$

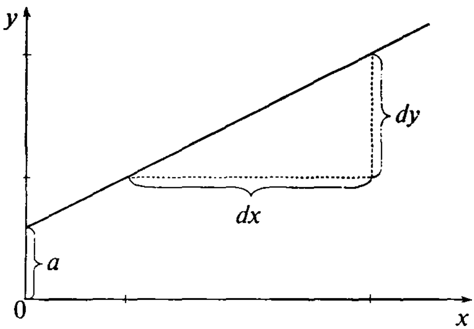


Рис. 3.3. Графическая оценка параметров линейной регрессии

Широкое применение в эконометрике находит линейная регрессия. Линейное уравнение вида $y_x = a_0 + a_1 x$ позволяет по заданным значениям фактора x иметь теоретические значения результативного признака. Построение линейной регрессии сводится к оценке ее параметров — a_0 и a_1. Оценка параметров линейной регрессии может производиться графическим способом (рис. 3.3).

Необходимо обратиться к полю корреляции, выбрать две точки, провести прямую линию, далее по графику определяют значения параметров: параметра a_0 как точку пересечения линии регрессии с осью Oy; параметра a_1 — исходя из угла наклона линии регрессии, как dy/dx, где dy — приращение результата y, а dx — приращение фактора x.

Оценка параметров линейной регрессии, как и любой другой, может достигаться также и другими методами, в частности методом наименьших квадратов.

3.3. Методы отбора факторов при построении регрессионных моделей

Выбор «оптимальных» факторов осуществляется на основе *содержательного* и *количественного* анализа тенденций социально-экономических процессов.

На этапе обоснования эконометрической модели эконометристы сталкиваются с проблемой отбора наиболее предпочтительного состава независимых факторов среди ряда альтернативных вариантов. В начале этой главы мы рассмотрели две ключевые стратегии и два подхода к решению указанной проблемы. Вместе с тем мы считаем необходимым конкретизировать и унифицировать изложенные выше подходы к решению проблемы спецификации. По нашему мнению, эффективное решение задачи отбора факторов достигается при совмещении подходов и стратегий.

Остановимся более подробно на процедурах отбора факторов. Итак, в основе «априорного» подхода лежат следующие предположения:

1) изначальное сильное влияние фактора на зависимую переменную должно подтверждаться определенными количественными характеристиками, например, парный линейный коэффициент корреляции, позволяющий говорить о наличии связи между переменными y и x;

2) если два и более фактора выражают одно и то же явление, то между ними также должна существовать сильная взаимосвязь. В таких ситуациях один из факторов целесообразно исключить из модели, чтобы одна и та же причина не учитывалась дважды.

Такой отбор основан на интуиции и эмпирических наблюдениях, он в большей степени ориентирован на содержательную сторону проблемы. *Содержательный* анализ позволяет решить вопрос о целесообразности включения в модель факторов, основываясь на допущениях экономической теории. Содержательный анализ решет проблему установления самого факта наличия взаимосвязей между явлениями.

Значительно усложняют проблему отбора факторов явления ложной корреляции, которые характеризуются высокими по абсолютной величине значениями коэффициентов парной корреляции у процессов, с содержательной точки зрения не связанных между собой, в тех случаях, когда тенденции рассматриваемых процессов совпали случайно.

Избежать ошибок ложной корреляции можно на основе качественного анализа проблемы, направленного на обоснование адекватного ей содержания и формы модели. При этом число факторов, включенных в модель, не должно быть слишком большим, а простота модели, в свою очередь, гарантирует ее адекватность, так как сложная модель может выражать второстепенные взаимосвязи между переменными в ущерб основным.

При подходе, основанном на статистическом анализе построенного варианта эконометрической модели (апостериорный подход), группу количественных характеристик образуют значения t-критерия Стьюдента, рассчитываемые для параметров уравнения регрессии. С помощью t-критерия проверяется гипотеза о значимости (существенности) влияния фактора на зависимую переменную, тем самым выявляются факторы, удаление которых целесообразно.

На основе апостериорного подхода можно предложить следующую процедуру построения модели:

1. В исходный вариант модели включаются все переменные, отобранные в ходе содержательного анализа.

2. Из модели удаляют незначимые факторы, характеризующиеся меньшими значениями t-критерия Стьюдента, и формируют новый вариант модели. Заметим, что удалять такие факторы необходимо последовательно, так как незначимость большинства из них обусловлена влиянием наихудшего, и на следующем шаге расчетов эти факторы окажутся значимыми.

3. Процесс построения модели завершается, когда в модели остаются значимые факторы, а сама модель удовлетворяет другим критериям качества. В противном случае формируется альтернативный вариант модели.

В случаях большого числа отобранных на этапе содержательного анализа факторов оптимально сочетать оба подхода. С помощью методов априорного отбора формируются альтернативные варианты включенных в модель наборов факторов, а с помощью методов апостериорного отбора эти наборы уточняются и соответствующие им варианты моделей сопоставляются по характеристикам их качества.

Таким образом, отбор факторов, включаемых в уравнение регрессии, является одним из важнейших этапов практического использования методов эконометрического моделирования. Подходы к отбору факторов могут быть различными, наиболее широкое распространение получили следующие методы:

- метод исключения (отсев факторов из полного набора);
- метод включения (дополнительное введение фактора);
- пошаговый отбор переменных (исключение ранее введенного или включение нового фактора).

3.4. Выбор формы уравнения множественной регрессии

Как и в парной регрессии, возможны разные классы аппроксимирующих функций для множественной регрессии, как линейные, так и нелинейные.

Наиболее широко используются линейная и степенная функции (ввиду четкой интерпретации параметров). В линейной множественной регрессии

$$y_x = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n \qquad (3.5)$$

параметры при x называются коэффициентами «чистой» регрессии. Параметры $a_1, a_2, \ldots, a_n$ характеризуют среднее изменение результата вследствие изменения соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне. При этом многие авторы указывают, что a_0 не имеет экономической интерпретации, что, по нашему мнению, неверно. Параметр a_0 характеризует усредненное влияние факторов, не включенных в модель.

В степенной функции

$$y_x = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot x_3^{a_3} \cdot \ldots \cdot x_n^{a_n} \qquad (3.6)$$

коэффициенты $a_1, a_2, a_3, ..., a_n$ являются коэффициентами эластичности, показывающими, на сколько процентов изменится в среднем результативный признак с изменением соответствующего фактора на 1% при неизменности действия других факторов. Этот вид уравнения регрессии получил наибольшее распространение в производственных функциях, исследованиях спроса и потребления.

В производственных функциях имеют смысл не только коэффициенты эластичности каждого фактора, но и их сумма – сумма эластичностей: $A = a_1 + a_2 + a_3 + ... + a_n$. Величина A фиксирует обобщенную характеристику эластичности производства, т. е. с ростом каждого фактора на 1% выпуск продукции увеличивается на величину, равную A.

Например, производственная функция имеет вид:

$$P = 2{,}8 \cdot x_1^{0{,}1} \cdot x_2^{0{,}7} \cdot x_3^{0{,}4},$$

где P — объем производства продукции, тыс. руб.;

x_1 — среднегодовая стоимость основных производственных фондов, тыс. руб.;

x_2 — затраты живого труда, чел.-час;

x_3 — переменные затраты на производство, руб.

Для данного уравнения сумма эластичностей (параметров уравнения) равна $A = 0{,}1 + 0{,}7 + 0{,}4 = 1{,}2\%$, следовательно, при увеличении каждого фактора модели на 1% объем производственной продукции увеличится на 1,2%.

В эконометрическом моделировании применяются и другие линеаризуемые функции для построения уравнения множественной регрессии:

— экспоненциальная: $y_x = e^{a_0 + x_1 a_1 + x_2 a_2 + ... + x_n a_n}$; (3.7)

— гиперболическая: $y_x = \dfrac{1}{a_0 + x_1 a_1 + x_2 a_2 + ... + x_n a_n}$, (3.8)

которая используется при синтезировании моделей с обратными связями.

Как отмечалось ранее, компьютерные программы обработки регрессионного анализа позволяют перебирать различные функции и выбрать наиболее подходящую (по наименьшей остаточной дисперсии, наибольшему коэффициенту детерминации) — *экспериментальный* метод выбора аппроксимирующей функции.

В том случае если эконометриста не устраивает предлагаемый программой набор функций, то можно использовать любые другие функции, приводимые к линейному виду. Например, дана функция:

$$y = a_0 + a_1 x_1 + a_2 x_2^{-2} + a_3\sqrt{x_3} + a_4 \lg x_4,$$

обозначим $x_1 = \beta_1$; $x_2^{-2} = \beta_2$; $\sqrt{x_3} = \beta_3$; $\lg x_4 = \beta_4$; в результате получим линейное уравнение множественной регрессии:

$$y = a_0 + a_1\beta_1 + a_2\beta_2 + a_3\beta_3 + a_4\beta_4.$$

Однако выше мы говорили, что чем сложнее функция, тем ее параметры экономически менее интерпретируемы. Кроме того, высокое число экзогенных и эндогенных переменных в модели требует большого числа наблюдений, вследствие чего полиномиальные функции высоких порядков используются редко.

3.5. Фиктивные переменные

В большинстве случаев рассматриваются регрессионные модели, в которых в качестве объясняющих переменных (регрессоров) выступают количественные переменные (производительность труда, себестоимость продукции, доход и т.п.). Однако на практике достаточно часто возникает необходимость исследовать влияние качественных признаков, имеющих два или несколько уровней (градаций). К числу таких признаков можно отнести пол (мужской, женский), образование (начальное, общее, высшее), фактор сезонности (зима, весна, лето, осень) и т.п.

Качественные признаки могут существенно влиять на структуру линейных связей между переменными и приводить к скачкообразному изменению параметров регрессионной модели. В этом случае говорят об исследовании регрессионных моделей с переменной структурой или о построении регрессионных моделей по неоднородным данным.

Например, нам надо изучить зависимость заработной платы работников не только от количественных факторов $x_1, x_2, x_3, \ldots, x_n$, но и от качественного признака Z_1 (например, фактор «пол работников»).

В принципе можно было получить оценки регрессионной модели

$$y_x = a_0 + a_1 x_{i1} + a_2 x_{i2} + \ldots + a_n x_{in} + \Delta_i, \; i = 1,2,\ldots,n$$

для каждого уровня качественного признака (т.е. выборочное уравнение регрессии для работников — отдельно для мужчин и для женщин), а затем изучить различия между ними.

Но есть и другой подход, позволяющий оценивать влияние значений количественных переменных и уровней качественных признаков с помощью одного уравнения регрессии. Этот подход связан с введением так называемых *фиктивных* (структурных, манекенных) переменных.

В качестве фиктивных переменных обычно используют *дихотомические (бинарные)* переменные, которые принимают всего два значения «0» и «1».

В этом случае первоначальная регрессионная модель заработной платы изменится и примет следующий вид:

$$y_x = a_0 + a_1 x_{i1} + a_2 x_{i2} + \ldots + a_n x_{in} + \alpha_1 Z_{i1} + \Delta_i, \; i = 1,2,\ldots,n, \quad (3.9)$$

где $Z_{i1} = \begin{cases} 1, \text{ если } i\text{-й работник мужского пола;} \\ 0, \text{ если } i\text{-й работник женского пола.} \end{cases}$

Таким образом, можно считать, что средняя заработная плата у мужчин на $\alpha_1 \cdot 1 = \alpha_1$ выше, чем у женщин, при неизменных значениях других параметров модели.

Если рассматриваемый качественный признак имеет несколько (k) уровней (градаций), то в принципе можно было бы ввести в регрессионную модель дискретную переменную, принимающую такое же количество значений (например, при исследовании зависимости заработной платы y от уровня образования Z можно рассматривать $k=3$ значения: $Z_{i1}=1$ при наличии начального образования, Z_{i2} — среднего, Z_{i3} — при наличии высшего образования). Однако обычно так не поступают из-за трудности содержательной интерпретации соответствующих коэффициентов регрессии, а вводят ($k-1$) бинарных переменных.

В рассматриваемом примере для учета фактора образования можно ввести в регрессионную модель $k-1=3-1=2$ бинарных переменных Z_{21} и Z_{22}:

$$y_x = a_0 + a_1 x_{i1} + a_2 x_{i2} + \ldots + a_n x_{in} + \alpha_1 Z_{i1} + \alpha_{21} Z_{i21} + \alpha_{22} Z_{i22} + \Delta_i, \tag{3.10}$$

где

$$Z_{i21} = \begin{cases} 1, \text{ если } i\text{-й работник имеет высшее образование;} \\ 0, \text{ — во всех остальных случаях;} \end{cases}$$

$$Z_{i22} = \begin{cases} 1, \text{ если } i\text{-й работник имеет среднее образование;} \\ 0, \text{ — во всех остальных случаях.} \end{cases}$$

Третьей бинарной переменной не требуется (Z_{23}): если *i-й* работник имеет начальное образование, это будет отражено парой значений $Z_{i21}=0$, $Z_{i22}=0$.

Более того, вводить третью бинарную переменную Z_{23} (со значением $Z_{i23}=1$, если *i-й* работник имеет начальное образование, $Z_{i23}=0$ во всех остальных случаях) нельзя, так как при этом для любого i-го работника $Z_{i21}+Z_{i22}+Z_{i23}=1$, т.е. при суммировании элементов столбцов общей матрицы соответствующих фиктивных переменных Z_{21}, Z_{22}, Z_{23} получили бы столбец, состоящий из одних единиц. А так как такой столбец уже есть, соответствующий свободному члену уравнения регрессии, то это означало бы линейную зависимость значений. Такая ситуация, когда сумма значений нескольких переменных, включенных в модель, равна постоянному члену (единице), получила название «ловушка» (*dummytrap*). Чтобы избежать такой ловушки, число вводимых бинарных переменных должно быть на единицу меньше числа уровней (градаций) качественного признака.

Рассмотренные выше регрессионные модели отражали влияние качественного признака (фиктивных переменных) только на значения переменной y, т.е. на свободный член уравнения регрессии. В более сложных моделях может быть отражена также зависимость фиктивных переменных на сами параметры при переменных регрессионных моделей. Например, при наличии в модели объясняющих переменных – количественной x_1 и фиктивных Z_{11}, Z_{12}, Z_{21}, Z_{22}, из которых Z_{11}, Z_{12} влияют только на значение коэффициента при x_1, а Z_{21}, Z_{22} – только на величину свободного члена уравнения, такая регрессионная модель будет иметь вид:

$$
\begin{aligned}
y_x = a_0 + a_1 x_{i1} + a_{11}(Z_{i11} x_{i1}) + a_{12}(Z_{i12} x_{i1}) + \\
+ \alpha_{21} Z_{i21} + \alpha_{22} Z_{i22} + \Delta_i , \\
i = 1,2,\ldots,n . \qquad (3.11)
\end{aligned}
$$

Модели подобного типа используются, например, при исследовании зависимости объема потребления y некоторого продукта от дохода потребителя x, когда одни качественные признаки (например фактор сезонности) влияют лишь на количество потребления продукта (свободный член уравнения регрессии), а другие (уровень доходности домашнего хозяйства) — на параметр a_1 при факторном признаке x, интерпретируемый как «склонность к потреблению».

3.6. Основные эконометрические модели и их типы

Эконометрическая модель есть инструмент исследования и измерения количественных связей между экономическими величинами. Она широко применяется в бизнесе, экономике, общественных науках, исследовании экономической активности и даже в исследовании политических процессов.

Эконометрические модели полезны для более полного понимания сущности происходящих процессов, их анализа. Модель, построенная и верифицированная (проверенная на достоверность) на основе уже имеющихся наблюденных значений, объясняющих перемены в будущем, или для других наборов значений объясняющих переменных.

В настоящее время все чаще применяются *комплексные эконометрические модели*, под которыми понимают систему регрессионных уравнений и тождеств, отражающих основные связи между макроэкономическими величинами во всех элементах процесса воспроизводства. Цель комплексной эконометрической модели состоит в анализе этих связей и прогнозировании развития народного хозяйства.

Комплексные эконометрические модели могут быть *краткосрочными* и *долгосрочными*. С помощью долгосрочных моделей изучают источники и тенденции экономического развития. Для краткосрочных моделей типичны детальное расчленение уравнений, переменных, временных интервалов, использование балансовых тождеств.

Эконометрические модели делятся на *статические* и *динамические*. В статической модели рассматриваются связи между явлениями, происходящими в данный момент, причем изменение во времени не играет роли. В динамической модели, напротив, взаимосвязи рассматриваются в развитии.

Эконометрические модели делятся *также по уровню агрегирования* (объединения) переменных. В агрегированных моделях изучаются связи между такими макроэкономическими показателями, как, например, национальный доход, потребление, основные фонды и т.д. Иногда применяются частично дезагрегированные модели, в которых расчленению подвергаются обычно производственные функции или функции потребления.

Иногда эконометрические модели классифицируют по *способу выражения переменных*: в постоянных или текущих ценах, годовые или квартальные, в абсолютных значениях показателей или приростах.

По целям применения эконометрические модели делят на *учебные, экспериментальные* и *операционные*. Соответственно этой классификации модели могут быть *аналитическими, имитационными* и *прогностическими*. И, наконец, модели различают по *числу входящих* в них *уравнений и переменных*.

Итак, можно выделить три основных типа (класса) эконометрических моделей, которые применяются для анализа и прогноза: модели временных рядов, регрессионные модели с одним уравнением, системы одновременных уравнений.

Модели временных рядов, к этому классу относят модели:

а) тренда:

$$y(t) = T(t) + \Delta_t, \tag{3.12}$$

где $T(t)$ — временной тренд заданного параметрического вида (например, линейный $T(t) = a_0 + a_1 t$);

Δ_t — случайная (стохастическая) компонента;

б) сезонности:

$$y(t) = S(t) + \Delta_t, \tag{3.13}$$

где $S(t)$ — периодическая (сезонная) компонента;

в) тренда и сезонности:

$$y(t) = T(t) + S(t) + \Delta_t \text{ (аддитивная) или} \tag{3.14}$$

$$y(t) = T(t) \cdot S(t) \cdot \Delta_t \text{ (мультипликативная).} \tag{3.15}$$

К моделям временных рядов относится множество более сложных моделей, таких как модели адаптивного прогноза, модели авторегрессии и скользящего среднего (*ARIMA*) и другие. Их общей чертой является то, что они объясняют поведение временного ряда исходя только из его предыдущих значений. Такие модели могут применяться, например, для изучения и прогнозирования объема продаж авиабилетов, краткосрочного прогноза процентных ставок и т.п.

Эконометрические модели с одним уравнением

В таких моделях зависимая (объясняемая) переменная y представляется в виде функции:

$$f(x,\beta) = f(x_1,\ldots,x_k,\beta_1,\ldots,\beta_p), \tag{3.16}$$

где $x_1,\ldots,x_k$ — независимые (объясняющие) переменные;

$\beta_1,\ldots,\beta_p$ — параметры.

В зависимости от вида функции $f(x,\beta)$ модели делятся на *линейные* и *нелинейные*. Например, можно исследовать спрос на мороженое как функцию от температуры воздуха, среднего уровня доходов и т.п. Область применения таких моделей значительно шире, чем моделей временных рядов.

Системы одновременных уравнений

Эти модели описываются системами уравнений. Системы могут состоять из тождеств и регрессионных уравнений, каждое из которых может кроме объясняющих переменных включать в себя также объясняемые переменные из других уравнений системы. Таким образом, имеется набор объясняемых переменных, связанных через уравнения системы. Примером может служить модель спроса и предложения. Системы одновременных уравнений требуют относительно более сложный математический аппарат. Они могут использоваться для моделей макроэкономики страны и в других случаях.

Модель спроса и предложения

Пусть Q^D — спрос на товар (*demand*); Q^S — предложение товара; P — цена товара (*pricelevel*); I — доход (*income*). Составим следующую систему уравнений «спрос— предложение»:

$$\begin{cases} Q^D = \alpha_0 + \alpha_1 P + \alpha_2 I + \Delta_1, & \text{(спрос)} \\ Q^S = \beta_0 + \beta_1 P + \Delta_2, & \text{(предложение)} \\ Q^D = Q^S. & \text{(равновесие)} \end{cases}$$

Таким образом, цена товара P и спрос на товар $Q^D = Q^S = Q$ определяются из уравнений модели, т.е. являются объясняемыми переменными, а величина дохода I — объясняющей переменной.

Разделение ролей между переменными в системе одновременных уравнений может быть проинтерпретировано следующим образом: переменные Q и P формируют свои значения внутри модели. Такие переменные называются *эндогенными*. Между тем переменная I считается в уравнении заданной, ее значения формируются вне модели. Такие переменные называются *экзогенными*.

3.7. Применение эконометрических моделей

Управление народным хозяйством требует комплексного анализа экономики. Эконометрические модели, построенные для этой цели, должны отвечать следующим условиям:

а) базироваться на положениях экономической теории;

б) содержать в себе информацию, необходимую для организации процесса воспроизводства;

в) давать широкий спектр альтернативных решений;

г) исходить из статистического анализа тенденций развития и взаимосвязей макроэкономических показателей;

д) быть гибкими и без труда применяться на практике.

Эконометрическая модель служит для определения и прогнозирования основных пропорций в экономике и как инструмент принятия экономических решений, обеспечивающих соблюдение этих пропорций.

Выбор варианта решения можно проводить с помощью так называемой имитационной модели, когда определяется множество вариантов решений и из них выбирается лучший в отношении принятого критерия. Для этой цели может оказаться полезным объединение регрессионной эконометрической и оптимизационной моделей.

Построение эконометрических моделей особенно необходимо в условиях рынка, насыщенного конкурирующими участниками, с медленными (товарными), среднего темпа (финансовыми) и быстрыми (информационными) потоками на нем.

С ростом объемов экономического программирования и индикативного планирования качества и затрат прогнозирование с вероятностным характером своих переменных становится все более важным этапом менеджерского проекта.

Бум прогнозирования пришелся на рубеж 50–60-х годов XX века, когда прогнозы использовались в политике капитальных вложений и научных исследованиях, в борьбе за рынки источников сырья, а государственное программирование активно воздействовало на эти тенденции, особенно в долговременных прогнозах.

Методологии прогнозов присущи общие черты. Все они в той или иной мере используют экстраполяцию прошлых тенденций в отношении как общенациональных, так и частичных показателей производства, народонаселения, технического прогресса. Общая черта эконометрических и эмпирических прогнозов – стремление на основе отдельных, частичных экономических показателей составить общую картину будущего экономического роста.

Одной из первых комплексных эконометрических моделей была модель Клейна–Гольдбергера. Она послужила фундаментом, на котором базировались некоторые краткосрочные модели комплексного развития. Эта модель состояла из 15 регрессионных уравнений и 5 тождеств и охватывала 40 макроэкономических показателей. Параметры модели были оценены на базе временных рядов за 20 лет.

Контрольные вопросы

1. Цели и задачи спецификации эконометрических моделей.
2. Организация процесса построения эконометрического моделирования.
3. Спецификация моделей парной регрессии.
4. Стратегии спецификации эконометрических моделей.
5. Критерии качества спецификации эконометрических моделей.
6. Методы выбора вида аппроксимирующей функции.
7. Основные классы аппроксимирующих функций.
8. Методы отбора факторов эконометрических моделей.
9. Содержательный и количественный анализ переменных.
10. Априорные и апостериорные подходы к отбору факторов.

☑ ТЕСТОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ ПРОВЕРКИ ЗНАНИЙ

1. На основании наблюдений за 100 домохозяйствами построено уравнение регрессии $y=145{,}65+0{,}825\cdot x$, где y — потребление, x — доход. Соответствуют ли знаки и значения коэффициентов регрессии теоретическим представлениям?

а) нет;
б) да;
в) частично соответствуют.

2. Парная регрессия представляет собой модель вида:

а) $y=f(x_1, x_2, ..., x_m)$;
б) $y=f(x)$;
в) $y=f(y_{t-1})$.

3. Уравнение парной регрессии характеризует связь между:

а) несколькими переменными;
б) двумя переменными;
в) тремя и более переменными.

4. Согласно содержанию регрессии наблюдаемая величина зависимой переменной складывается из теоретического значения зависимой переменной, найденного из уравнения регрессии:

а) скорректированного на величину стандартной ошибки;
б) и случайного отклонения;
в) и остаточной дисперсии.

5. Использование парной регрессии вместо множественной является примером ошибки:

а) выборки;
б) спецификации;
в) измерения.

6. Аналитический метод подбора вида уравнения регрессии основан на:
а) изучении поля корреляции;
б) изучении природы связи признаков;
в) сравнении величины остаточной дисперсии при разных моделях.

7. Экспериментальный метод подбора вида уравнения регрессии основан:
а) на изучении поля корреляции;
б) сравнении величины остаточной дисперсии при разных моделях;
в) изучении природы связи признаков.

8. Классический подход к оцениванию коэффициентов регрессии основан:
а) на графической оценке;
б) методе наименьших квадратов;
в) методе максимального правдоподобия.

9. Графический метод подбора вида уравнения регрессии основан:
а) на изучении природы связи признаков;
б) изучении поля корреляции;
в) сравнении величины остаточной дисперсии при разных моделях.

10. В большинстве случаев зависимости между экономическими переменными являются:
а) функциональными;
б) стохастическими;
в) строгими.

11. Если не наблюдается смена направленности связи признаков, параболу нелинейной функции можно заменить:
а) гиперболой;
б) степенной функцией;
в) логистической функцией.

12. Для обеспечения хорошего качества эконометрической модели в нее не должны включаться коллинеарные:

а) коэффициенты регрессии;
б) случайные величины;
в) факторы;
г) результаты.

13. Для отбора факторов множественной линейной модели регрессии рассматривается вопрос о взаимосвязи фактора и результата при неизменности прочих факторов, которые фиксируются, как правило, на среднем уровне. В этом случае используется:

а) матрица множественных коэффициентов корреляции;
б) коррелограмма для факторов модели;
в) автокорреляционная функция;
г) матрица частных коэффициентов корреляции.

14. На практике о наличии мультиколлинеарности обычно судят:

а) по величине математического ожидания;
б) количеству факторов в модели;
в) коэффициенту детерминации;
г) матрице парных коэффициентов корреляции.

15. Обобщенный метод наименьших квадратов используется для корректировки:

а) гетероскедастичности остатков в уравнении регрессии;
б) мультиколлинеарности между независимыми переменными;
в) параметров нелинейного уравнения регрессии;
г) точности определения коэффициента множественной корреляции.

ПРАКТИКУМ

Задача 3.1. На основе имеющихся данных определите набор независимых переменных, необходимых к включению в модель множественной линейной регрессии.

Таблица 3.1

Данные к задаче в разрезе муниципальных районов Ставропольского края за 2012 г.

№ района	y	x_1	x_2	x_3	x_4	x_5
1	188,2	3685	2907,1	17,5	7,6	2773
2	189,9	5460	3006,3	18,4	7,9	1990
3	190,4	2803	3133	20,1	10,8	2268
4	189,3	3164	2878,7	10,4	5,6	2041
5	190,8	3892	2943,5	22,0	5,2	3801
6	191,1	2831	2998,6	16,9	4,6	2867
7	205,8	4831	3062,9	16,6	3,9	3957
8	190,9	3932	2904,9	12,8	4,4	1753
9	191,2	5584	3228,2	24,9	5,1	3981
10	190,2	3749	2727,3	32,7	12,7	3646
11	192,6	5074	3061,4	17,7	4,3	3862
12	192	4644	2967,2	57,5	9,8	5208
13	191,3	3861	2977,4	18,8	4,3	4787
14	193,8	3530	2710,6	14,0	6,9	2057
15	191,4	3395	2966,1	19,6	13,6	2638
16	191,1	4713	3162,8	14,4	4,5	1382
17	192,1	5620	2860,3	17,0	10,3	2288
18	189,7	4003	2978	30,4	5,8	5314
19	192,6	3475	2879,6	16,5	4,3	3656
20	190,6	4317	2922,5	21,9	6,7	3283
21	193,4	5260	2986,2	30,9	6,6	4681
22	190,8	3832	2990,6	25,9	5,1	4349
23	191,5	3163	2804,7	14,8	4,1	2410
24	189,3	4275	2996,1	29,2	5	4940
25	190,4	3139	2882,2	12,7	7	3568
26	197,6	5186	2933,1	24,9	6,8	2680

y — фактическое потребление молочных продуктов на 1 человека в год, кг;

x_1 — среднемесячная заработная плата работающих, руб.;

x_2 — средний размер месячных пенсий всех пенсионеров, руб.;

x_3 — производство молока (заданной жирности), тыс. т;

x_4 — численность коров (в том числе) во всех категориях хозяйств на начало года, тыс. голов;

x_5 — средний надой молока на одну корову в сельхозпредприятиях, кг.

Решение

Для определения набора факторов, которые могут быть включены в модель, воспользуемся методом исключения переменных. Для этого воспользуемся средством «Линейная регрессия» *SPSSStatistics 20.* С этой целью необходимо в меню «Анализ» выбрать пункт «Регрессия», в котором находится инструмент «Линейная регрессия» (рис. 3.4).

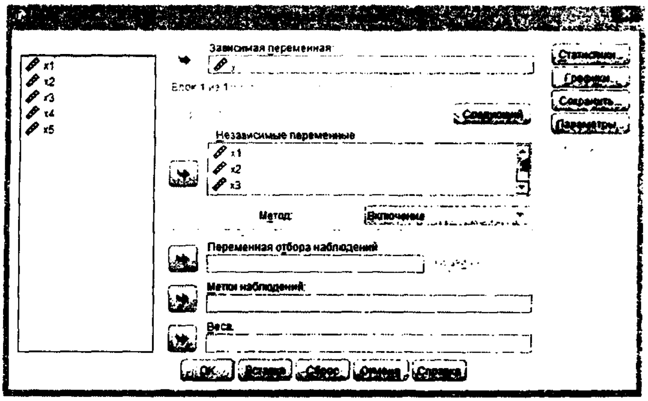


Рис. 3.4. Средство «Линейная регрессия» в *SPSSStatistics* 20

В диалоговом окне данного инструмента выберем метод исключения переменных, отметим набор всех переменных и нажмем «ОК». В результате получим несколько возможных моделей с различным набором факторов (рис. 3.5).

Исходя из полученных результатов в первоначальный вариант модели были включены все факторные переменные. Полученная модель имеет вид: $y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5$.

На следующем этапе из модели был исключен фактор x_2 как наименее значимый. В результате модель была преобразована в уравнение вида: $y = a_0 + a_1x_1 + a_3x_3 + a_4x_4 + a_5x_5$.

Далее из этого варианта модели был исключен фактор x_4 как незначимый для такого набора независимых переменных. Модель на этом этапе имеет вид: $y = a_0 + a_1x_1 + a_3x_3 + a_5x_5$.

Аналогично были построены модели следующих видов: $y = a_0 + a_1x_1 + a_3x_3$ и $y = a_0 + a_1x_1$.

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знч.
		B	Стд. Ошибка	Бета		
1	(Константа)	192,060	18,859		10,184	,000
	x1	,001	,001	,368	1,583	,129
	x2	-,002	,007	-,060	-,265	,794
	x3	-,066	,124	-,186	-,532	,601
	x4	-,136	,326	-,110	-,418	,680
	x5	,000	,001	,131	,410	,686
2	(Константа)	187,229	4,621		40,514	,000
	x1	,001	,001	,348	1,620	,120
	x3	-,067	,121	-,188	-,549	,589
	x4	-,120	,312	-,096	-,383	,705
	x5	,000	,001	,128	,410	,686
3	(Константа)	186,143	3,579		52,015	,000
	x1	,001	,001	,365	1,774	,090
	x3	-,094	,096	-,265	-,978	,339
	x5	,001	,001	,190	,718	,480
4	(Константа)	187,131	3,269		57,250	,000
	x1	,001	,001	,359	1,764	,091
	x3	-,049	,072	-,138	-,679	,504
5	(Константа)	186,696	3,169		58,916	,000
	x1	,001	,001	,321	1,659	,110

а. Зависимая переменная: y

Рис. 3.5. Результат спецификации модели линейной множественной регрессии средствами *SPSSStatistics 20*

Для полученных моделей экспериментальным способом были определены показатели статистической значимости, которые позволили сделать вывод, что оптимальным набором факторных переменных для даной эконометрической модели является линейная зависимость результативной переменной от независимой переменной x_1. В таком случае уравнение будет иметь вид $y = a_0 + a_1 x_1$.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 3.2. По данным таблицы 3.3 выполнить спецификацию эконометрической модели для представленных значений переменных.

Таблица 3.3

Данные о производстве мяса, заготовке кормов и падеже КРС по СХП в разрезе районов Ставропольского края в 2012 г.

№ района	Производство мяса (в живом весе) всех видов во всех категориях хозяйств, т	Заготовлено кормов в СХП в расчете на одну условную голову скота, ц корм. ед.	Падеж крупного рогатого скота в СХП, %
	Y	X_1	X_2
1	2	3	4
1	4038,0	22,0	0,5
2	5791,0	25,4	3,2
3	6580,0	22,8	1,3
4	4078,0	29,1	2,1
5	9322,0	24,5	1,1
6	5748,0	23,4	2,0
7	13195,0	23,9	2,3
8	7786,0	32,7	0,5
9	3586,0	37,1	2,6
10	6510,0	26,0	0,7

Окончание табл. 3.3

1	2	3	4
11	10307,0	24,5	1,2
12	8718,0	27,4	2,7
13	7105,0	45,0	2,9
14	9526,0	24,2	1,3
15	4844,0	22,0	0,5
16	11685,0	22,8	2,9
17	13195,0	23,9	2,9
18	7786,0	32,7	0,6

Задача 3.3. На основании данных таблицы 3.4 осуществите спецификацию эконометрической модели, сделайте вывод о наиболее существенных факторных переменных.

Таблица 3.4

Экономические показатели деятельности муниципальных районов региона за 2012 г.

№ района	Y	X_1	X_2
1	328,1	3685	1,026
2	2,6	5460	1,095
3	309,2	2803	0,835
4	244,1	3164	0,883
5	234,5	3892	0,963
6	414,5	2831	0,967
7	608,5	4831	1,087
8	20,8	3932	0,916
9	299,3	5584	1,066
10	567,2	3749	0,965
11	241,8	5074	1,009
12	112,5	4644	1,160
13	154,7	3861	1,157
14	111,5	3530	0,897
15	116,4	3395	0,944

Примечание. Y — прибыль (убыток) сельхозпредприятий, млн руб.; X_1 — среднемесячная заработная плата работающих, руб.; X_2 — соотношение среднесложившейся цены 1 ц зерна к средней по краю.

Задача 3.4. Построить уравнения множественной регрессии на основании таблицы 3.5.

Таблица 3.5

Показатели деятельности животноводства в разрезе районов края за 2012 г.

№ района	Коэффициент обеспеченности мясными продуктами населения	Доля животноводства в общем объеме произведенной с.-х. продукции, %	Производство мяса всех видов (в живом весе) во всех категориях хозяйств, т	Численность свиней на начало года, тыс. гол.
	Y	X_1	X_2	X_3
1	0,654	36,08	5411,0	14,0
2	0,660	62,39	4038,0	4,2
3	0,662	48,38	8283,0	30,2
4	0,658	30,32	5647,0	9,6
5	0,663	33,43	4253,0	11,8
6	0,664	28,22	5791,0	15,8
7	0,715	35,82	6580,0	21,3
8	0,664	39,68	4078,0	9,5
9	0,665	34,24	9322,0	23,1
10	0,661	37,99	7419,0	19,8
11	0,670	37,84	5748,0	30,8
12	0,667	40,99	13195,0	29,9
13	0,665	23,41	7786,0	30,7
14	0,674	32,02	5052,0	18,0
15	0,665	42,02	9472,0	10,6
16	0,664	32,45	3586,0	5,3
17	0,668	33,71	6510,0	2,4
18	0,660	23,79	10307,0	39,4
19	0,670	32,96	6239,0	39,0
20	0,663	30,93	8718,0	41,3

ГЛАВА

4 ОЦЕНКА ПАРАМЕТРОВ РЕГРЕССИОННЫХ УРАВНЕНИЙ

4.1. Метод наименьших квадратов (МНК)

В практике эконометрических исследований имеющиеся данные не всегда можно считать выборкой из многомерной нормальной совокупности, когда одна из рассматриваемых переменных не является случайной или когда линия регрессии явно не прямая и т.п. В этих случаях пытаются определить кривую (поверхность), которая дает наилучшее (в смысле метода наименьших квадратов) приближение к исходным данным. Соответствующие методы приближения получили название *регрессионный анализ.*

Методы и модели регрессионного анализа занимают центральное место в математическом аппарате эконометрики. *Задачами регрессионного анализа* являются:

а) установление формы зависимости между переменными;

б) оценка функции регрессии (или оценка параметров или коэффициентов уравнения регрессии);

в) оценка неизвестных значений (прогноз значений) зависимой переменной.

В регрессионном анализе рассматривается односторонняя зависимость случайной переменной y от одной (или нескольких) неслучайной независимой переменной x. Такая зависимость y от x может быть представлена в виде *модельного уравнения регрессии* (или просто *уравнения регрессии*).

Для точного описания уравнения регрессии необходимо знать условный закон распределения зависимой переменной y. В статистической практике такую информацию получить, как правило, не удается, так как обычно исследователь располагает лишь выборкой значений (x_i, y_i) ограниченного объема n. В этом случае речь может идти об *оценке* (приближенном выражении, аппроксимации) по выборке функции регрессии.

Классический подход к оцениванию параметров линейной регрессии с помощью метода наименьших квадратов (МНК) целесообразно рассмотреть на примере линейной парной регрессии:

$$y_x = a_0 + a_1 x \quad \text{или} \quad y = a_0 + a_1 x + \Delta, \tag{4.1}$$

где y — фактическое значение результативного признака;

y_x — теоретические значения результативного признака, найденные из уравнения регрессии путем подстановки в него фактических значений фактора x;

a_0, a_1 — параметры (или коэффициенты) уравнения регрессии;

Δ — случайная величина (или возмущение, ошибка), характеризующая отклонение фактического значения результативного признака от теоретического, найденного по уравнению регрессии.

Итак, пусть имеется два ряда эмпирических данных x $(x_1, x_2, \ldots, x_n)$ и y $(y_1, y_2, \ldots, y_n)$, соответствующие им точки с координатами (x_i, y_i), где $i = 1, 2, \ldots n$, отобразим на координатной плоскости. Такое изображение называется ***полем корреляции*** (рис. 4.1). По расположению эмпирических точек можно предположить наличие линейной корреляционной зависимости между переменными x и y.

Построение линейной регрессии сводится к оценке ее параметров a_0 и a_1. Оценки параметров линейной регрессии могут быть найдены различными методами (например ***метод наименьших модулей***, где $\sum_{i=1}^{n} |y_i - y_{x_i}|$). Однако существенно простым при проведении вычислительной процедуры и хорошим по статистическим свойствам оценки является МНК.

Согласно МНК неизвестные параметры a_0 и a_1 получают таким образом, чтобы сумма квадратов отклонений фактических значений y_i от значений y_{x_i}, найденных по уравнению регрессии, была минимальной:

$$S = \sum_{i=1}^{n} (y_i - y_{x_i})^2 \to \min. \tag{4.2}$$

Иначе говоря, из всего множества линий линия регрессии на графике выбирается таким образом, чтобы сумма квадратов расстояний по вертикали между точками и этой линией была минимальной (рис. 4.1):

$$\Delta_i = y_i - y_{x_i}, \qquad (4.3)$$

следовательно,

$$S = \sum_{i=1}^{n} \Delta_i^2 \to \min. \qquad (4.4)$$

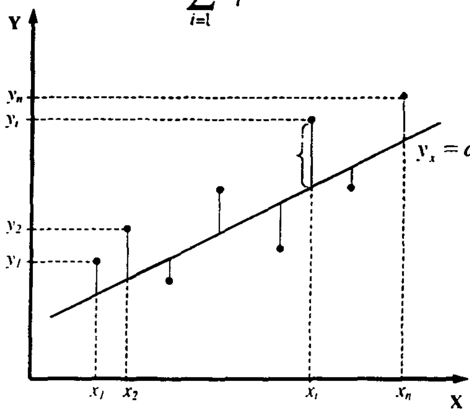


Рис. 4.1. Линейная регрессия

Чтобы найти минимум функции, надо вычислить частные производные по каждому из параметров a_0 и a_1 и приравнять их к нулю, тогда:

$$S = \sum_{i=1}^{n} \Delta_i^2 = \sum_{i=1}^{n} (y_i - y_{x_i})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x)^2; \qquad (4.5)$$

$$\frac{dS}{da_0} = -2\sum (y_i - a_0 - a_1 x_i) = 0 \text{ или } \sum y_i - na_0 - a_1 \sum x_i = 0; \qquad (4.6)$$

$$\frac{dS}{da_1} = -2\sum (y_i - a_0 - a_1 x_i) x_i = 0$$

$$\text{или } \sum x_i y_i - a_0 \sum x_i - a_1 \sum x_i^2 = 0. \qquad (4.7)$$

В результате преобразования получим следующую систему нормальных уравнений для оценки параметров a_0 и a_1:

$$\begin{cases} na_0 + a_1\sum x_i = \sum y_i, \\ a_0\sum x_i + a_1\sum x_i^2 = \sum x_i y_i. \end{cases} \tag{4.8}$$

Решая систему нормальных уравнений методом подстановки, последовательного исключения переменных либо методом определителей, найдем искомые оценки параметров a_0 и a_1.

Можно воспользоваться следующими готовыми формулами, которые получают, разделив обе части уравнения системы на n, подставляя выраженное значение:

$$a_0 = \bar{y} - a_1\bar{x}, \tag{4.9}$$

из первого уравнения системы в уравнение линейной парной регрессии, получим:

$$a_1 = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\bar{x}^2 - x^{-2}} = \frac{\operatorname{cov}(x, y)}{S_x^{\,2}}, \tag{4.10}$$

где $\operatorname{cov}(x,y)$ — ковариация признаков (или корреляционный момент):

$$\operatorname{cov}(x,y) = \overline{xy} - \bar{x}\cdot\bar{y} = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}\cdot\bar{y}; \tag{4.11}$$

$S_x^{\,2}$ – дисперсия признака x:

$$S_x^{\,2} = \bar{x}^2 - x^{-2} = \frac{\sum_{i=1}^{n} x_i^2}{n} - (\bar{x})^2 = \frac{\sum_{i=1}^{n}(x - \bar{x})^2}{n}. \tag{4.12}$$

Если между экономическими явлениями существуют нелинейные соотношения, то они выражаются с помощью нелинейных регрессий, которые подразделяются на два класса:

1) регрессии нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам:

а) полиномы разных степеней:

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \Delta;$$

$$y = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \Delta; \quad (4.13)$$

б) равносторонняя гипербола:

$$y = a_0 + \frac{a_1}{x} + \Delta; \quad (4.14)$$

2) регрессии нелинейные по оцениваемым параметрам:

а) нелинейные регрессии внутренне линейные:

— степенная $y = a_0 \cdot x^{a_1} \cdot \Delta$, (4.15)

— показательная $y = a_0 \cdot a_1^x \cdot \Delta$, (4.16)

— показательная $y = e^{a_0 + a_1x} \cdot \Delta$; (4.17)

б) нелинейные регрессии внутренне нелинейные, например:

$$y = a_0 + a_1x^{a_2} + \Delta \text{ или } y = a_0 \cdot (1 - \frac{1}{1 - x^{a_1}}) + \Delta. \quad (4.18)$$

Нелинейные регрессии по включенным переменным определяются, как и в линейной регрессии МНК, ибо эти функции линейны по параметрам. Так, для параболы второй степени:

$$y = a_0 + a_1x + a_2x^2 + \Delta,$$

заменяя переменные $x = x_1$, $x^2 = x_2$, получим двухфакторное уравнение линейной регрессии:

$$y = a_0 + a_1x_1 + a_2x_2 + \Delta, \quad (4.19)$$

а для полинома n-го порядка:

$$y = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \Delta$$

получим линейную модель множественной регрессии с «n» объясняющими переменными:

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n + \Delta. \quad (4.20)$$

Для оценки параметров линейной множественной регрессии также используется МНК. При его применении строится система нормальных уравнений (как и для случая линейной парной регрессии, приравнивая частные производные указанной суммы к нулю), решение которой и позволяет получить оценки параметров регрессии:

$$\begin{cases} na_0 + a_1\sum x_1 + a_2\sum x_2 + a_3\sum x_3 + \ldots + a_n\sum x_n = \sum y, \\ a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_1x_2 + a_3\sum x_1x_3 + \ldots + a_n\sum x_1x_n = \sum x_1y, \\ a_0\sum x_2 + a_1\sum x_1x_2 + a_2\sum x_2^2 + a_3\sum x_2x_3 + \ldots + a_n\sum x_2x_n = \sum x_2y, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0\sum x_n + a_1\sum x_1x_n + a_2\sum x_2x_n + a_3\sum x_3x_n + \ldots + a_n\sum x_n^2 = \sum x_ny. \end{cases} \quad (4.21)$$

Решение данной системы может быть осуществлено методом определителей:

$$a_0 = \frac{\Delta a_0}{\Delta},\ a_1 = \frac{\Delta a_1}{\Delta},\ a_2 = \frac{\Delta a_2}{\Delta},\ \ldots,\ a_n = \frac{\Delta a_n}{\Delta}, \quad (4.22)$$

где Δ — определитель системы;

$\Delta a_0, \Delta a_1, \Delta a_2, \ldots, \Delta a_n$ — частные определители.

При этом

$$\Delta = \begin{vmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 & \ldots & \sum x_n \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 & \sum x_1x_3 & \ldots & \sum x_1x_n \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 & \sum x_2x_3 & \ldots & \sum x_2x_n \\ \sum x_3 & \sum x_1x_3 & \sum x_2x_3 & \sum x_3^2 & \ldots & \sum x_3x_n \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \sum x_n & \sum x_1x_n & \sum x_2x_n & \sum x_3x_n & \ldots & \sum x_n^2 \end{vmatrix}, \quad (4.23)$$

а $\Delta a_0, \Delta a_1, \Delta a_2, \ldots, \Delta a_n$ получаются путем замены соответствующего столбца матрицы определителя системы данными правой части исходной системы нормальных уравнений.

Нелинейные регрессии внутренне линейные с помощью соответствующих преобразований также могут быть сведены к линейной функции. Например, степенная функция

$$y = a_0 \cdot x^{a_1} \cdot \Delta$$

путем логарифмирования обеих частей данного уравнения приводит его к линейному виду:

$$\lg(y) = \lg(a_0 \cdot x^{a_1} \cdot \Delta),$$

отсюда

$$\lg y = \lg a_0 + a_1 \lg x + \lg \Delta.$$

Или рассматривая множественную нелинейную регрессию

$$y = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot \ldots \cdot x_n^{a_n} \cdot \Delta, \qquad (4.24)$$

параметры которой также находятся МНК, преобразовывая исходный вид уравнения регрессии к линейному виду путем логарифмирования:

$$\lg y = \lg a_0 + a_1 \lg x_1 + a_2 \lg x_2 + \ldots + a_n \lg x_n + \lg \Delta.$$

Если уравнение регрессии ***внутренне нелинейно по оцениваемым параметрам***, то для оценки параметров используют достаточно сложные итеративные процедуры, успешность которых зависит от вида уравнений и особенностей применяемого итеративного подхода (например, метод прямого поиска, основанный на использовании итеративных процедур поиска минимума непосредственно рассматриваемой целевой функции, выраженной в виде суммы квадратов ошибки S). Модели внутренне нелинейные по параметрам могут иметь место в экономических исследованиях. Однако гораздо большее распространение получили модели, приводимые к линейному виду.

4.2. Предпосылки МНК

Использование МНК позволяет получить неизвестные параметры уравнения регрессии. При этом делаются определенные предпосылки относительно случайной составляющей Δ:

1. Математическое ожидание значений ошибки модели для всех моментов времени t равно нулю, т.е. для $t = 1, 2, \ldots, T$

$$M(\Delta_t) = 0, \; M(\Delta) = 0. \qquad (4.25)$$

2. Значение дисперсии ошибки является постоянной величиной для всех моментов времени $t = 1, 2, ..., T$:

$$S_{\Delta}^2 = \text{const}. \tag{4.26}$$

3. Значения ошибки, взятые в различные моменты времени, независимы между собой, т.е. ковариация ошибок Δ_t, Δ_{t+h}, где $h=1, 2, ...$ равна нулю:

$$\text{cov}(\Delta_t, \Delta_{t+h}) = 0. \tag{4.27}$$

4. Значение независимых факторов модели и ошибки, рассматриваемые в одни и те же моменты времени, являются независимыми, т.е. их ковариации равны нулю:

$$\text{cov}(x_t, \Delta_t) = 0. \tag{4.28}$$

5. Факторы x_{it}, $i=1, 2, 3, ...n$ независимы между собой в том смысле, что их выборочные парные коэффициенты корреляции не превышают некоторого порога p:

$$|r_{ij}| \le p, \tag{4.29}$$

где r_{ij} — выборочный коэффициент корреляции между факторами x_i и x_j, $i \ne j$.

6. Возмущение Δ есть нормально распределенная случайная величина (т.е. подчиняется закону нормального распределения).

Проверка условий, выполнение которых свидетельствует о «высоком» качестве полученных оценок параметров эконометрической модели, на практике обычно осуществляется с использованием ряда процедур и критериев на основе исходной и новой информации, полученной после построения модели. К исходной информации относятся данные наблюдения зависимой и независимой переменной (y и x). Новую информацию составляют значения y_x, найденные оценки параметров a_i, соответствующие им значения *фактической ошибки* (ε_t), $i = 1, 2, 3, ...n$, $t = 1, 2, ..., T$.

При этом следует иметь в виду, что новая информация обычно рассматривается как некоторая оценка (заменитель) истинных (но неизвестных) значений соответствующих характеристик. В этой связи совпадение свойств оценок с предполагаемыми теорией априорными свойствами соответствующих характеристик (часто, но не всегда) является определенной гарантией качества модели (оценок ее параметров). Особенно важную роль в определении качества модели играют значения ее фактической ошибки ε_t, $t = 1, 2, ..., T$.

Таким образом, проверка наличия стандартных свойств (предпосылок) у «истинной» ошибки эконометрической модели Δ проводится на основе анализа соответствующих свойств фактической ошибки ε_t. При этом наличие у ошибки ε_t каждого из этих свойств не всегда является доказательством присутствия соответствующего свойства у ошибки Δ. Вместе с тем если фактическая ошибка ε_t не обладает некоторым свойством, то можно говорить, что теоретические предпосылки эконометрической модели не подтверждены полученными результатами и качество уравнения недостаточно высоко.

Использование МНК позволяет получить несмещенные и эффективные оценки параметров эконометрических моделей. *Несмещенность* является желательным свойством и означает, что математическое ожидание значений ошибок равно нулю (*первая предпосылка МНК*). Использование МНК обеспечивает выполнение данного условия автоматически. Действительно, дифференцируя сумму квадратов ошибок по параметру a_0, получим:

$$\frac{dS}{da_0} = -2\sum_{i=1}^{n}(y_x - a_0 - a_1x_{1i} - \ldots - a_nx_{ni}) = 0. \qquad (4.30)$$

Из этого выражения автоматически вытекает, что

$$\sum_{i=1}^{n}(y_i - y_{xi}) = \sum_{i=1}^{n}\varepsilon_{ti} = 0. \qquad (4.31)$$

Для практических целей важна не только несмещенность, но и эффективность оценок. Оценки считаются *эффективными*, если они характеризуются наименьшей дисперсией.

Вместе с тем несмещенность оценок коэффициентов регрессии, полученных МНК, зависит от независимости значений независимых факторов модели и ошибки, что исследуется в рамках соблюдения *четвертой предпосылки МНК*. С этой целью строится график зависимостей фактических ошибок ε_t от факторов, включенных в регрессию x (рис. 4.2).

Если ошибки на графике расположены в виде горизонтальной полосы, то они независимы от значений x_j. Если же график показывает наличие зависимости ε_t от x_j, то модель неадекватна. Причины неадекватности могут быть разные. Может быть неправильна спецификация модели, и в нее необходимо ввести дополнительные члены от x_j, например x_j^2, или преобразовать значения y_j.

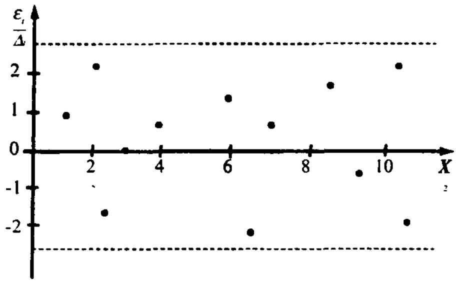


Рис. 4.2. Зависимость фактических ошибок от величины фактора x

Предпосылка о нормальном распределении ошибок позволяет проводить проверку параметров регрессии и корреляции с помощью t-критерия Стьюдента и F-критерия Фишера. Вместе с тем оценки регрессии, найденные с применением МНК, обладают хорошими свойствами даже при отсутствии нормального распределения ошибок, т.е. при нарушении *шестой предпосылки МНК*.

Совершенно необходимым для получения по МНК состоятельных оценок параметров регрессии является соблюдение второй и третьей предпосылок. *Состоятельность* оценок характеризует увеличение их точности с увеличением объема выборки.

В соответствии со *второй предпосылкой МНК* требуется, чтобы дисперсия остатков была *гомоскедастичной*. Это значит, что для каждого значения фактора от x_j ошибки ε_t имеют одинаковую дисперсию. Если это условие применения МНК не соблюдается, то имеет место *гетероскедастичность*. Однако необходимо отметить, что условие $S_{\Delta}^2 = \text{const}$ нельзя интерпретировать как постоянство значений $|\varepsilon_t|$ для $t = 1, 2, \ldots, T$. Оно лишь означает, что дисперсия истинной ошибки Δ является постоянной величиной на любом из отрезков рассмотренного временного интервала $(1, T)$. В этой связи проверка данной предпосылки МНК может быть идентична проверке гипотезы о постоянстве дисперсии фактической ошибки ε_t на различных отрезках интервала $(1, T)$. Такая проверка обычно проводится с использованием соответствующих тестов.

Наличие гетероскедастичности можно наглядно видеть из поля корреляции (рис. 4.3)

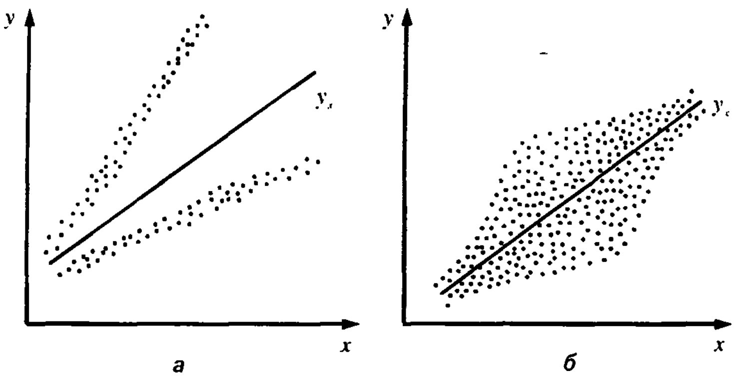


Рис. 4.3. Примеры гетероскедастичности:
а — дисперсия остатков растет по мере увеличения *x*;
б — дисперсия остатков достигает максимальной величины при средних значениях *x* и уменьшается при минимальных и максимальных значениях *x*

Используя трехмерное изображение, проиллюстрируем гомо- и гетероскедастичность ошибок на основании условия предпосылки о постоянстве дисперсии ошибок (рис. 4.4, рис. 4.5).

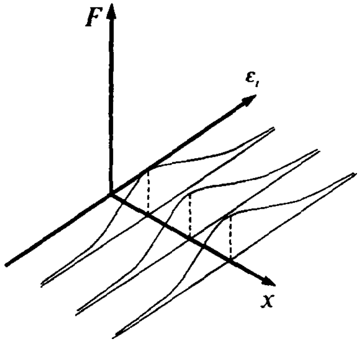


Рис. 4.4. Гомоскедастичность ошибок

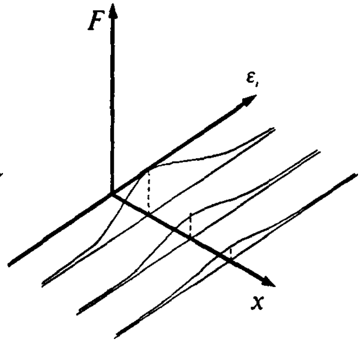


Рис. 4.5. Гетероскедастичность ошибок

Наличие гетероскедастичности может в отдельных случаях привести к смещенности оценок коэффициентов регрессии. Гетероскедастичность также сказывается на уменьшении эффективности оценок a_i. Практически при нарушении гомоскедастичности получаются неравенства:

$$S^2_{\varepsilon_i} \neq S^2_{\varepsilon_j} \neq S^2; \quad i \neq j \tag{4.32}$$

и можно записать:

$$S^2_{\varepsilon_i} = S^2 K_i. \tag{4.33}$$

При этом величина K_i может меняться при переходе от одного значения фактора x_i к другому. Это означает, что сумма квадратов отклонений (ошибок) для зависимости

$$y_x = a_0 + a_1 x$$

при наличии гетероскедастичности должна иметь вид

$$S = \sum \frac{1}{K_i}(y_x - a_0 - a_1 x)^2. \tag{4.34}$$

При минимизации этой суммы квадратов отдельные ее слагаемые взвешиваются: наблюдениям с наибольшей дисперсией придается пропорционально меньший вес. Иными словами, вклад каждого сочетания x_i и y_i в сумму квадратов ошибок должен быть дисконтирован, чтобы учесть влияние неоднородных элементов K_i.

Задача состоит в том, чтобы определить величину K_i и внести поправку в исходные переменные. С этой целью используют *обобщенный метод наименьших квадратов.*

Чтобы определить наличие или отсутствие гетероскедастичности, используются тесты. Все они берут в качестве гипотезы H_0 гипотезу об отсутствии гетероскедастичности.

Рассмотрим использование тестов на гетероскедастичность на примере теста *Голдфелда–Квандта* (разработан в 1965 г. Голдфелдом и Квандтом). Чтобы оценить нарушение гомоскедастичности, ими предложен параметрический тест, который включает следующие этапы:

а) первоначально необходимо проранжировать n наблюдений в порядке возрастания переменной x;

б) затем выбирают m первых и m последних наблюдений, исключая из рассмотрения C центральных наблюдений. Из экспериментальных расчетов для случая одного фактора в регрессионной модели рекомендовано, например, для $n = 30$ принимать $C = 8$, при $n = 60$ соответственно $C = 16$;

в) далее для разделенной совокупности из $(n - C)$ наблюдений на две группы (соответственно с малыми и большими значениями фактора x) определяется для каждой из групп уравнение регрессии;

г) в заключение определяется сумма квадратов фактических ошибок $\sum \varepsilon_i$ для первой (S_1) и второй (S_2) групп и находится их отношение:

$$R = S_2 : S_1. \tag{4.35}$$

Гипотеза о равенстве двух нормально распределенных совокупностей проверяется с помощью критерия *Фишера–Снедекора*.

Нулевая гипотеза о равенстве дисперсий двух наборов по m наблюдений (т.е. гипотеза об отсутствии гетероскедастичности) отвергается, если:

$$R=\frac{S_2}{S_1}>F_{\alpha,\,m-(p+1),\,p},\qquad (4.36)$$

где p — число параметров модельного уравнения.

Мощность теста, т.е. вероятность отвергнуть гипотезу об отсутствии гетероскедастичности, когда действительно гетероскедастичности нет, оказывается максимальной, если выбрать m порядка $n/3$.

Рассмотрим применение теста Голдфелда–Квандта на следующем *примере*: по данным n = 150 наблюдений о доходах (y), уровне образования (x_1) и возрасте индивидуумов (x_2) — можно ли считать при уровне значимости $\alpha=0{,}05$ линейную регрессионную модель y по x_1 и x_2 гетероскедастичной.

Для решения поставленной задачи разобьем по $m = n/3 = 150/3 = 50$ значений лиц с наименьшим и наибольшим уровнем образования x_1.

После определения параметров уравнения регрессии вычислим сумму квадратов ошибок:

$$S_1=\sum_{i=1}^{150}\varepsilon_t=894{,}1;\quad S_2=\sum_{i=101}^{150}\varepsilon_t=3918{,}2;$$

$$R=S_2:S_1=3918{,}2:894{,}1=4{,}38.$$

Так как $R=4{,}38>F_{0.05,48,48}=1{,}61$, то гипотеза об отсутствии гетероскедастичности отвергается, т.е. доходы более образованных людей действительно имеют существенно большую вариацию, подтверждая тем самым наличие гетероскедастичности.

При построении регрессионных моделей чрезвычайно важно соблюдение *третьей предпосылки МНК* – отсутствие *автокорреляции остатков* (ошибок). Проверка выполнимости данной предпосылки, свидетельствующей об отсутствии автокорреляционных взаимосвязей в ряду «истинной» ошибки модели Δ_t ($\operatorname{cov}(\Delta_i,\Delta_{i+h})=0$), на практике осуществляется тестированием ряда значений фактической ошибки ε_t.

Автокорреляция остатков означает наличие корреляции между фактическими ошибками (ε_t) текущих и предыдущих (последующих) наблюдений. Коэффициент корреляции между ε_i

и ε_j, где ε_i — ошибки текущих наблюдений, ε_j — остатки предыдущих наблюдений (где $j = i - 1$), может быть определен как:

$$r_{\varepsilon_i \varepsilon_j} = \frac{\text{cov}(\varepsilon_i, \varepsilon_j)}{S_{\varepsilon_i} S_{\varepsilon_j}}, \tag{4.37}$$

т.е. по обычной формуле *линейного коэффициента корреляции*. Если этот коэффициент окажется существенно отличным от нуля, то остатки автокоррелированы и функция плотности вероятности $F(\varepsilon)$ зависит от j-й точки наблюдения и от распределения значений фактических ошибок в других точках наблюдения.

Для регрессионных моделей со статистической информацией автокорреляция остатков может быть подсчитана, если наблюдения упорядочены по фактору x, как это имеет место в таблице 4.1. Коэффициент может быть найден по следующим рядам данных ($n - 1$).

Таблица 4.1

Исходные данные для определения коэффициента автокорреляции

ε_i	ε_2	ε_3	ε_4	ε_5	...	ε_{n-1}	ε_n
ε_{i-1}	ε_1	ε_2	ε_3	ε_4	...	ε_{n-2}	ε_{n-1}

Далее рассчитывается коэффициент автокорреляции по формуле линейного коэффициента корреляции:

$$r_{\varepsilon_i \varepsilon_{i-1}} = \frac{\text{cov}(\varepsilon_i, \varepsilon_{i-1})}{S_{\varepsilon_i} \cdot S_{\varepsilon_{i-1}}}, \tag{4.38}$$

где

$$\text{cov}(\varepsilon_i, \varepsilon_{i-1}) = \overline{\varepsilon_i \cdot \varepsilon_{i-1}} - \overline{\varepsilon}_i \cdot \overline{\varepsilon}_{i-1} = \frac{\sum \varepsilon_i \cdot \varepsilon_{i-1}}{n} - \frac{\sum \varepsilon_i}{n} \cdot \frac{\sum \varepsilon_{i-1}}{n};$$

S_{ε_i} и $S_{\varepsilon_{i-1}}$ — среднее квадратическое отклонение, которые находятся по формулам:

$$S_{\varepsilon_i} = \sqrt{\frac{\sum (\varepsilon_i - \overline{\varepsilon}_i)^2}{n}} = \sqrt{\overline{\varepsilon_i^2} - \overline{\varepsilon}_i^{\,2}}$$

$$\text{и } S_{\varepsilon_{i-1}} = \sqrt{\frac{\sum(\varepsilon_{i-1} - \bar{\varepsilon}_{i-1})^2}{n}} = \sqrt{\overline{\varepsilon_{i-1}^2} - \bar{\varepsilon}_{i-1}^{\,2}} \qquad (4.39)$$

Если коэффициент автокорреляции существенно отклоняется от нуля, то делают вывод о том, что случайные величины (Δ) в регрессионной модели не оказываются независимыми, в частности условие $r(\varepsilon_i, \varepsilon_j) = 0$ не выполняется, и такие модели называются моделями с наличием автокорреляции.

Отсутствие автокорреляции остаточных величин (ошибок, возмущений) обеспечивает состоятельность (состоятельность оценок характеризует увеличение их точности с увеличением объема выборки) и эффективность оценок коэффициента регрессии. Особенно актуально соблюдение данной предпосылки при построении регрессионных моделей по рядам динамики, где ввиду наличия тенденции последующие уровни динамического ряда зависят от своих предыдущих уровней.

Итак, модели, удовлетворяющие приведенным предпосылкам МНК 1–6, называются *классической нормальной моделью регрессии*. Если же среди приведенных не выполняется лишь предпосылка 6 о нормальном законе распределения ошибок Δ_i, то модель называется просто *классической линейной моделью регрессии*.

В соответствии с *теоремой Гаусса–Маркова* если регрессионная модель удовлетворяет предпосылкам 1–5, то оценка метода наименьших квадратов (a_i) является наиболее эффективной, т.е. обладает наименьшей дисперсией в классе линейных несмещенных оценок.

4.3. Мультиколлинеарность

Одним из основных условий построения уравнения множественной регрессии является независимость факторов x_i, включенных в модель, т.е. соблюдение *пятой предпосылки МНК*.

Высокая взаимная коррелированность (взаимозависимость) объясняющих (независимых) переменных называется *мультиколлинеарностью*. Она 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. Это приводит к невозможности решения соответствующей системы нормальных уравнений и получения оценок параметров регрессионной модели.

Однако в эконометрических исследованиях мультиколлинеарность чаще проявляется в *стохастической форме*, когда между хотя бы двумя объясняющими переменными существует тесная корреляционная связь. В результате получаются значительные средние квадраты отклонения коэффициентов регрессии a_0, a_1, a_2, ...,a_n, и оценка их значимости по t-критерию Стьюдента 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 в этом случае, 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.

Точных количественных критериев для определения наличия или отсутствия мультиколлинеарности не существует. Тем не менее имеются некоторые *эвристические подходы* по ее выявлению.

Один из таких подходов заключается в анализе корреляционной матрицы между объясняющими переменными $x_1, x_2, ..., x_n$ и выявлении пар переменных, имеющих высокие коэффициенты корреляции (обычно больше 0,8, $r_{xi}^{xj} \geq 0{,}8$). Если такие переменные существуют, то говорят о мультиколлинеарности между ними.

Например, пусть при изучении зависимости $y = f(x_1, x_2, x_3)$ матрица парных коэффициентов корреляции оказалась следующей:

	Y	X_1	X_2	X_3
Y	1			
X_1	0,8	1		
X_2	0,7	0,8	1	
X_3	0,6	0,5	0,2	1

Очевидно, что факторы x_1 и x_2 дублируют друг друга. Для устранения мультиколлинеарности одну из переменных исключают из рассмотрения. При этом какую переменную оставить, решают в первую очередь на основании экономических соображений. Если с экономической точки зрения ни одной из переменных нельзя отдать предпочтение, то оставляют ту переменную, которая имеет меньший коэффициент корреляции с зависимой переменной. В этом проявляется специфика множественной регрессии как метода исследования комплексного воздействия факторов в условиях их независимости друг от друга. В соответствии с этим в анализ целесообразнее включить фактор x_2, так как его корреляция с результатом y слабее, чем корреляция фактора x_1 с y ($r_{yx_2} < r_{yx_1}$).

Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии мультиколлинеарности факторов, когда более чем два фактора связаны между собой линейной зависимостью, т.е. имеет место совокупное воздействие факторов друг на друга.

Для оценки мультиколлинеарности факторов может использоваться подход, заключающийся в нахождении определителя матрицы парных коэффициентов корреляции между факторами.

Предположим, что для включающего три объясняющие переменные уравнения регрессии

$$y_x = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \Delta$$

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

$$Det|R| = \begin{vmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} \\ r_{x_3x_1} & r_{x_3x_2} & r_{x_3x_3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1.$$

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

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

Отсюда чем ближе к нулю определитель матрицы парных коэффициентов корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И наоборот, чем ближе к единице определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Для устранения или уменьшения мультиколлинеарности факторов используется ряд методов. Об одном из них мы уже говорили выше, и состоит он в том, что из двух объясняющих переменных, имеющих высокий коэффициент корреляции (больше 0,8), одну переменную исключают из рассмотрения.

При наличии более чем двух факторов, связанных между собой линейной зависимостью, можно использовать другой подход, когда через коэффициенты множественной детерминации (R^2) находят переменные, ответственные за мультиколлинеарность факторов. Для этого в качестве зависимой переменной рассматривается каждый из факторов. Чем ближе значение коэффициента множественной детерминации к единице (обычно уже больше 0,6), тем сильнее мультиколлинеарность факторов. Сравнивая между собой коэффициенты множественной детерминации факторов

$$(R^2_{x_1|x_2,x_3,...,x_n}; R^2_{x2|x_2,x_3,...,x_n}; ...),$$

можно выделить переменные, ответственные за мультиколлинеарность, следовательно, можно решать проблему отбора факторов, оставляя в уравнении факторы с минимальной величиной коэффициента множественной детерминации.

Для устранения мультиколлинеарности может быть использован переход от объясняющих переменных $x_1, x_2, ..., x_n$, связанных между собой достаточно тесной корреляционной зависимостью, к новым переменным, представляющим линейные комбинации исходных.

Еще одним из возможных методов устранения или уменьшения мультиколлинеарности является использование ***пошаговых процедур отбора*** наиболее информативных переменных. Например, 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 вместе 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 наибольший (скорректированный) коэффициент детерминации, и т.д.

Процедура введения новых переменных продолжается до тех пор, пока будет увеличиваться соответствующий (скорректированный) коэффициент детерминации (R^2), равный:

$$R^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-m-1}, \tag{4.40}$$

где m — число параметров при переменных x;

n — число наблюдений.

Кроме пошаговой процедуры ***присоединения*** используются также пошаговые процедуры *присоединения – удаления* и процедура *удаления* объясняющих переменных. Следует отметить, что какая бы пошаговая процедура ни использовалась, она не гарантирует определение оптимального (в смысле получения максимального коэффициента детерминации на одну степень свободы R^2) набора объясняющих переменных. Однако в большинстве случаев получаемые с помощью пошаговых процедур наборы переменных оказываются оптимальными или близкими к оптимальным.

4.4. Обобщенный метод наименьших квадратов (ОМНК)

Если при использовании классического метода наименьших квадратов обнаруживается наличие гетероскедастичности и автокорреляции ошибок, т.е. нарушаются предпосылки МНК 2 и 3, то рекомендуется использовать *обобщенный метод наименьших квадратов* (ОМНК).

При моделировании реальных экономических процессов довольно часто возникают такие ситуации, в которых условия классической линейной модели регрессии оказываются нарушенными, в частности, случайные возмущения (ошибки) модели имеют постоянную дисперсию и не коррелированы между собой. Так, например, при использовании зависимости расходов на потребление от уровня доходов семьи можно ожидать, что в более обеспеченных семьях вариация расходов выше, чем в малообеспеченных, т.е. дисперсии возмущения неодинаковы. При рассмотрении временных рядов сталкиваются с ситуацией, когда наблюдаемые в данный момент значения зависимой переменной коррелируют с их значениями в предыдущие моменты времени, т.е. наблюдается корреляция между возмущениями в разные моменты времени.

Сравнивая обобщенную модель с классической, можно отметить, что главное их отличие заключается в том, что в обобщенной модели ковариации и дисперсии объясняющих переменных x могут быть произвольными.

Сущность ОМНК заключается в том, что он применяется к преобразованным данным и позволяет получить оценки, которые обладают не только свойством несмещенности, но и имеют меньшие выборочные дисперсии. То есть, как и ранее, предполагается, что среднее значение возмущений (ошибок) равно нулю, а вот дисперсия их не остается неизменной для разных значений факторов, а пропорциональна величине K_i:

$$S_{\varepsilon_i}^2 = S^2 K_i,$$

где $S_{\varepsilon_i}^2$ — дисперсия ошибок при конкретном i-м значении фактора;

S^2 — постоянная дисперсия ошибки при соблюдении предпосылки о гомоскедастичности остатков;

K_i – коэффициент пропорциональности, меняющийся с изменением величины фактора, что и обусловливает непропорциональность остатков.

Так, для уравнения вида:

$$y_x = a_0 + a_1 x_i + \varepsilon_i \text{ при } S^2_{\varepsilon_i} = S^2 K_i$$

модель примет вид

$$y_x = a_0 + a_1 x_i + \sqrt{K_i} \cdot \varepsilon_i, \qquad (4.41)$$

где остаточные величины гетероскедастичны. Предполагая в них отсутствие автокорреляции, можно перейти к уравнению с гомоскедастичными остатками, поделив все переменные, зафиксированные в ходе i-го наблюдения на $\sqrt{K_i}$, тогда дисперсии остатков будут величиной постоянной. Таким образом, от регрессии y по x перейдем к регрессии новых переменных:

$$\frac{y}{\sqrt{K_i}} \text{ и } \frac{x}{\sqrt{K_i}}.$$

Исходные данные для полученного уравнения будут иметь вид:

$$y = \begin{vmatrix} \frac{y_1}{\sqrt{K_1}} \\ \frac{y_2}{\sqrt{K_2}} \\ \dots \\ \frac{y_n}{\sqrt{K_n}} \end{vmatrix}, \qquad x = \begin{vmatrix} \frac{x_1}{\sqrt{K_1}} \\ \frac{x_2}{\sqrt{K_2}} \\ \dots \\ \frac{x_n}{\sqrt{K_n}} \end{vmatrix}.$$

Оценка параметров нового уравнения с преобразованными переменными проводится с помощью МНК, для которого необходимо минимизировать сумму квадратов отклонений:

$$S = \sum \frac{1}{K_i} (y_i - a_0 - a_1 x_i)^2. \qquad (4.42)$$

Соответственно получим следующую систему нормальных уравнений:

$$\begin{cases} \sum \frac{y_i}{K_i} = a_0 \sum \frac{1}{K_i} + a_1 \sum \frac{x_i}{K_i}, \\ \sum \frac{x_i y_i}{K_i} = a_0 \sum \frac{x_i}{K_i} + a_1 \sum \frac{x_i^2}{K_i}. \end{cases} \quad (4.43)$$

Аналогичный подход возможен не только для уравнений парной регрессии, но и для множественной, например вида

$$y_x = a_0 + a_1 x_1 + a_2 x_2 + \Delta,$$

для которой дисперсия ошибок оказалась пропорциональна K_i^2, таким образом, при

$$S_{\varepsilon i}^2 = S^2 K_i^2$$

модель примет вид:

$$y_x = a_0 + a_1 x_{1i} + a_2 x_{2i} + K_i \varepsilon_i. \quad (4.44)$$

Для того чтобы получить уравнение, где ε_i гомоскедастичны, необходимо перейти к преобразованным переменным, разделив все члены исходного уравнения на коэффициент пропорциональности K_i. Уравнение с преобразованными переменными будет иметь следующий вид:

$$\frac{y_x}{K_i} = \frac{a_0}{K_i} + a_1 \frac{x_{1i}}{K_i} + a_2 \frac{x_{2i}}{K_i} + \varepsilon_i. \quad (4.45)$$

Однако при применении ОМНК в таком виде могут возникнуть определенные трудности, так как в практике эконометрического моделирования крайне редко бывает так, чтобы была известна постоянная дисперсия ошибки S^2 и соответственно коэффициент пропорциональности K_i для исходных данных (y, x_i), по которому можно было бы произвести их дисконтирование с целью корректировки гетероскедастичности. Использование тестов на гетероскедастичность (например, рассмотренный выше тест Голдфелда–Квандта) позволяет обнаружить лишь само наличие гетероскедастичности, но они не дают возможности проследить количественный характер зависимости дисперсий ошибок регрессии от значений регрессоров и, следовательно, не представляют каких-либо способов устранения гетероскедастичности. В такой ситуации для продвижения к цели, очевидно, необходимы некоторые дополнительные предположения относительно характера гетероскедастичности.

На практике при построении эконометрических моделей возможно применение методов смягчения проблемы гетероскедастичности. Одним из таких методов является *метод взвешенных наименьших квадратов* (ВНК). В соответствии с ним возможны два вида преобразования модели с целью устранения проблемы гетероскедастичности:

а) значения дисперсии фактической ошибки $S^2_{\varepsilon_i}$ известны для каждого наблюдения;

б) дисперсии фактических ошибок неизвестны.

Таким образом, применяя *обычный* МНК, неизвестные параметры регрессионной модели находят путем минимизации остаточной суммы квадратов $S=\sum_{i=1}^{n}(y_i-y_{x_i})^2$, используя *обобщенный* метод — минимизируются:

$$S=\sum_{i=1}^{n}\left[\frac{1}{K_i}(y_i-y_{x_i})^2\right],$$

наконец, в частном случае, применяя *взвешенный* МНК, – минимизируются:

$$S=\sum_{i=1}^{n}\left[\frac{1}{S_{\varepsilon_i}}(y_i-y_{x_i})^2\right].$$

Для случая применения ВНК, когда для каждого наблюдения известны значения дисперсии фактической ошибки $S^2_{\varepsilon_i}$, можно устранить гетероскедастичность, разделив каждое наблюдаемое значение на соответствующее ему значение дисперсии. Рассмотрим применение данного случая для ВНК на примере парной регрессии:

$$y_{x_i}=a_0+a_1x_i+\varepsilon_i.$$

Разделим обе части уравнения регрессии на известное $S_{\varepsilon_i} = \sqrt{S^2_{\varepsilon_i}}$:

$$\frac{y_{x_i}}{S_{\varepsilon_i}} = a_0 \frac{1}{S_{\varepsilon_i}} + a_1 \frac{x_i}{S_{\varepsilon_i}} + \frac{\varepsilon_i}{S_{\varepsilon_i}}.$$

Положив

$$\frac{y_{x_i}}{S_{\varepsilon_i}} = y_{x_i}^*; \quad \frac{x_i}{S_{\varepsilon_i}} = x_i^*; \quad \frac{\varepsilon_i}{S_{\varepsilon_i}} = v_i; \quad \frac{1}{S_{\varepsilon_i}} = z_i,$$

получим уравнение регрессии без свободного члена, но с дополнительной объясняющей переменной z_i и с «преобразованными» отклонениями v_i:

$$y_{x_i}^* = a_0 z_i + a_1 x_i^* + v_i.$$

Таким образом, ВНК включает следующие этапы:

а) значения каждой пары наблюдений (x_i, y_i) делят на известную величину S_{ε_i}. Тем самым наблюдениям с наименьшими дисперсиями придаются наибольшие «веса». Учет этого фактора увеличивает вероятность получения более точных оценок;

б) по МНК для преобразованных значений ($\frac{1}{S_{\varepsilon_i}}, \frac{x_i}{S_{\varepsilon_i}}, \frac{y_{x_i}}{S_{\varepsilon_i}}$) строится уравнение регрессии без свободного члена с гарантированными качествами оценок.

Но, как правило, на практике применение ВНК затруднено тем, что значения дисперсий фактических ошибок известны крайне редко. Следовательно, чтобы применить ВНК, необходимо сделать реалистические предположения о значениях дисперсий фактических ошибок ($S^2_{\varepsilon_i}$).

Например, может оказаться целесообразным предположить, что дисперсии $S^2_{\varepsilon_i}$ пропорциональны значениям x_i (рис. 4.6, *а*) или значениям x_i^2 (рис. 4.6, *б*).

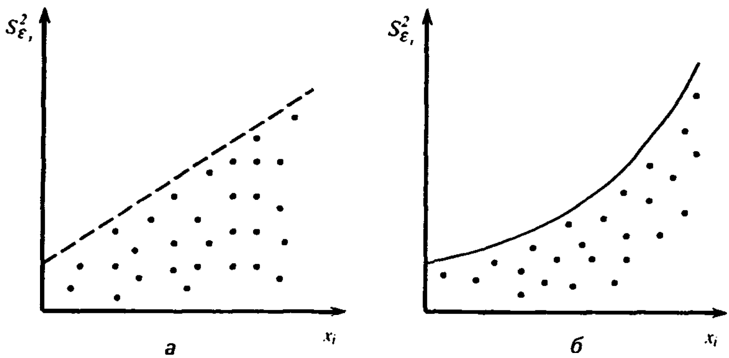


Рис. 4.6. Пропорциональность дисперсий фактических ошибок со значением факторного признака

Дисперсия $S^2_{\varepsilon_i}$ *пропорциональна значениям* x_i (рис. 4.6, *а*):

$$S^2_{\varepsilon_i} = S^2 x_i.$$

Тогда уравнение преобразуется делением его левой и правой частей на $\sqrt{x_i}$:

$$\frac{y_{x_i}}{\sqrt{x_i}} = a_0 \cdot \frac{1}{\sqrt{x_i}} + a_1 \cdot \frac{x_i}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}} \Rightarrow \frac{y_{x_i}}{\sqrt{x_i}} = a_0 \cdot \frac{1}{\sqrt{x_i}} + a_1 \cdot \sqrt{x_i} + v_i,$$

где

$$v_i = \frac{\varepsilon_i}{\sqrt{x_i}}.$$

Дисперсия $S^2_{\varepsilon_i}$ *пропорциональна значениям* x_i^2 (рис. 4.6, *б*).

В этом случае если зависимость $S^2_{\varepsilon_i}$ от x_i целесообразнее выразить не линейной функцией, а квадратической, то соответствующим преобразованием будет деление уравнения регрессии на x_i:

$$\frac{y_{x_i}}{x_i} = a_0 \cdot \frac{1}{x_i} + a_1 \cdot \frac{x_i}{x_i} + \frac{\varepsilon_i}{x_i} \Rightarrow \frac{y_{x_i}}{x_i} = a_0 \cdot \frac{1}{x_i} + a_1 + v_i,$$

где

$$v_i = \frac{\varepsilon_i}{x_i}.$$

Для применения описанных выше преобразований весьма значимы знания об «истинных» значениях дисперсий отклонений либо предположения, какими эти дисперсии могут быть. Во многих случаях дисперсии отклонений зависят не от включенных в уравнение регрессии объясняющих переменных, а от тех, которые не включены в модель, но играют существенную роль в исследуемой зависимости. В этом случае они должны быть включены в модель. В ряде случаев для устранения гетероскедастичности необходимо изменить спецификацию модели (например, линейную — на лог-линейную, мультипликативную — на аддитивную и т.п.).

В заключение отметим, что наличие гетероскедастичности не позволяет получить эффективные оценки, что зачастую приводит к необоснованным выводам по их качеству. Обнаружение гетероскедастичности является достаточно трудоемкой проблемой, и для ее решения разработано несколько методов (тестов). В случае установления наличия гетероскедастичности ее корректировка также является достаточно серьезной проблемой. Одним из возможных решений является метод взвешенных наименьших квадратов (при этом необходимы определенная информация либо обоснованные предположения о величинах дисперсий отклонений).

На практике имеет смысл применить несколько методов определения гетероскедастичности и способов ее корректировки.

Контрольные вопросы

1. Регрессионный эконометрический анализ: понятия и задачи.
2. Метод наименьших квадратов.
3. Нелинейные регрессии по включенным переменным.
4. Нелинейные регрессии по оцениваемым параметрам.
5. Нелинейные регрессии внутренне нелинейных по оцениваемым параметрам.

6. Общие понятия и применение фиктивных переменных.
7. Дихотомические фиктивные переменные.
8. Предпосылки метода наименьших квадратов.
9. Критерии несмещенности, эффективности и состоятельности оценки параметров.
10. Гомоскедастичность и гетероскедастичность остатков.
11. Тестирование моделей на гетероскедастичность (тест Голдфелда-Квандта).
12. Автокорреляция (авторегрессия) остатков.
13. Мультиколлинеарность переменных.
14. Критерий определения мультиколлинеарности.
15. Методы устранения мультиколлинеарности.
16. Обобщенный метод наименьших квадратов.
17. Взвешенный метод наименьших квадратов.

ТЕСТОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ ПРОВЕРКИ ЗНАНИЙ

1. Суть МНК состоит в минимизации суммы квадратов:

а) коэффициентов регрессии;
б) отклонений точек наблюдений от уравнения регрессии;
в) значений зависимой переменной;
г) отклонений точек эмпирического уравнения регрессии от точек теоретического уравнения регрессии.

2. Коэффициент уравнения регрессии показывает:

а) на сколько процентов изменится результат при изменении фактора на 1%;
б) на сколько единиц изменится результат при изменении фактора на 1 единицу;
в) на сколько процентов изменится фактор при изменении результата на 1%;
г) на сколько единиц изменится фактор при изменении результата на 1 единицу;
д) во сколько раз изменится результат при изменении фактора на 1 единицу.

3. Найдите предположение, не являющееся предпосылкой классической модели:
а) случайное отклонение имеет нулевое математическое ожидание;
б) случайное отклонение не обладает нормальным распределением;
в) случайное отклонение имеет постоянную дисперсию;
г) отсутствует автокорреляция случайных отклонений;
д) случайное отклонение независимо от объясняющих переменных.

4. Величина коэффициента регрессии показывает:
а) среднее изменение результата с изменением фактора на один процент;
б) среднее изменение результата с изменением фактора на одну единицу;
в) изменение результата в процентах с изменением фактора на один процент.

5. «Объясненная» сумма квадратов отклонений отражает влияние на разброс y:
а) прочих факторов;
б) изучаемого фактора x;
в) изучаемого фактора x и прочих факторов.

6. Остаточная сумма квадратов отклонений отражает влияние на разброс y:
а) изучаемого фактора x;
б) прочих факторов;
в) изучаемого фактора x и прочих факторов.

7. Если фактор не оказывает влияния на результат, то линия регрессии на графике:
а) параллельна оси Oy;
б) параллельна оси Ox;
в) является биссектрисой первой четверти декартовой системы координат.

8. При применении метода наименьших квадратов методом определителей решается:
   а) корреляционная матрица;
   б) уравнение регрессии;
   в) система нормальных уравнений;
   г) система неравенств.

9. Метод наименьших квадратов позволяет оценить ... уравнений регрессии:
   а) переменные;
   б) переменные и случайные величины;
   в) параметры и переменные;
   г) параметры.

10. Для линейной регрессионной модели $y = a_0 + a_1 x + \varepsilon$ математическое ожидание случайного отклонения равно нулю. Это утверждение является:
   а) критерием Стьюдента;
   б) условием линеаризации;
   в) одной из основных предпосылок метода наименьших квадратов;
   г) нарушением предпосылок метода наименьших квадратов.

11. Остаточная сумма квадратов равна нулю в том случае, когда:
   а) значения y, рассчитанные по уравнению регрессии, равны среднему значению y;
   б) y связан с x функционально;
   в) общая дисперсия y обусловлена влиянием прочих факторов.

12. Общая сумма квадратов отклонений совпадает с остаточной, когда:
   а) прочие факторы не влияют на результат;
   б) фактор x не оказывает влияния на результат;
   в) фактор x и прочие факторы в равной степени влияют на результат.

13. Уравнение регрессии статистически значимо, если:
а) остаточная сумма квадратов отклонений значимо больше «объясненной» суммы квадратов отклонений;
б) «объясненная» сумма квадратов отклонений значимо больше остаточной суммы квадратов отклонений;
в) «объясненная» и остаточная суммы квадратов отклонений равны.

14. Число степеней свободы связано:
а) с числом определяемых по совокупности констант;
б) числом единиц совокупности n и числом определяемых по совокупности констант;
в) числом единиц совокупности n.

15. Факторная сумма квадратов отклонений в парной регрессии имеет число степеней свободы, равное:
а) $n - 1$;
б) 1;
в) $n - 2$.

ПРАКТИКУМ

Задача 4.1. По данным 26 муниципальных районов региона о стоимости продукции сельского хозяйства (Y) и стоимости основных фондов (X) оценить параметры парных уравнений регрессии следующего вида:

а) линейной $y_x = a_0 + a_1 x$;

б) степенной $y_x = a_0 \cdot x^{a_1}$;

в) показательной $y_x = a_0 \cdot a_1^{x}$;

г) логарифмической $y_x = a_0 + a_1 \lg x$.

Таблица 4.2

Данные о стоимости основных фондов и продукции сельского хозяйства в разрезе районов Ставропольского края в 2012 г., млн руб.

№ района	X	Y	№ района	X	Y
1	201,6	1011,3	14	655,1	1902,1
2	439,3	1919	15	863,4	1730
3	1857,8	2195,9	16	818,8	2566,4
4	1612	1296	17	492,05	2323
5	786,3	2425,3	18	814,1	2042,3
6	1415	3657	19	1717,4	2323
7	3551	5768	20	1157	1683,2
8	450,7	880	21	1441	1529
9	1732	5018	22	2114,7	3955,7
10	3669,1	4678,8	23	192,5	2685
11	1903	2814	24	1247,5	3021,7
12	2021,6	1424,6	25	1253,1	2232
13	4600,3	5874	26	1873,5	6259

Решение

а) *линейная регрессия* $y_x = a_0 + a_1x$.

Для определения параметров уравнения линейной парной регрессии необходимо решить следующую систему нормальных уравнений, полученную МНК:

$$\begin{cases} na_0 + a_1\sum x_i = \sum y_i, \\ a_0\sum x_i + a_1\sum x_i^2 = \sum x_i y_i. \end{cases}$$

Кроме того, для оценки параметров можно воспользоваться средствами анализа данных *Excel*. Для этого сформируем таблицу исходных значений переменных *x* и *y*. Далее необходимо во вкладке «Данные» открыть инструмент «Анализ данных». В появившемся диалоговом окне выберем пункт «Регрессия», нажмем «ОК» (рис. 4.7).

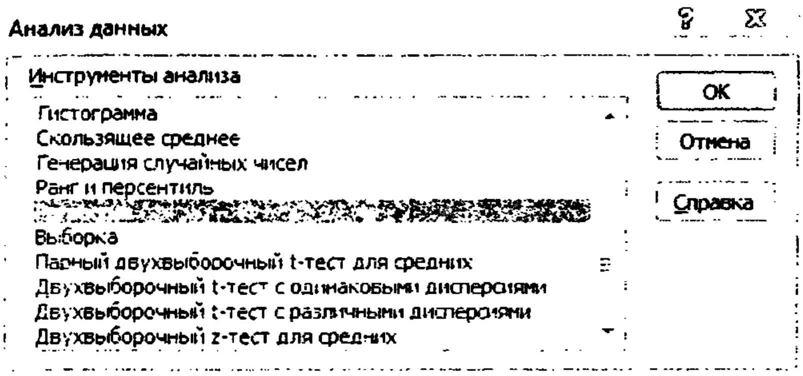


Рис. 4.7. Инструмент «Регрессия» пакета анализа данных *Excel*

В открывшемся диалоговом окне «Регрессия» необходимо указать в качестве входного интервала *Y* ряд значений переменной результативной переменной, а в качестве входного интервала *X* — ряд значений факторной переменной. Для вычисления теоретических значений переменной y_x отметим пункт «Остатки» (рис. 4.8), нажмем «OK».

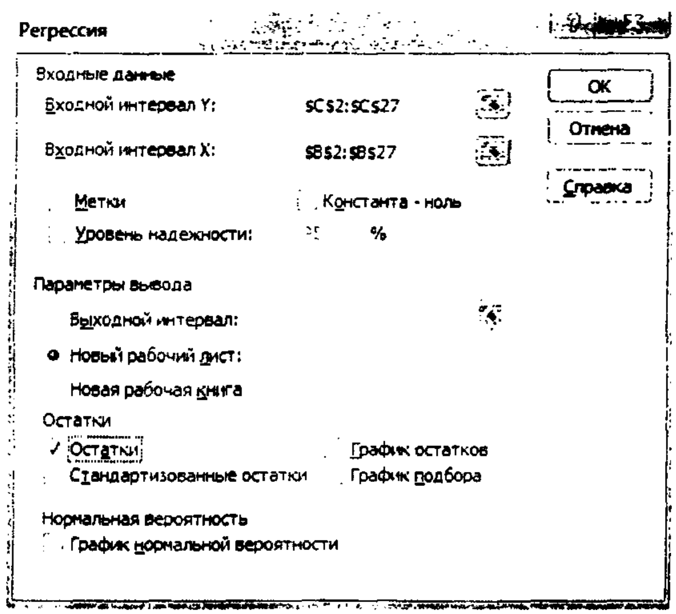


Рис. 4.8. Настройки инструмента «Регрессия» пакета анализа данных *Excel*

В результате получим новый рабочий лист *Excel*, содержащий результаты построения парной линейной регрессии (рис. 4.9).

ВЫВОД ИТОГОВ	
Регрессионная статистика	
Множественный R	0,715547825
R-квадрат	0,51200869
Нормированный R-квадрат	0,491675718
Стандартная ошибка	1098,433749
Наблюдения	26
Дисперсионный анализ	
	df
Регрессия	1
Остаток	24
Итого	25
	Коэффициенты
Y-пересечение	1289,794303
Переменная X 1	1,020572725

Рис. 4.9. Итоги оценки параметров в *Excel*

Таким образом, уравнение линейной парной регрессии имеет вид:

$$y_x = 1289{,}8 + 1{,}02 \cdot x.$$

Теоретические значения результативной переменной (y_x) найдем в столбце «Предсказанное *Y*» (рис. 4.10).

ВЫВОД ОСТАТКА		
Наблюдение	*ПредсказанноеY*	*Остатки*
1	1495,538624	-484,23862
2	1738,128869	180,87113
3	3185,811925	-989,91192
4	2934,955037	-1638,955
5	2092,267763	333,03224
6	2733,902121	923,09788
7	4913,846433	854,15357
8	1749,763404	-869,7634
9	3057,423819	1960,5762
10	5034,376125	-355,57613
11	3231,941833	-417,94183
12	3352,981812	-1928,3818
13	5984,73387	-110,73387
14	1958,368561	-56,268561
15	2170,953955	-440,95395
16	2125,436391	440,96361
17	1791,964104	531,0359
18	2120,639697	-78,339697
19	3042,52345	-719,52345
20	2470,59424	-787,39424
21	2760,437023	-1231,437
22	3447,997175	507,70283
23	1486,251408	1198,7486
24	2562,956113	458,74389
25	2568,671323	-336,67132
26	3201,834924	3057,1651

Рис. 4.10. Теоретические значения результативной переменной y_x в *Excel*

б) *степенная парная регрессия* $y_x = a_0 \cdot x^{a_1}$.

Путем логарифмирования обеих частей исходного вида уравнения приведем его к линейному виду:

$$\lg(y) = \lg(a_0 \cdot x^{a_1}),$$

отсюда

$$\lg y = \lg a_0 + a_1 \lg x.$$

Таблица 4.3

Расчетная таблица для модели степенной парной регрессии

№ района	x	y	X	Y	XY	X^2	y_x
1	201,6	1011,3	2,30	3,00	6,92	5,31	1241,55
2	439,3	1919	2,64	3,28	8,68	6,98	1687,50
3	1857,8	2195,9	3,27	3,34	10,92	10,69	2978,38
4	1612	1296	3,21	3,11	9,98	10,29	2816,42
5	786,3	2425,3	2,90	3,38	9,80	8,38	2122,55
6	1415	3657	3,15	3,56	11,23	9,93	2675,43
7	3551	5768	3,55	3,76	13,35	12,60	3844,44
8	450,7	880	2,65	2,94	7,81	7,04	1704,62
9	1732	5018	3,24	3,70	11,98	10,49	2897,23
10	3669,1	4678,8	3,56	3,67	13,08	12,71	3894,32
11	1903	2814	3,28	3,45	11,31	10,75	3006,73
12	2021,6	1424,6	3,31	3,15	10,43	10,93	3079,21
13	4600,3	5874	3,66	3,77	13,80	13,42	4257,29
14	655,1	1902,1	2,82	3,28	9,24	7,93	1975,24
15	863,3	1730	2,94	3,24	9,51	8,62	2202,17
16	818,8	2566,4	2,91	3,41	9,93	8,49	2156,69
17	492,05	2323	2,69	3,37	9,06	7,25	1764,60
18	814,1	2042,3	2,91	3,31	9,63	8,47	2151,80
19	1717,4	2323	3,23	3,37	10,89	10,46	2887,58
20	1157	1683,2	3,06	3,23	9,88	9,38	2471,43
21	1441	1529	3,16	3,18	10,06	9,98	2694,69
22	2114,7	3955,7	3,33	3,60	11,96	11,06	3134,32
23	192,5	2685	2,28	3,43	7,83	5,22	1219,16
24	1247,5	3021,7	3,10	3,48	10,77	9,59	2545,86
25	1253,1	2232	3,10	3,35	10,37	9,60	2550,36
26	1873,5	6259	3,27	3,80	12,42	10,71	2988,28
Итого	38880	73214,3	79,53	88,17	270,88	246,27	66947,84

Введем новые условные переменные, заменяя в уравнении переменные $\lg y = Y$, $\lg a_0 = A_0$, $\lg x = X$, получим однофакторное уравнение линейной регрессии:

$$Y = A_0 + a_1 X,$$

параметры которого получим из следующей системы:

$$\begin{cases} nA_0 + a_1 \sum X = \sum Y, \\ A_0 \sum X + a_1 \sum X^2 = \sum YX. \end{cases}$$

Для ее решения сначала в расчетной таблице 4.3 вычислим необходимые значения сумм.

Решая систему относительно A_0 и a_1, имеем:

$$\begin{cases} 26A_0 + 79{,}53a_1 = 88{,}17, \\ 79{,}53A_0 + 246{,}27a_1 = 270{,}88. \end{cases}$$

Отсюда $A_0 = 2{,}186$, $a_1 = 0{,}394$.

Проведем потенцирование для параметра a_0:

$a_0 = 10^{A_0}$, $a_0 = 153{,}46$.

Известные значения параметров подставим в уравнение степенной регрессии:

$y_x = 153{,}46 \cdot x^{0{,}394}$.

Определим теоретические (расчетные) значения результативного показателя в таблице 4.3.

в) *показательная регрессия* $y_x = a_0 \cdot a_1^{\,x}$.

Путем логарифмирования обеих частей уравнения приведем его к линейному виду

$$\lg(y) = \lg(a_0 \cdot a_1^{\,x}),$$

отсюда

$$\lg y = \lg a_0 + x \lg a_1.$$

Введем новые условные переменные: $\lg y = Y$, $\lg a_0 = A_0$, $\lg a_1 = A_1$. Получим однофакторное уравнение линейной регрессии:

$$Y = A_0 + A_1 x,$$

параметры которого получим из системы, рассчитав для этого в таблице 4.4 необходимые значения сумм:

$$\begin{cases} nA_0 + A_1\sum x = \sum Y, \\ A_0\sum x + A_1\sum x^2 = \sum Yx. \end{cases}$$

Решая систему относительно A_0 и A_1, имеем:

$$\begin{cases} 26A_0 + 38880A_1 = 88{,}17, \\ 38880A_0 + 87309842{,}1A_1 = 135918{,}73. \end{cases}$$

Отсюда $A_0 = 3{,}491$, $A_1 = -0{,}00007$.

Проведем потенцирование для параметров a_0 и a_1:

$a_0 = 10^{A_0}$, , $a_0 = 3097{,}4$,

$a_1 = 10^{A_1}$, $a_1 = 0{,}9998$.

Таблица 4.4

Расчетная таблица для модели показательной парной регрессии

№ п/п	x	y	Y	Y_x	x^2	y_x
1	201,6	1011,3	3,00	605,78	40642,6	2975,0
2	439,3	1919	3,28	1442,25	192984,5	2836,8
3	1857,8	2195,9	3,34	6208,05	3451420,8	2136,1
4	1612	1296	3,11	5017,52	2598544,0	2243,7
5	786,3	2425,3	3,38	2661,44	618267,7	2646,6
6	1415	3657	3,56	5041,82	2002225,0	2333,9
7	3551	5768	3,76	13355,40	12609601,0	1522,4
8	450,7	880	2,94	1327,08	203130,5	2830,4
9	1732	5018	3,70	6409,32	2999824,0	2190,5
10	3669,1	4678,8	3,67	13466,09	13462294,8	1486,9
11	1903	2814	3,45	6564,06	3621409,0	2116,8
12	2021,6	1424,6	3,15	6375,51	4086866,6	2067,2
13	4600,3	5874	3,77	17338,23	21162760,1	1234,2
14	655,1	1902,1	3,28	2148,23	429156,0	2717,0
15	863,3	1730	3,24	2795,41	745286,9	2606,2
16	818,8	2566,4	3,41	2791,55	670433,4	2629,5
17	492,05	2323	3,37	1656,26	242113,2	2807,1

Окончание табл. 4.4

№ п/п	x	y	Y	Y_x	x^2	y_x
18	814,1	2042,3	3,31	2694,77	662758,8	2632,0
19	1717,4	2323	3,37	5780,85	2949462,8	2196,9
20	1157	1683,2	3,23	3732,64	1338649,0	2457,5
21	1441	1529	3,18	4588,73	2076481,0	2321,8
22	2114,7	3955,7	3,60	7607,05	4471956,1	2029,1
23	192,5	2685	3,43	660,07	37056,3	2980,4
24	1247,5	3021,7	3,48	4341,61	1556256,3	2413,4
25	1253,1	2232	3,35	4196,25	1570259,6	2410,7
26	1873,5	6259	3,80	7112,75	3510002,3	2129,4
Итого	38880	73214,3	88,17	135918,73	87309842,1	60951,4

Известные значения параметров подставим в уравнение показательной регрессии:

$$y_x = 3097{,}4 \cdot 0{,}9998^x.$$

Определим теоретические (расчетные) значения результативного показателя в таблице 4.4.

г) *логарифмическая регрессия* $y_x = a_0 + a_1 \lg x$.

Таблица 4.5

Расчетная таблица для модели логарифмической парной регрессии

№ района	x	y	X	yX	X^2	y_x
1	201,6	1011,3	2,30	2330,5	5,3	779,7
2	439,3	1919	2,64	5071,5	7,0	1692,7
3	1857,8	2195,9	3,27	7178,4	10,7	3383,1
4	1612	1296	3,21	4156,7	10,3	3216,7
5	786,3	2425,3	2,90	7022,7	8,4	2375,2
6	1415	3657	3,15	11522,3	9,9	3063,9
7	3551	5768	3,55	20478,4	12,6	4142,5
8	450,7	880	2,65	2335,4	7,0	1722,8
9	1732	5018	3,24	16251,0	10,5	3300,9
10	3669,1	4678,8	3,56	16677,9	12,7	4180,9
11	1903	2814	3,28	9228,3	10,8	3411,3

Окончание табл. 4.5

№ района	x	y	X	yX	X^2	y_x
12	2021,6	1424,6	3,31	4709,3	10,9	3482,1
13	4600,3	5874	3,66	21515,2	13,4	4446,0
14	655,1	1902,1	2,82	5356,9	7,9	2161,2
15	863,3	1730	2,94	5079,6	8,6	2484,7
16	818,8	2566,4	2,91	7476,4	8,5	2422,6
17	492,05	2323	2,69	6253,5	7,2	1825,7
18	814,1	2042,3	2,91	5944,5	8,5	2415,9
19	1717,4	2323	3,23	7514,6	10,5	3291,0
20	1157	1683,2	3,06	5156,2	9,4	2827,9
21	1441	1529	3,16	4829,6	10,0	3085,3
22	2114,7	3955,7	3,33	13153,7	11,1	3534,9
23	192,5	2685	2,28	6133,7	5,2	725,5
24	1247,5	3021,7	3,10	9355,3	9,6	2916,2
25	1253,1	2232	3,10	6914,7	9,6	2921,5
26	1873,5	6259	3,27	20483,5	10,7	3392,9
Итого	38880	73214,3	79,53	232129,9	246,3	73203,1

Рассмотрим алгоритм параметризации модели логарифмической парной регрессии:

$$y_x = a_0 + a_1 \lg x.$$

Введем новую условную переменную $\lg x = X$, получим однофакторное уравнение линейной регрессии:

$$y_x = a_0 + a_1 X,$$

параметры которого получим из системы:

$$\begin{cases} na_0 + a_1 \sum X = \sum y, \\ a_0 \sum X + A_1 \sum X^2 = \sum yX. \end{cases}$$

Решая систему относительно a_0 и a_1, рассчитав необходимые суммы в таблице 4.5, получим:

$$\begin{cases} 26a_0 + 79{,}53a_1 = 73214{,}3, \\ 79{,}53a_0 + 246{,}3a_1 = 232129{,}9. \end{cases}$$

Отсюда $a_0 = -5440{,}6 a_1 = 2699{,}2$.

Известные значения параметров подставим в уравнение логарифмической регрессии:

$$y_x = -5440{,}6 + 2699{,}2 \lg x.$$

Определим теоретические (расчетные) значения результативного показателя в таблице 4.5.

В заключение отобразим графически результаты решения задачи. Для этого в единой системе координат построим по исходным данным (x_i, y_i) поле корреляции и в соответствии с синтезированными моделями линии регрессии. Координатами для их отображения являются фактические данные по факторной переменной (x) и соответствующие значения теоретических значений результативной (y_x).

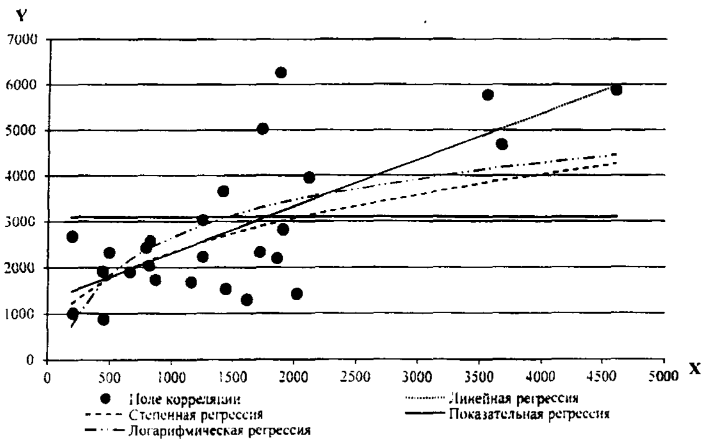


Рис. 4.11. Исходное поле корреляции и полученные линии регрессии

Задача 4.2. По данным о производстве мяса всех видов во всех категориях хозяйств (Y) и заготовке кормов в сельхозпредприятиях в расчете на одну условную голову скота (X) с помощью программного продукта *Statistica* 6.0 оценить параметры уравнений регрессии следующего вида:

а) линейной $y_x = a_0 + a_1 x$;

б) степенной $y_x = a_0 \cdot x^{a_1}$;

в) логарифмической $y_x = a_0 + a_1 \ln x$.

Таблица 4.6

Данные о производстве мяса всех видов и заготовке кормов в сельхозпредприятиях в разрезе районов Ставропольского края в 2012 г.

№ района	Производство мяса всех видов во всех категориях хозяйств, т (в живом весе)	Заготовлено кормов в сельхозпредприятиях в расчете на одну условную голову скота, ц корм. ед.
	Y	X
1	4038,0	22,0
2	5791,0	25,4
3	6580,0	22,8
4	4078,0	29,1
5	9322,0	24,5
6	5748,0	23,4
7	13195,0	23,9
8	7786,0	32,7
9	3586,0	37,1
10	6510,0	26,0
11	10307,0	24,5
12	8718,0	27,4
13	7105,0	45,0
14	9526,0	24,2
15	4844,0	22,0
16	11685,0	22,8

Решение

Для определения параметров уравнений регрессии воспользуемся инструментами анализа программы *Statistica* 6.0. Для этого в «Рабочей книге» создадим папку «Задача 4.2», в которую поместим таблицу 4.6 с исходными значениями x и y.

В спецификациях переменных обозначим переменную x текстовой меткой «производство мяса», а переменную y — текстовой меткой «обеспеченность населения мясными продуктами» (рис. 4.12). Редактор спецификаций переменных находится в меню «Данные».

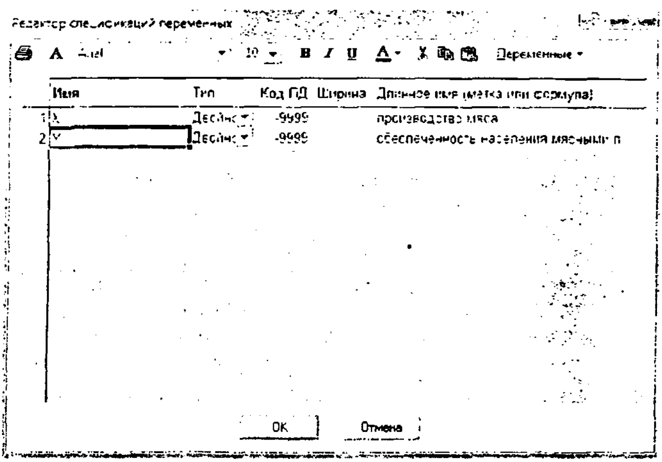


Рис. 4.12. Редактор спецификаций переменных в *Statistica* 6.0

а) линейная регрессия $y_x = a_0 + a_1 x$.

Для оценки параметров линейной регрессии обратимся к модулю «Множественная регрессия», расположенному в меню «Анализ». В диалоговом окне «Множественная регрессия» необходимо указать зависимую (y) и независимую (x) переменные регрессии (рис. 4.13), нажмем «ОК».

Во вкладке «Быстрый» выберем пункт «Итоговая таблица регрессии» (рис. 4.14). В столбце «B» находятся значения параметров регрессии. Следовательно, уравнение линейной регрессии имеет вид $y_x = 101\,139 - 100{,}3x$.

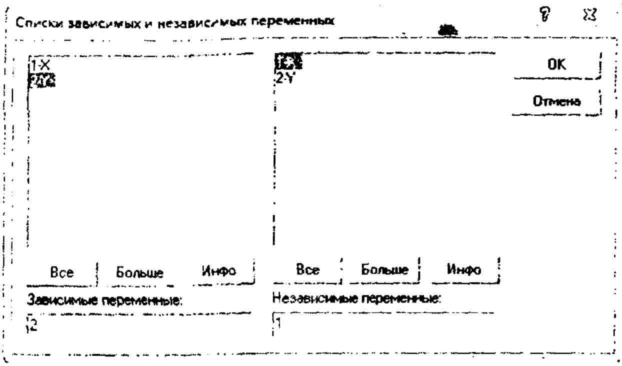


Рис. 4.13. Диалоговое окно «Списки зависимых и независимых переменных» программы *Statistica 6.0*

	Итоги регрессии для зависимой переменной: Y (Таблица 4.6 R= ,22253344 R2= ,04952113 Скорректир. R2= --- F(1.14)=,72942 p<,40745 Станд. ошибка оценки: 2863.4					
N=16	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(14)	p-уров.
Св.член			10138 66	3255,656	3.114166	0.007613
X	-0,222533	0,260560	-100,28	117,411	-0,854059	0,407449

Рис. 4.14. Итоговая таблица регрессии в *Statistica* 6.0

В окне «Результаты множественной регрессии» необходимо открыть вкладку «Остатки/предсказанные/наблюдаемые значения», в которой выбераем пункт «Анализ остатков» → «Диаграммы рассеяния» → «Две переменные». В результате получим диаграмму, содержащую эмпирические (наблюдаемые) значения переменных x и y и линию регрессии $y_x = 101\ 139 - 100{,}3x$ (рис. 4.15).

Обратим внимание, что на диаграмме пунктирными линиями указаны доверительные интервалы линейной регрессии, в пределах которых эконометрическое моделирование считается статистически значимым.

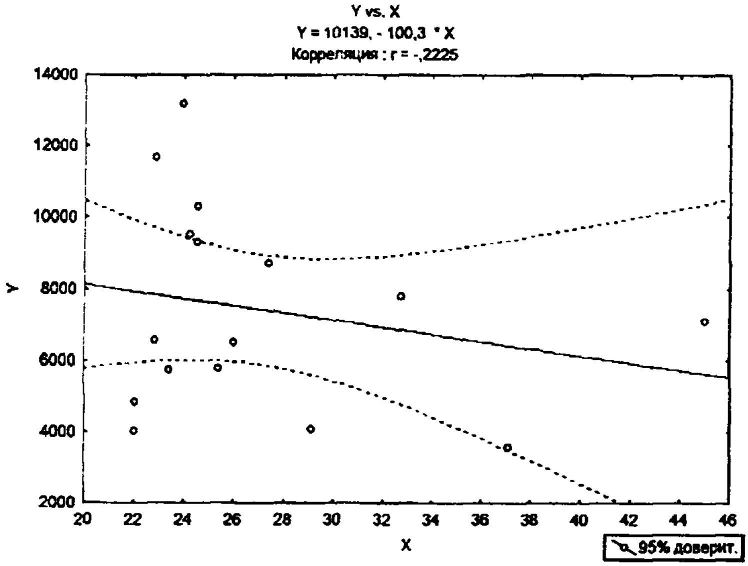


Рис. 4.15. Графическая интерпретация регрессионного анализа средствами *Statistica 6.0*

Кроме того, во вкладке «Быстрый» можно рассчитать теоретические значения y_x и значения остатков (рис. 4.16).

	Предсказанные значения и остатки Зависимая перемен.: Y		
Набл. No.	Наблюд. Значение	Предск. Значение	Остатки
1	4038,00	7932,583	-3894,58
2	5791,00	7591,644	-1800,64
3	6580,00	7852,362	-1272,36
4	4078,00	7220,621	-3142,62
5	9322,00	7681,892	1640,11
6	5748,00	7792,196	-2044,20
7	13195,00	7742,058	5452,94
8	7786,00	6859,626	926,37
9	3586,00	6418,411	-2832,41
10	6510,00	7531,478	-1021,48
11	10307,00	7681,892	2625,11
12	8718,00	7391,091	1326,91
13	7105,00	5626,228	1478,77
14	9526,00	7711,975	1814,03
15	4844,00	7932,583	-3088,58
16	11685,00	7852,362	3832,64

Рис. 4.16. Фактические, теоретические значения и остатки линейной регрессии в *Statistica 6.0*

б) степенная регрессия $y_x = a_0 * x^{a_1}$.

Для оценки параметров степенной регрессии обратимся к инструменту «Нелинейное оценивание», который находится в меню «Анализ» → «Углубленные методы анализа». В окне «Оцениваемая функция» необходимо указать уравнение степенной регрессии: $Y = a_0 * x^{a_1}$. В результате получим значения параметров для степенной регрессии (рис. 4.17). Таким образом, уравнение регрессии имеет вид: $y_x = 31\,430{,}77 * x^{-0{,}44}$.

Модель Y=a0*X**a1 (Таблица 4 6 в Workbook2)
Зав Пер Y
Уров. значимости 95.0% (альфа=0.050)

	Оценка	Стандарт. ошиб.	t-знач. cc = 14	p-уров.	Ниж. Дов Предел	Вер. Дов Предел
a0	31430.77	57237.51	0.549129	0.591568	-91331.5	154193.0
a1	-0.44	0.56	-0.788049	0.443807	-1.6	0.8

Рис. 4.17. Оценка параметров степенной регрессии в *Statistica 6.0*

Для графического изображения фактических значений переменных x и y, а также линии степенной регрессии построим таблицу «Наблюдаемые, предсказанные и значения остатков» и добавим к этой таблице значения независимой переменной x (рис. 4.18):

Модель Y=a0*X**a1 (Таблица 4 6 в Workbock2)
Зав. Пер. Y

	X	Наблюд.	Предсказанные	Остатки
1	22	4038.00	8032.452	-3994.45
2	25.4	5791.00	7538.791	-1747.79
3	22.8	6580.00	7906.814	-1326.81
4	29.1	4078.00	7099.612	-3021.61
5	24.5	9322.00	7659.791	1662.21
6	23.4	5748.00	7816.681	-2068.68
7	23.9	13195.00	7744.078	5450.92
8	32.7	7786.00	6743.371	1042.63
9	37.1	3586.00	6377.910	-2791.91
10	26	6510.00	7461.504	-951.50
11	24.5	10307.00	7659.791	2647.21
12	27.4	8718.00	7290.766	1427.23
13	45	7105.00	5856.989	1248.01
14	24.2	9526.00	7701.558	1824.44
15	22	4844.00	8032.452	-3188.45
16	22.8	11685.00	7906.814	3778.19

Рис. 4.18. Исходные данные для графического изображения линии степенной регрессии в *Statistica 6.0*

Воспользовавшись инструментом «Диаграммы рассеяния» в панели инструментов «Графика», построим поле корреляции, содержащее эмпирические значения переменных x и y.

К этому графику добавим линию степенной регрессии, указав в качестве координат по оси ординат теоретические значения y_x, вычисленные из уравнения степенной регрессии (рис. 4.19).

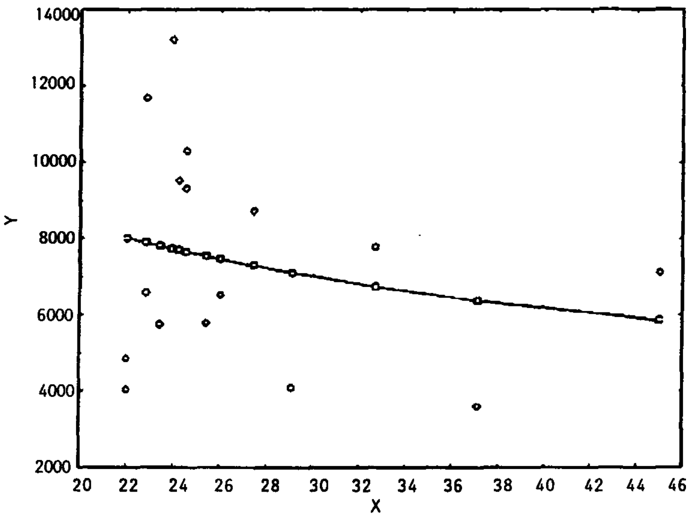


Рис. 4.19. Линия степенной регрессии в *Statistica 6.0*

в) логарифмическая регрессия $y_x = a_0 + a_1 \ln x$.

Для оценки параметров логарифмической регрессии обратимся к инструменту «Нелинейное оценивание». В окне «Оцениваемая функция» необходимо указать уравнение логарифмической регрессии.

В окне результатов анализа найдем значения параметров, обратившись к кнопке «Оценки параметров модели». Как видим, значение $a_0 = 17\,746{,}66$; $a_1 = -3149{,}62$. Следовательно, уравнение логарифмической регрессии имеет вид $y_x = 17\,746{,}66 - 3149{,}62 \ln x$ (рис. 4.20).

	Модель: y=a0+a1*log(x) (Таблица данных в Workbook1) Зав. Пер. : y Уров. значимости: 95.0% (альфа=0.050)					
	Оценка	Стандарт ошиб.	t-знач. cc = 14	p-уров.	Ниж. Дов Предел	Вер. Дов Предел
a0	17746.66	12020.31	1.476389	0.161976	-8034.3	43527.66
a1	-3149.62	3661.88	-0.860111	0.404217	-11003.6	4704.32

Рис. 4.20. Оценка параметров логарифмической регрессии в *Statistica 6.0*

Далее обратимся ко вкладке «Наблюдаемые, предсказанные и значения остатков». В результате получим фактические, теоретические значения зависимой переменной y и значения остатков $(y - y_t)$. Предварительно выделив наблюдения с 1 по 26 по трем столбцам, воспользуемся инструментом «Блоковые статистики» $\rightarrow$ «По столбцам» $\rightarrow$ «Суммы». Получим таблицу, изображенную на рисунке 4.21.

	Модель: y=a0+a1*log(x) (Таблица данных Зав. Пер. : y		
	Наблюд.	Предсказанные	Остатки
1	4038.00	8011.037	-3973.04
2	5791.00	7558.415	-1767.42
3	6580.00	7898.539	-1318.54
4	4078.00	7130.101	-3052.10
5	9322.00	7672.041	1649.96
6	5748.00	7816.726	-2068.73
7	13195.00	7750.135	5444.86
8	7786.00	6762.739	1023.26
9	3586.00	6365.125	-2779.12
10	6510.00	7484.880	-974.88
11	10307.00	7672.041	2634.96
12	8718.00	7319.693	1398.31
13	7105.00	5757.104	1347.90
14	9526.00	7710.846	1815.15
15	4844.00	8011.037	-3167.04
16	11685.00	7898.539	3786.46
СУММА	118819.00	118819.000	0.00

Рис. 4.21. Фактические, теоретические значения и остатки логарифмической регрессии в *Statistica 6.0*

Из этой таблицы видно, что суммы эмпирических и теоретически найденных из уравнения логарифмической регрессии значения переменной y не отличаются, что свидетельствует о высокой точности вычислений.

В заключение построим диаграмму, содержащую 26 точек эмпирических пар значений переменных x и y, а также график найденного уравнения логарифмической регрессии (рис. 4.22).

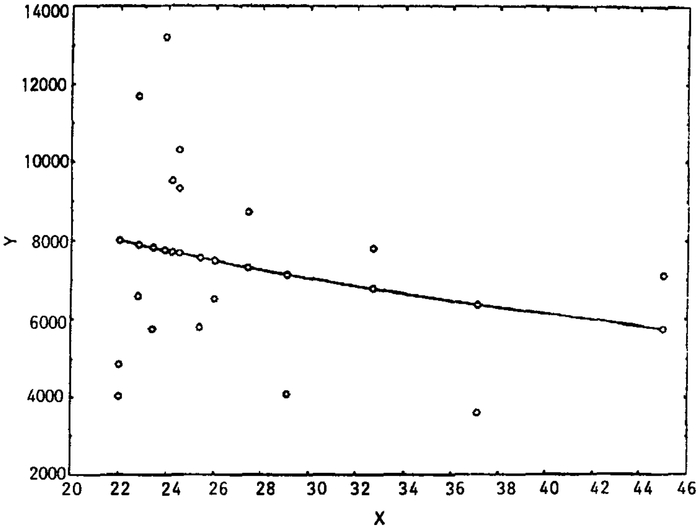


Рис. 4.22. Линия логарифмической регрессии в *Statistica 6.0*

Задача **4.3.** По данным таблицы 4.7 определите параметры следующих уравнений множественной регрессии:

а) линейной: $y_x = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$;

б) степенной: $y_x = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} + x_3^{a_3} + x_4^{a_4}$;

в) логарифмической:

$$y_x = a_0 + a_1 \cdot \ln x_1 + a_2 \cdot \ln x_2 + a_3 \cdot \ln x_3 + a_4 \cdot \ln x_4.$$

Таблица 4.7

Данные к задаче в разрезе районов края за 2012 г.

№ района	Y	X_1	X_2	X_3	X_4
1	188,2	8685	8907,1	33,7	0,9
2	189,9	8866	7006,3	28,2	7,8
3	190,4	7938	9133,0	23,8	4,8
4	189,3	11089	8878,7	15,2	1,4
5	190,8	10413	9943,5	20,5	1,2

Окончание табл. 4.7

№ района	Y	X_1	X_2	X_3	X_4
6	191,1	11573	8998,6	21,0	0,6
7	205,8	13599	10062,9	15,5	0,0
8	190,9	11636	6904,9	5,1	5,7
9	191,2	15168	7228,2	29,9	0,5
10	190,2	12363	8727,3	58,4	3,9
11	192,6	11398	7061,4	1,4	0,3
12	192,0	17029	8967,2	49,2	4,1
13	191,3	20416	9977,4	23,0	1,3
14	193,8	12530	6710,6	18,1	0,1
15	191,4	6903	6966,1	23,0	4,5
16	191,1	10343	10162,8	20,8	0,1
17	192,1	9847	8860,3	24,5	0,4
18	189,7	16582	9978,0	30,7	3,0
19	192,6	12612	6879,6	18,5	0,9
20	190,6	10549	8922,5	26,8	1,2
21	193,4	14274	9986,2	30,7	1,0
22	190,8	18749	9990,6	24,1	2,3
23	191,5	10163	10804,7	16,2	0,0
24	189,3	16077	8996,1	12,9	1,8
25	190,4	8945	7882,2	13,2	1,6
26	197,6	18360	6933,1	50,3	2,3

Примечание. Y – фактическое потребление молочных продуктов на 1 человека в год, кг; X_1 – среднемесячная заработная плата работающих, руб.; X_2 – средний размер месячных пенсий всех пенсионеров, руб.; X_3 – производство молока (заданной жирности), тыс. т; X_4 – численность коров во всех категориях хозяйств на начало года, тыс. голов.

Решение

Оценку параметров уравнений регрессии осуществим с помощью программного продукта *SPSSStatistics 20.* Для этого вначале сформируем таблицу исходных наблюдений в окне ввода данных.

а) *линейная регрессия.*

Для оценки параметров линейной регрессии обратимся к инструменту «Анализ» → «Регрессия» → «Линейная».

В диалоговом окне «Линейная регрессия» отметим в качестве зависимой переменной переменную y, а в качестве независимых — совокупность переменных x_1, x_2, x_3, x_4 (рис. 4.23).

В качестве метода включения независимых переменных в уравнение укажем метод принудительного включения.

Во вкладке «Сохранить» отметим пункт «Нестандартизированные предсказанные значения», нажмем «Продолжить».

Для отображения результата моделирования нажмем «ОК».

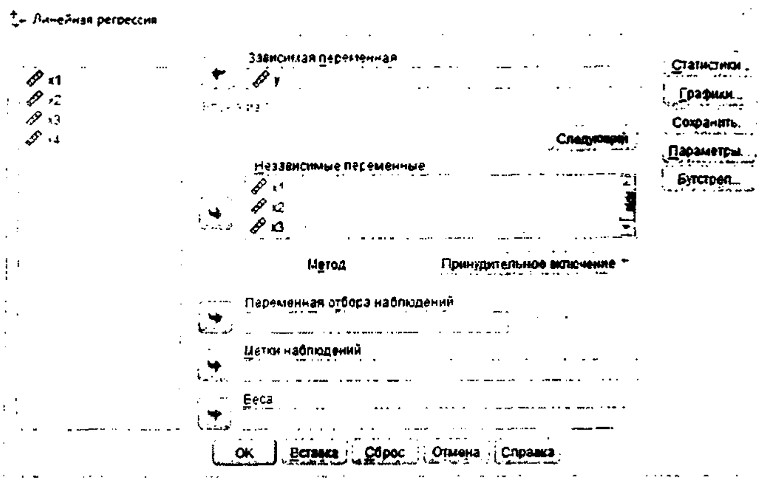


Рис. 4.23. Опции модуля «Линейная регрессия» в *SPSSStatistics 20*

В окне вывода результатов найдем таблицу «Коэффициенты» (рис. 4.24). В столбце «В» данной таблицы указаны значения искомых параметров:

$$a_0 = 192{,}97;\ a_1 = 0{,}000;\ a_2 = 0{,}000;\ a_3 = 0{,}003;\ a_4 = -0{,}482.$$

Следовательно, уравнение множественной линейной регрессии имеет вид $y_x = 192{,}97 + 0{,}003x_3 - 0{,}482x_4$.

Обратим внимание, что в окне ввода данных в столбце «PRE_1» отображены теоретические значения результативной переменной y, найденные по уравнению линейной регрессии.

Коэффициенты[a]

Модель		Нестандартизованные коэффициенты B	Нестандартизованные коэффициенты Стд. ошибка	Стандартизованные коэффициенты Бета	t	Знч.
1	(Константа)	192.974	5.440		35.475	.000
	x1	,000	.000	.211	,957	.350
	x2	,000	.001	-,121	-.552	,587
	x3	,003	.059	,011	,048	.962
	x4	-.482	.383	-.288	-1.259	.222

Рис. 4.24. Оценка параметров линейной регрессии в *SPSSStatistics 20*

Откроем меню «Анализ» → «Описательные статистики» → «Описательные». В окне «Переменные» укажем фактические и теоретические значения результативной переменной (рис. 4.25). Во вкладке «Параметры» отметим пункт «Сумма»; нажмем «ОК».

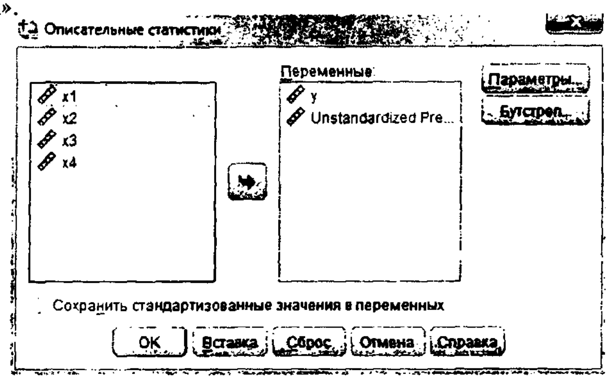


Рис. 4.25. Окно «Описательные статистики» *SPSSStatistics 20*

б) *степенная регрессия.*

Для оценки параметров степенной регрессии воспользуемся инструментом «Анализ» → «Регрессия» → «Нелинейная». В диалоговом окне «Нелинейная регрессия» укажем следующие опции:

- в качестве зависимой переменной укажем переменную y;
- во вкладке «Параметры» добавим параметры a_0, a_1, a_2, a_3 и a_4, присвоив каждому из них начальное значение «0» (рис. 4.26), нажмем «Продолжить»;

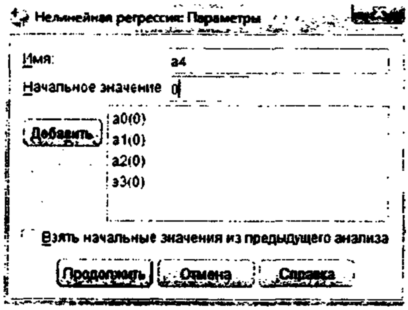


Рис. 4.26. Ввод имен параметров степенной регрессии в *SPSSStatistics 20*

- в окне «Выражение, задающее модель» необходимо указать уравнение степенной регрессии вида «*a*0**x*1***a1***x2****a2***x3****a3***x4****a4*» (рис. 4.27);

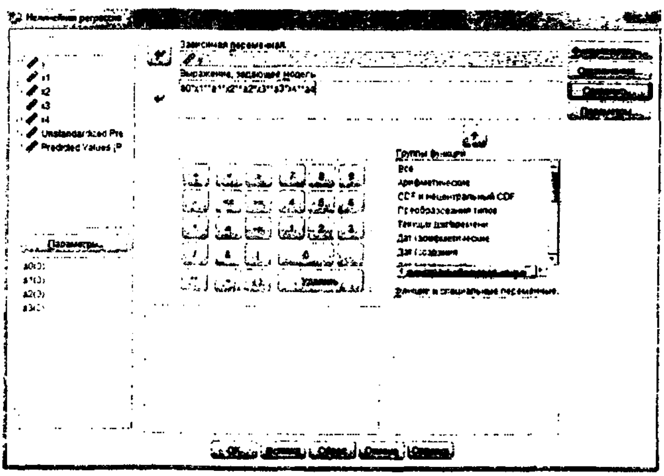


Рис. 4.27. Опции модуля «Нелинейная регрессия» в *SPSSStatistics 20*

- во вкладке «Сохранить» отметим пункт «Предсказанные значения»; нажмем «Продолжить», затем «ОК».

В окне вывода результатов найдем таблицу «Оценки параметров» (рис. 4.28).

Оценки параметров

Параметр	Оценка	Стд. ошибка	Доверительный интервал 95 %	
			Нижняя граница	Верхняя граница
a0	192,974	5,440	181,661	204,287
a3	,003	,059	-,120	,125
a4	-,482	,383	-1,278	,314

Рис. 4.28. Оценка параметров степенной регрессии в *SPSSStatistics 20*

Таким образом, уравнение степенной регрессии имеет вид $y_x = 192{,}97 \cdot x_3^{0{,}003} \cdot x_4^{-0{,}482}$.

В окне ввода данных в столбце «PRED_» отображены теоретические значения результативной переменной y, найденные по уравнению степенной регрессии.

Откроем меню «Анализ» → «Описательные статистики» → «Описательные». В окне «Переменные» укажем фактические и теоретические значения результативной переменной (рис. 4.29). Во вкладке «Параметры» отметим пункт «Сумма»; нажмем «ОК».

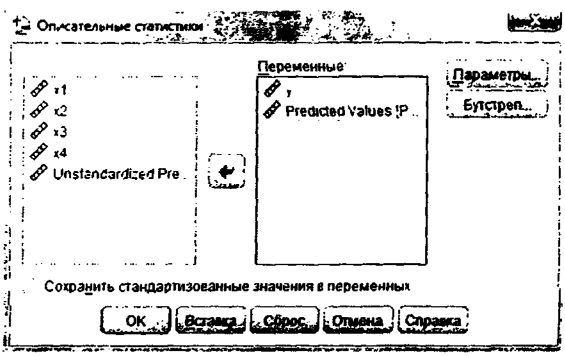


Рис. 4.29. Окно «Описательные статистики» в *SPSSStatistics 20*

в) *логарифмическая регрессия.*

Аналогично вычислению параметров степенной регрессии можно найти параметры логарифмической регрессии. Для этого в окне инструмента «Нелинейная регрессия» необходимо ввести логарифмическую функцию: «*a0*+*a1**LN(*x1*)+*a2**LN(*x2*)+ +*a3**LN(*x3*)+*a4**LN(*x4*)» (рис. 4.30):

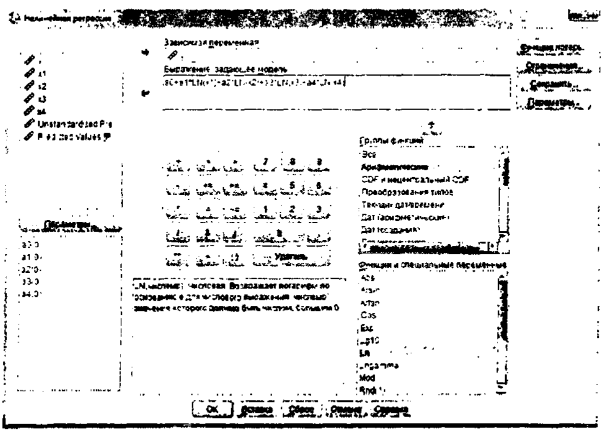


Рис. 4.30. Ввод логарифмической функции в *SPSSStatistics 20*

Анализируя результаты оценки параметров логарифмической регрессии, имеем уравнение следующего вида: $y_x = 228{,}38 + 2{,}73 \cdot \ln x_1 - 7{,}12 \cdot \ln x_2 + 0{,}56 \cdot \ln x_3 - 0{,}49 \cdot \ln x_4$ (рис. 4.31).

Оценки параметров

Параметр	Оценка	Стд. ошибка	Доверительный интервал 95 %	
			Нижняя граница	Верхняя граница
a0	228.377	20.341	186.075	270.679
a1	2.730	1.089	.466	4.994
a2	-7.119	2.209	-11.713	-2.524
a3	,559	,442	-,359	1.478
a4	-,486	,272	-1,052	,081

Рис. 4.31. Оценка параметров логарифмической регрессии в *SPSSStatistics 20*

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 4.4. Для линейного парного уравнения регрессии при двенадцати наблюдениях известны следующие значения: $\sum x = 15$, $\sum x^2 = 85$, $\sum yx = 125$, $\sum y = 58$, $\sum y^2 = 120$. Определите параметры уравнения регрессии.

Задача 4.5. Для парного уравнения регрессии, синтезированного на основе равносторонней гиперболы, при десяти наблюдениях известны следующие значения:

$$\sum \frac{1}{x} = 15, \sum \frac{1}{x^2} = 85, \sum \frac{y}{x} = 125, \sum y = 58, \sum \frac{1}{y^2} = 120.$$

Определите параметры уравнения регрессии.

Задача 4.6. Для уравнения линейной парной регрессии при пятнадцати наблюдениях известны следующие значения: $\sum x = 15$, $\sum x^2 = 85$, $\sum yx = 125$, $\sum y = 58$, $\sum y^2 = 120$. Рассчитайте параметры уравнения регрессии методом определителей.

Задача 4.7. Для парного уравнения регрессии, аппроксимированного на основе степенной функции, для девяти наблюдений известны следующие значения сумм: $\sum \lg x = 15$, $\sum (\lg x)^2 = 85$, $\sum \lg y \cdot \lg x = 125$, $\sum \lg y = 58$, $\sum (\lg y)^2 = 120$. Определите параметры уравнения регрессии.

Задача 4.8. Для парного уравнения регрессии, аппроксимированного на основе показательной функции, для десяти наблюдений известны следующие значения: $\sum x = 132$, $\sum x^2 = 1183$, $\sum x \cdot \lg y = 1152$, $\sum \lg y = 58{,}3$, $\sum (\lg y)^2 = 151{,}43$. Определите параметры уравнения регрессии.

Задача 4.9. Для парного уравнения регрессии, аппроксимированного в соответствии с логарифмической функцией, для восемнадцати наблюдений известны следующие значения:

$\sum \lg x = 35{,}6$, $\sum (\lg x)^2 = 85{,}2$, $\sum y \cdot \lg x = 323$, $\sum y = 58$, $\sum y^2 = 120$. Найдите параметры уравнения регрессии методом определителей.

Задача 4.10. По данным об индексе физического объема инвестиций (Y) и среднесложившихся цен на зерно (X) оценить параметры уравнений регрессии следующего вида:

а) линейной $y_x = a_0 + a_1 x$;

б) степенной $y_x = a_0 \cdot x^{a_1}$.

Таблица 4.8

Данные об индексе физического объема инвестиций и стоимости сельскохозяйственной продукции в 2012 г. в разрезе районов Ставропольского края

№ района	Индекс физического объема инвестиций, в % к предыдущему году	Среднесложившиеся цены на зерно, реализованное по всем каналам, руб./ц
	Y	X
1	200,0	648,6
2	47,6	612,9
3	68,9	549,8
4	96,8	645,5
5	26,4	694,1
6	143,5	686,3
7	177,8	689,6
8	49,6	683,2
9	127,2	747,2
10	84,6	673,7
11	269,7	643,2
12	68,5	751,7
13	161,4	787,9
14	98,5	591,9

Задача 4.11. По данным о прибыли сельхозпредприятий (Y) и соотношении производственной себестоимости зерна к средней по краю (X) оценить параметры уравнений регрессии следующего вида:

а) линейной $y_x = a_0 + a_1 x$;

б) равносторонней гиперболы $y_x = a_0 + \frac{a_1}{x}$.

Таблица 4.9

Данные о соотношении производственной себестоимости зерна к средней по краю и прибыли (убытка) сельхозпредприятий в 2012 г. в разрезе районов края

№ района	Соотношение производственной себестоимости 1 ц зерна к средней по краю, %	Прибыль (убыток) сельхозпредприятий, млн руб.
	X	*Y*
1	116,38	328,1
2	181,36	2,6
3	108,19	309,2
4	117,23	244,1
5	92,09	234,5
6	107,06	414,5
7	125,99	608,5
8	100,00	20,8
9	100,85	299,3
10	148,31	567,2
11	94,35	241,8
12	94,92	112,5

Задача 4.12. По данным об обеспеченности населения мясными продуктами (Y) и объемах производства всех видов мяса в сельхозпредприятиях (X) оценить параметры уравнений регрессии следующего вида:

а) линейной $y_x = a_0 + a_1 x$; ;

б) показательной $y_x = a_0 \cdot a_1{}^x$; ;

в) логарифмической $y_x = a_0 + a_1 \lg x$; ;

г) степенной $y_x = a_0 \cdot x^{a_1}$;

Таблица 4.10

Данные об обеспеченности мясными продуктами населения региона и о производстве мяса всех видов в разрезе районов Ставропольского края в 2012 г.

№ района	Коэффициент обеспеченности мясными продуктами населения	Производство мяса всех видов во всех категориях хозяйств, т (в живом весе)
	Y	X
1	0,654	6264,2
2	0,660	5512,3
3	0,658	11251
4	0,663	7300
5	0,664	5824
6	0,670	8986
7	0,665	24855
8	0,674	5,71
9	0,665	9073
10	0,664	9729,3
11	0,668	4537
12	0,672	53982
13	0,663	13851
14	0,666	7260
15	0,662	13053

Задача 4.13. Используя данные таблицы 4.11, оцените параметры линейного уравнения множественной регрессии при помощи метода наименьших квадратов и рассчитайте теоретические уровни результативного показателя.

Таблица 4.11

Исходные данные о производстве мясной продукции в разрезе районов Ставропольского края в 2012 г.

№ района	Производство мяса всех видов во всех категориях хозяйств, т (в живом весе)	Заготовлено кормов в сельхозпредприятиях в расчете на одну условную голову скота, ц корм. ед.	Падеж крупного рогатого скота в сельхозпредприятиях, %
	Y	X_1	X_2
1	2	3	4
1	6264,2	25	0,5

Окончание табл. 4.11

1	2	3	4
2	5512,3	12,5	3,2
3	11251	17,4	1,3
4	7300	14,9	2,1
5	5824	23	1,1
6	8986	22,9	2,0
7	24855	11,85	2,3
8	5713	12,7	0,5
9	9073	14	2,6
10	9729,3	22,9	0,7
11	4537	16,9	1,2
12	53982	16	2,7
13	13851	25	2,9
14	7260	9,85	1,3
15	13053	26	0,5
16	4527	21,7	2,9
17	15464	24,3	2,9
18	6297	41,6	0,6

Задача 4.14. Используя данные таблицы 4.12, оцените при помощи метода наименьших квадратов параметры следующего уравнения нелинейной множественной регрессии: $y_x = a_0 \cdot a_1^{x_1} \cdot a_2^{x_2}$, рассчитайте теоретические уровни результативного показателя.

Таблица 4.12

Данные о прибыли сельхозпредприятий, среднемесячной заработной плате и соотношении среднесложившейся цены зерна к средней по краю в 2012 г. в разрезе районов Ставропольского края

№ района	Прибыль (убыток) сельхозпредприятий, млн руб.	Среднемесячная заработная плата работающих, руб.	Соотношение среднесложившейся цены 1 ц зерна к средней по краю
	Y	X_1	X_2
1	2	3	4
1	328,1	10685	1,026
2	2,6	8866	1,095
3	309,2	7938,3	0,835

Окончание табл. 4.12

1	2	3	4
4	244,1	11089	0,883
5	328,1	10685	1,026
2	2,6	8866	1,095
3	309,2	7938,3	0,835
4	244,1	11089	0,883
5	234,5	10413	0,963
6	414,5	11573	0,967
7	608,5	13599	1,087
8	20,8	11636	0,916
9	299,3	15168	1,066
10	567,2	12363	0,965
11	241,8	11398	1,009
12	112,5	17028,8	1,160
13	2,6	20416	1,157
14	309,2	8866	0,897
15	244,1	7938,3	0,944

Задача 4.15. По данным таблицы 4.13 определите параметры следующих уравнений множественной регрессии:

а) линейной $y_x = a_0 + a_1x_1 + a_2x_2 + a_3x_3$;

б) степенной $y_x = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot x_3^{a_3}$.

Рассчитайте теоретические уровни результативного показателя.

Таблица 4.13

Данные к задаче в разрезе районов края за 2012 г.

№ района	Коэффициент обеспеченности мясными продуктами населения	Доля животноводства в общем объеме произведенной с.-х. продукции, %	Производство мяса всех видов во всех категориях хозяйств, т (в живом весе)	Численность свиней во всех категориях хозяйств на начало года, тыс. голов
	Y	X_1	X_2	X_3
1	2	3	4	5
1	0,654	36,08	6264,2	4,8
2	0,660	62,39	5512,3	1,7
3	0,662	48,38	11251	0,0

Окончание табл. 4.13

1	2	3	4	5
4	0,658	30,32	7300	5,6
5	0,663	33,43	5824	0,2
6	0,664	28,22	8986	4,4
7	0,715	35,82	24855	1,4
8	0,664	39,68	5713	9,5
9	0,665	34,24	9073	2,7
10	0,661	37,99	9729,3	12,4
11	0,670	37,84	4537	4,3
12	0,667	40,99	53982	8,4
13	0,665	23,41	13851	41,5
14	0,674	32,02	7260	0,5
15	0,665	42,02	13053	2,5
16	0,664	32,45	3586,0	0,1
17	0,668	33,71	6510,0	0,7
18	0,660	23,79	10307,0	15,5
19	0,670	32,96	6239,0	19,2
20	0,663	30,93	8718,0	1,5

Задача 4.16. По данным таблицы 4.14 определите параметры следующих уравнений множественной регрессии:

а) линейной $y_x = a_0 + a_1x_1 + a_2x_2 + a_3x_3$;

б) логарифмической $y_x = a_0 + a_1 \lg x_1 + a_2 \lg x_2 + a_3 \lg x_3$,

рассчитайте теоретические уровни результативного показателя.

Таблица 4.14

Данные к задаче в разрезе районов края за 2012 г.

№ района	Объем сельскохозяйственной продукции СХО, млн руб.	Среднегодовая стоимость основных фондов СХО, млн руб.	Численность работающих в сельскохозяйственном производстве, тыс. чел.	Среднее значение температуры воздуха, °С
	Y	X_1	X_2	X_3
1	**2**	**3**	**4**	**5**
1	191,9	646,3	2,0	10,4
2	219,59	718,8	0,5	10,5
3	129,6	1177,6	4,7	11,2

Окончание табл. 4.14

1	2	3	4	5
4	242,3	776,1	2,5	11,3
5	365,7	1216,8	2,2	11
6	576,8	864,2	3,3	11,2
7	880	580,4	3,4	11
8	501,8	549,8	1,2	10,2
9	467,8	10761,8	2,7	10,2
10	281,4	1250,0	4,3	11,4
11	587,4	1579,3	1,6	11
12	190,2	2052,6	4,5	10,5
13	173,0	944,4	3,2	11,2
14	256,64	894,0	1,9	11,7
15	232,3	886,7	2,5	11,3
16	395,7	13770,5	0,9	11,2
17	268,5	9624,9	1,8	11,7
18	302,1	3589,9	4,2	11,2

ГЛАВА

5 ИДЕНТИФИКАЦИЯ ПАРНЫХ ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ

5.1. Статистическая корректность эконометрической модели

Первоначально построенное уравнение (или их система) крайне редко удовлетворяет эконометриста по тем или иным характеристикам. Поэтому важнейшей задачей является оценка результатов моделирования. В эконометрике принята устоявшаяся схема проверки результатов анализа.

Проверка статистической корректности построенной эконометрической модели проверяется при помощи следующих характеристик:

1) стандартная ошибка уравнения регрессии;
2) общее качество уравнения регрессии;
3) стандартная ошибка параметров уравнения;
4) выполнимость предпосылок МНК:
   4.1) оценка автокорреляции остатка;
   4.2) оценка мультиколлинеарности переменных;
5) корректность модели в целом.

Результаты модели вместе с ее экономической интерпретацией служат основой усовершенствования эконометрической модели.

Расчет стандартной ошибки уравнения основан на разложении общей суммы квадратов отклонений результативного признака y от среднего значения $\bar{y}$ на две составляющие — факторную сумму квадратов (объясненную) и остаточную (необъясненную):

$$\sum(y-\bar{y})^2=\sum(y-y_x)^2+\sum(y_x-\bar{y})^2, \qquad (5.1)$$

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

$\sum(y-y_x)^2$ — остаточная сумма квадратов отклонений;

$\sum(y_x-\bar{y})^2$ — факторная сумма квадратов отклонений.

Отклонение общей суммы квадратов индивидуальных значений результативного признака y от среднего значения $\bar{y}$ вызвано двумя группами причин: влиянием изучаемого фактора x и прочих не рассматриваемых факторов.

Если фактор не оказывает влияния на результат, то линия регрессии на графике параллельна оси Ox и $\bar{y} = y$. В этом случае дисперсия результативного признака обусловлена воздействием прочих факторов, а общая сумма квадратов отклонений совпадает с остаточной. В случае если прочие факторы не влияют на результат, то результативный признак y функционально связан с изучаемым фактором x, а остаточная сумма квадратов отклонений равна нулю.

Так как не все точки поля корреляции лежат на линии регрессии, то имеет место разброс, обусловленный влиянием фактора x и вызванный действием прочих причин (необъясненная вариация). Если факторная сумма квадратов отклонений будет больше остаточной, то фактор x оказывает существенное влияние на изучаемый признак y.

Сумма квадратов отклонений связана с *числом степеней свободы* (числом свободы независимого варьирования признака). Число степеней свободы должно показывать, сколько независимых отклонений из n возможных требуется для образования данной суммы квадратов.

Для общей суммы квадратов $\sum(y-\bar{y})^2$ требуется $(n-1)$ независимых отклонений, так как по совокупности из n единиц после расчета среднего уровня свободно варьируют лишь $(n-1)$ число отклонений. Если $\sum(y-\bar{y})=0$, то применительно ко множественной регрессии число степеней свободы должно быть уменьшено на число независимых переменных m (для простой регрессии $m=1$).

Таким образом, стандартная (средняя квадратическая) ошибка с поправкой на число степеней свободы рассчитывается по формуле

$$S_y = \sqrt{\frac{\sum(y-\bar{y})^2}{n-1}}. \qquad (5.2)$$

Стандартная ошибка уравнения с поправкой на число степеней свободы характеризует абсолютную величину разброса случайной составляющей регрессионного уравнения. Факторная сумма квадратов для расчетов использует теоретические значения результативного признака y_x, найденные по уравнению регрессии.

5.2. Идентификация парной линейной регрессионной модели

Построенное уравнение регрессии должно дополняться расчетом показателя тесноты связи. При использовании линейной регрессии в качестве такого показателя выступает *линейный коэффициент корреляции* (r_{xy}):

$$r_{xy} = a_1 \cdot \frac{S_x}{S_y} = \frac{\operatorname{cov}(x,y)}{S_x \cdot S_y} = \frac{\overline{yx} - \overline{y} \cdot \overline{x}}{S_x \cdot S_y}. \qquad (5.3)$$

Графическая интерпретация коэффициента корреляции подтверждает, что зависимость между переменными тем теснее, чем точки корреляционного поля находятся ближе к линии регрессии (рис. 5.1).

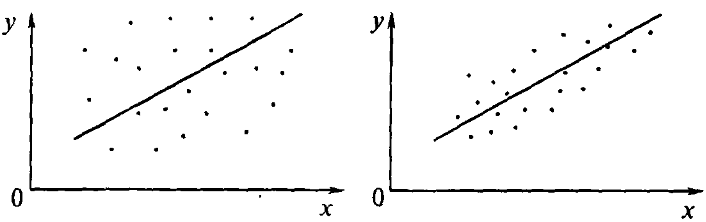


Рис. 5.1. Графическая характеристика тесноты связи между линией регрессии и полем корреляции

Корреляционная связь между переменными называется прямой, если $r_{xy} > 0$, и обратной, если $r_{xy} < 0$. При обратной связи увеличение одной переменной ведет к уменьшению результативного признака.

Для практических расчетов наиболее удобна формула

$$r_{xy} = \frac{n \cdot \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i}{\sqrt{(n \cdot \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2) \cdot (n \cdot \sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2)}}, \qquad (5.4)$$

так как по ней коэффициент корреляции для линейной зависимости находится из данных наблюдений, а на значение r_{xy} не оказывает влияния погрешность округления.

Коэффициент корреляции обладает следующими свойствами:

1. Коэффициент корреляции принимает значения от -1 до 1, т. е. $-1 \le r_{xy} \le 1$. Чем ближе по абсолютной величине коэффициент корреляции к единице, тем связь теснее. Для характеристики силы связи используют *шкалу Чеддока*:

Показатель тесноты связи	0,1–0,3	0,3–0,5	0,5–0,7	0,7–0,9	0,9–0,99
Характеристика силы связи	слабая	умеренная	заметная	высокая	весьма высокая

2. При значении коэффициента корреляции, равном ± 1, связь представлена функциональной зависимостью. При этом все наблюдаемые значения располагаются на линии регрессии (рис. 5.2).

3. При $r_{xy} = 0$ корреляционная связь между признаками отсутствует. При этом, как отмечалось выше, линия регрессии параллельна оси Ox (рис. 5.3).

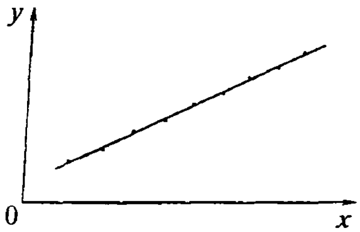


Рис. 5.2. Функциональная линейная зависимость

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции r_{xy}^2, который называется коэффициентом детерминации. Он характеризует долю вариации результативного

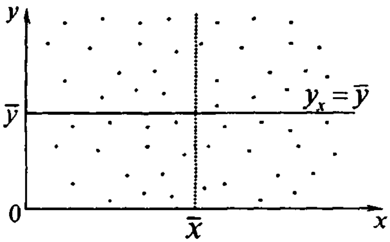


Рис. 5.3. Корреляционная связь отсутствует

признака y, объясняемую регрессией, в общей дисперсии результативного признака:

$$r^2_{xy} = \frac{S^2_{y_x}}{S_y}. \tag{5.5}$$

Величина $1 - r^2$ характеризует долю дисперсии y, вызванную влиянием остальных не учтенных в модели факторов.

Обратившись к формуле 5.3 линейного коэффициента корреляции, получим коэффициент детерминации, равный:

$$r^2_{xy} = (a_1 \cdot \frac{S_x}{S_y})^2 = a_1^2 \cdot \frac{S_x^2}{S_y^2}, \tag{5.6}$$

следовательно,

$$S^2_{y_x} = a_1^2 \cdot S_x^2. \tag{5.7}$$

Соответственно, сумма квадратов отклонений, обусловленных линейной регрессией, составит:

$$\sum (y_x - \bar{y})^2 = a_2^1 \cdot \sum (x - \bar{x})^2. \tag{5.8}$$

Поскольку при заданном объеме наблюдений по x и y факторная сумма квадратов при линейной регрессии зависит только от одной константы — коэффициента регрессии a_1, то данная сумма квадратов имеет одну степень свободы. Соответственно, факторная сумма квадратов отклонений имеет число степеней свободы, равное 1.

Существует равенство между числом степеней свободы общей, факторной и остаточной суммами квадратов. Число степеней свободы остаточной суммы квадратов при линейной регрессии составляет $n-2$. Число степеней свободы для общей суммы квадратов определяется числом единиц наблюдений, и поскольку используется средняя величина по данным выборки, то теряется одна степень свободы $(n-1)$. В результате имеем равенства:

$$\sum(y-\bar{y})^2=\sum(y_x-\bar{y})^2+\sum(y-y_x)^2,$$
$$(n-1)=1+(n-2).$$

Разделив каждую сумму квадратов на соответствующее число степеней свободы, получим дисперсии на одну степень свободы:

- общая дисперсия $S_y^2=\dfrac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}$; (5.9)
- факторная дисперсия $S_{y_x}^2=\dfrac{\sum_{i=1}^{n}(y_{x_i}-\bar{y})^2}{1}$; (5.10)
- остаточная дисперсия $S_\Delta^2=\dfrac{\sum_{i=1}^{n}(y_i-y_{x_i})^2}{n-2}$. (5.11)

В общем виде схему дисперсионного анализа для всех видов зависимостей можно представить следующим образом:

Таблица 5.1

Схема дисперсионного анализа

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на одну степень свободы
Факторная	$\sum_{i=1}^{n}(y_{x_i}-\bar{y})^2$	m	$S_{y_x}^2=\dfrac{\sum_{i=1}^{n}(y_{x_i}-\bar{y})^2}{m}$

Окончание табл. 5.1

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на одну степень свободы
Остаточная	$\sum_{i=1}^{n}(y_i - y_{x_i})^2$	$n-(m+1)$	$S_{\Delta}^2 = \frac{\sum_{i=1}^{n}(y_i - y_{x_i})^2}{n-(m+1)}$
Общая	$\sum_{i=1}^{n}(y_i - \bar{y})^2$	$(n-1)$	$S_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$

В таблице 5.1 m — число оцениваемых параметров уравнения регрессии при независимых переменных; n — число наблюдений.

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получаем величину F-критерия Фишера (F-отношения):

$$F = \frac{S_{y_x}^2}{S_{\Delta}^2}, \tag{5.12}$$

где F — критерий для проверки нулевой гипотезы H_0:

$$S_{y_x}^2 = S_{\Delta}^2.$$

Если нулевая гипотеза справедлива, то факторная и остаточная дисперсии не отличаются друг от друга. Для H_0 необходимо опровержение, чтобы факторная дисперсия превышала остаточную в несколько раз, т. е. выполнялась бы гипотеза H_1:

$$S_{y_x}^2 > S_{\Delta}^2.$$

Величина F-критерия тесно связана с коэффициентом детерминации. Факторную сумму квадратов отклонений можно представить как

$$\sum(y_x - \bar{y})^2 = r^2 \cdot S_y^2 \cdot n, \tag{5.13}$$

а остаточную сумму квадратов — как:

$$\sum (y-y_x)^2 = (1-r^2)\cdot S_y^2 \cdot n. \qquad (5.14)$$

Тогда значение F-критерия Фишера для парной линейной регрессии можно записать:

$$F = \frac{r^2}{1-r^2}\cdot (n-2). \qquad (5.15)$$

Английский статистик Снедекор разработал таблицы критических значений F-критерия для различных уровней существенности нулевой гипотезы и различных степеней свободы. Табличное значение F-критерия — это максимальная величина отношения дисперсий, которая имеет место при случайном их расхождении для данного уровня вероятности наличия нулевой гипотезы.

Вычисленное значение F-критерия признается достоверным (отличающимся от единицы), если оно больше табличного. В этом случае нулевая гипотеза (H_0) отвергается и делается вывод о существенности изучаемой связи: $F_{факт} > F_{табл}$.

Если величина $F_{факт}$ оказывается меньше $F_{табл}$, то вероятность нулевой гипотезы выше заданного уровня ($p = 0{,}05$ или $p = 0{,}01$) и она не может быть отклонена без риска неверного вывода о наличии связи. В этом случае уравнение регрессии признается статистически незначимым.

В линейной регрессии обычно оценивается значимость не только уравнения в целом, но и отдельных его параметров. Для этих целей по каждому параметру определяется его стандартная ошибка: m_{a_0} и m_{a_1}.

Стандартная ошибка коэффициента регрессии определяется по формуле

$$m_{a_1} = \sqrt{\frac{\sum (y-y_x)^2/(n-2)}{\sum (x-\bar{x})^2}}. \qquad (5.16)$$

Стандартная ошибка параметра a_0 определяется по формуле

$$m_{a_0} = \sqrt{\frac{\sum (y-y_x)^2}{(n-2)}\cdot \frac{\sum x^2}{n\cdot \sum (x-\bar{x})^2}}. \qquad (5.17)$$

Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т. е. определяется фактическое значение t-критерия Стьюдента:

$$t_{a_1}=\frac{a_1}{m_{a_1}} \quad \text{или} \quad t_{a_1}=a_1\cdot\frac{\sqrt{n-2}\cdot S_x}{S_\Delta}, \tag{5.18}$$

которое затем сравнивается с табличным (критическим) при определенном уровне значимости и числе степеней свободы $(n-2)$. Если фактическое значение t-критерия Стьюдента превышает табличное ($t_{\text{крит}}$), то гипотезу о несущественности коэффициента регрессии можно отклонить.

Процедура оценивания существенности параметра a_0 не отличается от рассмотренной выше для коэффициента регрессии(a_1), t-критерий Стьюдента равен:

$$t_{a_0}=\frac{a_0}{m_{a_0}} \quad \text{или} \quad t_{a_0}=a_0\cdot\frac{\sqrt{n-2}}{S_\Delta}. \tag{5.19}$$

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

$$m_r=\sqrt{\frac{1-r^2}{n-2}}. \tag{5.20}$$

Фактическое значение t-критерия Стьюдента определяется как

$$t_r=\frac{r}{m_r} \quad \text{или} \quad t_r=r\cdot\sqrt{\frac{n-2}{1-r^2}}. \tag{5.21}$$

Данная формула показывает, что в парной линейной регрессии $t_r^2=F$, следовательно, оба способа проверки значимости модели для линейной парной регрессии равносильны. Кроме того, $t_r^2=F$, так как

$$t_{a_1}^2=\frac{a_1^2}{m_{a_1}^2}=\frac{a_1^2}{\dfrac{\sum(y-y_x)^2/(n-2)}{\sum(x-\bar{x})^2}}=\frac{a_1^2\cdot\sum(x-\bar{x})^2}{\sum(y-y_x)^2/(n-2)},$$

из формулы 5.8

$$a_1^2 = \frac{\sum (y_x - \bar{y})^2}{\sum (x - \bar{y})^2},$$

тогда

$$t_{a_1}^2 = \frac{a_1^2 \cdot \sum (x - \bar{x})^2}{\sum (y - y_x)^2 / (n-2)} = \frac{\sum (y_x - \bar{y})^2}{\sum (y - y_x)^2 / (n-2)} = \frac{S_{y_x}^2}{S_\Delta^2} = F.$$

Следовательно, $t_{a_1}^2 = t_r^2$.

В случае если величина коэффициента корреляции близка к ± 1, то распределение его оценок отличается от нормального или распределения Стьюдента, так как величина коэффициента корреляции ограничена $[-1;+1]$. Для решения этой проблемы Фишером было предложено ввести вспомогательную величину Z для оценки существенности коэффициента корреляции (r), связанную со следующим отношением:

$$Z = \frac{1}{2} \cdot \ln \frac{1 - r^2}{n - 2}. \quad (5.22)$$

При изменении коэффициента корреляции от -1 до $+1$ величина Z изменяется от $-\infty$ до $+\infty$, что соответствует нормальному распределению. Математический анализ доказывает, что распределение величины Z мало отличается от нормального даже при близких к единице значениях коэффициента корреляции.

Стандартная ошибка величины Z определяется по формуле

$$m_Z = \frac{1}{\sqrt{n-3}}. \quad (5.23)$$

Величину Z можно не рассчитывать, а воспользоваться готовыми таблицами Z-преобразований, в которых приведены значения величины Z для соответствующих значений коэффициента корреляции r.

Для проверки нулевой гипотезы H_0 необходимо сравнить фактическое значение $t_Z = Z / m_Z$ с табличным значением (при уровне значимости $p = 0{,}05$ или $p = 0{,}01$), если $t_Z > t_{крит}$, то коэффициент корреляции значимо отличается от нуля.

5.3. Статистическое изучение парной нелинейной регрессионной эконометрической модели

Применительно к нелинейной парной зависимости уравнение регрессии, так же как и в линейной зависимости, дополняется показателем корреляции — *индексом корреляции* (R):

$$R = \sqrt{1 - \frac{S_{\Delta}^2}{S_y^2}}, \tag{5.24}$$

где S_{Δ}^2 — остаточная дисперсия, определяемая исходя из уравнения регрессии;

S_y^2 — общая дисперсия результативного признака.

Так как $S_y^2 = \dfrac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}$, а $S_{\Delta}^2 = \dfrac{\sum_{i=1}^{n}(y_i - y_{x_i})^2}{n}$,

то индекс корреляции можно выразить как

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - y_{x_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}. \tag{5.25}$$

Следует отметить, что сила связи между признаками оценивается исходя из характеристики по шкале Чеддока, а направление согласуется со знаком коэффициента регрессии.

Если нелинейное относительно объясняемой переменной уравнение регрессии при линеаризации принимает форму линейного уравнения парной регрессии, то для оценки тесноты связи может быть использован линейный коэффициент корреляции, величина которого в этом случае совпадает с индексом корреляции:

$$R_{xy} = r_{yz}, \tag{5.26}$$

где z — преобразованная величина факторного признака, например: $z = 1/x$; $z = \lg x$.

Так, в уравнении равносторонней гиперболы вида

$$y_x = a_0 + a_1 \frac{1}{x},$$

заменив $\frac{1}{x} = z$, имеем $y_x = a_0 + a_1 \cdot z$, для которого коэффициент корреляции равен:

$$r_{yz} = a_1 \cdot \frac{S_z}{S_y}.$$

Возведем данное выражение в квадрат, получим:

$$r_{yz}^2 = a_1^2 \cdot \frac{S_z^2}{S_y^2} = a_1^2 \frac{\sum_{i=1}^{n}(z_i - \bar{z})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

Так как $a_1^2 \cdot \sum (z - \bar{z})^2 = \sum (y_z - \bar{y})^2$, то

$$r_{yz}^2 = \frac{\sum_{i=1}^{n}(y_z - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

Если (как было показано выше)

$$\sum (y - \bar{y})^2 = \sum (y_z - \bar{y})^2 + \sum (y - y_z)^2,$$

$$\sum (y_z - \bar{y})^2 = \sum (y - \bar{y})^2 - \sum (y - y_z)^2,$$

то $r_{yz}^2 = \dfrac{\sum (y - \bar{y})^2 - \sum (y - y_z)^2}{\sum (y - \bar{y})^2}$,

т. е. приходим к формуле индекса корреляции:

$$r_{yz} = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - y_{z_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Далее, заменив z на $\frac{1}{x}$, получим $y_z = y_x$, соответственно $r_{yz} = R_{yx}$.

Иначе обстоит дело, когда преобразования уравнения в линейную форму связаны с зависимой переменной. В этом случае линейный коэффициент корреляции по преобразованным значениям признаков дает лишь приближенную оценку тесноты связи и численно не совпадает с индексом корреляции. Так, для степенной функции $y_x = a_0 \cdot x^{a_1}$ после перехода к линейному уравнению $\lg y = \lg a_0 + a_1 \cdot \lg x$ линейный коэффициент корреляции находится не для фактических значений x и y, а для их логарифмов $\lg x$ и $\lg y$, т. е. $r_{\lg y \lg x}$.

Оценка существенности индекса корреляции проводится так же, как и оценка надежности коэффициента корреляции.

Индекс детерминации используется для проверки существенности в целом уравнения нелинейной регрессии по F-критерию Фишера:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-(m+1)}{m}, \tag{5.27}$$

где R^2 — индекс детерминации;

n — число наблюдений;

m — число параметров при переменных x.

Величина m характеризует число степеней свободы для факторной суммы квадратов, а $n-(m+1)$ — число степеней свободы для остаточной суммы квадратов.

Для степенной функции $y_x = a_0 \cdot x^{a_1}$ $m=1$ и формула F-критерия принимает вид, как и при линейной парной зависимости:

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

Для параболы второй степени $y_x = a_0 + a_1 x + a_2 x^2$ $m=2$ и F-критерий рассчитывается по формуле

$$F = \frac{R^2}{1-R^2} \cdot \frac{(n-3)}{2}.$$

В практических расчетах индекс детерминации R^2_{yx} можно сравнить с коэффициентом детерминации r^2_{yx} для обоснования возможности применения линейной функции. Чем больше кривизна линии регрессии, тем величина коэффициента детерминации r^2_{yx} меньше индекса детерминации R^2_{yx}. Близость этих показателей означает, что нет необходимости усложнять форму уравнения регрессии и можно использовать линейную функцию. Практически если величина

$$\Delta_d = (R^2_{yx} - r^2_{yx}) \le 0{,}1,$$

то предположение о линейной форме связи значимо. В противоположном случае необходимо оценить существенность различия через t-критерий Стьюдента:

$$t_{\Delta_d} = \frac{R^2_{yx} - r^2_{yx}}{m_{|R-r|}}, \qquad (5.28)$$

где $m_{|R-r|}$ — ошибка разности между R^2_{yx} и r^2_{yx}, рассчитываемая по формуле

$$m_{|R-r|} = 2 \cdot \sqrt{\frac{(R^2 - r^2) - (R^2 - r^2)^2 \cdot (2 - (R^2 + r^2))}{n}}. \qquad (5.29)$$

Если $t_{\text{факт}} > t_{\text{табл}}$, то различия между рассматриваемыми показателями корреляции существенны и замена нелинейной регрессии уравнением линейной функции невозможна. На практике если величина $t_{\text{факт}} < 2$, то различия между R^2_{yx} и r^2_{yx} несущественны и, следовательно, возможно применение линейной регрессии.

Анализируя значения коэффициентов регрессии на практике, затруднительно сделать вывод о мере влияния их на исследуемый показатель, поскольку факторы измерены в различных

Таблица 5.2

Коэффициенты эластичности математических функций

Вид функции, y_x	**Первая производная,** y'_x	**Коэффициент эластичности** $Э = f'(x) \cdot \frac{x}{y}$
Парабола второго порядка $y_x = a_0 + a_1x + a_2o^2$	$a_1 + 2a_2x$	$Э = \frac{(a_1 + 2a_2x) \cdot x}{a_0 + a_1x + a_2x^2}$
Равносторонняя гипербола $y_x = a_0 + \frac{a_1}{x}$	$-\frac{a_1}{x^2}$	$Э = \frac{-a_1}{a_0 \cdot x + a_1}$
Показательная $y_x = a_0 \cdot a_1^x$	$\lg a_1 \cdot a_0 \cdot a_1^x$	$Э = x \cdot \lg a_1$
Логарифмическая $y_x = a_0 + a_1 \cdot \lg x$	$\frac{a_1}{x}$	$Э = \frac{a_1}{a_0 + a_1 \cdot \lg x}$
Обратная $y_x = \frac{1}{a_0 + a_1x}$	$\frac{-a_1}{(a_0 + a_1x)^2}$	$Э = \frac{-a_1 \cdot x}{a_0 + a_1x}$

единицах. Различия в единицах измерения факторов устраняются с помощью коэффициентов эластичности. Как отмечалось выше, в степенных функциях параметр уравнения a_1 является коэффициентом эластичности. О правомерности такого толкования параметра a_1 для степенной функции вида $y_x = a_0 \cdot x^{a_1}$ можно судить, рассмотрев формулу расчета коэффициента эластичности:

$$Э = f'(x) \cdot \frac{x}{y}, \tag{5.30}$$

где $f'(x)$ — первая производная, характеризующая соотношение прироста результата и фактора для соответствующей формы связи.

Для степенной функции $f'(x)=a_0 \cdot a_1 \cdot x^{a_1-1}$. Соответственно, коэффициент эластичности окажется равным:

$$Э = a_0 \cdot a_1 \cdot x^{a_1-1} \cdot \frac{x}{a_0 \cdot x^{a_1}} = \frac{a_0 \cdot a_1 \cdot x^{a_1}}{a_0 \cdot x^{a_1}} = a_1.$$

Коэффициент эластичности определяется и при наличии других форм связи. Так, для линейной регрессии он обычно рассчитывается по формуле

$$Э = \frac{a_1 \cdot x}{a_0 + a_1 \cdot x} = a_1 \cdot \frac{x}{y}.$$

Формулы расчета коэффициентов эластичности для наиболее распространенных типов уравнений регрессии приведены в таблице 5.2.

5.4. Оценка адекватности модели

Величина отклонений эмпирических и расчетных значений $(y - y_x)$ по каждому наблюдению представляет собой *ошибку аппроксимации.* Их число соответствует объему совокупности. В отдельных случаях ошибка аппроксимации может оказаться равной нулю. Отклонения $(y - y_x)$ несравнимы между собой, исключая величину, равную нулю. Для сравнения используют величины отклонений, выраженные в процентах к фактическим значениям. Так как отклонение $(y - y_x)$ может быть положительным и отрицательным, то ошибку аппроксимации принято определять в процентах по модулю.

Отклонения $(y - y_x)$ можно рассматривать как абсолютную ошибку аппроксимации, а

$$\left|\frac{y - y_x}{y}\right| \cdot 100$$

как относительную ошибку аппроксимации. Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют *среднюю ошибку аппроксимации* как среднюю арифметическую простую:

$$\bar{A} = \frac{1}{n} \cdot \sum \left| \frac{y - y_x}{y} \right| \cdot 100. \qquad (5.31)$$

Возможно и другое определение средней ошибки аппроксимации как

$$\bar{A} = \frac{100}{\bar{y}} \cdot \sqrt{\frac{\sum (y - y_x)^2}{n}}, \qquad (5.32)$$

однако формула 5.31 используется чаще.

5.5. Верификация регрессионных моделей

При подстановке в уравнение регрессии $y_x = a_0 + a_1 x$ соответствующего значения x можно определить предсказываемое (прогнозируемое) значение y_p как вариант точечного прогноза y_k при $x_p = x_k$. Однако точечный прогноз явно не будет отвечать реальным событиям. Поэтому его необходимо дополнить расчетом стандартной ошибки m_{y_x} и соответственно интервальной оценкой прогнозного значения:

$$y_x - m_{y_x} \le y_p \le y_x + m_{y_x}.$$

Для построения формулы, позволяющей определить величину стандартной ошибки m_{y_x}, необходимо обратиться к уравнению регрессии:

$$y_x = a_0 + a_1 x.$$

Подставим в это уравнение выражение параметра a_0:

$$a_0 = \bar{y} - a_1 \bar{x},$$

тогда уравнение регрессии примет вид

$$y_x = \bar{y} - a_1 \bar{x} + a_1 x = \bar{y} + a_1 \cdot (x - \bar{x}).$$

Отсюда следует, что стандартная ошибка m_{y_x} зависит от ошибки $\bar{y}$ и ошибки коэффициента регрессии a_1:

$$m_{yx}^2 = m_{\bar{y}}^2 + m_{a_1}^2 \cdot (x - \bar{x})^2.$$

Из теории выборочного исследования статистических данных известно:

$$m_{\bar{y}}^2 = \frac{S^2}{n},$$

однако, по нашему мнению, в качестве меры оценки следует использовать остаточную дисперсию на одну степень свободы S^2. Соответственно, получим формулу для расчета среднего значения переменной y:

$$m_{\bar{y}}^2 = \frac{S^2}{n}.$$

Ошибка коэффициента регрессии a_1 определяется по формуле

$$m_{a_1}^2 = \frac{S^2}{\sum(x-\bar{x})^2}.$$

Считая, что предсказываемое значение фактора $x_p = x_k$, получим формулу расчета стандартной ошибки прогнозируемого значения результативного признака по уравнению регрессии:

$$m_{y_x}^2 = \frac{S^2}{n} + \frac{S^2}{\sum(x-\bar{x})^2} \cdot (x-\bar{x})^2 = S^2 \cdot \left[\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x-\bar{x})^2}\right]. \quad (5.33)$$

Рассмотренная формула стандартной ошибки предсказываемого значения при заданном $x_p = x_k$ характеризует ошибку линии регрессии. Величина стандартной ошибки m_{y_x} достигает максимума при $x_p = x_k = \bar{x}$, возрастает по мере удаления x_p от $\bar{x}$ в любом направлении, т. е. чем больше разность x_p от $\bar{x}$, тем больше m_{y_x}. Наилучших результатов прогноза следует ожидать, если факторный признак x_p находится в центре области наблюдений x, и нельзя ожидать хороших результатов прогноза при удалении x_p от $\bar{x}$. Если значение x_p оказывается за пределами наблюдаемых значений факторных признаков x, используемых при построении модели, то результаты прогноза ухудшаются по мере отклонения x_p от области наблюдаемых значений фактора x.

Графически доверительные границы для y_x представляют собой гиперболы, расположенные по обе стороны от линии регрессии (рис. 5.4).

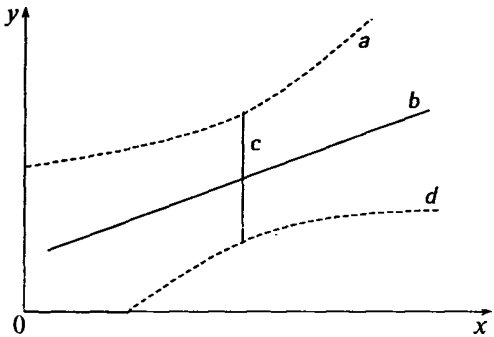


Рис. 5.4. Доверительный интервал линии регрессии: a — верхняя доверительная граница y_p; b — линия регрессии $y_x = a_0 + a_1x$; c — доверительный интервал для y_p при заданных x_p; d — нижняя граница для y_p.

Фактические значения y варьируют около среднего значения y_x. Индивидуальные значения y отличаются от y_x на величину случайной ошибки, дисперсия которой оценивается как остаточная дисперсия на одну степень свободы S^2. Поэтому ошибка индивидуального предсказываемого значения y_p должна включать не только стандартную ошибку m_{y_x}, но и случайную ошибку.

Средняя ошибка прогнозируемого индивидуального значения у при $x_p = x_k$ составит:

$$m_{y_i(x_k)} = S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x - \bar{x})^2}}. \tag{5.34}$$

Тогда с надежностью $p = 0{,}05$ или $p = 0{,}01$ доверительный интервал для y_p при заданном $x_p = x_k$ составит:

$$y_p = y_x \pm t_k \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x - \bar{x})^2}}, \tag{5.35}$$

где t_k — критическое (табличное) значение t-критерия Стьюдента для соответствующего уровня значимости $\alpha = 1 - p$ и числа степеней свободы $(n - 2)$.

При прогнозировании на основе уравнения регрессии необходимо помнить, что величина прогноза зависит не только от величины стандартной ошибки индивидуального значения y_p, но и от точности прогноза значения фактора x. Его величина может быть задана на основе анализа других моделей исходя из конкретной экономической ситуации и анализа динамики данного фактора.

Рассмотренная формула стандартной ошибки индивидуального значения признака $m_{y_i(x_k)}$ может использоваться для оценки существенности различия предсказываемого значения исходя из регрессионной модели и выдвинутой гипотезы развития событий. Для этой цели используется односторонний t-критерий Стьюдента:

$$t = \frac{y_x - y_p}{m_{y_i(x_k)}}. \tag{5.36}$$

Полученное фактическое значение одностороннего t-критерия Стьюдента сравнивается с табличным, и если $t_{\text{факт}} > t_{\text{табл}}$, то предполагаемое значение значимо (существенно) отличается от прогнозируемого по модели, и наоборот.

Контрольные вопросы

1. Характеристики статистической корректности эконометрических моделей.
2. Стандартная ошибка уравнения регрессии.
3. Оценка существенности коэффициентов регрессии.
4. Расчет коэффициентов корреляции для линейного уравнения парной связи.
5. Оценка коэффициентов детерминации для линейного уравнения парной связи.
6. Дисперсионный анализ.
7. t-критерий Стьюдента для оценки значимости коэффициента корреляции.
8. Оценка значимости модели по F-критерию Фишера.
9. Прогнозирование по модели парной линейной регрессии.
10. Оценка адекватности линейной парной регрессии.

☑ ТЕСТОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ ПРОВЕРКИ ЗНАНИЙ

1. Суть коэффициента детерминации R^2 состоит в следующем:
а) коэффициент свидетельствует о значимости коэффициентов регрессии;
б) коэффициент определяет тесноту связи между признаками;
в) коэффициент определяет долю общего разброса значений y, объясненного уравнением регрессии;
г) коэффициент свидетельствует о наличии/отсутствии автокорреляции.

2. Уравнение парной регрессии дополняется коэффициентом парной корреляции, потому что:
а) это требуется для получения оценок коэффициентов регрессии;
б) это необходимо для расчета величины остаточной дисперсии;
в) необходимо знать тесноту связи в линейной форме.

3. Коэффициент детерминации характеризует:
а) соотношение факторной и остаточной дисперсий;
б) долю остаточной дисперсии в общей дисперсии результативного признака;
в) долю факторной дисперсии в общей дисперсии результативного признака.

4. F-критерий характеризует:
а) долю факторной дисперсии в общей дисперсии результативного признака;
б) долю остаточной дисперсии в общей дисперсии результативного признака;
в) соотношение факторной и остаточной дисперсий.

5. Оценка значимости уравнения регрессии в целом дается с помощью:

а) коэффициента детерминации;
б) стандартной ошибки регрессии;
в) F-критерия Фишера.

6. Для оценки значимости коэффициентов регрессии рассчитывают:

а) F-критерий Фишера;
б) коэффициент детерминации;
в) t-статистику Стьюдента.

7. Коэффициент регрессии a_1 пропорционален:

а) стандартному отклонению x;
б) стандартному отклонению y;
в) коэффициенту корреляции.

8. Уравнение регрессии потребления материалов (y) от объема производства (x), построенное по 15 наблюдениям, имеет вид

$$y = 5 + 5x + \varepsilon.$$
$$(4,\ 0)$$

В скобках — фактическое значение t-критерия. Коэффициент детерминации для этого уравнения равен:

а) 0,575;
б) 0,439;
в) 0,552;
г) 0,648.

9. По совокупности 15 предприятий торговли изучается зависимость между ценой (x) на товар A и прибылью (y) торгового предприятия. При оценке регрессионной модели были получены следующие результаты:

$$\sum (y - y_x)^2 = 32000;$$

$$\sum (y - \bar{y})^2 = 40000.$$

Индекс корреляции, фактическое значение F-критерия следующие:

а) $R = 0{,}64$, $F_{\text{факт}} = 6{,}15$;

б) $R = 0{,}830$, $F_{\text{факт}} = 2{,}78$;

в) $R = 0{,}447$, $F_{\text{факт}} = 3{,}25$;

г) $R = 0{,}8$, $F_{\text{факт}} = 5{,}12$.

10. Статистическая надежность оценки коэффициентов регрессии увеличивается:

а) с уменьшением числа степеней свободы;

б) не зависит от числа степеней свободы;

в) с увеличением числа степеней свободы.

11. Добавление новой объясняющей переменной:

а) иногда уменьшает значение коэффициента детерминации;

б) не оказывает влияния на значение коэффициента детерминации;

в) никогда не уменьшает значение коэффициента детерминации.

12. Проверка статистической корректности уравнения регрессии включает:

а) проверку статистической значимости коэффициентов уравнения и выполнимости предпосылок МНК;

б) вычисление доверительных интервалов зависимой переменной и проверку общего качества уравнения;

в) проверку статистической значимости коэффициентов уравнения, общего качества уравнения, выполнимости предпосылок МНК.

13. Если коэффициент детерминации равен нулю, то:

а) величина зависимой переменной Y линейно зависит от независимых переменных X_i;

б) нельзя сделать вывод о линейной зависимости Y от независимых переменных X_i;

в) величина зависимой переменной Y линейно не зависит от независимых переменных X_i.

14. Если коэффициент детерминации равен нулю, то F-критерий Фишера:
а) равен единице;
б) больше или равен единице;
в) равен нулю.

15. Уравнение регрессии является качественным, если:
а) t-статистики, F-статистика, DW-статистика высокие;
б) коэффициент детерминации больше 0,8;
в) t-статистики, F-статистика больше критических значений, предпосылки МНК соблюдены.

ПРАКТИКУМ

Задача 5.1. По данным задачи 4.1 осуществить идентификацию представленных уравнений парной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод о наиболее приемлемой форме модели для описания исходной криволинейной зависимости изменения стоимости продукции сельского хозяйства (Y) и стоимости основных фондов сельхозтоваропроизводителей (X).

Решение

а) *линейная регрессия* $y_x = a_0 + a_1 x$.

Для оценки практической значимости синтезированной модели парной линейной регрессии вида

$$y_x = 1289{,}8 + 1{,}02 \cdot x$$

обратимся к результатам регрессионного анализа, проведенного в процессе решения задачи 4.1 (рис. 5.5).

Проанализируем полученные результаты.

В соответствии с величиной линейного коэффициента корреляции $r_{xy} = 0{,}72$, отраженного в строке «Множественный R»,

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,715547825
R-квадрат	0,51200869
Нормированный R	0,491675718
Стандартная ошиб	1098,433749
Наблюдения	26

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	30382549,95	30382549,95	25,18120362	3,97064E-05
Остаток	24	28957360,81	1206556,7		
Итого	25	59339910,76			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	1289,794303	372,6926199	3,460745489	0,002029997	520,5945408	2058,994065
Переменная X 1	1,020572725	0,203378816	5,018087645	3,97064E-05	0,600819479	1,440325971

Рис. 5.5. Результаты регрессионного анализа парной линейной регрессии, проведенного средствами *Excel*

связь между переменными модели можно охарактеризовать, используя шкалу Чеддока, как высокую, прямую.

В строке «R-квадрат» указано значение коэффициента детерминации $r_{xy}^2 = 0{,}512$. Таким образом, вариация стоимости продукции сельского хозяйства (Y) по синтезированной модели парной линейной регрессии на 51,2% объясняется изменением стоимости основных фондов сельхозтоваропроизводителей (X).

Для оценки значимости коэффициента корреляции и модели в целом рассчитаем фактическое значение t-критерия Стьюдента по формуле (5.21):

$$t_r = 0{,}72 \cdot \sqrt{\frac{26-2}{1-0{,}512}} = 5{,}05.$$

Из специальных таблиц критических значений t-критерия Стьюдента найдем $t_{табл}$ для уровня значимости 0,05 и числа степеней свободы ($n - 2$), которое составит 2,0639. Так как $t_r > t_{табл}$, делаем вывод о существенности полученного коэффициента корреляции и модели парной линейной регрессии в целом.

Далее осуществим проверку типичности параметров a_0 и a_1. В столбце «t-статистика» указаны значения $t_{a_0} = 3{,}46$ и $t_{a_1} = 5{,}02$.

Сравним полученные значения с $t_{табл}$, взятым при уровне значимости 0,05 и числе степеней свободы ($n - 2$), равным 2,0639. В обоих случаях $t_{факт} > t_{табл}$ делаем вывод, что значения параметров a_0 и a_1 не случайно отклоняются от нуля и отражают реальную природу взаимосвязей между рассматриваемыми переменными в рамках разработанной модели парной линейной регрессии.

В заключение для оценки адекватности уравнения регрессии рассчитаем показатель средней ошибки аппроксимации по формуле (5.31), предварительно определив сумму относительных ошибок аппроксимации.

Таким образом,

$$\bar{A} = \frac{1}{26} \cdot 947{,}45 = 36{,}44\,\%.$$

б) *степенная парная регрессия* $y_x = a_0 \cdot x^{a_1}$.

Для оценки значимости синтезированной модели степенной парной регрессии (задача 4.1) $y_x = 153{,}46 \cdot x^{0{,}394}$ целесообразно воспользоваться инструментом «Нелинейная регрессия» программного продукта *SPSSStatistics 20.* Для этого в данной программе в окне ввода данных отразим исходные значения переменных x и y.

На следующем этапе обратимся к меню «Анализ» → «Регрессия» → «Нелинейная» в появившемся диалоговом окне (рис. 5.6).

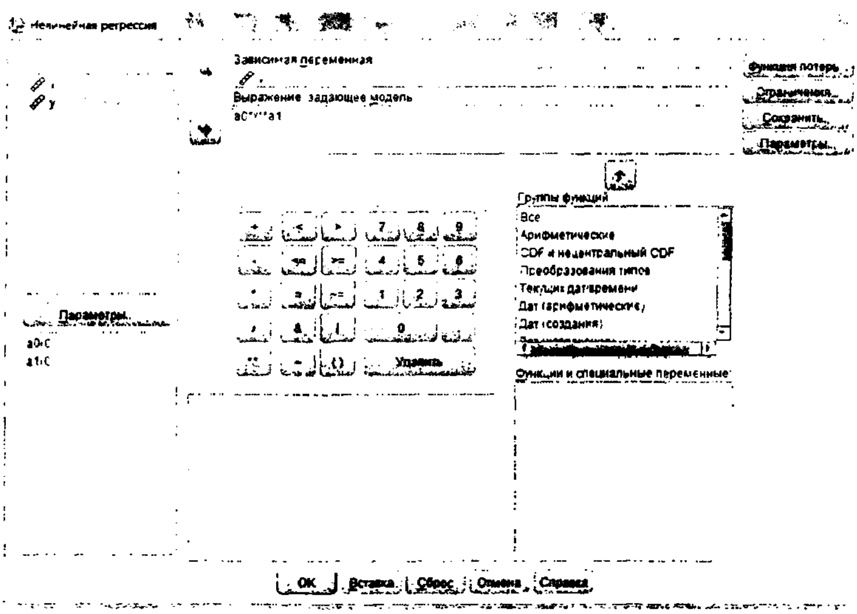


Рис. 5.6. Окно «Нелинейная регрессия» программного продукта *SPSSStatistics 20*

Выполним следующие действия:

- в поле «Зависимая переменная» введем переменную y;
- в поле «Параметры» укажем имена параметров a_0 и a_1, присвоив им начальные значения, равные нулю;

- в поле «Выражение, задающее модель», введем выражение «a0*x**a1»;
- нажмем «ОК».

В появившемся окне вывода результатов найдем таблицу «Дисперсионный анализ» (рис. 5.7).

Дисперсионный анализ[a]

Источник	Сумма квадратов	ст.св.	Средние квадраты
Регрессия	234376508,884	2	117188254,442
Остаток	31130083,586	24	1297086,816
Нескорректированный итог	265506592,470	26	
Скорректированный итог	59339910,759	25	

Зависимая переменная: yR-квадрат=1 —
(остаточная сумма квадратов / скорректированная сумма квадратов) = 0,475.

Рис. 5.7. Дисперсионный анализ уравнения степенной регрессии в *SPSSStatistics 20*

В этой таблице указано значение индекса детерминации $R^2 = 0{,}475$, или 47,5%.

Тогда индекс корреляции составит: $R = \sqrt{0{,}475} = 0{,}69$.

Таким образом, связь по модели в соответствии со шкалой Чеддока можно охарактеризовать как заметную.

Для оценки статистической значимости полученной модели парной степенной регрессии определим фактическое значение F-критерия Фишера по формуле (5.3):

$$F = \frac{0{,}475}{1-0{,}475} \cdot (26-2) = 21{,}7.$$

Полученное значение сопоставим с табличным критическим значением, которое при уровне значимости 0,05 и соответствующем числе степеней свободы составит 4,26. Так как $F_{факт} > F_{табл}$, то делаем вывод о существенности изучаемой связи по данной модели.

Для расчета значения средней ошибки аппроксимации выполним следующие действия:

- в окне ввода данных выполним команду «Преобразовать» → «Вычислить переменную»;
- в появившемся диалоговом окне укажем имя вычисляемой переменной, например «*A*»;
- в поле «Числовое выражение» укажем формулу для расчета относительных ошибок аппроксимации: «ABS((y-PRE_1)/y))*100» (рис. 5.8), нажмем «OK».

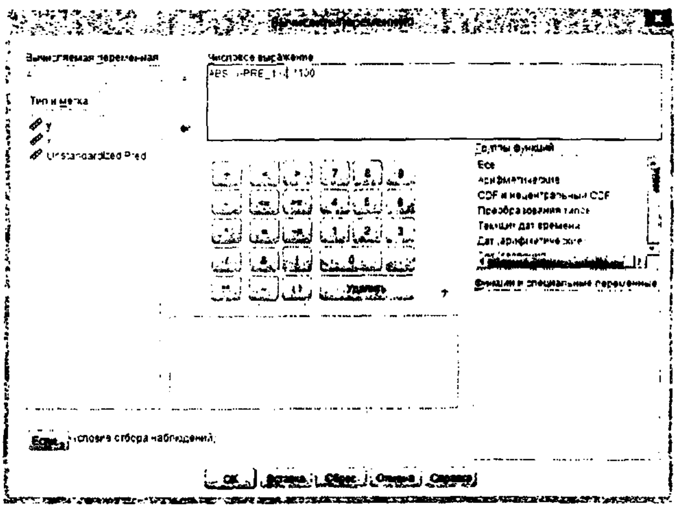


Рис. 5.8. Расчет относительных ошибок аппроксимации в *SPSS Statistics 20*

Далее выполним следующие действия:

- откроем пункт меню «Анализ» → «Описательные статистики» → «Описательные»;
- в появившемся диалоговом окне в поле «Переменные» переместим вновь вычисленную переменную *A*;
- во вкладке «Параметры» отметим «Среднее», нажмем «OK».

В результате в окне вывода результатов появится таблица «Описательные статистики», в которой отражено значение средней ошибки аппроксимации как арифметической средней совокупности относительных ошибок аппроксимации. Таким образом,

$$\bar{A} = 37{,}15\%.$$

Аналогичным образом были рассчитаны показатели для моделей показательной и логарифмической парных регрессий.

Так, для синтезированной в задаче 4.1 показательной парной регрессии

$$y_x = 3097{,}4 \cdot 0{,}9998^x$$

показатели корреляционного анализа составили:

$$R = 0{,}732; \; R^2 = 0{,}536; \; F_{\text{факт}} = 27{,}73; \; \bar{A} = 21{,}86\%.$$

Для модели логарифмической парной регрессии

$$y_x = -5440{,}6 + 2699{,}2\lg x$$

показатели составили:

$$R = 0{,}748; \; R^2 = 0{,}561; \; F_{\text{факт}} = 30{,}67; \; \bar{A} = 24{,}16\%.$$

Таким образом, все полученные модели парной регрессии являются статистически значимыми, но наилучшей (с точки зрения полученных характеристик) для описания зависимости стоимости продукции сельского хозяйства (Y) в рассматриваемых районах региона от вариации стоимости основных фондов сельхозтоваропроизводителей (X) следует признать уравнение парной логарифмической регрессии, так как для данной модели получены наименьшая величина средней ошибки аппроксимации ($\bar{A} = 24{,}16\%$) и наибольшее значение показателей корреляции и детерминации ($R = 0{,}748$ и $R^2 = 0{,}561$).

Задача 5.2. Оцените статистическую значимость регрессионных моделей, полученных в результате решения задачи 4.2.

Решение

а) *линейная регрессия* $y_x = 10\,139 - 100{,}3x$.

Для анализа показателей статистической значимости представленной линейной регрессии обратимся к инструменту

«Множественная регрессия», который применялся в задаче 4.2. Для этого откроем таблицу 4.6 исходных данных, сделаем ее активной. Далее в меню «Анализ» откроем инструмент «Множественная регрессия», укажем зависимую и независимую переменные, нажмем «ОК». В результате получим окно результатов множественной регрессии (рис. 5.9).

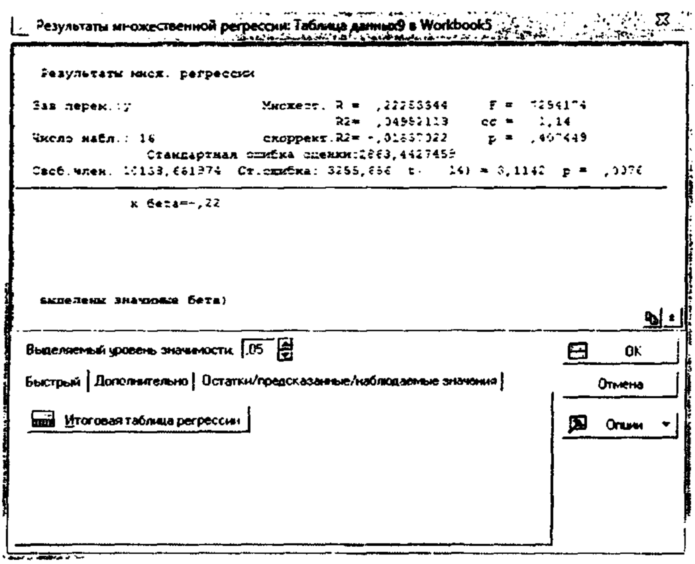


Рис. 5.9. Окно вывода результатов оценки линейной регрессии в программе *Statistica 6.0*

Рассмотрим основные показатели, необходимые для оценки статистической значимости парной линейной регрессии.

Для оценки тесноты связи между независимой (x) и зависимой (y) переменной используется линейный коэффициент корреляции r_{xy}. В нашем случае r_{xy} = 0,22, что в соответствии со шкалой Чеддока характеризует тесноту связи между переменными как слабую.

Для оценки качества выбранной регрессии используется коэффициент детерминации r^2. В нашем случае r_2 = 0,0495. Значит, изменение результативного признака y на 4,9% объясняется влиянием факторного признака x.

Значение F-критерия Фишера $F = 0{,}73$ ниже табличного значения (4,6), следовательно, можно сделать вывод о несущественности изучаемой связи.

Для нахождения t-критериев Стьюдента для параметров a_0 и a_1 можно воспользоваться окном «Итоговая таблица регрессии» (рис. 5.10). В нашем случае $t_{a_0} = 3{,}11$; $t_{a_1} = -0{,}85$. Учитывая, что критическое значение $t_{\text{кр.}} = 2{,}14$ при уровне значимости $p = 0{,}05$, сделаем следующий вывод: свободный член уравнения регрессии a_0 статистически значим, а коэффициент регрессии a_1 является не адекватным исследуемому процессу.

	Итоги регрессии для зависимой переменной: y (Таблица данных9 R= ,22253344 R2= ,04952113 Скорректир. R2= — F(1,14)=,72942 p<,40745 Станд. ошибка оценки: 2863,4					
N=16	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(14)	p-уров.
Св.член			10138,66	3255,656	3,114168	0,007615
x	-0,222533	0,260560	-100,28	117,411	-0,854059	0,407449

Рис. 5.10. Итоговая таблица линейной регрессии в *Statistica 6.0*

Для вычисления средней ошибки аппроксимации как характеристики общего качества модели можно воспользоваться окном «Предсказанные значения и остатки», добавив в него новую переменную для расчета относительных ошибок аппроксимации по формуле

$$A = \left|\frac{y - y_x}{y}\right| \cdot 100\%,$$

и вычислим из них среднюю арифметическую ошибку аппроксимации по формуле (5.7) (рис. 5.11).

В результате имеем: $\bar{A} = 37{,}6\%$.

б) степенная регрессия $y_x = 31430{,}77 \cdot x^{-0{,}44}$.

Рассчитаем индекс корреляции по формуле (5.12).

Для расчета данного показателя табличным способом вычислим значения необходимых сумм (рис. 5.12).

Набл. No.	Предсказанные значения и остатки (Таблица данных4) Зависимая перемен.: y y Значение	yx Значение	y-yx	A =Abs((y-yx)/y)*100
1	4038,00	7932,583	-3894,58	96,45
2	5791,00	7591,644	-1800,64	31,09
3	6580,00	7852,362	-1272,36	19,34
4	4078,00	7220,621	-3142,62	77,06
5	9322,00	7681,892	1640,11	17,59
6	5748,00	7792,196	-2044,20	35,56
7	13195,00	7742,058	5452,94	41,33
8	7786,00	6859,626	926,37	11,90
9	3586,00	6418,411	-2832,41	78,99
10	6510,00	7531,478	-1021,48	15,69
11	10307,00	7681,892	2625,11	25,47
12	8718,00	7391,091	1326,91	15,22
13	7105,00	5626,228	1478,77	20,81
14	9526,00	7711,975	1814,03	19,04
15	4844,00	7932,583	-3088,58	63,76
16	11685,00	7852,362	3832,64	32,80
СРЕДНЕЕ набл. 1-16				37,631587

Рис. 5.11. Расчет средней ошибки аппроксимации для линейной регрессии в *Statistica 6.0*

	Модель: y=a0*x**a1 (Таблица данных9) Зав. Пер. : y Y	Yx	$(Y-Yx)^2$ =(Y-Yx)**2	$(Y-Yc)^2$ Y-7426,188)'
1	4038,00	8032,452	15955645,8	11479817,9
2	5791,00	7538,791	3054772,6	2673839,8
3	6580,00	7906,814	1760434,6	716034,131
4	4078,00	7099,612	9130139,8	11210362,9
5	9322,00	7659,791	2762937,9	3594103,14
6	5748,00	7816,681	4279441,4	2816314,96
7	13195,00	7744,078	29712553,9	33279191,9
8	7786,00	6743,371	1087075,0	129464,675
9	3586,00	6377,910	7794760,0	14747043,9
10	6510,00	7461,504	905359,0	839400,451
11	10307,00	7659,791	7007714,1	8299077,78
12	8718,00	7290,766	2036997,9	1668778,24
13	7105,00	5856,989	1557531,3	103161,731
14	9526,00	7701,558	3328588,8	4409210,44
15	4844,00	8032,452	10166225,4	6667694,87
16	11685,00	7906,814	14274691,6	18137479,7
СРЕДНЕЕ	7426,188			
СУММА			114814868,9	120770976

Рис. 5.12. Расчетные данные для вычисления показателей статистической значимости степенной регрессии в *Statistica 6.0*

Тогда

$$R=\sqrt{1-\frac{114\,814\,868{,}9}{120\,770\,976}}=0{,}22.$$

Таким образом, в соответствии со шкалой Чеддока связь между переменными можно охарактеризовать как слабую.

Индекс детерминации $R^2=0{,}22^2=0{,}048$ указывает, что вариация результативного признака на 4,8% объяснена вариацией факторного признака.

Для оценки статистической значимости модели в целом рассчитаем F-критерий Фишера:

$$F=\frac{0{,}048}{1-0{,}048}\cdot(16-2)=0{,}71.$$

Сравнив полученное значение с табличным (4,6), делаем вывод о несущественности изучаемой связи по модели степенной регрессии.

Средняя ошибка аппроксимации для данной модели составит 37,58%.

в) логарифмическая регрессия $y_x = 17\,746{,}66 - 3149{,}62\ln x$.

Для логарифмической регрессии показатели статистической значимости модели были рассчитаны аналогичным образом и составили:

$R=0{,}22;$

$R^2=0{,}05;$

$F=\dfrac{0{,}05}{1-0{,}05}\cdot(16-2)=0{,}74;$

$\bar{A}=37{,}6\%.$

Данные для расчета перечисленных показателей отражены на рисунке 5.13.

Таким образом, можно сделать вывод, что уравнения линейной, степенной и логарифмической регрессий в данном случае являются не адекватными исследуемому явлению. В этом случае можно говорить о несущественности влияния количества заго-

	Модель: y=a0+a1*Log(x) (Таблица данных9 в Workbook5) Зав. Пер. : y				
	y	yx	(y-yx)^2 =(y-yx)**2	(y-yc)^2 =(y-7426,188)**2	A =Abs(y-yx)/y*100
1	4038,00	8011,037	15785025,9	11479817,9	98,3912176
2	5791,00	7558,415	3123757,06	2673839,8	30,5200373
3	6580,00	7898,539	1738544,75	716034,131	20,0385846
4	4078,00	7130,101	9315322,14	11210362,9	74,8430914
5	9322,00	7672,041	2722363,58	3594103,14	17,6996209
6	5748,00	7816,726	4279626,75	2816314,96	35,9903597
7	13195,00	7750,135	29646553	33279191,9	41,2646065
8	7786,00	6762,739	1047063,11	129464,675	13,1423198
9	3586,00	6365,125	7723533,56	14747043,9	77,4992918
10	6510,00	7484,880	950390,944	839400,451	14,9751147
11	10307,00	7672,041	6943007,13	8299077,78	25,5647488
12	8718,00	7319,693	1955261,53	1668778,24	16,0393056
13	7105,00	5757,104	1816823,69	103161,731	18,9710911
14	9526,00	7710,846	3294783,23	4409210,44	19,0547321
15	4844,00	8011,037	10030125,7	6667694,87	65,3806228
16	11685,00	7898,539	14337287,9	18137479,7	32,4044598
СУММА			114709470	120770976	
СРЕДНЕЕ					37,6112003

Рис. 5.13. Расчетные данные для вычисления показателей статистической значимости логарифмической регрессии в *Statistica 6.0*

товленных кормов в сельхозпредприятиях на производство мяса всех видов во всех категориях хозяйств. Следовательно, для улучшения качества модели необходимо включить в уравнение регрессии одну или несколько дополнительных управляющих (независимых) переменных, которые являлись бы существенными для исследуемого результативного признака.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 5.3. Рассчитайте t-критерий Стьюдента для параметра a_0 уравнения парной регрессии, равного 15,4, если известно, что число узловых точек равно 10, среднее квадратическое отклонение факторного признака — 3,16; остаточная дисперсия результативного признака — 5,7, общая дисперсия — 12,4.

Задача 5.4. Рассчитайте t-критерий Стьюдента для параметра a_1 уравнения парной регрессии, равного 0,4, если известно, что число узловых точек равно 12, среднее квадратическое отклонение факторного признака — 2,76; общая дисперсия результативного признака — 6,2, остаточная дисперсия — 1,8.

Задача 5.5. Рассчитайте t-критерий Стьюдента для параметра a_1 уравнения парной регрессии, равного 1,4, если известно, что число наблюдений равно 12, среднее квадратическое отклонение факторного признака — 4,6, общая дисперсия результативного признака — 5,7, остаточная дисперсия — 1,4. Дайте характеристику значимости параметра.

Задача 5.6. Рассчитайте t-критерий Стьюдента для параметра a_0 уравнения парной регрессии, равного 16,4, если известно, что число узловых точек равно 14, среднее квадратическое отклонение факторного признака — 2,6, общая дисперсия результативного признака — 3,9, остаточная дисперсия — 1,2. Дайте характеристику значимости параметра.

Задача 5.7. Рассчитайте коэффициент корреляции для парной прямолинейной зависимости при двенадцати узловых точках, если известно, что $\sum x = 15$, $\sum x^2 = 85$, $\sum yx = 95$, $\sum y = 58$, $\sum y^2 = 320$, $\sum yx^2 = 95$, $\sum y^2x^2 = 95$. Дайте характеристику силе связи.

Задача 5.8. Рассчитайте коэффициент детерминации для парной прямолинейной зависимости при двенадцати узловых точках, если известно, что $\sum x = 15$, $\sum x^2 = 85$, $\sum yx = 95$, $\sum y = 58$, $\sum y^2 = 320$, $\sum yx^2 = 95$, $\sum y^2x^2 = 95$. Сделайте вывод относительно полученного результата.

Задача 5.9. Рассчитайте коэффициент корреляции для парной криволинейной зависимости при двенадцати узловых точках, если известно, что $\sum x = 15$, $\sum x^2 = 85$, $\sum yx = 95$, $\sum y = 58$, $\sum y^2 = 320$, остаточная дисперсия результативного признака равна 7,6, а общая дисперсия — 15,2. Дайте характеристику силе связи.

Задача 5.10. Рассчитайте коэффициент детерминации для парной криволинейной зависимости при двенадцати узловых точках, если известно, что $\sum x = 15$, $\sum x^2 = 85$, $\sum yx = 95$, $\sum y = 58$, $\sum y^2 = 320$, остаточная дисперсия результативного признака равна 7,6, а общая дисперсия — 15,2. Сделайте вывод относительно полученного результата.

Задача 5.11. Рассчитайте t-критерий Стьюдента для коэффициента корреляции уравнения парной регрессии, равного 0,4, если известно, что число узловых точек равно 12, среднее квадратическое отклонение факторного признака — 4,6, общая дисперсия результативного признака — 5,7, остаточная дисперсия — 1,4. Дайте характеристику значимости показателя.

Задача 5.12. Рассчитайте t-критерий Стьюдента для коэффициента корреляции уравнения парной регрессии, равного 0,8, если известно, что число узловых точек равно 14, среднее квадратическое отклонение факторного признака — 3,6, общая дисперсия результативного признака — 8,7, остаточная дисперсия — 2,4. Дайте характеристику значимости показателя.

Задача 5.13. Оцените значимость индекса детерминации на основе F-критерия Фишера, если известно, что парный коэффициент корреляции равен 0,7, индекс детерминации — 0,49, число узловых точек — 14, относительная ошибка аппроксимации — 8,7%, факторная дисперсия результативного признака — 3,6.

Задача 5.14. Оцените значимость индекса детерминации на основе F-критерия Фишера, если известно, что парный коэффициент корреляции равен 0,6, индекс детерминации — 0,36, число наблюдений — 12, относительная ошибка аппроксимации — 9,7%, факторная дисперсия результативного признака — 5,6.

Задача 5.15. Рассчитайте среднюю ошибку аппроксимации для уравнения регрессии, построенного по двенадцати наблюдениям, если известно, что сумма абсолютных ошибок аппроксимации равна 125,3, а относительных — 72,3%.

Задача 5.16. Рассчитайте среднюю ошибку аппроксимации для уравнения регрессии, построенного по двенадцати наблюдениям, если известно, что сумма абсолютных ошибок аппроксимации равна 325,3, а относительных — 86,3%.

Задача 5.17. По данным задачи 5.10 осуществить идентификацию представленных уравнений парной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод о наиболее приемлемой форме модели для описания исходной зависимости изменения индекса физического объема инвестиций (Y) и среднесложившихся цен на зерно (X). Для выбранной формы модели оцените доверительные интервалы теоретических значений результативной переменной с вероятностью 95%.

Таблица 5.3

Данные об индексе физического объема инвестиций и среднесложившиеся цены на зерно в 2012 г. в разрезе районов Ставропольского края

№ района	Индекс физического объема инвестиций, в % к предыдущему году	Среднесложившиеся цены на зерно, реализованное по всем каналам, руб./ц
	Y	X
1	200,0	648,6
2	47,6	612,9
3	68,9	549,8
4	96,8	645,5
5	26,4	694,1
6	143,5	686,3
7	177,8	689,6
8	49,6	683,2
9	127,2	747,2
10	84,6	673,7
11	269,7	643,2
12	68,5	751,7
13	161,4	787,9
14	98,5	591,9

Задача 5.18. По данным задачи 4.11 осуществить идентификацию представленных уравнений парной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод о наиболее приемлемой форме модели для описания исходной зависимости изменения прибыли сельхозпредприятий (*Y*) и соотношения производственной себестоимости зерна к средней по краю (*X*). Для выбранной формы модели оцените доверительные интервалы теоретических значений результативной переменной с вероятностью 99%.

Таблица 5.4

Данные о соотношении производственной себестоимости зерна к средней по краю и прибыли (убытка) сельхозпредприятий в 2012 г. в разрезе районов Ставропольского края

№ района	Соотношение производственной себестоимости 1 ц зерна к средней по краю, %	Прибыль (убыток) сельхозпредприятий, млн руб.
	X	*Y*
1	116,38	328,1
2	181,36	2,6
3	108,19	309,2
4	117,23	244,1
5	92,09	234,5
6	107,06	414,5
7	125,99	608,5
8	100,00	20,8
9	100,85	299,3
10	148,31	567,2
11	94,35	241,8
12	94,92	112,5

Задача 5.19. По данным задачи 4.12 осуществить идентификацию представленных уравнений парной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод о наиболее приемлемой форме модели для описания исходной зависимости изменения обеспеченности населения мясными продуктами (*Y*) и объемах производства всех видов мяса в

сельхозпредприятиях (X). Для выбранной формы модели оцените доверительные интервалы теоретических значений результативной переменной с вероятностью 95%.

Таблица 5.5

Данные об обеспеченности мясными продуктами населения региона и о производстве мяса всех видов в разрезе районов Ставропольского края в 2012 г.

№ района	Коэффициент обеспеченности мясными продуктами населения	Производство мяса всех видов во всех категориях хозяйств, т (в живом весе)
	Y	X
1	0,654	6264,2
2	0,660	5512,3
3	0,658	11251
4	0,663	7300
5	0,664	5824
6	0,670	8986
7	0,665	24855
8	0,674	5,71
9	0,665	9073
10	0,664	9729,3
11	0,668	4537
12	0,672	53982
13	0,663	13851
14	0,666	7260
15	0,662	13053

Задача 5.20. Осуществить идентификацию уравнений парной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод о наиболее приемлемой форме модели для описания исходной зависимости изменения стоимости сельскохозяйственной продукции СХО (Y) и среднегодовой стоимости основных фондов (X). Для выбранной формы модели оцените доверительные интервалы теоретических значений результативной переменной с вероятностью 90%.

Таблица 5.6

Данные о среднегодовой стоимости основных фондов и стоимости сельскохозяйственной продукции СХО в 2012 г. в разрезе районов Ставропольского края, млн руб.

№ района	Среднегодовая стоимость основных фондов СХО	Объем сельскохозяйственной продукции СХО
	X	*Y*
1	718,8	141,9
2	1216,8	64,1
3	864,2	145,7
4	580,4	89,3
5	549,8	49,6
6	1250,0	116,3
7	1579,3	110,0
8	2052,6	66,9
9	894,0	138,6
10	886,7	159,5
11	495,6	129,0
12	321,0	162,6
13	1377,7	87,5
14	520,4	84,3

Задача 5.21. Осуществить идентификацию уравнений парной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод о наиболее приемлемой форме модели для описания исходной зависимости изменения производства мяса всех видов (*Y*) и заготовки кормов в сельхозпредприятиях (*X*). Для выбранной формы модели оцените доверительные интервалы теоретических значений результативной переменной с вероятностью 90%.

Таблица 5.7

Данные о производстве мяса всех видов и заготовке кормов в сельхозпредприятиях в разрезе районов Ставропольского края в 2012 г.

№ района	Производство мяса всех видов во всех категориях хозяйств, т (в живом весе)	Заготовлено кормов в сельхозпредприятиях в расчете на одну условную голову скота, ц корм. ед.
	Y	*X*
1	2	3
1	4038,0	22,0

Окончание табл. 5.7

1	2	3
2	5791,0	25,4
3	6580,0	22,8
4	4078,0	29,1
5	9322,0	24,5
6	5748,0	23,4
7	13195,0	23,9
8	7786,0	32,7
9	3586,0	37,1
10	6510,0	26,0
11	10307,0	24,5
12	8718,0	27,4
13	7105,0	45,0
14	9526,0	24,2
15	4844,0	22,0
16	11685,0	22,8

ГЛАВА

6 ИДЕНТИФИКАЦИЯ МНОГОФАКТОРНЫХ ЭКОНОМЕТРИЧЕСКИХ МОДЕЛЕЙ

6.1. Идентификация множественной регрессии

Практическая значимость уравнения множественной регрессии оценивается при помощи тех же показателей, что и парной — показателей корреляции и детерминации. Показатель множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым результативным признаком, оценивает тесноту совместного влияния факторов на результат. Независимо от формы связи показатель множественной корреляции может быть рассчитан как *индекс множественной корреляции*:

$$R_{yx_1,x_2,\ldots,x_n} = \sqrt{1-\frac{S_\varepsilon^2}{S_y^2}} = \sqrt{1-\frac{\sum(y-y_{x_1,x_2,\ldots,x_n})^2}{\sum(y-\bar{y})^2}}. \qquad (6.1)$$

При правильном включении факторов в регрессионную модель величина индекса множественной корреляции существенно отличается от коэффициента корреляции парной зависимости. Если дополнительно включаемые в модель факторы малозначимы, то индекс множественной корреляции будет практически совпадать с индексом парной корреляции. Следовательно, сравнивая эти показатели корреляции, эконометрист делает вывод о целесообразности включения в модель того или иного фактора.

При линейной зависимости признаков формула индекса корреляции может быть представлена следующим выражением:

$$R_{yx_1,x_2,\ldots,x_n} = \sqrt{\beta_{x_i} \cdot r_{yx_i}}, \qquad (6.2)$$

где β_{x_i} — стандартные (нормированные) коэффициенты регрессии;

r_{yx_i} — парные коэффициенты корреляции результата с соответствующим фактором.

Между частными коэффициентами корреляции и β-коэффициентами существует следующая засисимость:

$$\begin{cases} r_{x_1y} = \beta_1 + r_{x_1x_2} \cdot \beta_2 + r_{x_1x_3} \cdot \beta_3, \\ r_{x_2y} = \beta_1 \cdot r_{x_1x_2} + \beta_2 + r_{x_2x_3} \cdot \beta_3, \\ r_{x_3y} = \beta_1 \cdot r_{x_1x_3} + \beta_2 \cdot r_{x_2x_3} + \cdot \beta_3. \end{cases} \tag{6.3}$$

Формула индекса множественной корреляции для линейной регрессии получила название линейного коэффициента множественной корреляции, или совокупного коэффициента корреляции.

Методика построения индекса детерминации для множественной зависимости аналогична нахождению коэффициента детерминации для парной регрессии:

$$(R_{yx_1,x_2,\dots,x_n})^2 = R^2_{yx_1,x_2,\dots,x_n}. \tag{6.4}$$

В рассмотренных показателях множественной корреляции используется остаточная дисперсия, которая имеет систематическую ошибку в сторону приуменьшения, тем более значимую, чем больше параметров определяется в уравнении регрессии при заданном объеме наблюдений n. Если число параметров при x_i равно m и стремится к n, то остаточная дисперсия будет равна нулю, а показатель корреляции приблизится к единице даже при слабой связи факторов с результатом. В целях повышения точности при характеристике тесноты связи используют *скорректированный* индекс множественной корреляции, содержащий поправку на число степеней свободы:

$$R = \sqrt{1 - \frac{\sum (y - y_x)^2 : (n - (m+1))}{\sum (y - \bar{y})^2 : (n-1)}}. \tag{6.5}$$

Так как $$\frac{\sum (y - y_x)^2}{\sum (y - \bar{y})^2} = 1 - R^2,$$

скорректированный индекс детерминации можно записать:

$$R^2 = 1-(1-R^2)\cdot\frac{n-1}{n-(m+1)}. \qquad (6.6)$$

Величина индекса множественной детерминации используется для оценки качества регрессионной модели. Низкое значение показателя множественной корреляции означает, что в регрессионную модель не включены существенные факторы, и рассматриваемая форма связи не отражает реальные соотношения между переменными. Следовательно, требуются дальнейшие исследования по улучшению качества модели и увеличению ее практической значимости.

Одним из методов отбора факторов для множественной регрессии является анализ *стандартизированных коэффициентов* регрессии (β-коэффициентов), характеризующих наиболее крупные резервы улучшения изучаемого признака, так как учитывается степень варьирования факторов.

Стандартизированные коэффициенты регрессии определяются по формуле

$$\beta_i = a_1\cdot\frac{S_x}{S_y}. \qquad (6.7)$$

Чем выше значение β-коэффициента, тем выше варьируемость рассматриваемого фактора.

Другим способом является анализ частных коэффициентов корреляции (для линейных зависимостей) и частных индексов детерминации (для нелинейных).

Применительно ко множественной регрессии рассчитываются частные коэффициенты эластичности, определяемые на основе частных уравнений регрессии:

$$Э_{yx_i} = a_i\cdot\frac{x_i}{y_{x_i\cdot x_1,x_2,\ldots,x_{i-1},x_{i+1},\ldots,x_n}}, \qquad (6.8)$$

где a_i — коэффициент регрессии для фактора x_i в уравнении множественной регрессии;

$y_{x_i\cdot x_1,x_2,\ldots,x_{i-1},x_{i+1},\ldots,x_n}$ — частное уравнение регрессии.

Также в уравнении регрессии могут быть рассчитаны средние частные коэффициенты эластичности:

$$\bar{Э}_{yx_i} = a_i \cdot \frac{\bar{x}_i}{\bar{y}_{x_i}}. \tag{6.9}$$

6.2. Частные регрессия и корреляция

Частные уравнения регрессии характеризуют изолированное влияние факторного признака на результативный, так как другие факторы закреплены на неизменном уровне. Эффекты влияния других факторов присоединены в них к свободному члену уравнения множественной регрессии.

Для линейного уравнения множественной регрессии

$y_x = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n$

могут быть построены частные уравнения регрессии:

$$\begin{cases} y_{x_1 \cdot x_2, x_3, \ldots, x_n} = f(x_1), \\ y_{x_2 \cdot x_1, x_3, \ldots, x_n} = f(x_2), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ y_{x_n \cdot x_1, x_2, \ldots, x_{n-1}} = f(x_n). \end{cases} \tag{6.10}$$

В частных уравнениях регрессии изучается влияние соответствующего фактора x на результативный признак при закреплении других учитываемых во множественной регрессии факторов на среднем уровне. Таким образом, частные уравнения регрессии имеют вид:

$$\begin{cases} y_{x_1 \cdot x_2, x_3, \ldots, x_n} = A_1 + a_1 x_1, \\ y_{x_2 \cdot x_1, x_3, \ldots, x_n} = A_2 + a_2 x_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_{x_n \cdot x_1, x_2, \ldots, x_{n-1}} = A_n + a_n x_n, \end{cases} \tag{6.11}$$

где
$$\begin{cases} A_1 = a_0 + a_2\bar{x}_2 + a_3\bar{x}_3 + \ldots + a_n\bar{x}_n, \\ A_2 = a_0 + a_1\bar{x}_1 + a_3\bar{x}_3 + \ldots + a_n\bar{x}_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ A_n = a_0 + a_1\bar{x}_1 + a_2\bar{x}_2 + \ldots + a_{n-1}\bar{x}_{n-1}. \end{cases} \tag{6.12}$$

Частные коэффициенты (индексы) корреляции характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других, включенных в уравнение регрессии. Показатели частной корреляции представляют собой отношение сокращения остаточной дисперсии за счет дополнительного включения в анализ нового фактора к остаточной дисперсии, имевшей место до введения его в модель. Чем больше доля сокращения остаточной дисперсии за счет дополнительного включения фактора в остаточной вариации до его введения, тем теснее связь между результатом и новым фактором при постоянном действии фактора x_i. Квадратный корень из этой величины называют индексом частной корреляции, показывающим тесноту связи результативного признака и включенного фактора.

Для двухфакторной модели регрессии чистое влияние на результат фактора x_1 определяется как
$$r_{yx_1 \cdot x_2} = \sqrt{\frac{S^2_{yx_2} - S^2_{yx_1x_2}}{S^2_{x_2}}},$$

а фактор x_2
$$r_{yx_2 \cdot x_1} = \sqrt{\frac{S^2_{yx_1} - S^2_{yx_1x_2}}{S^2_{yx_1}}},$$

где
$$S^2_{yx_1} = \frac{\sum (y - y_{x_1})^2}{n};$$
$$S^2_{yx_2} = \frac{\sum (y - y_{x_2})^2}{n};$$
$$S^2_{yx_1x_2} = \frac{\sum (y - y_{x_1x_2})^2}{n}.$$

Если выразить остаточную дисперсию через показатель детерминации:

$$S_{\Delta}^2 = S_y^2 \cdot (1 - r^2),$$

то коэффициент частной корреляции принимает вид

$$r_{yx_1 \cdot x_2} = \sqrt{\frac{S_{yx_2}^2 - S_{yx_1 \cdot x_2}^2}{S_{yx_2}^2}} = \sqrt{1 - \frac{S_{yx_1 x_2}^2}{S_{yx_2}^2}} = \sqrt{1 - \frac{1 - R_{yx_1 x_2}^2}{1 - r_{yx_2}^2}},$$

соответственно

$$r_{yx_2 \cdot x_1} = \sqrt{1 - \frac{1 - R_{yx_2 x_1}^2}{1 - r_{yx_1}^2}}.$$

Рассмотренные выше показатели частной корреляции называются коэффициентами (индексами) частной корреляции первого порядка, так как они фиксируют тесноту связи двух переменных при элиминировании (закреплении влияния) одного фактора.

Если рассматриваются регрессии с числом факторов n, то возможны частные коэффициенты корреляции первого, второго и т. д. до ($n - 1$) порядка. Следовательно, влияние фактора x_1 можно оценивать через коэффициент корреляции:

- *первого порядка* $r_{yx_1 \cdot x_2}$ — при постоянном действии фактора x_2;
- *второго порядка* $r_{yx_1 \cdot x_2 x_3}$ — при постоянном действии факторов x_2 и x_3;
- *($n - 1$) порядка* $r_{yx_1 \cdot x_2 x_3 \ldots x_n}$ — при постоянном действии всех факторов, включенных в модель.

Соответственно, коэффициент парной корреляции называют коэффициентом нулевого порядка.

Аналитический интерес представляют показатели частной корреляции разных порядков, однако на практике предпочтение отдают показателям частной корреляции наивысшего порядка, так как они дополняют уравнение множественной регрессии.

Для n-го множества факторов уравнения регрессии вида

$$y_x = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n$$

коэффициент частной корреляции, измеряющий влияние на результативный признак y фактора x_i при неизменном влиянии других, определяется по формуле

$$r_{yx_i \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_n} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2 \ldots x_i \ldots x_n}}{1 - R^2_{yx_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_n}}}, \qquad (6.13)$$

где $R^2_{yx_1 x_2 \ldots x_i \ldots x_n}$ — множественный коэффициент детерминации всех факторов с результатом;

$R^2_{yx_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_n}$ — множественный коэффициент детерминации без введения в модель фактора x_i.

При $i = 1$ коэффициент частной корреляции принимает вид

$$r_{yx_1 \cdot x_2 x_3 \ldots x_n} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2 \ldots x_n}}{1 - R^2_{yx_2 x_3 \ldots x_n}}}.$$

Коэффициент частной корреляции более высоких порядков можно определить через коэффициенты частной корреляции более низких порядков по рекуррентной формуле

$$r_{yx_i \cdot x_1 x_2 \ldots x_n} = \frac{r_{yx_i \cdot x_1 x_2 \ldots x_{n-1}} - r_{yx_n \cdot x_1 x_2 \ldots x_{n-1}} \cdot r_{x_i x_n \cdot x_1 x_2 \ldots x_{n-1}}}{\sqrt{(1 - r^2_{yx_n \cdot x_1 x_2 \ldots x_{n-1}}) \cdot (1 - r^2_{x_i x_n \cdot x_1 x_2 \ldots x_{n-1}})}}. \qquad (6.14)$$

В двухфакторной модели при $i = 1$ данная формула принимает вид

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r^2_{yx_2}) \cdot (1 - r^2_{x_1 x_2})}}.$$

Соответственно, при $i = 2$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{(1 - r^2_{yx_1}) \cdot (1 - r^2_{x_1 x_2})}}.$$

При трехфакторной регрессионной модели и $i = 1$ имеем формулу для расчета

$$r_{yx_1 \cdot x_2 x_3} = \frac{r_{yx_1 \cdot x_2} - r_{yx_3 \cdot x_2} \cdot r_{x_1 x_3 \cdot x_2}}{\sqrt{(1 - r^2_{yx_3 \cdot x_2}) \cdot (1 - r^2_{x_1 x_3 \cdot x_2})}};$$

при $i = 2$

$$r_{yx_2 \cdot x_1 x_3} = \frac{r_{yx_2 \cdot x_1} - r_{yx_3 \cdot x_1} \cdot r_{x_2 x_3 \cdot x_1}}{\sqrt{(1 - r^2_{yx_3 \cdot x_1}) \cdot (1 - r^2_{x_2 x_3 \cdot x_1})}};$$

при $i = 3$

$$r_{yx_3 \cdot x_1 x_2} = \frac{r_{yx_3 \cdot x_1} - r_{yx_2 \cdot x_1} \cdot r_{x_2 x_3 \cdot x_1}}{\sqrt{(1 - r^2_{yx_2 \cdot x_1}) \cdot (1 - r^2_{x_2 x_3 \cdot x_1})}}.$$

Зная частные коэффициенты корреляции (различных порядков), можно определить совокупный коэффициент множественной корреляции по формуле

$$R_{yx_1 x_2 \ldots x_n} = \sqrt{1 - (1 - r^2_{yx_1}) \cdot (1 - r^2_{yx_2 \cdot x_1}) \cdot (1 - r^2_{yx_3 \cdot x_1 x_2}) \cdot \ldots \cdot (1 - r^2_{yx_n \cdot x_1 x_2 \ldots x_{n-1}})}.$$

В приведенной формуле подкоренное выражение характеризует совокупное действие всех исследуемых факторов, так как из единицы вычитается доля остаточной вариации результативного признака $(1 - r^2)$, обусловленная последовательно включаемыми в анализ факторами.

В эконометрике частные коэффициенты корреляции обычно не имеют самостоятельного значения. В основном их используют на стадии формирования модели. Так, строя многофакторную модель, например, методом исключения факторов, на первом шаге отбирают фактор с наименьшей и несущественной (по t-критерию Стьюдента) величиной показателя частной корреляции. Исключив его из модели, строят новое уравнение регрессии. Процедура продолжается до тех пор, пока все частные коэффициенты корреляции не будут существенно отличаться от нуля. При этом множественные коэффициенты детерминации на двух смежных шагах построения регрессионной модели почти не отличаются $R^2_{n+1} \approx R^2_n$, если исключается несущественный фактор.

6.3. Оценка статистической значимости уравнения множественной регрессии

Значимость уравнения множественной регрессии в целом оценивается, так же как и в парной, с помощью F-критерия Фишера. При этом оценивается значимость не только уравнения в целом, но и фактора, дополнительно включенного в модель. Необходимость такой оценки связана с тем, что не каждый фактор, вошедший в модель, существенно увеличивает долю объясненной вариации результативного признака. Кроме того, при наличии в модели нескольких факторов они могут вводиться в модель в разной последовательности. Ввиду корреляции между факторами значимость одного и того же фактора может быть разной в зависимости от последовательности его введения в модель. Мерой для оценки включения в модель фактора служит частный F-критерий (F_{x_i}), построенный на сравнении прироста факторной дисперсии на одну степень свободы по регрессионной модели в целом:

$$F_{x_i} = \frac{\Delta S^2_{y_x}}{S^2_\Delta}$$

или

$$F_{x_i} = \frac{R^2_{yx_1x_2...x_i...x_n} - R^2_{yx_1x_2...x_{i-1}x_{i+1}...x_n}}{1 - R^2_{yx_1x_2...x_i...x_n}} \cdot \frac{n-(m+1)}{1}, \quad (6.15)$$

где $R^2_{yx_1x_2...x_i...x_n}$ — коэффициент множественной детерминации для модели с полным набором факторов;

$R^2_{yx_1x_2...x_{i-1}x_{i+1}...x_n}$ — коэффициент множественной детерминации без включения в модель фактора x_i.

Так как прирост факторной суммы квадратов обусловлен дополнительным включением в модель одного исследуемого фактора, то число степеней свободы для него равно 1.

Фактическое значение частного F-критерия сравнивается с табличным при соответствующем уровне значимости и числе степеней свободы. Если фактическое значение F_{x_i} больше $F_{табл}$, то дополнительное включение фактора x_i в модель статистически обосновано и коэффициент чистой регрессии a_i при факторе x_i значим. Если фактическое значение меньше табличного, то дополнительное включение в модель фактора x_i не увеличивает существенно долю объясненной вариации признака y, следовательно, его включение нецелесообразно, а коэффициент регрессии a_i при факторе x_i статистически незначим.

С помощью частного F-критерия можно проверить значимость всех коэффициентов регрессии, если предположить, что фактор x_i вводится в множественное уравнение регрессии последним:

$$F_{x_1} = \frac{R^2 - r^2_{yx_2}}{1 - R^2} \cdot (n - m - 1),$$
$$F_{x_2} = \frac{R^2 - r^2_{yx_1}}{1 - R^2} \cdot (n - m - 1). \qquad (6.16)$$

Частный F-критерий, оценивая значимость коэффициентов регрессии, позволяет определить t-критерий для коэффициента регрессии a_i при x_i факторе:

$$t_{a_i} = \sqrt{F_{x_i}}. \qquad (6.17)$$

Если уравнение содержит больше двух факторов, то может быть показана значимость последовательного добавления к уравнению регрессии соответствующего фактора. Так, для уравнения вида

$$y_x = a_0 + a_1x_1 + a_2x_2 + a_3x_3$$

определяют последовательно F-критерий для уравнения с одним фактором x_1, далее F-критерий для дополнительно включенного фактора x_2 (для перехода к двухфакторной модели), а затем для x_3, оценка значимости которого дается после включения в модель факторов x_1 и x_2. В этом случае F-критерий для дополнительного включения фактора x_2 после x_1 является *после-*

довательным, а F-критерий для дополнительного включения в модель фактора x_3 — частным, так как оценивает значимость фактора исходя из предположения, что он включен в модель последним. С t-критерием Стьюдента связан частный F-критерий, а последовательный анализируется на стадии формирования модели.

Оценка значимости коэффициентов чистой регрессии по t-критерию Стьюдента может быть проведена без расчетов частных F-критериев:

$$t_{a_i} = \frac{a_i}{m_{a_i}},$$

где

$$m_{a_i} = \frac{S_y \cdot \sqrt{1 - R^2_{yx_1 \ldots x_n}}}{S_{x_i} \cdot \sqrt{1 - R^2_{x_i x_1 \ldots x_n}}} \cdot \frac{1}{\sqrt{n - (m+1)}};$$

$$a_i = \frac{S_y}{S_{x_i}} \cdot \sqrt{\frac{R^2_{yx_1 \ldots x_n} - R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_n}}{1 - R^2_{x_i x_1 \ldots x_n}}}.$$

Проверим, что

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

На основе указанного соотношения получим:

$$t_{a_i} = \frac{S_y}{S_{x_i}} \cdot \sqrt{\frac{R^2_{yx_1 \ldots x_n} - R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_n}}{1 - R^2_{x_i x_1 \ldots x_n}}} : \frac{S_y \cdot \sqrt{1 - R^2_{yx_1 \ldots x_n}}}{S_{x_i} \cdot \sqrt{1 - R^2_{x_i x_1 \ldots x_n}}} \cdot \frac{1}{\sqrt{n - (m+1)}} =$$

$$= \sqrt{\frac{R^2_{yx_1 \ldots x_n} - R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_n}}{1 - R^2_{x_i x_1 \ldots x_n}}} \cdot \frac{1}{\sqrt{n - (m+1)}} = \sqrt{F_{x_i}}.$$

Взаимосвязь частного коэффициента корреляции, частного F-критерия и t-критерия Стьюдента для коэффициентов чистой

регрессии используется в процедурах отбора факторов. Отбор факторов методом исключения осуществляется не только по частным коэффициентам корреляции, исключая фактор с наименьшим незначимым значением частного коэффициента корреляции, но и по величинам t_{a_i} и F_{x_i}. Частный F-критерий широко используется и при построении модели методами включения и пошагового отбора переменных.

Как отмечалось выше, проверка статистической корректности эконометрической модели предполагает оценку автокорреляции остатка и мультиколлинеарности переменных, которая изложена в главе 4.

Корректность модели в целом в большей степени предполагает оценку систем уравнений и будет показана в главе 10. Между тем оценить качество синтезированной модели в целом можно, основываясь на минимальности отклонения фактических значений результативного признака от теоретических, рассчитанных по уравнению регрессии.

Контрольные вопросы

1. Оценка практической значимости модели множественной регрессии.
2. Расчет индекса множественной корреляции.
3. Методика построения индекса множественной детерминации.
4. Скорректированный индекс детерминации.
5. Средние частные коэффициенты эластичности.
6. Частные уравнения регрессии.
7. Методы отбора факторов для множественной регрессии.
8. Стандартизованные β-коэффициенты.
9. Порционные коэффициенты детерминации.
10. Индексы частной корреляции.
11. Оценка значимости уравнения множественной регрессии с помощью F-критерия Фишера.
12. Оценка адекватности моделей множественной регрессии.

ТЕСТОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ ПРОВЕРКИ ЗНАНИЙ

1. Критические значения критерия Стьюдента определяются:

а) по уровню значимости и одной степени свободы;
б) трем и более степеням свободы;
в) двум степеням свободы;
г) уровню незначимости.

2. Среди нелинейных эконометрических моделей рассматривают следующие классы нелинейных уравнений:

а) внешне нелинейные;
б) внутренне линейные;
в) внутренне нелинейные;
г) внешне линейные.

3. В эконометрическую модель множественной регрессии включаются факторы:

а) коллинеарные;
б) неколлинеарные;
в) существенные;
г) несущественные.

4. Выберите дисперсии, которые участвуют в расчете значения критерия Фишера:

а) остаточная;
б) факторная;
в) независимая;
г) неопределенная.

5. Значение индекса корреляции, рассчитанного для нелинейного уравнения регрессии, характеризует тесноту связи:

а) случайной;
б) нелинейной;
в) обратно пропорциональной;
г) линейной.

6. Значение коэффициента детерминации близко к нулю. Следовательно, остаточная сумма квадратов стремится:
   а) к средней сумме квадратов отклонений;
   б) общей сумме квадратов отклонений;
   в) факторной сумме квадратов отклонений;
   г) объясненной регрессией сумме квадратов отклонений.

7. Индекс корреляции для нелинейных форм связи изменяется в пределах:
   а) (0;1);
   б) [0;1);
   в) [0;4];
   г) [0;1].

8. Какие статистические гипотезы выдвигаются при проверке статистической значимости построенной модели?
   а) альтернативная о статистической значимости;
   б) независимая о статистической независимости;
   в) нулевая о статистической незначимости;
   г) зависимая о статистической зависимости.

9. Квадрат частного коэффициента корреляции $r_{yx_j \cdot x_1 x_2 \dots x_{j-1} x_{j+1} \dots x_k}, j \in (1;k)$ представляет собой:
   а) долю дисперсии y, объясненную переменной x_j после удаления эффекта от действия переменных $\{x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_k\}$;
   б) долю дисперсии y, объясненную переменными $\{x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_k\}$;
   в) долю дисперсии y, объясненную добавлением переменной x_j к набору факторных переменных $\{x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_k\}$;
   г) долю дисперсии y, объясненную переменными $\{x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_k\}$.

10. Коэффициент детерминации может принимать значения в интервале:

а) от −1 до 0;
б) от 0 до 100;
в) от −1 до 1;
г) от 0 до 1.

11. Малое значение t-статистики для коэффициента регрессии в парной линейной регрессии соответствует:

а) отсутствию статистической связи результирующего и факторного признака;
б) отсутствию статистической связи между коэффициентом регрессии и факторным признаком;
в) наличию сильной статистической связи результирующего и факторного признака;
г) наличию статистической связи результирующего признака и свободного члена регрессионной зависимости.

12. При отборе факторов в модель множественной регрессии необходимо:

а) включать в модель факторы, имеющие достаточно тесную связь с другими объясняющими переменными;
б) избегать включения в модель факторов, имеющих достаточно тесную связь с зависимой переменной;
в) избегать одновременного включения в модель коллинеарных факторов;
г) включать в модель факторы, имеющие достаточно тесную связь с зависимой переменной.

13. Коэффициент эластичности показывает:

а) на сколько единиц изменится фактор при изменении результата на 1 единицу;
б) на сколько единиц изменится результат при изменении фактора на 1 единицу;
в) на сколько процентов изменится результат при изменении фактора на 1%;

г) во сколько раз изменится результат при изменении фактора на одну единицу;
д) на сколько процентов изменится фактор при изменении результата на 1%.

14. Следующая регрессия построена по месячным данным за 6 лет:

$$y = -12{,}23 + 0{,}91 \cdot x_1 - 2{,}1 \cdot x_2,$$

где y — потребление; x_1 — располагаемый доход; x_2 — процентная банковская ставка по вкладам.

Модель характеризуется следующими показателями: $R^2 = 0{,}976$, $F = 3{,}2$. Построенная модель характеризуется следующим качеством и направлением влияния объясняющих переменных:

а) качество модели низкое, направление влияния совпадает;
б) качество модели высокое, но направление влияния не совпадает;
в) качество модели высокое, направление влияния совпадает;
г) качество модели низкое, направление влияния совпадает.

15. Имеется уравнение, полученное МНК:

$$y_t = 1{,}12 - 0{,}0098 \cdot x_{t_1} - 5{,}62 \cdot x_{t_2} + 0{,}044 \cdot x_{t_3}.$$

Так как регрессионная (факторная) сумма квадратов составила 110,32; остаточная сумма квадратов – 21,43; индекс множественной детерминации равен:
а) 0,67;
б) 0,45;
в) 0,84.

ПРАКТИКУМ

Задача 6.1. По данным задачи 4.3 произведите идентификацию представленных уравнений множественной регрессии.

Решение

Для оценки статистической значимости полученных уравнений регрессии воспользуемся программным продуктом SPSS 20.

а) линейная регрессия

$$y_x = 192{,}9 + 0{,}001x_1 - 0{,}002x_2 - 0{,}028x_3 - 0{,}2x_4.$$

Осуществим идентификацию данного уравнения линейной множественной регрессии в *Excel*.

Для этого выполним следующие действия:

- сформируем таблицу исходных данных (табл. 4.7);
- в меню «Данные» откроем инструмент «Анализ данных», выберем пункт «Регрессия», нажмем «ОК»;
- в появившемся диалоговом окне в поле «Входной интервал *Y*» укажем значения результативной переменной, а в поле «Входной интервал *X*» — совокупность значений факторных признаков (рис. 6.1), нажмем «ОК»

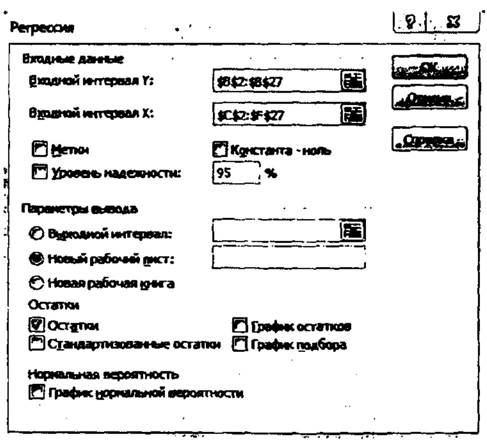


Рис. 6.1. Диалоговое окно «Регрессия» в *Excel*

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,377439166
R-квадрат	0,142460324
Нормированный R-кв	-0,020880567
Стандартная ошибка	3,4030953
Наблюдения	26

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	4	40,40240537	10,10060134	0,87216571	0,497032666
Остаток	21	243,20221	11,58105762		
Итого	25	283,6046154			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	192,9067096	18,37114957	10,50052469	0,00000000082	154,7018126	231,111607
Переменная X 1	0,001376272	0,000876158	1,57080271	0,131174541	-0,000445799	0,00319834
Переменная X 2	-0,001626471	0,006370965	-0,255294231	0,800979578	-0,014875619	0,01162268
Переменная X 3	-0,027641689	0,080054702	-0,345285011	0,73331472	-0,194124555	0,13884118
Переменная X 4	-0,202124202	0,277486231	-0,728411644	0,474406728	-0,779188409	0,37494001

Рис. 6.2. Вывод итогов регрессии в *Excel*

В результате получим новый рабочий лист, в котором отражены основные показатели статистической значимости построенного уравнения линейной множественной регрессии (рис. 6.2).

Произведем интерпретацию полученных результатов.

В соответствии со значением $R = 0,377$ можно сделать вывод об умеренной тесноте связи между набором факторных признаков и результативной переменной y.

Значение $R^2 = 0,142$ свидетельствует о том, что вариация результативной переменной y лишь на 14,2% обусловлена воздействием совокупности включенных в уравнение линейной множественной регрессии управляющих переменных.

В столбце «F» найдем значение F-критерия Фишера $F = 0,872$. Сравнив его с критическим значением (2,84), сделаем вывод о статистической неадекватности построенной модели. Состоятельность такого вывода подтверждает и уровень значимости F-критерия $p = 0,497$, в соответствии с которым вероятность получения ошибочного результата по построенной модели составляет 49,7%.

Оценку практической значимости представленных нелинейных уравнений множественной регрессии целесообразно выполнить с помощью *SPSSStatistics* 20.

б) степенная регрессия $y_x = 191,95 \cdot x_1^{0,03} \cdot x_2^{-0,028} \cdot x_3^{-0,003} \cdot x_4^{-0,01}$.

Дисперсионный анализ[a]

Источник	Сумма квадратов	ст.св.	Средние квадраты
Регрессия	956974,426	5	191394,885
Остаток	237,794	21	11,324
Нескорректированный итог	957212,220	26	
Скорректированный итог	283,605	25	

Зависимая переменная: yR-квадрат = 1 – (остаточная сумма квадратов / скорректированная сумма квадратов) = 0,162.

Рис. 6.3. Дисперсионный анализ уравнения степенной множественной регрессии

Для оценки статистической значимости степенной регрессии обратимся к таблице «Дисперсионный анализ» в окне вывода результатов построения степенной регрессии в задаче 4.3 (рис. 6.3).

В соответствии с данными указанной таблицы $R^2 = 0,162$, что свидетельствует о том, что вариация результативной переменной y на 16,2% обусловлена воздействием совокупности включенных в уравнение степенной множественной регрессии управляющих переменных.

Следовательно, $R = \sqrt{0,162} = 0,4$. Таким образом, результативная переменная y находится в умеренной связи с набором включенных в уравнение степенной регрессии факторных переменных.

Для оценки статистической значимости модели в целом по формуле (5.27) рассчитаем значение F-критерия Фишера:

$$F = \frac{0,162}{1-0,162} \cdot \frac{26-4-1}{4} = 1,01.$$

Сравнивая его с критическим значением (2,84), следует признать уравнение степенной регрессии статистически незначимым.

в) логарифмическая регрессия

$$y_x = 189,896 + 5,73 \cdot \ln x_1 - 5,01 \cdot \ln x_2 - 0,66 \cdot \ln x_3 - 1,96 \cdot \ln x_4.$$

Для идентификации логарифмической регрессии аналогично проанализируем таблицу «Дисперсионный анализ» в окне вывода результатов построения логарифмической регрессии в задаче 4.3 (рис. 6.4).

Дисперсионный анализ[a]

Источник	Сумма квадратов	ст.св.	Средние квадраты
Регрессия	956974,085	5	191394,817
Остаток	238,135	21	11,340
Нескорректированный итог	957212,220	26	
Скорректированный итог	283,605	25	

Зависимая переменная: yR-квадрат = 1 — (остаточная сумма квадратов / скорректированная сумма квадратов) = 0,160.

Рис. 6.4. Дисперсионный анализ уравнения логарифмической множественной регрессии

В соответствии с данными указанной таблицы $R^2 = 0{,}16$, что свидетельствует о том, что вариация результативной переменной y на 16,0% обусловлена воздействием совокупности включенных в уравнение логарифмической множественной регрессии экзогенных переменных.

Следовательно, $R = \sqrt{0{,}16} = 0{,}4$. Таким образом, результативная переменная y находится в умеренной связи с набором включенных в уравнение логарифмической регрессии факторных переменных.

Для оценки статистической значимости модели в целом по формуле (6.14) рассчитаем значение F-критерия Фишера:

$$F = \frac{0{,}16}{1-0{,}16} \cdot \frac{26-4-1}{4} = 0{,}998.$$

Сравнивая его с критическим значением (2,84), следует признать уравнение степенной регрессии статистически неадекватным.

Итак, на основе проведенной оценки статистической значимости представленных уравнений множественной регрессии можно сделать вывод, что ни одно из уравнений не является статистически значимым. На практике это означает, что в целях улучшения модели необходимо включить в анализ дополнительные факторные переменные.

Задача 6.2. На основе имеющихся данных осуществите поиск оптимального уравнения линейной множественной регрессии, описывающего зависимость результативной переменной y от предлагаемого набора факторов, найдите коэффициенты регрессии. На основе показателей статистической адекватности выявите наиболее адекватную изучаемому явлению эконометрическую модель.

y — фактическое потребление молочных продуктов на 1 человека в год, кг;

x_1 — среднемесячная заработная плата работающих, руб.;

x_2 — средний размер месячных пенсий всех пенсионеров, руб.;

x_3 — производство молока (заданной жирности), тыс. т;

Таблица 6.1

Данные к задаче в разрезе муниципальных районов Ставропольского края за 2010 г.

№ района	y	x_1	x_2	x_3	x_4	x_5
1	188,2	3685	2907,1	17,5	7,6	2773
2	189,9	5460	3006,3	18,4	7,9	1990
3	190,4	2803	3133	20,1	10,8	2268
4	189,3	3164	2878,7	10,4	5,6	2041
5	190,8	3892	2943,5	22,0	5,2	3801
6	191,1	2831	2998,6	16,9	4,6	2867
7	205,8	4831	3062,9	16,6	3,9	3957
8	190,9	3932	2904,9	12,8	4,4	1753
9	191,2	5584	3228,2	24,9	5,1	3981
10	190,2	3749	2727,3	32,7	12,7	3646
11	192,6	5074	3061,4	17,7	4,3	3862
12	192	4644	2967,2	57,5	9,8	5208
13	191,3	3861	2977,4	18,8	4,3	4787
14	193,8	3530	2710,6	14,0	6,9	2057
15	191,4	3395	2966,1	19,6	13,6	2638
16	191,1	4713	3162,8	14,4	4,5	1382
17	192,1	5620	2860,3	17,0	10,3	2288
18	189,7	4003	2978	30,4	5,8	5314
19	192,6	3475	2879,6	16,5	4,3	3656
20	190,6	4317	2922,5	21,9	6,7	3283
21	193,4	5260	2986,2	30,9	6,6	4681
22	190,8	3832	2990,6	25,9	5,1	4349
23	191,5	3163	2804,7	14,8	4,1	2410
24	189,3	4275	2996,1	29,2	5	4940
25	190,4	3139	2882,2	12,7	7	3568
26	197,6	5186	2933,1	24,9	6,8	2680

x_4 — численность коров (в том числе) во всех категориях хозяйств на начало года, тыс. голов;

x_5 — средний надой молока на одну корову в сельхозпредприятиях, кг.

Решение

Для начала оценим параметры и статистическую значимость линейной множественной регрессии, включающей весь набор предлагаемых факторных признаков x. Уравнение имеет вид $y_x = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + a_5 x_5$.

Используя инструмент «Множественная регрессия», получим уравнение $y_x = 192{,}15 + 0{,}0016 \cdot x_1 - 0{,}0021 \cdot x_2 - 0{,}048 \cdot x_3 - 0{,}21 \cdot a_4 x_4 + 0{,}0005 \cdot x_5$.

	Итоги регрессии для зависимой переменной: y (Таблица к задаче R= ,38568303 R2= ,14875140 Скорректир. R2= — F(5,20)=,69898 p<,63054 Станд. ошибка оценки: 3,4743					
N=26	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(20)	**р-уров.**
Св.член			192,1493	18,71015	10,26979	0,000000
x1	0,406591	0,270548	0,0016	0,00105	1,50284	0,148507
x2	-0,072127	0,237968	-0,0021	0,00679	-0,30310	0,764947
x3	-0,135078	0,522004	-0,0478	0,18482	-0,25877	0,798459
x4	-0,169012	0,517514	-0,2098	0,64245	-0,32658	0,747375
x5	0,204959	0,286557	0,0005	0,00074	0,71525	0,482723

Рис. 6.5. Итоги множественной регрессии

Обратим внимание на значение множественного коэффициента корреляции, указанного в окне итогов регрессии $R = 0{,}39$, что свидетельствует об умеренной степени связи между полным набором факторных и результативной переменными.

Очевидно, что полученная модель требует более детального рассмотрения. Используя инструмент «Парные и частные корреляции» из меню «Анализ», найдем парные коэффициенты корреляции (рис. 6.6).

Анализируя полученную матрицу, отметим, что независимые переменные x_3 и x_5 не следует включать в модель совместно, поскольку они имеют заметную корреляционную связь между собой. Поэтому одну из этих переменных необходимо исключить из рассмотрения для выполнения предпосылки МНК о независимости переменных, включаемых в модель в качестве внешних.

Переменная	Корреляции (Таблица данных13) Отмеченные корреляции значимы на уровне p <,05000 N=26 (Построчное удаление ПД)					
	y	x1	x2	x3	x4	x5
y	1,00	0,32	0,10	-0,04	-0,19	0,07
x1	0,32	1,00	0,36	0,28	-0,05	0,15
x2	0,10	0,36	1,00	0,08	-0,23	0,15
x3	-0,04	0,28	0,08	1,00	0,33	0,66
x4	-0,19	-0,05	-0,23	0,33	1,00	-0,13
x5	0,07	0,15	0,15	0,66	-0,13	1,00

Рис. 6.6. Парные коэффициенты корреляции

На следующем этапе необходимо решить вопрос, какую из переменных следует исключить из модели. Количественными характеристиками для принятия такого решения являются парные коэффициенты корреляции между этими переменными и результативной переменной y. Исходя из полученных значений можно сделать вывод об исключении переменной x_5 из представленной модели.

Таким образом, необходимо оценить коэффициенты и статистическую значимость уравнения регресси вида

$$y_x = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4.$$

Сделать это можно с помощью инструмента «Множественная регрессия». В результате получим уравнение

$y_x = 192{,}91 + 0{,}0014x_1 - 0{,}0016x_2 - 0{,}028x_3 - 0{,}2x_4$ (рис. 6.7).

	Итоги регрессии для зависимой переменной: y (Таблица данных13) R= ,37743917 R2= ,14246032 Скорректир. R2= --- F(4,21)=,87217 p<,49703 Станд. ошибка оценки: 3,4031					
N=26	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(21)	p-уров.
Св.член			192,9067	18,37115	10,50052	0,000000
x1	0,354175	0,225474	0,0014	0,00088	1,57080	0,131175
x2	-0,057022	0,223360	-0,0016	0,00637	-0,25529	0,800980
x3	-0,078072	0,226107	-0,0276	0,08005	-0,34529	0,733315
x4	-0,162817	0,223524	-0,2021	0,27749	-0,72841	0,474407

Рис. 6.7. Результаты оценки уравнения регрессии

$$y_x = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$$

На рисунке 6.8 представлены результаты идентификации полученного уравнения регрессии. На основе множественного коэффициента корреляции $R = 0{,}35$ можно сделать вывод об умеренной силе связи между исследуемыми переменными. Множественный коэффициент детерминации $R^2 = 0{,}12$ свидетельствует о том, что вариация зависимой переменной y на 12% объясняется совокупным воздействием факторов x_1, x_2, x_3, x_4. Фактическое значение F-критерия Фишера меньше критического, что позволяет говорить о статистической незначимости полученного уравнения регрессии.

	Итоговые статистики
Статистика	Значение
Множест. R	0.347554
Множест R2	0.120794
Скорр. R2	0.000902
F(3,22)	1.007524
p	0,408151
Стд Ош. Оценки	3,366593

Рис. 6.8. Идентификация уравнения регрессии
$y_x = 192{,}91 + 0{,}0014x_1 - 0{,}0016x_2 - 0{,}028x_3 - 0{,}2x_4$

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 6.3. Рассчитайте множественный индекс корреляции для прямолинейной зависимости при двадцати шести наблюдениях, если известно, что $\sum(x - x_i)^2 = 45$, $\sum x^2 = 5$, $\sum yx = 95$, $\sum(y - y_i)^2 = 48$, $\sum y^2 = 320$, $\sum y^2x^2 = 115$, остаточная дисперсия результативного признака равна 7,6, а общая дисперсия — 15,2. Дайте характеристику силе связи.

Задача 6.4. Рассчитайте множественный индекс детерминации для криволинейной зависимости при тридцати узловых точках, если известно, что $\sum(x - x_i)^2 = 55$, $\sum x^2 = 85$, $\sum yx = 115$, $\sum(y - y_i)^2 = 38$, $\sum y^2 = 187$, $\sum y^2x^2 = 115$, остаточная дисперсия результативного признака равна 5,6, а общая дисперсия — 15,2. Сделайте вывод относительно полученного результата.

Задача 6.5. Рассчитайте индекс корреляции для множественной криволинейной зависимости при двадцати наблюдениях, если известно, что $\sum(x - x_i)^2 = 35$, $\sum x^2 = 65$, $\sum yx = 95$, $\sum(y - y_i)^2 = 68$, $\sum y^2 = 157$, $\sum y^2x^2 = 145$, остаточная сумма квадратов отклонений равна 5,6, а общая дисперсия результативного признака — 12,2. Дайте характеристику силе связи.

Задача 6.6. Рассчитайте индекс детерминации для множественной криволинейной зависимости при двадцати четырех узловых точках, если известно, что $\sum(x - x_i)^2 = 65$, $\sum x^2 = 45$, $\sum yx = 95$, $\sum(y - y_i)^2 = 58$, $\sum y^2 = 187$, $\sum y^2x^2 = 115$, остаточная сумма квадратов отклонений равна 4,6, а общая дисперсия результативного признака — 14,2. Сделайте вывод относительно полученного результата.

Задача 6.7. Оцените значимость индекса детерминации на основе F-критерия Фишера, если известно, что индекс множественной корреляции для двухфакторной модели равен 0,7, индекс детерминации — 0,49, число узловых точек — 24, относительная ошибка аппроксимации — 8,7%, факторная дисперсия результативного признака — 3,6.

Задача 6.8. Оцените значимость индекса детерминации на основе F-критерия Фишера, если известно, что индекс множественной корреляции для трехфакторной модели равен 0,8, индекс детерминации — 0,64, число узловых точек — 28, относительная ошибка аппроксимации — 8,7%, факторная дисперсия результативного признака — 3,6.

Задача 6.9. Оцените значимость индекса детерминации на основе F-критерия Фишера, если известно, что индекс множественной корреляции для двухфакторной модели равен 0,6, индекс детерминации — 0,36, число наблюдений — 22, относительная ошибка аппроксимации — 9,7%, факторная дисперсия результативного признака — 5,6. Сделайте вывод.

Задача 6.10. Оцените значимость индекса детерминации на основе F-критерия Фишера, если известно, что индекс множественной корреляции для трехфакторной модели равен 0,9, индекс детерминации — 0,81, число наблюдений — 31, относительная ошибка аппроксимации — 9,7%, факторная дисперсия результативного признака — 5,6. Сделайте вывод.

Задача 6.11. Определите значение скорректированного индекса множественной корреляции для двухфакторной модели, построенной по 27 наблюдениям, если известно, что индекс множественной корреляции равен 0,67, индекс детерминации — 0,45.

Задача 6.12. Определите значение скорректированного индекса множественной корреляции для двухфакторной модели, построенной по 24 наблюдениям, если известно, что индекс множественной корреляции равен 0,77, индекс детерминации — 0,59.

Задача 6.13. Определите значение скорректированного индекса множественной корреляции для трехфакторной модели, построенной по 34 наблюдениям, если известно, что индекс множественной корреляции равен 0,7, индекс детерминации — 0,49.

Задача 6.14. Рассчитайте стандартизированные коэффициенты регрессии (β-коэффициенты) для двухфакторной модели, построенной по 30 наблюдениям, если известно, что свободный член модели равен 13,4, параметр a_0= 1,9, a_1= –0,98, факторная дисперсия – 5,8, общая дисперсия — 13,4, остаточная дисперсия результативного признака — 7,6, среднее квадратичное отклонение фактора x_1 = 2,7, а среднее квадратичное отклонение фактора x_2= 1,9.

Задача 6.15. Рассчитайте стандартизированные коэффициенты регрессии (β-коэффициенты) для двухфакторной модели, построенной по 30 наблюдениям, если известно, что свободный член модели равен 4,8, параметр a_0= 1,9, a_1= –0,98, факторная дисперсия — 3,8, общая дисперсия — 9,4, остаточная дисперсия результативного признака — 5,6, среднее квадратичное отклонение фактора x_1= 1,7, а среднее квадратичное отклонение фактора x_2= 1,1.

Задача 6.16. Для двухфакторной модели определите частные коэффициенты регрессии, если известно, что $r_{yx_1}=0{,}58$, $r_{yx_2}=0{,}32$, $r_{x_1x_2}=0{,}17$, $r^2_{yx_2}=0{,}1$, $r^2_{x_1x_2}=0{,}03$, $r^2_{yx_1}=0{,}34$. Сделайте выводы.

Задача 6.17. Для двухфакторной модели определите частные коэффициенты регрессии, если известно, что $r_{yx_1}=0{,}48$, $r_{yx_2}=0{,}32$, $r_{x_1x_2}=0{,}27$, $r^2_{yx_2}=0{,}1$, $r^2_{x_1x_2}=0{,}07$, $r^2_{yx_1}=0{,}23$. Сделайте выводы.

Задача 6.18. Рассчитайте среднюю ошибку аппроксимации для уравнения множественной регрессии, построенного по 46 наблюдениям, если известно, что абсолютная ошибка аппроксимации равна 125,3, относительная ошибка аппроксимации — 72,3 %. Сделайте выводы.

Задача 6.19. Рассчитайте среднюю ошибку аппроксимации для уравнения множественной регрессии, построенного по 32 наблюдениям, если известно, что абсолютная ошибка аппроксимации равна 325,3, относительная ошибка аппроксимации — 86,3 %. Сделайте выводы.

Задача 6.20. По данным задачи 4.10 осуществить идентификацию представленных уравнений парной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод о наиболее приемлемой форме модели для описания исходной зависимости изменения индекса физического объема инвестиций (*Y*) и среднесложившихся цен на зерно (*X*). Для выбранной формы модели оцените доверительные интервалы теоретических значений результативной переменной с вероятностью 95%.

Таблица 6.2

Данные об индексе физического объема инвестиций и среднесложившиеся цены на зерно в 2012 г. в разрезе районов Ставропольского края

№ района	Индекс физического объема инвестиций, в % к предыдущему году	Среднесложившиеся цены на зерно, реализованное по всем каналам, руб./ц
	Y	*X*
1	200,0	648,6
2	47,6	612,9
3	68,9	549,8
4	96,8	645,5
5	26,4	694,1
6	143,5	686,3
7	177,8	689,6
8	49,6	683,2
9	127,2	747,2
10	84,6	673,7
11	269,7	643,2
12	68,5	751,7
13	161,4	787,9
14	98,5	591,9

Задача 6.21. По данным задачи 4.11 осуществить идентификацию представленных уравнений парной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод о наиболее приемлемой форме модели для описания исходной зависимости изменения прибыли сельхозпредприятий (*Y*) и соотношения производственной себестоимости зерна к средней по краю (*X*). Для выбранной формы модели оцените доверительные интервалы теоретических значений результативной переменной с вероятностью 99%.

Таблица 6.3

Данные о соотношении производственной себестоимости зерна к средней по краю и прибыли (убытка) сельхозпредприятий в 2012 г. в разрезе районов Ставропольского края

№ района	Соотношение производственной себестоимости 1 ц зерна к средней по краю, %	Прибыль (убыток) сельхозпредприятий, млн руб.
	X	*Y*
1	116,38	328,1
2	181,36	2,6
3	108,19	309,2
4	117,23	244,1
5	92,09	234,5
6	107,06	414,5
7	125,99	608,5
8	100,00	20,8
9	100,85	299,3
10	148,31	567,2
11	94,35	241,8
12	94,92	112,5

Задача 6.22. По данным задачи 4.12 осуществить идентификацию представленных уравнений парной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод о наиболее приемлемой форме модели для описания исходной зависимости изменения обеспеченности населения мясными продуктами (*Y*) и объемах производства всех видов мяса в сельхозпредприятиях (*X*). Для выбранной формы модели оцените доверительные интервалы теоретических значений результативной переменной с вероятностью 95%.

Таблица 6.4

Данные об обеспеченности мясными продуктами населения региона и о производстве мяса всех видов в разрезе районов Ставропольского края в 2012 г.

№ района	Коэффициент обеспеченности мясными продуктами населения	Производство мяса всех видов во всех категориях хозяйств, т (в живом весе)
	Y	X
1	0,654	6264,2
2	0,660	5512,3
3	0,658	11251
4	0,663	7300
5	0,664	5824
6	0,670	8986
7	0,665	24855
8	0,674	5,71
9	0,665	9073
10	0,664	9729,3
11	0,668	4537
12	0,672	53982
13	0,663	13851
14	0,666	7260
15	0,662	13053

Задача 6.23. Осуществить идентификацию уравнений парной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод о наиболее приемлемой форме модели для описания исходной зависимости изменения стоимости сельскохозяйственной продукции СХО (Y) и среднегодовой стоимости основных фондов (X). Для выбранной формы модели оцените доверительные интервалы теоретических значений результативной переменной с вероятностью 90%.

Таблица 6.5

Данные о среднегодовой стоимости основных фондов и стоимости сельскохозяйственной продукции СХО в 2012 г. в разрезе районов Ставропольского края, млн руб.

№ района	Среднегодовая стоимость основных фондов СХО	Объем сельскохозяйственной продукции СХО
	X	*Y*
1	718,8	141,9
2	1216,8	64,1
3	864,2	145,7
4	580,4	89,3
5	549,8	49,6
6	1250,0	116,3
7	1579,3	110,0
8	2052,6	66,9
9	894,0	138,6
10	886,7	159,5
11	495,6	129,0
12	321,0	162,6
13	1377,7	87,5
14	520,4	84,3

Задача 6.24. Осуществить идентификацию уравнений парной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод о наиболее приемлемой форме модели для описания исходной зависимости изменения производства мяса всех видов (*Y*) и заготовки кормов в сельхозпредприятиях (*X*). Для выбранной формы модели оцените доверительные интервалы теоретических значений результативной переменной с вероятностью 90%.

Таблица 6.6

Данные о производстве мяса всех видов и заготовке кормов в сельхозпредприятиях в разрезе районов Ставропольского края в 2012 г.

№ района	Производство мяса всех видов во всех категориях хозяйств, т (в живом весе)	Заготовлено кормов в сельхозпредприятиях в расчете на одну условную голову скота, ц корм.ед.
	Y	X
1	4038,0	22,0
2	5791,0	25,4
3	6580,0	22,8
4	4078,0	29,1
5	9322,0	24,5
6	5748,0	23,4
7	13195,0	23,9
8	7786,0	32,7
9	3586,0	37,1
10	6510,0	26,0
11	10307,0	24,5
12	8718,0	27,4
13	7105,0	45,0
14	9526,0	24,2
15	4844,0	22,0
16	11685,0	22,8

Задача 6.25. По данным задачи 4.13 осуществить идентификацию полученного линейного уравнения множественной регрессии. В соответствии с величиной средней ошибки аппроксимации сделать вывод об адекватности модели для описания исходной зависимости, оценить доверительные интервалы теоретических значений результативной переменной с вероятностью 99%.

Таблица 6.7

Данные о производстве мяса, заготовке кормов и падеже КРС по СХП в разрезе районов Ставропольского края в 2012 г.

№ района	Производство мяса (в живом весе) всех видов во всех категориях хозяйств, т	Заготовлено кормов в СХП в расчете на одну условную голову скота, ц корм. ед.	Падеж крупного рогатого скота в СХП, %
	Y	X_1	X_2
1	6264,2	25	0,5
2	5512,3	12,5	3,2
3	11251	17,4	1,3
4	7300	14,9	2,1
5	5824	23	1,1
6	8986	22,9	2,0
7	24855	11,85	2,3
8	5713	12,7	0,5
9	9073	14	2,6
10	9729,3	22,9	0,7
11	4537	16,9	1,2
12	53982	16	2,7
13	13851	25	2,9
14	7260	9,85	1,3
15	13053	26	0,5
16	4527	21,7	2,9
17	15464	24,3	2,9
18	6297	41,6	0,6

ГЛАВА 7 ЭКОНОМЕТРИЧЕСКИЙ АНАЛИЗ ВОСПРОИЗВОДСТВЕННОГО ПРОЦЕССА

7.1. Анализ производства и издержек

Эконометрический анализ производства и издержек основывается на изучении причинных связей между объемом конечного продукта и производственными факторами. Производственная функция обобщает процесс производства, при этом необходимо различать *экономические* и *технологические* производственные функции. Технологические производственные функции описывают конкретные этапы производства, экономическая производственная функция, напротив, принимает во внимание лишь экономические аспекты. Разные технологические процессы, зависящие от одинакового набора факторов, с экономической точки зрения могут считаться равноценными. Отсюда следует, что при экономическом анализе производства его технологические особенности не являются решающими. С другой стороны, технологический процесс все же накладывает ограничения на выбор экономических решений.

Обозначая продукт (конечный результат) через y, а производственные факторы через x_i ($i=1, 2,\ldots, n$), связь между ними можно представить в следующем виде:

$$y=f(x_1,x_2,\ldots,x_n). \tag{7.1}$$

Это выражение называется *производственной функцией*, оно характеризует связь между производственными факторами и величиной продукта.

При построении производственной функции решающими являются выбор производственных факторов и определение размерности входящих величин, так, чтобы можно было численно оценить влияние производственных факторов на объем продукта. Вид производственной функции может быть неизменным или, напротив, меняться, если изменения в производстве ведут к нарушению установленных связей между производственными факторами и его объемом. Выбор производственных факторов в производственной функции зависит от задач исследования. Следует рассматривать лишь те факторы, о которых имеется статистическая информация.

Зависимой переменной в производственной функции является объем продукта (товаров или услуг). Это может быть валовой или чистый продукт или доход. Выбор зависимой переменной определяется целями анализа и статистическими возможностями.

Часто рассматривают производственную функцию, в которой зависимая переменная определяется двумя факторами: основные производственные фонды (F) и рабочая сила (Z), т.е. строится производственная функция вида

$$y = f(F, Z). \tag{7.2}$$

Измерение фактора F вызывает на практике некоторые трудности: статистика описывает состояние основных фондов на моменты времени, это состояние зависит от срока службы основных фондов, интенсивности их использования и износа и т.д.

Фактор Z представляет собой затраты живого труда и оценивается обычно числом занятых. Также используются данные об отработанных часах или заработной плате. Проблемы выражения фактора Z возникают лишь при измерении квалификации рабочей силы.

На практике разрабатываются производственные функции, в которые помимо указанных независимых переменных F (основные фонды) и Z (рабочая сила) включаются и другие факторы:

1) уровень производства в предыдущем периоде, Y_{t-1}. Этот фактор вытекает из технологических особенностей производства и отражает непрерывность производственного процесса. Фактор Y_{t-1} имеет важное значение при краткосрочном и среднесрочном анализе;

2) действие НТП выражается обычно определенным типом трендовой переменной, этот фактор следует принимать во внимание при среднесрочном и долгосрочном анализе;

3) переменные, корректирующие действие основных факторов, как, например, индекс использования основных средств производства, индекс квалификации рабочей силы;

4) влияние производства смежных отраслей, влияние импорта. Эти факторы имеют особое значение, если в качестве зависимой переменной рассматривается производство конкретного вида продукта или валовой национальный доход;

5) влияние специфических факторов в отдельных отраслях, например в сельском хозяйстве: обеспеченность кормами, удобрениями, индекс, учитывающий погодно-климатические условия, и т.д.

7.2. Типы производственных функций

В настоящее время известен целый ряд производственных функций и изучены их свойства, однако нет однозначных правил выбора их специфических форм. Свойства и вид производственных функций зависят прежде всего от цели, задачи и условий, в которых функция определяется. С этой точки зрения множество всех производственных функций можно разделить по количеству используемых переменных, по виду функций и по их свойствам.

По *количеству переменных* производственные функции делятся на:

- однофакторные, которые выражают, например, зависимость производства либо от одних основных средств производства, либо только от рабочей силы;
- двухфакторные, которые обычно рассматривают зависимость конечного продукта от основных фондов и рабочей силы одновременно;
- многофакторные, которые, кроме указанных выше двух факторов F и Z, учитывают влияние на производство других факторов, таких как, например, технический процесс, и т.д.

По *аналитическому виду* производственные функции можно разделить следующим образом:

а) линейные производственные функции с разным числом факторов; например, линейная производственная функция вида

$$y = a_0 + a_1 F + a_2 Z. \tag{7.3}$$

Параметр a_0 – свободный член, параметры a_1 и a_2 выражают производительность факторов F и Z, т.е. показывают абсолютный прирост производства, когда один из факторов остается неизменным, а другой возрастает на единицу. Линейные функции часто применяют в краткосрочных и среднесрочных эконометрических моделях;

б) степенные производственные функции, среди которых наиболее известна производственная функция Кобба–Дугласа (CDPF). Эта функция имеет вид

$$y = aF^{\alpha}Z^{\beta}. \tag{7.4}$$

Параметры α и β выражают эластичность уровня производства y по отношению к факторам F и Z, т.е. показывают относительный прирост продукции (в процентах), связанный с относительными приростами основных фондов и рабочей силы;

в) более сложные типы производственных функций, среди которых наиболее известна функция с постоянной эластичностью замены, так называемая функция *CES*. Она имеет следующий вид:

$$y = a\left[bF^{-S} + (1-b)Z^{-S}\right]^{-1/S}, \tag{7.5}$$

где S выражает эластичность основных фондов и занятости.

Производственные функции различают по их *свойствам*, связанным с такими характеристиками, как:

а) *однородность производственных функций*. Данное понятие включает в себя следующее свойство: равномерное увеличение всех производственных факторов вызывает пропорциональное увеличение продукта. Это свойство можно выразить математически:

функция $f(x_1, x_2, ..., x_n)$ однородна в степени h, если

$$f(\lambda x_1, \lambda x_2, ..., \lambda x_n) = \lambda^h f(x_1, x_2, ..., x_n).$$

Таким образом, каждая независимая переменная принимает значение $\lambda x_i (i = 1, 2, \ldots, n)$, значение функции возрастает в l^h раз. Величина h показывает степень использования производственных факторов или их эффективность. Когда $h = 1$, эффективность производства будет равна 1; при $h > 1$ говорят, что производственные факторы обладают растущей эффективностью; при $h < 1$ эффективность факторов снижается;

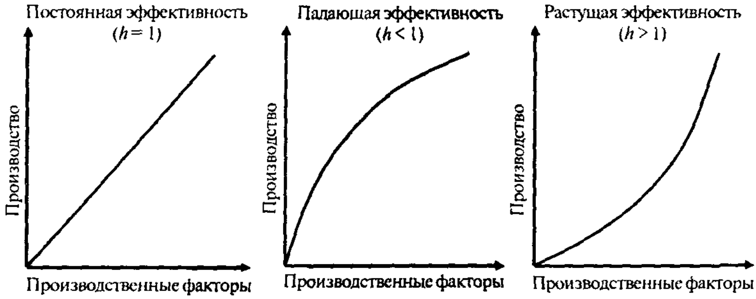


Рис. 7.1. Три типа однородной производственной функции

б) *предельная производительность* факторов. Наряду с часто используемыми в экономическом анализе рядами, содержащими абсолютные и относительные величины, применяют показатель *предельная величина*. Смысл этой величины и ее связь с абсолютными и относительными (или средними) значениями можно понять из следующего примера (табл. 7.1).

Пусть имеется ряд значений числа занятых (Z) и соответствующие значения валового продукта (Y). Относительный продукт (или средняя производительность труда) рассчитывается как отношение валового продукта к числу занятых (Y/Z).

Предельный продукт (предельная производительность труда) определяется как отношение прироста валового продукта (ΔY) к приросту рабочей силы (ΔZ), т.е.

$$M = \frac{\Delta Y}{\Delta Z}, \tag{7.6}$$

где $\Delta Y = Y_t - Y_{t-1}$; $\Delta Z = Z_t - Z_{t-1}$.

Таблица 7.1

Значение валового, относительного и предельного продукта

Число занятых, млн чел (Z)	Валовой продукт, млрд руб. (Y)	Относительный продукт (средняя производительность труда) (Y/Z)	Предельный продукт (предельная производительность труда) (M)
0	0	0	0
1	5	5	5
2	13	6,5	8
3	23	7,67	10
4	38	9,5	15
5	50	10	12
6	60	10	10
7	68	9,71	8
8	75	9,38	7
9	81	9	6
10	86	8,6	5
11	90	8,18	4

Из таблицы видно, что при увеличении числа занятых (Z) валовой продукт (Y) растет, средняя производительность труда растет до тех пор, пока число занятых не достигло 5 млн человек, затем некоторое время держится на уровне, а при дальнейшем увеличении числа занятых (Z) падает. Предельный продукт (M) ведет себя аналогично Y/Z, но быстрее достигает максимального уровня;

в) *эластичность* объема производства по отношению к производственным факторам. Эластичностью экономического показателя называется его способность реагировать в большей или меньшей степени на изменение другого показателя.

Эластичность объема производства (Y) по некоторому фактору определяется как отношение темпов прироста Y к темпам прироста этого фактора. Например, коэффициент эластичности Y по основным фондам определяется как

$$E_{Y,F} = \lim \frac{\dfrac{\Delta Y}{Y}}{\dfrac{\Delta F}{F}}. \qquad (7.7)$$

Отсюда следует:

$$E_{Y,F}=\lim\frac{\frac{dY}{Y}}{\frac{dF}{F}}=\frac{F}{Y}\cdot\frac{dY}{dF}. \tag{7.8}$$

Поскольку часто рассматриваются производственные функции, содержащие несколько факторов, то следует исследовать эластичность по всем этим факторам. В этом случае вводится понятие частной эластичности. Для двухфакторной производственной функции вида $Y=f(F,Z)$ частные коэффициенты эластичности по факторам F и Z определяются следующим образом:

$$E_{Y.F}=\frac{F}{Y}\cdot\frac{dY}{dF}, \tag{7.9}$$

$$E_{Y.Z}=\frac{Z}{Y}\cdot\frac{dY}{dZ}, \tag{7.10}$$

где $\frac{dY}{dF}$ и $\frac{dY}{dZ}$ — частные производные функции двух переменных;

г) ***замещение факторов***. Понятие ***взаимозаменяемость*** основывается на предположении, что производственные факторы могут замещать друг друга, и показывает, как при неизменной величине продукта можно изменять соотношение между факторами. Так, например, в случае производственной функции с двумя независимыми переменными F и Z можно поставить вопрос, насколько должно изменяться число занятых при некотором изменении объема основных фондов, чтобы величина произведенного продукта осталась неизменной.

Оценка заменяемости F и Z называется ***предельной нормой замены*** и определяется в виде следующего отношения:

$$S=\frac{\frac{dY}{dZ}}{\frac{dY}{dF}}=\frac{dF}{dZ}. \tag{7.11}$$

На рисунке 7.2 изображены ***изокванты*** производственной функции. На осях F и Z откладывается множество возможных значений производственных факторов. Каждая точка в пространстве

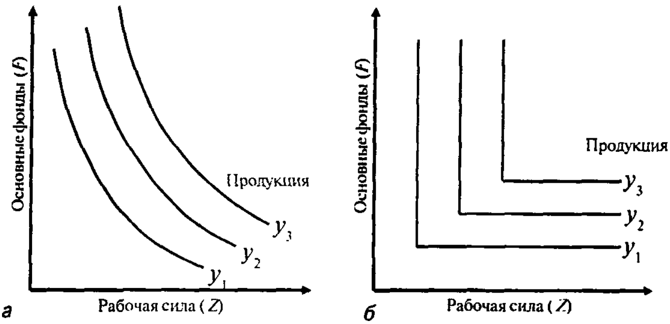


Рис. 7.2. Изокванты для двух типов производственных функций:
а — производственная функция с взаимозаменяемыми факторами;
б — производственная функция с дополняющими факторами

переменных Y, F, Z показывает значение продукта, произведенного с помощью комбинации факторов F и Z.

Изокванты производственной функции, как правило, выпуклы, а S растет с увеличением F и уменьшением Z (рис. 7.2, *а*). Последнее отражает тот факт, что со временем становится все труднее заменять Z фактором F.

Если производственные факторы можно заменить лишь в фиксированных пропорциях, то говорят, что производственная функция обладает нулевой предельной нормой замены (рис. 7.2, *б*).

7.3. Производственная функция Кобба–Дугласа

Функция Кобба–Дугласа (*CDPF*) принадлежит к наиболее известным, широко применяемым производственным функциям. Классическая производственная функция Кобба–Дугласа имеет вид

$$y = aF^{\alpha}Z^{1-\alpha}, \qquad (7.12)$$

где все параметры a, α, $1-\alpha$ положительны.

Сумма параметров классической *CDPF* или степень ее однородности равна 1. Это означает, что при увеличении обоих производственных факторов на 1 конечный продукт тоже возрастет на 1. Таким образом, эффективность производственных факторов в этом случае постоянна.

Практические исследования *CDPF* показали, что предположения о линейной однородности в реальной экономике выполняются редко. Поэтому была предложена производственная функция общего вида:

$$y = aF^{\alpha}Z^{\beta}. \tag{7.13}$$

Сумма параметров ($\alpha + \beta$), в отличие от предыдущего случая, может быть как меньше, так и больше единицы. Если сумма параметров больше 1, то говорят, что темпы роста продукции выше темпов роста факторов, и, наоборот, если сумма параметров меньше 1, то темпы роста продукции ниже темпов роста факторов (это утверждение верно, если темпы роста производственных факторов одинаковы).

Предположим, что каждый производственный фактор вырос на $m\%$, тогда значения этих факторов будут равны соответственно:

$$F(1+\frac{m}{100}) \text{ и } Z(1+\frac{m}{100}). \tag{7.14}$$

Величина конечного продукта вычисляется следующим образом:

$$y = a\left[F(1+\frac{m}{100})\right]^{\alpha}\left[Z(1+\frac{m}{100})\right]^{\beta} =$$

$$= aF^{\alpha}Z^{\beta}(1+\frac{m}{100})^{\alpha}(1+\frac{m}{100})^{\beta} = aF^{\alpha}Z^{\beta}(1+\frac{m}{100})^{\alpha+\beta}. \tag{7.15}$$

Следовательно, при $\alpha + \beta > 1$ конечный продукт возрастает больше, чем на $m\%$, при $\alpha + \beta < 1$ — меньше, чем на $m\%$, и при $\alpha + \beta = 1$ конечный продукт возрастает ровно на $m\%$.

Производственная функция Кобба–Дугласа нелинейна, однако легко приводится к линейному выражению путем логарифмирования:

$$\lg(y) = \lg(aF^{\alpha}Z^{\beta}),$$

$$\lg(y) = \lg a + \alpha\lg F + \beta\lg Z. \tag{7.16}$$

Параметры производственной функции в линейной форме можно оценивать на основе как временных, так и структурных (пространственных) рядов.

Производственная функция Кобба–Дугласа позволяет учитывать динамику с помощью множителя e^{rt}, в этом случае она имеет вид

$$y_t = aF_t^{\alpha} Z_t^{\beta} e^{rt}, \qquad (7.17)$$

где e – основание натурального логарифма;

t – число лет;

r – годовая норма автономного роста выпуска продукции, обусловленного иными факторами, чем количество труда и накопление основного капитала (масштаб производства, технический прогресс, квалификация и т.д.).

7.4. Функции издержек

Предметом эконометрического анализа издержек является изучение влияния издержек производства на объем производства и другие его технико-экономические показатели. При этом могут приниматься во внимание как *общие издержки* производства, так и *относительные* (или удельные), которые можно определить как отношение суммарных издержек к объему производства.

Статистические данные в анализе издержек должны удовлетворять следующим условиям:

1) удельные издержки должны определяться фактическими издержками производственных факторов. При этом особое значение имеет правильное распределение издержек во времени;
2) данные об издержках должны на разных временных интервалах быть соизмеримыми (выражены, например, в одинаковых ценах);
3) статистические данные об издержках не должны содержать в себе издержек, которые связаны с производством некондиционных товаров.

Чаще всего рассматривается функция издержек следующего вида:

$$N = f(Y) + \sum_{i=1}^{n} a_i x_i, \qquad (7.18)$$

где N — суммарные издержки;

Y — количество произведенной продукции;

x_i — другие переменные.

Если поделить обе части уравнения на Y, получим функцию издержек в следующей форме:

$$Q = \frac{1}{Y}\left[f(Y) + \sum_{i=1}^{n} a_i x_i\right], \qquad (7.19)$$

где $Q = \frac{N}{Y}$ — удельные издержки.

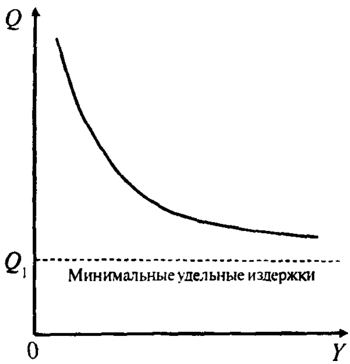


Рис. 7.3. Связь между удельными издержками (Q) и объемом продукции (Y)

Если отвлечься от влияния факторов x_i, то имеет смысл рассмотреть следующие функции издержек:

$$N = f(Y); \qquad (7.20)$$

$$Q = \frac{f(Y)}{Y}. \qquad (7.21)$$

Линейная функция издержек имеет вид

$$N = a_0 + a_1 Y.$$

Функция удельных издержек в этом случае будет убывающей. График этой функции приведен на рисунке 7.3.

7.5. Анализ спроса и предложения

Целью эконометрических моделей является изучение основных зависимостей процесса воспроизводства. Это означает, что отдельные проблемы, связанные со спросом и предложением, могут быть исследованы с помощью этих моделей. Предложение определяется сферой материального производства и может быть изучено на основе производственных функций. Спрос, напротив, определяется в сфере конечного потребления и зависит от величины дохода, уровня цен и т.п.

Функция спроса выражает зависимость спроса от экономических (доходы, цены) и внеэкономических (потребительские привычки) факторов. Функции спроса могут быть *макроэкономическими*, если они охватывают всю сферу потребления, и *микроэкономическими*, описывающими спрос индивидуальных потребителей. При разработке функций учитываются демографические и социальные аспекты.

Впервые определение функции спроса дал французский экономист А. Курно. Его функция спроса D показывала зависимость спроса от цены p, т.е.

$$D = f(p). \tag{7.22}$$

Функция убывает с ростом p (цены). Аналогично была определена и функция предложения:

$$S = \varphi(p). \tag{7.23}$$

Общий для обеих функций фактор p оказывает в них противоположное влияние, и кривые предложения и спроса движутся в противоположных направлениях. Точка их пересечения определяет $\bar{p}$, называемое рыночной или равновесной ценой (рис. 7.4).

Часто товары классифицируют по степени реакции спроса на них в зависимости от изменения цен. В этом случае говорят, что товары обладают более высокой или низкой *эластичностью* спроса.

Существуют факторы, а также их комбинации, относительно которых имеет смысл изучать эластичность спроса. Чаще всего рассматривается эластичность спроса по следующим трем факторам:

а) эластичность спроса по цене (обычно с отрицательным знаком);

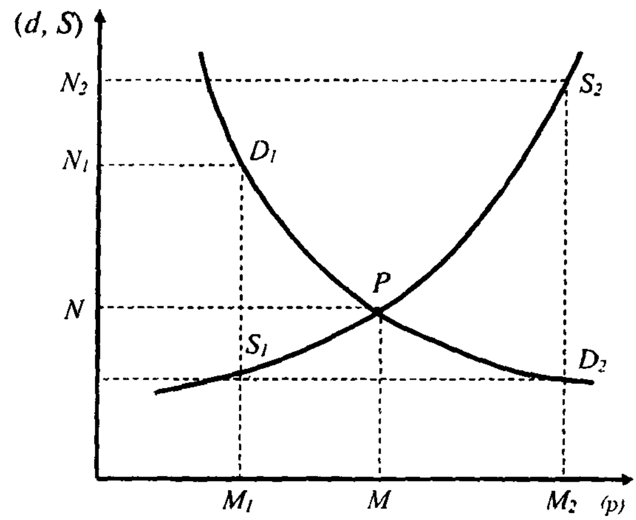


Рис. 7.4. Функции спроса *(D)* и предложения *(S)* в зависимости от цены *(p)*

б) эластичность спроса по доходу (обычно с положительным знаком);

в) перекрестная эластичность. Эта эластичность спроса данного вида товаров в зависимости от цены другого вида товаров (может быть как положительной, так и отрицательной).

Рассмотрим на следующем примере определение эластичности спроса. Предположим, что цена товара, равная в начальный момент времени $p=10$, выросла на 2 денежные единицы, т.е. $\Delta p = 2$, относительное изменение цены тогда составит:

$$\frac{\Delta p}{p} = \frac{2}{10}100 = 20\%.$$

Пусть цене $p=10$ соответствовала величина спроса $d=100$, и пусть при новой цене $p_1 = p + \Delta p = 12$ спрос определяется величиной $d_1 = 90$. Тогда $\Delta d = d_1 - d = -10$, а относительное изменение:

$$\frac{\Delta d}{d} = \frac{-10}{100}100 = -10\%.$$

Коэффициент эластичности, вычисляемый как отношение относительных величин приростов спроса и цены, составит:

$$E_{d,p} = \frac{\Delta d}{d} : \frac{\Delta p}{p} = \frac{-10}{100} : \frac{2}{10} = -\frac{1}{2}. \qquad (7.24)$$

Таким образом, коэффициент эластичности спроса по цене показывает, что при изменении цены на 1% величина спроса изменится на 0,5%.

Аналогично можно определить эластичность предложения по цене:

$$E_{S,p} = \frac{\Delta S}{S} : \frac{\Delta p}{p} = \frac{\Delta Sp}{\Delta pS}. \qquad (7.25)$$

В реальной экономике вид зависимостей меняется со временем вследствие непрерывной модификации структуры спроса, производства, влияния социальных и экономических факторов и т.д. Это означает, что точки равновесия кривых спроса, предложения и доходов непрерывно сдвигаются. В связи с этим возникает необходимость введения фактора времени в функции предложения и спроса:

$$D_t = f(p_t); \qquad (7.26)$$

$$S_t = \varphi(p_{t-1}). \qquad (7.27)$$

Данная модель называется ***паутинообразной*** (рис. 7.5).

Предположим, что в начальный момент времени t_0 на рынке наблюдается нехватка товаров, в этих условиях производитель заинтересован в увеличении объема товара в следующий момент времени t_1 с x_0 до x_1. Это множество товаров не может быть теперь реализовано по старой цене p_0, а продажная цена должна быть p_1.

Учитывая вновь сложившуюся ситуацию, производитель товара в следующий момент времени t_2 выбросит на рынок товар в меньшем объеме x_2. Кривая спроса снова будет под кривой предложения, и, следовательно, товар реализуется по более высокой цене p_2. Таким образом, полученная паутинообразная модель показывает непрерывное колебание вокруг равновесной точки цен и производства товаров и отражает их циклическое движение.

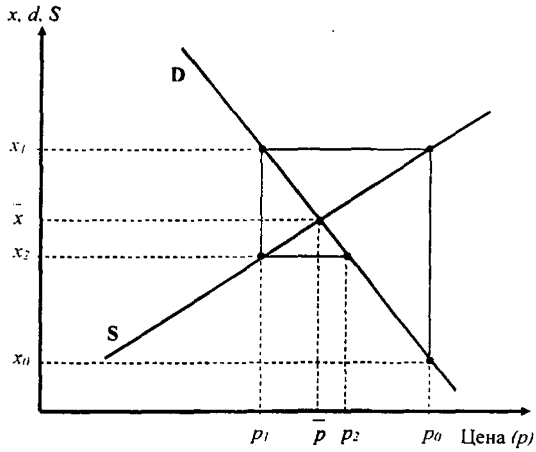


Рис. 7.5. Паутинообразная модель динамического равновесия спроса и предложения

7.6. Анализ инвестиций и основных фондов

Увеличение основных средств ведет к росту продукции. Если предположить, что рост основных средств обеспечивается инвестициями, то можно ввести следующую инвестиционную функцию:

$$I = aY - bK, \tag{7.28}$$

где I — инвестиции;

Y — объем продукции;

K — размер основных фондов.

Инвестиционная функция данного вида выражает положительную зависимость инвестиций от величины продукта (или дохода) и отрицательную связь инвестиций с основными фондами. Иногда инвестиции рассматриваются как функции от прибыли. В этом случае инвестиционная функция записывается как

$$I_t = I_0 + cR_{t-1}, \tag{7.29}$$

где I_t — инвестиции в текущий момент времени;

I_0 — значение инвестиций в начальный момент времени;

R_{t-1} — прибыль, полученная в предыдущем временном интервале.

Временной сдвиг показывает, что существует лаг между принятием решения об инвестировании и его реализацией. Фактор прибыли выступает как стимулятор инвестиционного процесса и вместе с тем является основным источником финансирования инвестиций.

Спрос на основные фонды или ожидаемый их объем определяется ожидаемой величиной продукта, уровнем цен и тенденциями научно-технического прогресса. В связи с этим можно ввести следующую функцию:

$$K_t = aY_t^b p_t^c e^n, \tag{7.30}$$

где Y_t — ожидаемый объем продукции;

p_t — индекс цен;

e^n — трендовая составляющая, характеризующая научно-технический прогресс.

Инвестиции являются источником прироста основных средств производства. Однако только часть инвестиций, вкладываемых в текущий момент времени, обусловливает прирост основных фондов в этот же момент времени. Другая часть инвестиций направляется на рост основных фондов в будущем, таким образом, они реализуются с некоторым запозданием, обусловленным спецификой воспроизводства этих фондов.

Предположим, что прирост основных фондов ΔK_t зависит от всей совокупности инвестиций, сделанных в предыдущие моменты времени, причем влияние более отдаленных инвестиционных вложений убывает в геометрической прогрессии:

$$\Delta K_t = a + bI_t + bqI_{t-1} + bq^2 I_{t-2} + \ldots + bq^n I_{t-n}. \tag{7.31}$$

Соотношение такого вида называется распределительным запаздыванием, оно было введено Койком.

Зависимость прироста основных фондов от текущих инвестиций и прироста за предыдущий момент времени имеет вид

$$\Delta K_t = a(1-q) + bI_t + q\Delta K_{t-1}. \tag{7.32}$$

7.7. Эконометрические модели экономического роста

Очевидно, что глубокая теоретическая проработка причин и условий экономического роста на разных уровнях системы должна дополняться методологией его оценки и характеристикой меры влияния каждого фактора. Макроэкономические методы исследования детерминант экономического роста базируются на аппарате математического моделирования эмпирических процессов. Большинство практических исследований основано на стандартном уравнении экономического роста, в котором к стандартным экономическим переменным (инвестиции, физический и человеческий капитал, живой и овеществленный труд) добавляют потенциально значимые переменные исследуемых детерминант, политических, социальных, географических. Базовое уравнение регрессии может быть представлено в следующей форме:

$$Y_i = a_0 + \sum a_i \cdot x_i + \sum b_i \cdot k_i + \sum c_i \cdot z_i + \varepsilon_i, \qquad (7.33)$$

где Y_i — темп прироста национального продукта (ВВП, ВНП);

a_0 — постоянный член уравнения;

a_i — коэффициент при экономической переменной;

x_i — экономические переменные;

b_i — коэффициент при дополнительных переменных;

k_i — дополнительные переменные (политические, социальные, географические);

c_i — коэффициент при фиктивной переменной, отражающей региональный эффект;

z_i — фиктивная переменная, отражающая региональный эффект;

ε_i — случайная составляющая, ошибка упрощения.

Современная наука и практика концентрируются на исследовании внутренних источников постоянного и устойчивого экономического роста, поэтому в настоящее время наибольшее распространение получили так называемые эндогенные теории роста, среди которых выделяют модели эндогенного развития технологии и роста населения, модели экономического роста с человеческим капиталом и модели инновационного экономического роста.

Модели эндогенных изменений технологий предполагают, что высокий уровень численности и развития населения стимулирует изменения технологий, т. е. при неизменной величине ресурсов рост населения ведет к нарастанию технологических изменений. Априорные предпосылки модели состоят в следующем: рост экономики определяют накопленные знания, технический прогресс является функцией размера населения. В свою очередь технический прогресс и рост объема экономики ведут к росту населения, причем раньше, чем к росту выпуска продукции на душу населения. Таким образом, выпуск (объем производства) зависит от технологии, населения (его труда) и земли. Уравнение эндогенного развития технологии и роста населения может быть представлено в виде производственной функции:

$$Y = AL^{\alpha}T^{1-\alpha}, \tag{7.34}$$

где Y – совокупный выпуск;

A – уровень используемой технологии;

L – население;

T — земля.

Модели экзогенного экономического роста с человеческим капиталом рассматривают как фактор производства не только знания, персонифицированные в каждом конкретном работнике, но и процесс накопления человеческого капитала, который, по мнению авторов модели, аналогичен процессу накопления физического капитала. Поскольку человеческий капитал способен накапливаться и амортизировать, то можно выделить и формы инвестирования: образование, повышение квалификации, забота о здоровье, миграция, а также поиск информации о ценах и доходах. Таким образом, данное моделирование предполагает синтез производственной функции с включением в нее нейтрального технического прогресса (модель Харроду):

$$Y_t = K_t^{\alpha}H_t^{\beta}\left[A_tL_t\right]^{1-\alpha-\beta}, \tag{7.35}$$

где Y — совокупный выпуск;

K — физический капитал;

H — человеческий капитал;

L — труд.

Модели инновационного экономического роста характеризуют возможность существования устойчивого роста с постоянным темпом прироста на основе внедрения технического прогресса. Технический прогресс, по мнению авторов модели, выражается видовым расширением производственных (промежуточных) продуктов, каждый из которых упрощенно идентифицируется с определенной технологией. В качестве математической модели предлагается производственная функция Спенса–Диксита–Стиглица, включающая эффект расширения разнообразия промежуточных (инновационных) продуктов:

$$Y = AL^{1-\alpha}\sum x_j^{\alpha}, \tag{7.36}$$

где Y — совокупный выпуск;

A — параметр производительности сектора конечной продукции;

L — объем труда;

x_j — количество используемого типа промышленных товаров;

α — эластичность выпуска по промежуточному товару.

Приведенные макроэкономические модели позволяют сделать вывод о возможности постоянного устойчивого роста на основе эндогенных факторов, рассматриваемых, как правило, в качестве результата целенаправленной деятельности человека (труд, капитал, инновации, технологии, научно-технический прогресс и т.д.).

Контрольные вопросы

1. **Анализ производства и издержек.**
2. **Производственные функции и их типы.**
3. **Свойства производственных функций и их виды.**
4. **Производственная функция Кобба-Дугласа.**
5. **Функции издержек.**
6. **Эконометрический анализ спроса и предложения.**

7. Анализ инвестиций и основных фондов.
8. Исследование детерминант экономического роста.
9. Модели эндогенных изменений технологий.
10. Модели инновационного экономического роста.

☑ ТЕСТОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ ПРОВЕРКИ ЗНАНИЙ

1. В производственной функции Кобба–Дугласа параметр β соответствует коэффициенту:

а) корреляции;
б) вариации;
в) эластичности;
г) детерминации.

2. Пусть истинной моделью является $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ (x_1 и x_2 — существенные факторы), однако мы не имеем статистических данных по переменной x_1. Но другая переменная z выступает идеальным заменителем для нее в том смысле, что имеется строгая (функциональная) линейная связь $x_1 = \lambda + \mu z$, где λ и μ являются постоянными, но неизвестными величинами. Если мы построим регрессию $\hat{y} = a + cz + b_2 x_2$, то коэффициент детерминации R^2 по этому уравнению будет:

а) таким же, как и при построении регрессии с использованием x_1;
б) значительно больше, чем при построении регрессии с использованием x_1;
в) статистически незначимым;
г) значительно меньше, чем при построении регрессии с использованием x_1.

3. Для оценки заработной платы некоего работника используется следующая модель: $Y_i = \alpha + \beta_1 x_i + \beta_2 D_i + \beta_3 C_i + \beta_4 S_i + \varepsilon_i$, где Y_i — заработная плата i-го работника; X_i — общий стаж его работы на данном предприятии; D_i — переменная, принимающая значение 1, если работник с высшим образованием, и 0 в

противном случае; C_i — количество детей у работника; S_i — переменная, принимающая значение 1, если работник мужчина, и 0, если женщина; W_i — количество должностей, которые сменил работник на различных предприятиях в течение последнего года. Тогда фиктивными переменными в данной модели являются:

а) X_i;
б) W_i;
в) D_i;
г) S_i.

4. В эконометрической модели множественной регрессии $t_y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j + \ldots + \beta_k x_k$ стандартизованный коэффициент регрессии β_j показывает, на сколько:

а) средних квадратичных отклонений y_y изменяется y при увеличении x_j на одну единицу своего измерения;
б) единиц своего измерения изменяется y при увеличении x_j на одну единицу своего измерения;
в) средних квадратичных отклонений y_y изменяется y при увеличении x_j на одно среднеквадратическое отклонение y_{xj}, если остальные факторы, входящие в уравнение регрессии, считать неизменными;
г) единиц своего измерения изменяется y при увеличении x_j на одно среднеквадратическое отклонение y_{xj}.

5. Приведенное выражение $\sum_{i=1}^{n}(Y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}))^2 \to \min$ представляет собой ... для линейной двухфакторной модели регрессии:

а) исходное положение метода наименьших квадратов;
б) условие отсутствия автокорреляции остатков;
в) систему нормальных уравнений;
г) теорему Гаусса–Маркова.

6. Эмпирический коэффициент b-регрессии $y = a + bx + \varepsilon$ является состоятельной оценкой теоретического коэффициента β регрессии $y = a + \beta x + \varepsilon$ при условии, что:

а) b сходится по вероятности к β при числе наблюдений, стремящемся к 0;
б) математическое ожидание оценки b равно нулю;
в) дисперсия оценки b равна 1;
г) b сходится по вероятности к β при числе наблюдений, стремящемся к бесконечности.

7. Если эффективность производства растет по мере его укрупнения и оно описывается производственной функцией Кобба–Дугласа, то параметры модели удовлетворяют соотношению:
а) $\alpha+\beta=1$;
б) $\alpha+\beta<1$;
в) $\alpha+\beta>1$;
г) $\alpha+\beta=0$.

8. Получены две производственные функции Кобба–Дугласа, имеющие равные значения параметров «альфа» и «бета», но различающиеся по параметру A. Первое производство следует считать более эффективным, если:
а) $A_1<A_2$;
б) $A_1=A_2$;
в) $A_1>A_2$.

9. Нижеприведенная модель производственной функции получена по 26 наблюдениям, уровень значимости $p=0{,}05$:

$$y=0{,}46l+032k,\ R^2=0{,}41;$$
$$t=\ (1{,}81)\quad(2{,}87),$$

где y, l, k — темпы прироста объема выпуска, затрат труда и капитал.

Верным является следующий вывод:
а) надо исключить фактор k, так как он оказался статистически незначимым;
б) модель имеет удовлетворительные статистики, поэтому нет смысла ее совершенствовать;
в) надо исключить фактор l, так как он оказался статистически незначимым.

10. При гетероскедастичности вероятнее всего, что t-статистики коэффициентов регрессии и F-статистика будут:
а) заниженные;
б) точные;
в) завышенные.

ПРАКТИКУМ

Задача 7.1. По данным таблицы 7.1 постройте двухфакторную производственную функцию, описывающую изменение объема производства продукции сельского хозяйства Ставропольского края, обусловленное воздействием стоимости основных фондов и среднегодовой численности работников сельскохозяйственных организаций.

Таблица 7.2

Исходные данные к задаче 7.1

№ района	Валовая продукция сельхозпредприятий района, млн руб.	Среднегодовая численность работников сельхозорганизаций, чел.	Стоимость основных фондов на конец года, млн руб.
	Y	X_1	X_2
1	2	3	4
1	1530	1009	844,2
2	564	376	242,7
3	1428	3274	1053,3
4	996	1364	1100,6
5	1074	1509	835,2
6	2495	2419	2244,0
7	2785	2737	2554,3
8	758	683	532,7
9	2422	1629	2031,7
10	2370	2428	3545,3
11	1418	1590	1437,9
12	5631	3919	4029,2
13	3401	2258	5887,6

Окончание табл. 7.2

1	2	3	4
14	801	1047	708,4
15	844	1616	1067,3
16	1071	927	920,4
17	577	726	561,8
18	5261	4581	3713,8
19	1802	2427	1356,1
20	1881	2412	1306,1
21	2059	2552	2395,5
22	2068	3001	1207,3
23	564	887	170,9
24	2001	2273	1996,1
25	515	1108	400,3
26	8843	4710	6571,4

Решение

Экономическая производственная функция с двумя факторами может быть описана в линейной или нелинейной формах. Проведем оценку параметров и статистической значимости этих форм зависимости производства продукции от численности работников и стоимости основных фондов сельхозорганизаций.

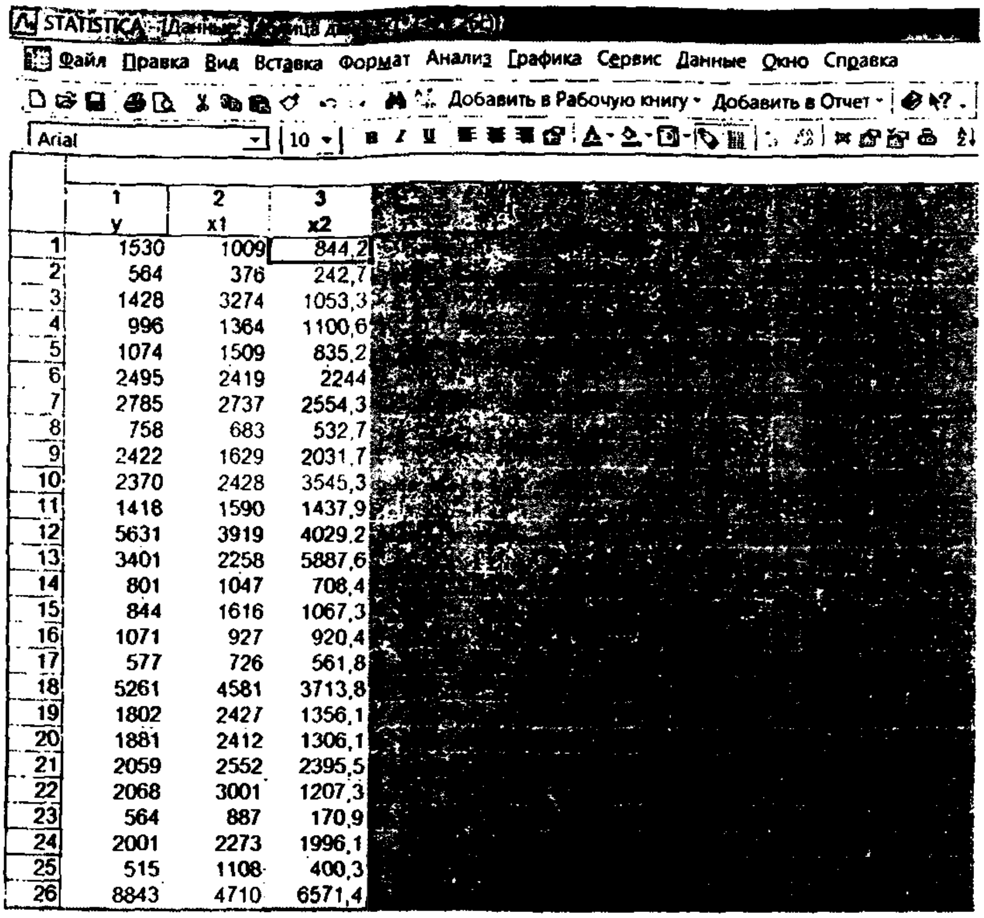
а) *линейная производственная функция вида*

$$y = a_0 + a_1 x_1 + a_2 x_2.$$

Для оценки производственной функции линейного вида воспользуемся инструментом «Нелинейное оценивание» пакета *Statistica* 6.0. Для начала необходимо сформировать исходную таблицу, содержащую значения представленных переменных (рис. 7.6).

Далее откроем инструмент «Нелинейное оценивание» в строке «Анализ» → «Углубленные методы анализа». В появившемся диалоговом окне следует выбрать пункт «Регрессия пользователя – метод наим. квадратов МНК» (рис. 7.7).

В новом диалоговом окне требуется ввести линейную производственную функцию заданного вида. Для корректного



	1 y	2 x1	3 x2
1	1530	1009	844,2
2	564	376	242,7
3	1428	3274	1053,3
4	996	1364	1100,6
5	1074	1509	835,2
6	2495	2419	2244
7	2785	2737	2554,3
8	758	683	532,7
9	2422	1629	2031,7
10	2370	2428	3545,3
11	1418	1590	1437,9
12	5631	3919	4029,2
13	3401	2258	5887,6
14	801	1047	708,4
15	844	1616	1067,3
16	1071	927	920,4
17	577	726	561,8
18	5261	4581	3713,8
19	1802	2427	1356,1
20	1881	2412	1306,1
21	2059	2552	2395,5
22	2068	3001	1207,3
23	564	887	170,9
24	2001	2273	1996,1
25	515	1108	400,3
26	8843	4710	6571,4

Рис. 7.6. Таблица исходных данных для решения задачи в *Statistica 6.0*

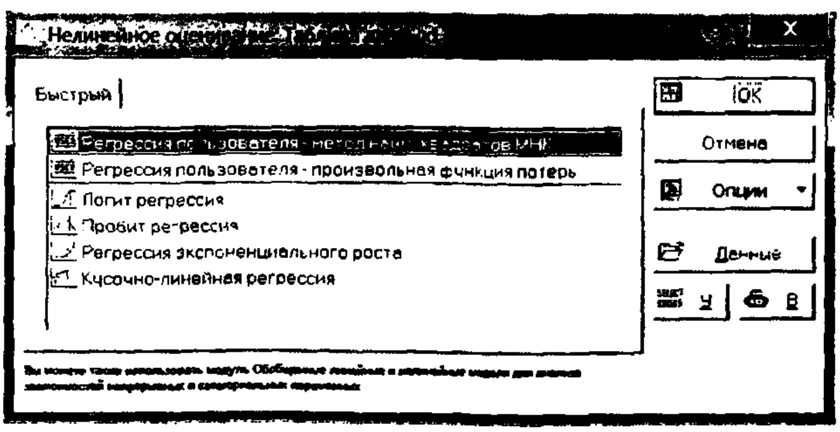


Рис. 7.7. Диалоговое окно «Нелинейное оценивание» в *Statistica 6.0*

представления данной функции в *Statistica 6.0* в поле «Оцениваемая функция» введите следующее выражение: $y=a+b*x1+c*x2$ (рис. 7.8).

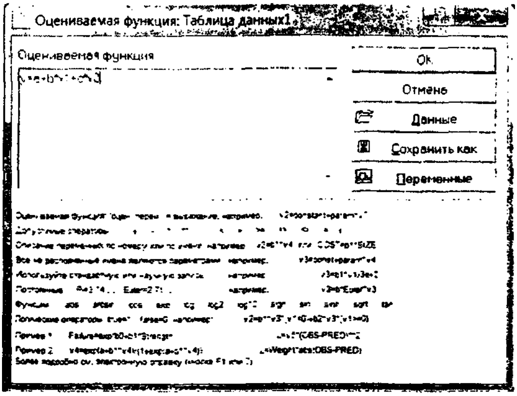


Рис. 7.8. Диалоговое окно «Оцениваемая функция» в *Statistica 6.0*

Далее переходим к окну «Результаты». Во вкладке «Быстрый» следует выбрать диалоговое окно «Оценки параметров модели» (рис. 7.9).

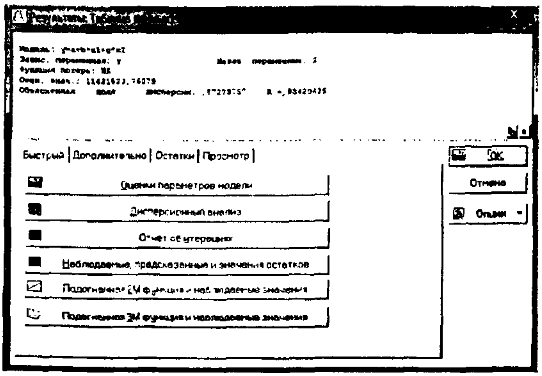


Рис. 7.9. Диалоговое окно «Результаты» в *Statistica 6.0*

Для линейной производственной функции результаты оценки параметров модели в *Statistica* 6.0 следующие. Производственная функция представлена уравнением вида

$$y = -530{,}53 + 0{,}697 \cdot x_1 + 0{,}65 \cdot x_2 \text{ (рис. 7.10).}$$

Модель: y=a+b*x1+c*x2 (Таблица данных1)
Зав. Пер. : y
Уров. значимости: 95.0% (альфа=0.050)

	Оценка	Стандарт ошиб.	t-знач. cc = 23	p-уров.	Ниж. Дов Предел	Вер. Дов Предел
a	-530,531	286,5988	-1,85113	0,077035	-1123,41	62,34386
b	0,697	0,1809	3,85544	0,000805	0,32	1,07152
c	0,650	0.1274	5,10254	0,000036	0,39	0.91372

Рис. 7.10. Оценки параметров линейной производственной функции в *Statistica 6.0*

Статистическая значимость представленной линейной функции достаточно высока. Об этом свидетельствует значение коэффициента множественной корреляции $R = 0{,}93$, что определяет степень зависимости стоимости валовой продукции сельхозпредприятий Ставропольского края и совокупности двух факторов – численности работников сельхозпредприятий и стоимости основных фондов — как высокую.

Значение множественного коэффициента детерминации свидетельствует о том, что совокупность выбранных факторов на 87,3% объясняет изменение стоимости валовой продукции сельхозтоваропризводителей региона в 2013 году (рис. 7.9).

Что касается оценки статистической адекватности найденных параметров уравнения регрессии, они отличаются высоким качеством. Об этом свидетельствуют значения t-критериев Стьюдента для каждого из коэффициентов регрессии (рис. 7.10).

Результаты анализа содержат также теоретические значения результативной переменной (рис. 7.11).

В заключение для наглядного представления полученных результатов можно воспользоваться графическим инструментом «Подогнанная 3М функция и наблюдаемые значения» во вкладке «Быстрый» диалогового окна «Результаты». Полученная диаграмма представлена на рисунке 7.12.

Модель: y=a+b*x1+c*x2 (Таблица данных1)
Зав. Пер. : y

	Наблюд.	редсказанные	Остатки
1	1530,000	721,945	808,05
2	564,000	-110,538	674,54
3	1428,000	2437,390	-1009,39
4	996,000	1136,201	-140,20
5	1074,000	1064,770	9,23
6	2495,000	2615,277	-120,28
7	2785,000	3038,773	-253,77
8	758,000	292,090	465,91
9	2422,000	1926,345	495,66
10	2370,000	3467,581	-1097,58
11	1418,000	1513,095	-95,09
12	5631,000	4821,935	809,07
13	3401,000	4871,855	-1470,85
14	801,000	660,155	140,84
15	844,000	1290,284	-446,28
16	1071,000	714,303	356,70
17	577,000	340,995	236,00
18	5261,000	5078,527	182,47
19	1802,000	2043,596	-241,60
20	1881,000	2000,629	-119,63
21	2059,000	2806,521	-747,52
22	2068,000	2347,135	-279,14
23	564,000	199,129	364,87
24	2001,000	2352,294	-351,29
25	515,000	502,386	12,61
26	8843,000	7026,326	1816,67

Рис. 7.11. Теоретические значения результативной переменной и остатки в *Statistica 6.0*

В результате построенная линейная производственная функция позволяет сделать вывод, что увеличение среднегодовой численности работников сельхозорганизаций отдельного муниципального района на 1 человека приводит к росту стоимости его валовой продукции на 697 тыс. руб., а увеличение стоимости основных фондов на 1 млн руб. вызывает рост стоимости валовой продукции на 650 тыс. руб.

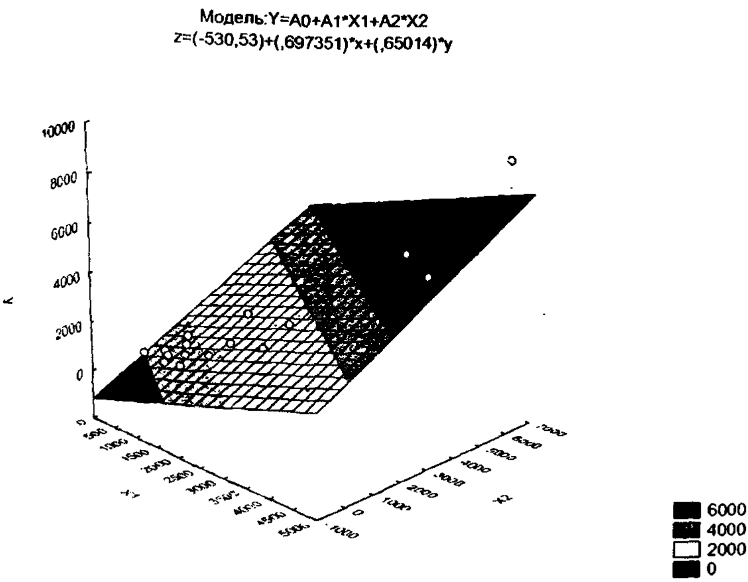


Рис. 7.12. Пространственное изображение результатов моделирования линейного вида производственной функции в *Statistica 6.0*

б) *нелинейная производственная функция вида* $y = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2}$.

Для оценки параметров и статистической значимости степенной производственной функции воспользуемся аналогичными инструментами *Statistica* 6.0. В отличие от линейной производственной функции, в диалоговом окне «Оцениваемая функция» следует ввести выражение: y=a*x1**b*x2**c. (рис. 7.13).

Результаты оценки параметров степенной производственной функции представлены на рисунках 7.14, 7.15. Функция имеет вид $y = 0{,}0193 \cdot x_1^{0{,}913} \cdot x_2^{0{,}597}$.

Полученная эконометрическая модель отличается высокой статистической корректностью, понятие о которой дают показатели множественной корреляции и детерминации (рис. 7.14).

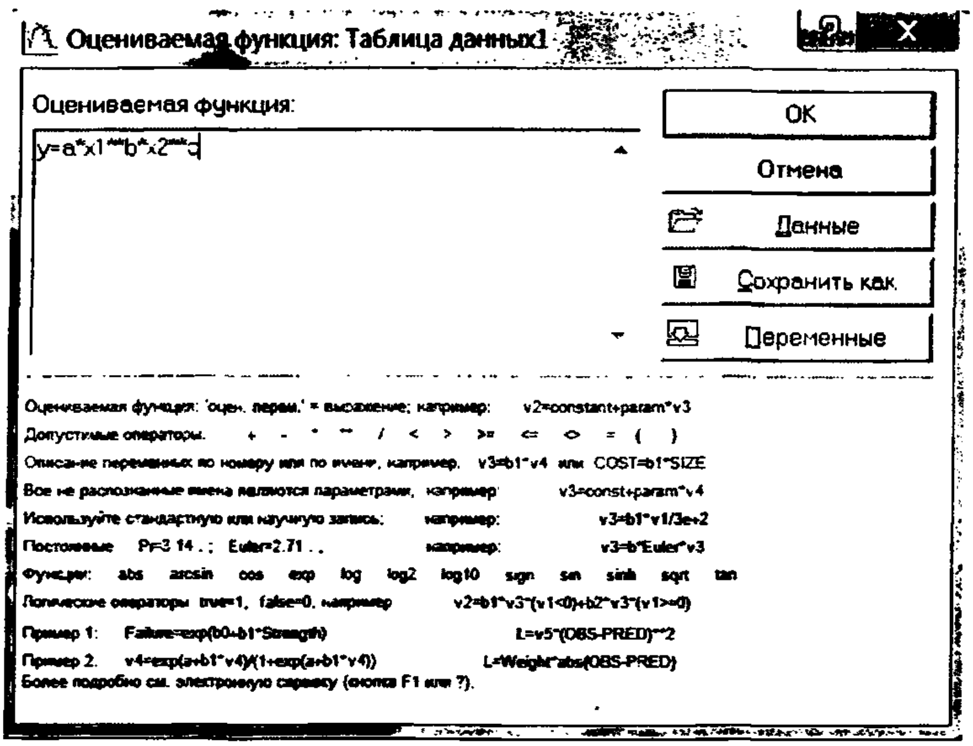


Рис. 7.13. Степенная функция в *Statistica 6.0*

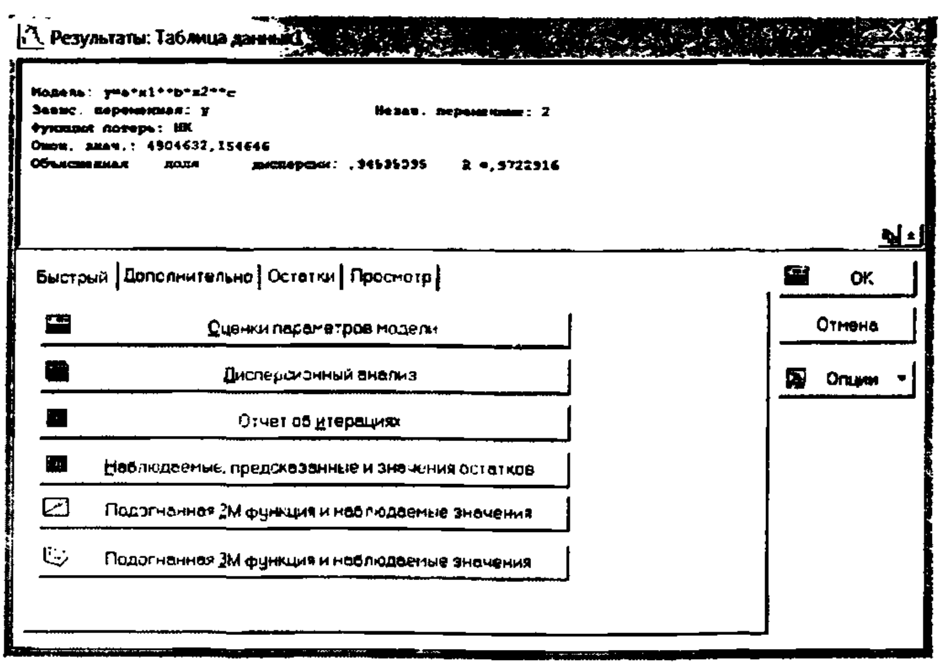


Рис. 7.14. Окно результатов оценки степенной производственной функции в *Statistica 6.0*

Значение коэффициента множественной корреляции $R=0,97$ определяет тесноту связи между исследуемыми экономическими переменными как весьма высокую.

Значение множественного коэффициента детерминации свидетельствует о том, что совокупность выбранных факторов на 94,5% объясняет изменение стоимости валовой продукции сельхозтоваропризводителей региона в 2013 году.

Найденные параметры уравнения регрессии отличаются высоким качеством: на рисунке 7.15 представлены значения t-статистик Стьюдента для этих параметров.

Модель: y=a*x1**b*x2**c (Таблица данных1)
Зав. Пер. : y
Уров. значимости: 95.0% (альфа=0.050)

	Оценка	Стандарт ошиб.	t-знач. сс = 23	p-уров.	Ниж. Дов Предел	Вер. Дов Предел
a	0,019345	0,015046	1,285715	0,211337	-0,011780	0,050471
b	0,912628	0,124945	7,304252	0,000000	0,654160	1,171096
c	0,597192	0,078523	7,605293	0,000000	0,434755	0,759630

Рис. 7.15. Коэффициенты степенной функции в *Statistica 6.0*

Результаты анализа содержат также теоретические значения результативной переменной (рис. 7.16).

На рисунке 7.17 представлена графическая интерпретация полученных результатов для степенной функции, построенная при помощи *Statistica* 6.0.

В результате построенная степенная производственная функция позволяет сделать вывод, что увеличение среднегодовой численности работников на 1% приводит к росту стоимости его валовой продукции на 0,913%, а увеличение стоимости основных фондов на 1% вызывает рост стоимости валовой продукции на 0,597% (при сохранении влияния прочих факторов).

в) *производственная функция Кобба–Дугласа вида* $y=a_0 \cdot F^{\alpha} \cdot Z^{1-\alpha}$.

Построение производственной функции Кобба–Дугласа предполагает введение ограничения для оцениваемых коэффициентов уравнения: $a_0 + a_1 = 1$. Для выполнения данного условия мож-

	Модель: y=a*x1**b*x2**c (Таблица данных1) Зав. Пер. : y		
	Наблюд.	Предсказанные	Остатки
1	1530,000	596,568	933,432
2	564,000	115,109	448,891
3	1428,000	1993,327	-565,327
4	996,000	920,304	75,696
5	1074,000	855,866	218,134
6	2495,000	2375,638	119,362
7	2785,000	2872,925	-87,925
8	758,000	317,381	440,619
9	2422,000	1560,601	861,399
10	2370,000	3132,367	-762,367
11	1418,000	1241,739	176,261
12	5631,000	5233,784	397,216
13	3401,000	3968,761	-567,761
14	801,000	555,682	245,318
15	844,000	1054,769	-210,769
16	1071,000	581,404	489,596
17	577,000	346,397	230,603
18	5261,000	5748,275	-487,275
19	1802,000	1763,861	38,139
20	1881,000	1714,999	166,001
21	2059,000	2593,812	-534,812
22	2068,000	1997,382	70,618
23	564,000	204,319	359,681
24	2001,000	2092,881	-91,881
25	515,000	416,132	98,868
26	8843,000	8289,956	553,044

Рис. 7.16. Теоретические значения объема производства в *Statistica 6.0*

но воспользоваться средствами регрессионного анализа *SPSSStatistics* 20.

Для этого сформируем таблицу исходных данных, обозначив в качестве результативной переменной валовую продукцию (*Y*), а в качестве факторных — соответственно среднегодовую численность работников (*Z*) и стоимость основных фондов (*F*).

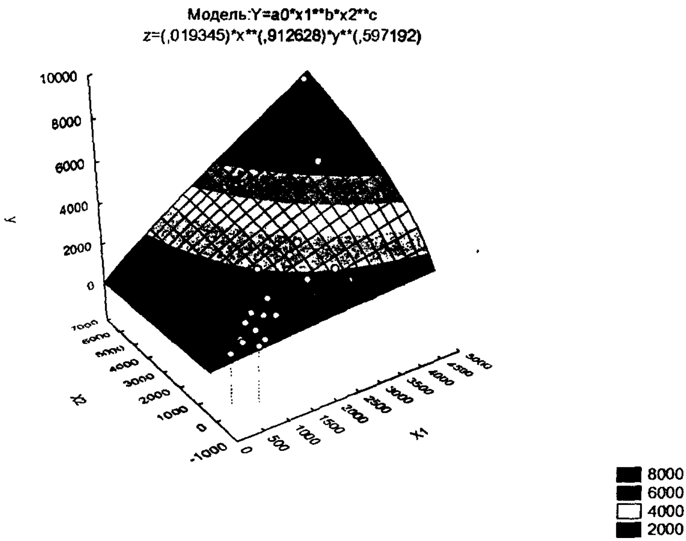


Рис. 7.17. Графическое изображение результатов моделирования степенного вида производственной функции в *Statistica* 6.0

Воспользуемся инструментом «Нелинейная регрессия» из меню «Анализ» (рис. 7.18).

На следующем этапе необходимо задать уравнение функции Кобба–Дугласа, которая является разновидностью степенной функции с наличием ограничений (рис. 7.19).

Для определения необходимого ограничения необходимо открыть вкладку «Ограничения», в которой следует указать, что сумма коэффициентов эконометрической модели должна равняться единице (рис. 7.20).

Продолжив построение модели, получим следующие результаты.

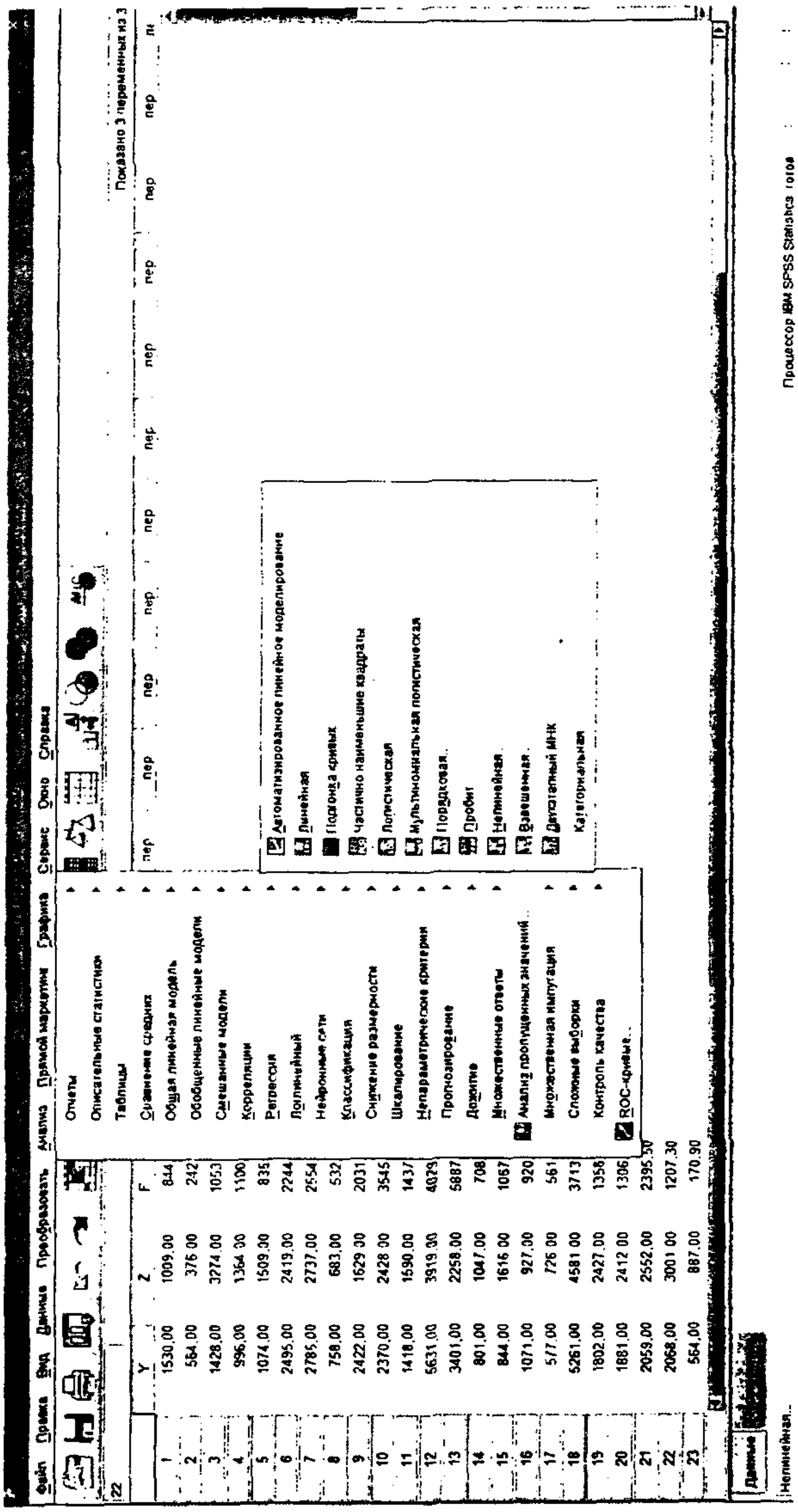


Рис. 7.18. Инструмент «Нелинейная регрессия» в *SPSSStatistics 20*

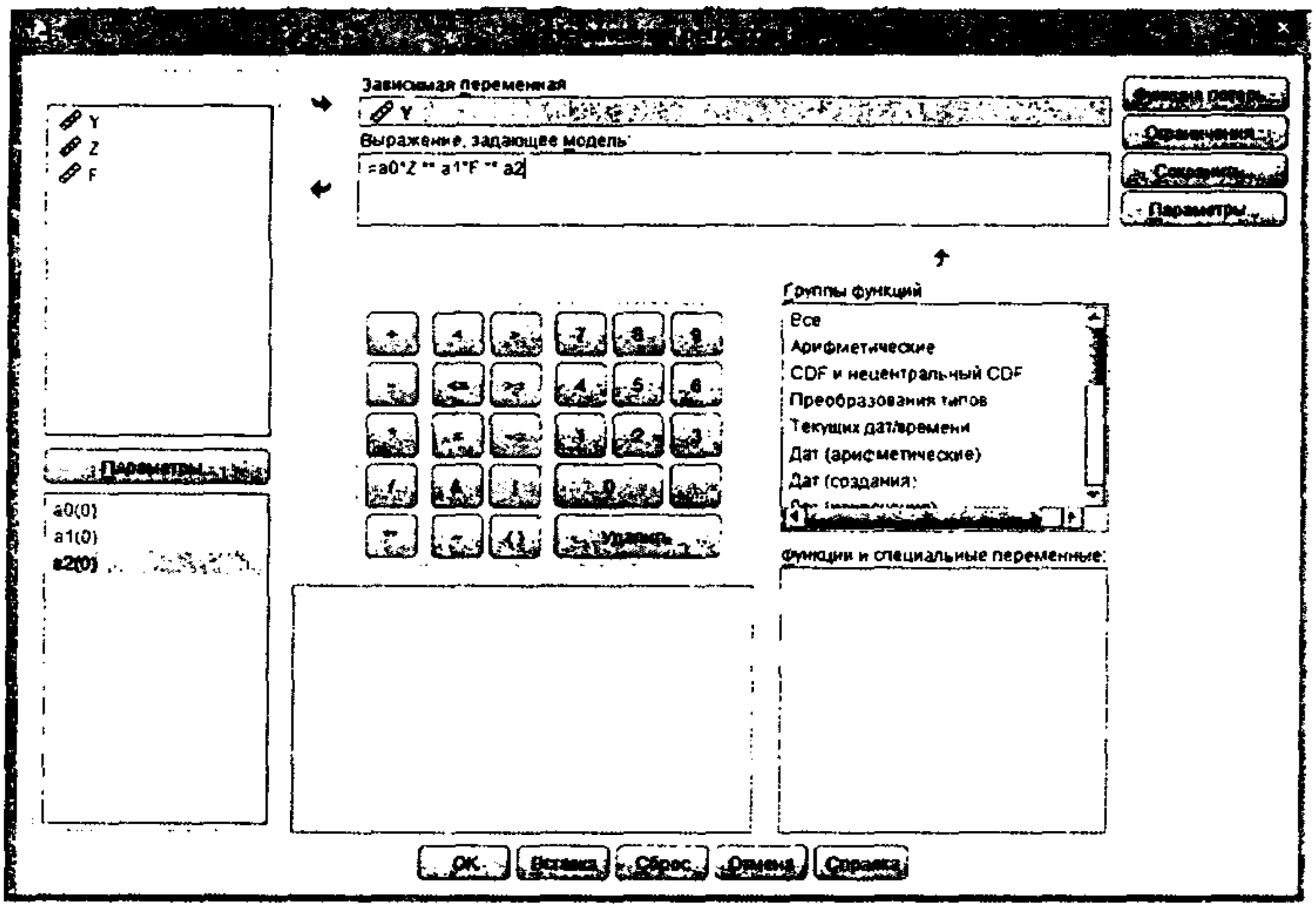


Рис. 7.19. Диалоговое окно «Нелинейная регрессия» *SPSSStatistics 20*

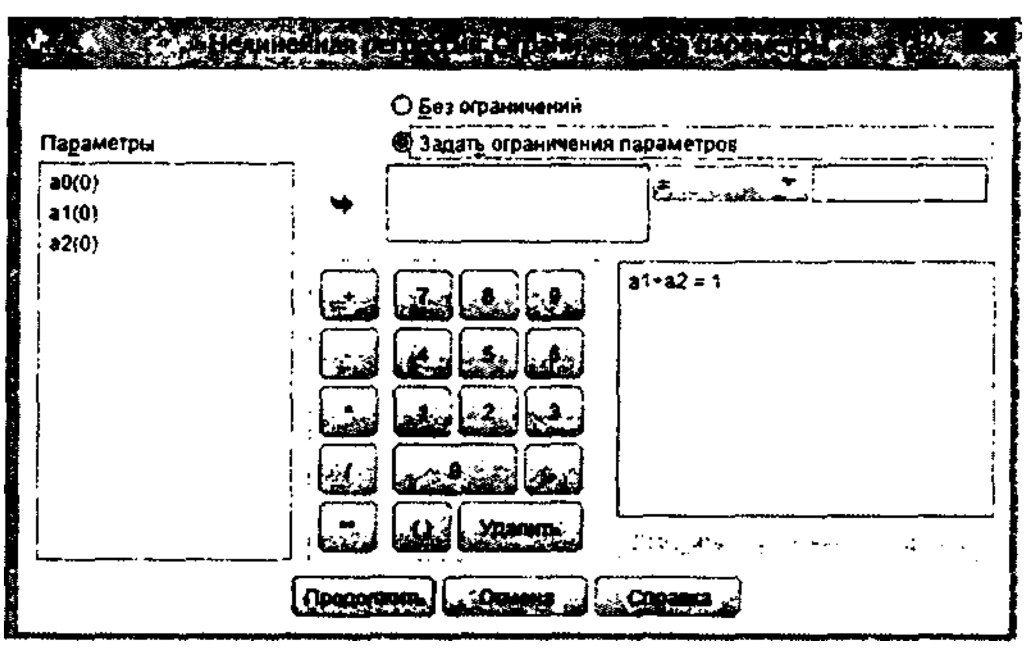


Рис. 7.20. Ограничение функции Кобба-Дугласа *вSPSSStatistics 20*

Оценки параметров

Параметр	Оценка	Стд. Ошибка	Доверительный интервал 95 %	
			Нижняя граница	Верхняя граница
a0	1,191	1,170	-1,230	3,612
a1	,418	,168	,071	,764
a2	,582	,109	,358	,807

Дисперсионный анализ[a]

Источник	Сумма квадратов	ст.св.	Средние квадраты
Регрессия	195570836,397	2	97785418,198
Остаток	11196788,603	24	466532,858
Нескорректированный итог	206767625,000	26	
Скорректированный итог	89747806,500	25	

Зависимая переменная: yR-квадрат = 1 — (остаточная сумма квадратов / скорректированная сумма квадратов) = 0,875.

Таким образом, можно сделать вывод, что функция Кобба–Дугласа для данного набора эмпирических данных выглядит следующим образом: $Y = 1{,}191 \cdot Z^{0,42} \cdot F^{0,58}$.

В таблице 7.3 представим обобщенные характеристики полученных производственных функций.

Таблица 7.3

Производственные функции линейного и степенного вида

Функция	R	R^2
$y = -530{,}53 + 0{,}697 \cdot x_1 + 0{,}65 \cdot x_2$	0,93	0,873
$y = 0{,}0193 \cdot x_1^{0,913} \cdot x_2^{0,597}$	0,97	0,945
$Y = 1{,}191 \cdot Z^{0,42} \cdot F^{0,58}$	0,94	0,875

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 7.2. По данным таблицы 7.4 постройте производственную функцию Кобба–Дугласа, описывающую изменение объема производства зерновых культур в Ставропольском крае, обусловленное воздействием представленных факторов.

Таблица 7.4

Данные о производстве зерновых культур в муниципальных районах Ставропольского края

№ района	Объем производства продукции сельского хозяйства, млн руб.	Среднемесячная заработная плата работающих, руб.	Стоимость основных фондов, млн руб.
	Y	X_1	X_2
1	2	3	4
1	331,9	3685	1195,1
2	323,7	5460	559,9
3	859,2	2803	1607,9
4	614,9	3164	1095,4
5	587,8	3892	1424,6
6	931	2831	1683,2
7	201,6	4831	1011,3
8	384,6	3932	1050,1
9	814,1	5584	2042,3
10	741,5	3749	2403,5
11	953,8	5074	1529
12	1253,1	4644	2435,9
13	533,8	3861	2441,9
14	655,1	3530	1278,5
15	515,3	3395	1648
16	457,6	4713	1020,6
17	242,6	5620	1490,4
18	1873,5	4003	3082,1
19	397,7	3475	1482,8
20	760,9	4317	2010
21	1247,5	5260	2034,4

Окончание табл. 7.4

1	2	3	4
22	720,1	3832	2091,4
23	255,4	3163	1024,5
24	1148,9	4275	2048,4
25	450,7	3139	1151,7
26	1092,6	5186	3063,9

Задача 7.3. По данным таблицы 7.5 осуществите спецификацию, оценку параметров и проверку статистической корректности производственной функции, описывающей зависимость выручки от реализации продукции животноводства в Ставропольском крае от ее себестоимости и затрат труда.

Таблица 7.5

Данные о реализации и себестоимости продукции животноводства в муниципальных районах Ставропольского края

№ района	Объем сельскохозяйственной продукции района, млн руб.	Среднегодовая численность работников сельхозорганизаций, чел.	Основные фонды на конец года, млн руб.
	Y	X_1	X_2
1	2	3	4
1	1908,1	1009	646,3
2	1333,5	376	718,8
3	1699,0	3274	1177,6
4	2033,2	1364	776,1
5	2416,2	1509	1216,8
6	1262,1	2419	864,2
7	2451,6	2737	580,4
8	2878,3	683	549,8
9	1782,3	1629	10761,8
10	2902,2	2428	1250,0
11	3158,0	1590	1579,3
12	1478,8	3919	2052,6
13	1969,7	2258	944,4
14	1219,5	1047	894,0

Окончание табл. 7.5

1	2	3	4
15	1796,9	1616	886,7
16	3787,3	927	13770,5
17	1878,7	726	9624,9
18	2432,1	4581	3589,9
19	2248,8	2427	495,6
20	2496,9	2412	3151,2
21	1208,6	2552	1398,3
22	2441,6	3001	1867,9
23	1432,1	887	321,0
24	4046,8	2273	1377,7
25	1908,1	1108	520,4
26	1333,5	4710	1854,1

ГЛАВА

8 ЭКОНОМЕТРИЧЕСКИЙ АНАЛИЗ ДИНАМИКИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

8.1. Классификация и компонентный анализ рядов динамики

При анализе социально-экономических явлений часто используют ежегодные, ежеквартальные, ежемесячные, ежедневные данные. Для рационального анализа следует систематизировать моменты получения соответствующих статистических данных. Упорядоченные статистические данные по времени их получения называют *рядами динамики* (временными, хронологическими рядами).

В динамическом ряду процесс экономического развития изображается в виде совокупности «прерывов» непрерывного, позволяющих детально проанализировать особенности развития с помощью характеристик, отображающих изменение параметров системы во времени.

Составляющими элементами ряда динамики являются показатели уровней ряда, которые принято обозначать через y и периоды или моменты (даты) времени, к которым относятся уровни (обозначаются через t).

Пусть исследуется показатель y. Его значение в текущий период (момент) времени t обозначается y_t, в последующие моменты y_{t+1}; y_{t+2}; $y_{t+3}, \ldots, y_{t+n}$, а в предыдущие моменты y_{t-1}; y_{t-2}; $y_{t-3}, \ldots, y_{t-n}$.

Временные ряды классифицируются по следующим признакам:

а) в зависимости от способа выражения уровней ряды динамики подразделяются на ряды *абсолютных*, *относительных* и *средних* величин;

б) в зависимости от того, как уровни ряда (т. е. показатели) отражают состояние изучаемого явления либо в определенный момент времени — ряд динамики *моментный*, либо за определенный интервал времени — ряд динамики *интервальный*;

в) в зависимости от расстояния между уровнями ряды подразделяются на ряды динамики с *равноотстоящими уровнями* и *неравноотстоящими уровнями* во времени;

г) в зависимости от наличия основной тенденции изучаемого процесса хронологические ряды подразделяются на *стационарные* и *нестационарные*. Если математическое ожидание значения признака и дисперсия постоянны, не зависят от времени, то процесс считают стационарным и ряды динамики считают *стационарными*. Однако эконо-

Рис. 8.1. Классификация временных рядов

мические процессы во времени обычно не являются стационарными, так как содержат основную тенденцию развития, но их можно преобразовать в стационарные путем исключения тенденции;

д) по числу показателей выделяют *изолированные* и *комплексные* (многомерные) ряды динамики.

Как известно, ряд динамики подвержен влиянию *факторов эволюционного* и *осциллятивного* характера, а также находится под влиянием факторов разного воздействия.

Влияния *эволюционного* характера — это изменения, определяющие некое общее направление развития, многолетнюю эволюцию, которая пробивает себе дорогу через другие систематические и случайные колебания. Такие изменения динамического ряда называются *тенденцией развития*, или *трендом*.

Влияния *осциллятивного* характера — это циклические (конъюнктурные) и сезонные колебания. *Циклические* (или периодические) состоят в том, что значение изучаемого признака в течение какого-то времени возрастает, достигая определенного максимума, затем понижается, достигая определенного минимума, вновь возрастает до прежнего значения и т.д. Циклические колебания в экономических процессах примерно соответствуют так называемым циклам конъюнктуры. *Сезонные* колебания — это колебания, периодически повторяющиеся в некоторое определенное время каждого года, месяца, дня или его часа.

Нерегулярные колебания для социально-экономических явлений условно делят на две группы:

а) спорадически наступающие изменения, вызванные, к примеру, войной или экологической катастрофой;

б) собственно, *случайные колебания*, являющиеся результатом действия большого количества относительно слабых второстепенных факторов.

Следовательно, первоначальные значения ряда динамики подвергаются воздействию, состоящему из четырех компонентов:

а) основная тенденция (тренд) — T;

б) циклическая или конъюнктурная — K;

в) сезонная — S;

г) случайное колебание — Δ_t (рис. 8.2).

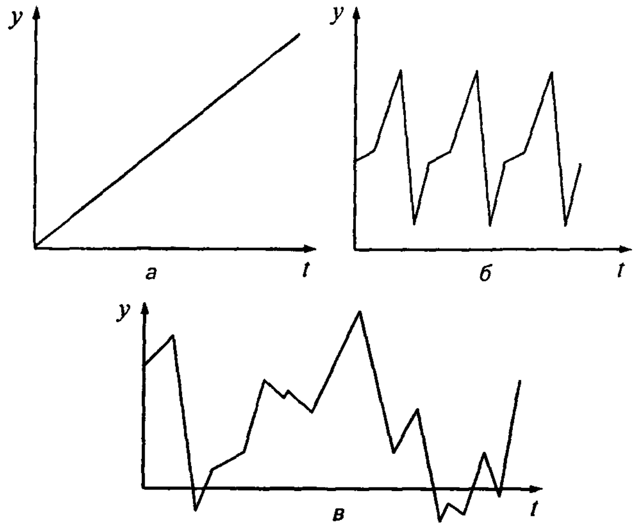


Рис. 8.2. Компоненты ряда динамики:
а — возрастающая тенденция; *б* — сезонная компонента;
в — случайная компонента

Если ряд динамики разделить на различные компоненты, то он представляется в следующем виде:

$$y = f(T,\ K,\ S,\ \Delta_t). \qquad (8.1)$$

Как известно, в зависимости от взаимосвязи этих компонент между собой строят *аддитивную* и *мультипликативную модели* ряда динамики.

Аддитивная модель временного ряда

$$y = T + K + S + \Delta_t \qquad (8.2)$$

характеризуется главным образом тем, что характер циклических и сезонных флюктуаций (колебаний) остается постоянным (рис. 8.3).

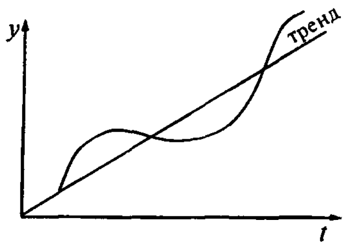


Рис. 8.3. Сочетание составляющих ряда динамики при аддитивной связи

Мультипликативная модель ряда динамики

$$y = T \cdot K \cdot S \cdot \Delta, \tag{8.3}$$

характеризуется постоянством циклических и сезонных флюктуаций только по отношению к тренду (рис. 8.4).

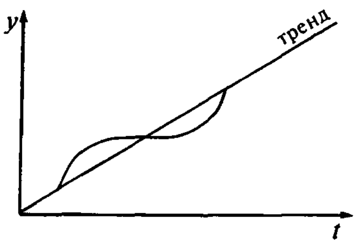


Рис. 8.4. Сочетание составляющих ряда динамики при мультипликативной связи

8.2. Методология регрессионного анализа тенденции временного ряда

Тренд представляет собой долговременную компоненту временного ряда, характеризующую основную тенденцию развития явления. При этом остальные компоненты рассматриваются как мешающие процедуре его определения. При наличии ряда наблюдаемых явлений для различных моментов (периодов) времени следует найти подходящую трендовую кривую, сглаживающую остальные колебания.

В социально-экономических рядах динамики наблюдают основную тенденцию трех видов:

а) среднего уровня;

б) дисперсии;

в) автокорреляции.

Тенденция среднего уровня аналитически выражается с помощью математической функции, вокруг которой варьируют фактические уровни исследуемого явления. В таком случае значения тренда в отдельные моменты времени являются математическим ожиданием ряда динамики. Часто тенденцию среднего уровня называют *детерминированной составляющей исследуемого явления*, и соответствующий ряд динамики выражается уравнением

$$y_t = f(t) + \Delta_t. \qquad (8.4)$$

Тенденция дисперсии представляет собой тенденцию изменения отклонений между эмпирическими уровнями и детерминированной компонентой ряда.

Тенденция автокорреляции характеризует изменения связи между отдельными уровнями ряда динамики.

Перед выделением и анализом тренда следует проверить гипотезу о его существовании. Отсутствие основной тенденции означает неизменность среднего уровня хронологического ряда во времени.

Для проверки наличия тренда известно около десятка критериев, различающихся по мощности и сложности математического аппарата.

Наиболее доступный из них — метод проверки существенности разности средних двух частей одного и того же ряда, основанный на t-критерии Стьюдента.

Ряд динамики разбивают на две равные или почти равные (что не принципиально) части. Проверяется гипотеза о существовании разности средних $H_0: \bar{y}_1 = \bar{y}_2$.

Так как число уровней анализируемого ряда зачастую незначительно, то можно воспользоваться методом проверки гипотезы H_0, разработанным для малых выборок.

За основу проверки берут критическое (табличное) значение t-критерия Стьюдента. При $t_{\text{факт}} = t_{\text{крит}}$ гипотеза об отсутствии тренда отвергается, а при $t_{\text{факт}} = t_{\text{крит}}$ гипотеза H_0 принимается.

В случае равенства или при несущественном различии дисперсий двух исследуемых совокупностей ($S_1^2 = S_2^2$) исчисляется отношение средних с помощью выражения

$$t = \frac{\bar{y}_1 - \bar{y}_2}{S \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}, \tag{8.5}$$

где S — среднее квадратическое отклонение разности средних.

Значение критического t-критерия Стьюдента берется с учетом заданного уровня значимости и числа степеней свободы, равного $n_1 + n_2 - 2$. Необходимое значение S определяется на основе средней взвешенной величины дисперсий отдельных совокупностей:

$$S = \sqrt{\frac{(n_1 - 1)^2 \cdot S_1^2 + (n_2 - 1)^2 \cdot S_2^2}{n_1 + n_2 - 2}}. \tag{8.6}$$

Проверка временного ряда на наличие тренда может быть также осуществлена с помощью F-критерия Фишера, основанного на сравнении расчетного отношения с табличным:

$$F = \frac{S_1^2}{S_2^2}, \tag{8.7}$$

где $S_1^2 > S_2^2$.

Если расчетное значение F-критерия меньше табличного при заданном уровне вероятности, то гипотеза о равенстве дисперсий принимается.

Для полиноминальных моделей характерно отсутствие прямой связи между абсолютными приростами и приростами уровней рядов динамики.

Функцией, отражающей процесс роста явления, может быть и экспонента, характеризующая прирост и зависящая от величины основания функции.

Отдельные уравнения выражают различные типы динамики. Монотонное возрастание или убывание процесса характеризуют функции: линейная, параболическая, степенная, экспоненциальная простая (показательная) и производная от нее логарифмическая линейная, сложная экспоненциальная и производная от нее логарифмическая парабола, гиперболическая и комбинации их видов.

Для моделирования динамических рядов, проявляющих быстрое развитие в начале ряда и замедляющих его развитие к концу, т. е. тех, которые характеризуются стремлением к некоторой предельной величине, применяются *логистические функции*.

Методика спецификации и параметризации эконометрических моделей, построенных по данным временного ряда, сходна с аналогичными этапами эконометрического моделирования, практическое применение которых рассмотрено в предыдущих главах, на примерах пространственных статистических совокупностей.

8.3. Методы измерения устойчивости тенденций динамики

Наиболее простым показателем устойчивости тенденции временного ряда является коэффициент рангов Спирмена:

$$K_S = 1 - \frac{6 \cdot \sum_{i=1}^{n} d^2}{n^3 - n},$$

где d — разность рангов уровней изучаемого ряда P_y и рангов номеров периодов времени P_t;

n — число таких периодов (моментов).

Для определения коэффициентов рангов Спирмена величины уровней изучаемого явления y_i нумеруются в порядке возрастания, а при наличии одинаковых уровней им присваивается определенный ранг, равный частному от деления сумм рангов, приходящихся на эти значения, на число этих равных значений.

При наличии дробных рангов необходима поправка к формуле коэффициентов рангов Спирмена:

$$K_S = 1 - \frac{6 \cdot \sum_{i=1}^{n} d^2 - A}{n^3 - n - 12A}, \tag{8.8}$$

где $A = \frac{1}{12} \sum_{j=1}^{12} \left(A_j^3 - A_j \right)$;

j — номера связок по порядку;

A_j — число одинаковых рангов в j-й связке.

При малой вероятности совпадения уровней и достаточном их числе эта поправка несущественна.

Коэффициенты рангов Спирмена могут принимать значения от 0 до ±1. Если каждый уровень ряда исследуемого периода выше предыдущего, то ранги уровней ряда и номера периодов времени совпадают $K_S = 1$, что означает полную устойчивость факта роста уровней ряда, непрерывность роста.

Чем ближе K_S к 1, тем ближе рост уровней к непрерывному, выше устойчивость роста. При $K_S = 0$ рост совершенно неустойчив, а при отрицательных значениях чем ближе K_S к -1, тем устойчивее снижение изучаемого показателя.

Коэффициент устойчивости роста можно получить и по другой формуле:

$$K_S = \frac{12 \cdot \sum_{i=1}^{n} P_{t_i} \cdot P_{y_i}}{n^3 - n} - \frac{3 \cdot (n+1)}{n - 1}. \tag{8.9}$$

Этот вариант расчета несколько сокращает вычисления. Коэффициент рангов Спирмена здесь применен как новая функция, его нельзя трактовать как меру связи изучаемого явления со временем. Преимуществом коэффициента корреляции рангов как показателя устойчивости является то, что для его вычисления не требуется аналитического выравнивания динамического ряда.

Необходимо помнить, что даже при полной устойчивости роста (снижения) в ряду динамики может быть колеблемость уровней и коэффициент их устойчивости будет ниже 100%. При слабой колеблемости, но еще более слабой тенденции, напротив, возможен высокий коэффициент устойчивости уровней, но близкий к нулю коэффициент устойчивости изменения.

Обычно эти показатели изменяются взаимосвязанно: большей устойчивости уровней соответствует большая устойчивость изменения.

Недостатком коэффициента устойчивого роста является его слабая чувствительность к изменениям скорости роста уровней ряда, он может показать устойчивый рост при незначительно отличающихся от нуля прироста уровней.

В качестве характеристики устойчивости изменения можно применить индекс корреляции:

$$R=\sqrt{1-\frac{\sum\left(y_i-y_x\right)^2}{\sum\left(y_i-\bar{y}\right)^2}}. \qquad (8.10)$$

Индекс корреляции показывает степень сопряженности колебаний исследуемых показателей с совокупностью факторов, изменяющих их во времени. Приближение индекса корреляции к 1 означает большую устойчивость изменения уровней временного ряда.

8.4. Моделирование сезонных и циклических колебаний временного ряда

Для оценки влияния факторов *осциллятивного* характера на изменение уровней временного ряда, вызывающих циклические (конъюнктурные) и сезонные колебания, в методологическом аппарате эконометрики существует достаточно большое количество приемов и методов.

Периодические колебания, имеющие определенный и постоянный период, равный годовому промежутку, называют *сезонными колебаниями*, или *сезонной волной*, а временной ряд называют *тренд-сезонным*, или *сезонным рядом динамики*.

Сезонные колебания характеризуются специальными показателями, которые называются *индексами сезонности* (I_s). Совокупность этих показателей отражает сезонную волну. Индекс сезонности — это процентное соотношение фактических внутригодовых уровней и постоянной или переменной средней.

Для выявления сезонных колебаний, как правило, берут данные за несколько периодов, чтобы выявить устойчивую сезонную волну, не отражающую случайные условия одного из них.

Для исчисления индексов сезонности применяют различные методы. Если ряд динамики не содержит ярко выраженной тенденции развития, то индексы сезонности вычисляются непосредственно по эмпирическим данным без их предварительного выравнивания.

Для каждого периода рассчитывают среднюю величину уровня (как минимум за три периода) $\bar{y}_i$:

$$\bar{y}_1 = \frac{y_{1_1} + y_{1_2} + y_{1_3}}{3},$$

затем из них вычисляется средний уровень всего ряда $\bar{y}$:

$$\bar{y} = \frac{\bar{y}_1 + \bar{y}_2 + \ldots + \bar{y}_n}{n} \text{ или } \bar{y} = \frac{\frac{1}{2}\bar{y}_1 + \bar{y}_2 + \ldots + \bar{y}_{n-1} + \frac{1}{2}\bar{y}_n}{n-1}.$$

И в заключение определяется процентное отношение средних для каждого месяца к общему среднему за месяц уровню ряда:

$$I_s = \frac{\bar{y}_i}{\bar{y}} \cdot 100. \qquad (8.11)$$

Для наглядного представления о сезонной волне рекомендуется графически изобразить полученные данные.

Функцию, заданную в каждой точке изучаемого интервала времени, можно представить бесконечным рядом синусоидальных и косинусоидальных функций. Нахождение конечной суммы уровней с использованием функций косинусов и синусов времени называется *гармоническим анализом*.

Аппроксимация динамики экономических явлений при помощи ряда Фурье состоит в выборе таких гармонических колебаний, наложение которых друг на друга отразит периодические колебания фактических уровней динамического ряда.

С помощью ряда Фурье можно представить динамику явлений в виде некоторой функции времени, в которой слагаемые расположены по убыванию периодов:

$$y_t = a_0 + \sum_{k=1}^{m} (a_k \cos kt + b_k \sin kt). \qquad (8.12)$$

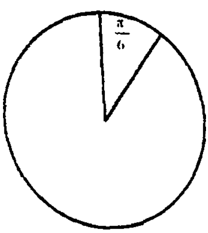
В уравнении Фурье величина k определяет гармонику ряда и может быть взята целым числом (обычно от 1 до 4).

При решении уравнения параметры определяются на основе положений метода наименьших квадратов. Определяя для функции частные производные и приравнивая их к нулю, получают систему нормальных уравнений, параметры которых вычисляются по формулам:

$$a_0 = \frac{\sum y_i}{n};$$

$$a_k = \frac{2}{n} \sum y_i \cos kt;$$

$$b_k = \frac{2}{n} \sum y_i \sin kt.$$



Последовательные значения t обычно определяются от 0 с увеличением (приростом), равным $2\pi/n$, где n — число уровней ряда динамики.

При анализе ряда внутригодовой динамики по месяцам значение n принимается равным 12.

Представляя месячные периоды как части окружности, ряд внутригодовой динамики можно записать в виде

Периоды (t)	0	$\pi/6$	$\pi/3$	$\pi/2$	$2\pi/3$	$5\pi/6$	π	$7\pi/6$	$4\pi/3$	$3\pi/2$	$5\pi/3$	$11\pi/6$
Уровни (y)	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}

Для определения в каждом конкретном случае t находят значения синусов и косинусов разных гармоник, которые для удобства расчетов представлены в таблице 8.1.

Таблица 8.1

Коэффициент гармонического анализа месячных наблюдений для расчета параметров a_k и b_k

t	$\cos t$	$\cos 2t$	$\cos 3t$	$\cos 4t$	$\sin t$	$\sin 2t$	$\sin 3t$	$\sin 4t$
0	1	1	1	1	0	0	0	0
?/6	0,866	0,5	0	–0,5	0,5	0,866	1	0,866
?/3	0,5	–0,5	–1	–0,5	0,866	0,866	0	–0,866
?/2	0	–1	0	1	1	0	–1	0
2?/3	–0,5	–0,5	1	–0,5	0,866	–0,866	0	0,866
5?/6	–0,866	0,5	0	–0,5	0,5	–0,866	1	–0,866
?	–1	1	–1	1	0	0	0	0
7?/6	–0,866	0,5	0	–0,5	–0,5	0,866	–1	0,866
4?/3	–0,5	–0,5	1	–0,5	–0,866	0,866	0	–0,866
3?/2	0	–1	0	1	–1	0	1	0
5?/3	0,5	–0,5	–1	–0,5	–0,866	–0,866	0	0,866
11?/6	0,866	0,5	0	–0,5	–0,5	–0,866	–1	–0,866

Ряд Фурье с двумя гармониками:

$$y_t = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t, \qquad (8.13)$$

где $a_2 = \frac{1}{6}\sum y \cos 2t$; $\quad b_2 = \frac{1}{6}\sum y \sin 2t$.

Эффективно вычислять ряды Фурье можно, когда порядок гармоник является делителем числа $2N$, при этом некоторые из возможных значений тригонометрических функций не требуются. Поэтому на практике часто используются значения $2N$, равные 12, 24 и 60.

8.5. Моделирование тенденции ряда динамики при наличии структурных изменений

От сезонных и циклических колебаний необходимо отличать единовременные изменения характера тенденций временного ряда, вызванные структурными изменениями в экономике или некоторыми другими факторами. В этом случае с некоторого момента времени происходит изменение характера динамики изучаемого показателя, что ведет к изменению параметров тренда, описывающего эту динамику.

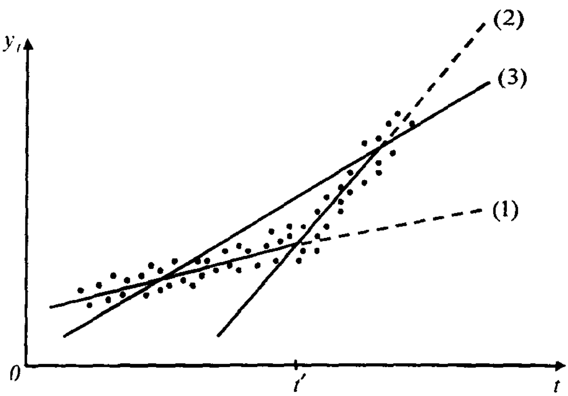


Рис. 8.5. Изменение характера тенденции временного ряда

Момент времени t' сопровождается существенным изменением ряда факторов, оказывающих значительное воздействие на изучаемый показатель y. Часто такие изменения вызваны факторами глобального характера, ведущими к изменению структуры экономики. Если изучаемый ряд динамики включает в себя соответствующий период или момент времени, то одной из задач его изучения становится решение вопроса о значимости влияния структурных изменений на характер этой тенденции.

Если это влияние значимо, то для моделирования тенденции изучаемого ряда динамики используют кусочно-линейные модели регрессии. Для этого разделяют исходную совокупность на две подсовокупности (до момента времени t' и после). Затем строят отдельно по каждой подсовокупности уравнения линейной регрессии (на рис. 8.5 этим уравнениям соответствуют прямые (1) и (2)).

Если же структурные изменения незначительно влияют на характер тенденции изучаемого ряда динамики y_t, то ее можно описывать на основе единого для всей совокупности данных уравнения тренда (на рис. 8.5 этому уравнению соответствует прямая (3)).

У описанных выше подходов есть как положительные, так и отрицательные стороны. Так, при построении кусочно-линейной модели происходит уменьшение остаточной суммы квадратов по сравнению с единым для всей совокупности уравнением тренда. Однако разделение исходной совокупности на две части ведет к потере числа наблюдений, а следовательно, к снижению числа степеней свободы в каждом уравнении кусочно-линейной модели.

При построении общего для всей совокупности уравнения тренда, напротив, сохраняется число наблюдений исходной совокупности, однако остаточная сумма квадратов по этому уравнению выше, чем по кусочно-линейной модели.

Выбор одной из двух моделей зависит от соотношения между сокращением остаточной дисперсии и потерей числа степеней свободы при переходе от общего уравнения регрессии к кусочно-линейной модели.

Для оценки этого соотношения был предложен статистический тест Чоу. Практическое применение этого теста предполагает расчет параметров уравнений тренда, графики которых отображены на рисунке 8.5.

Чоу предложил гипотезу о структурной стабильности тенденции изучаемого ряда динамики.

Остаточную сумму квадратов по кусочно-линейной модели ($S_{ост}^{кл}$) можно найти как сумму:

$$S_{ост}^{кл} = S_{ост}^{1} + S_{ост}^{2}. \qquad (8.14)$$

Ей соответствует число степеней свободы:

$$(n_1 - m_1) + (n_2 - m_2) = (n - m_1 - m_2).$$

Сокращение остаточной суммы квадратов при переходе от единого уравнения тренда к кусочно-линейной модели можно выразить:

$$\Delta S_{ост} = S_{ост}^{3} - S_{ост}^{кл}. \qquad (8.15)$$

Число степеней свободы, соответствующее $\Delta S_{ост}$, будет равно:

$$n - m_3 - (n - m_1 - m_2) = m_1 + m_2 - m_3.$$

В соответствии с предложенной Чоу методикой определяется фактическое значение F-критерия по следующим дисперсиям на одну степень свободы:

$$F_{факт} = \frac{D_{\Delta}}{D_{кл}} = \frac{\dfrac{\Delta S_{ост}}{m_1 + m_2 - m_3}}{\dfrac{S_{ост}^{кл}}{n - m_1 - m_2}}. \qquad (8.16)$$

Полученное значение $F_{факт}$ сравнивают с табличным (критическим), полученным по таблицам распределения Фишера для уровня значимости α и числа степеней свободы $(m_1 + m_2 - m_3)$ и $(n - m_1 - m_2)$.

Если $F_{факт} > F_{крит}$, то гипотеза о структурной стабильности тенденций отклоняется, а влияние структурных изменений на динамику изучаемого показателя признают значимым. В этом случае моделирование временного ряда следует осуществлять на основе кусочно-линейной модели. Если $F_{факт} > F_{крит}$, то основания для принятия нулевой гипотезы о структурной стабильности тенденции существенны. Моделирование тенденции следует осуществлять с помощью единого для всей изучаемой совокупности уравнения тренда.

Если гипотеза о структурной стабильности тенденции ряда y_t отклоняется, дальнейший анализ может заключаться в изучении вопроса о причинах этих структурных различий и детальном изучении характера изменения тенденции. Эти причины обусловливают различия оценок параметров уравнений графика (1) и (2), которые представлены на рисунке 8.6.

Возможны некоторые сочетания изменения численных оценок параметров этих уравнений.

1. Изменение численной оценки свободного члена уравнения тренда a_0^2 по сравнению с a_0^1 при условии, что различия между a_1^2 и a_1^1 статистически незначимы. Геометрически это означает, что прямые (1) и (2) параллельны.

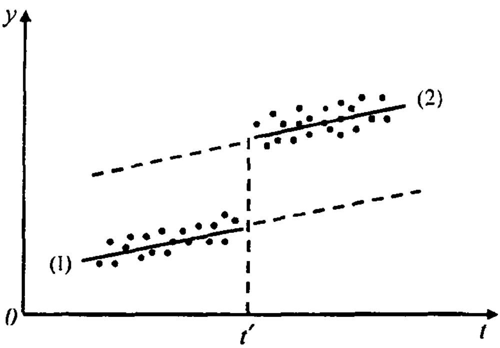


Рис. 8.6. Изменение тенденции ряда динамики при различии между a_0^1 и a_0^2

В этом случае можно говорить о скачкообразном изменении уровней ряда y_t в момент времени t' при неизменном среднем абсолютном приросте за период.

2. Изменение численной оценки параметра a_1^2 по сравнению с a_1^1 при условии, что различия между a_0^1 и a_0^2 статистически не-

значимы. Геометрически это означает, что прямые (1) и (2) пересекают ось ординат в одной точке.

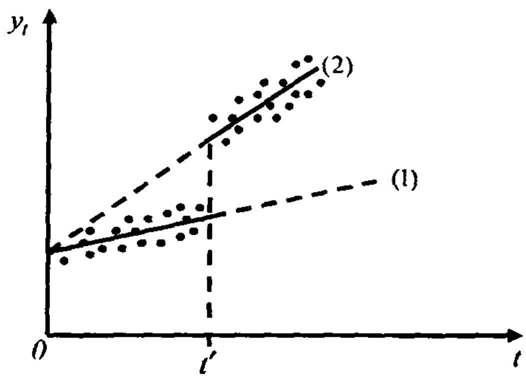


Рис. 8.7. Изменение тенденции временного ряда при статистически значимом различии между a_1^2 и a_1^1

В этом случае изменение тенденции связано с изменением среднего абсолютного прироста ряда динамики, начиная с момента времени t' при неизменном уровне ряда в момент времени $t = 0$.

3. Изменение численных оценок параметров a_0^1 и a_0^2; a_1^1 и a_1^2.

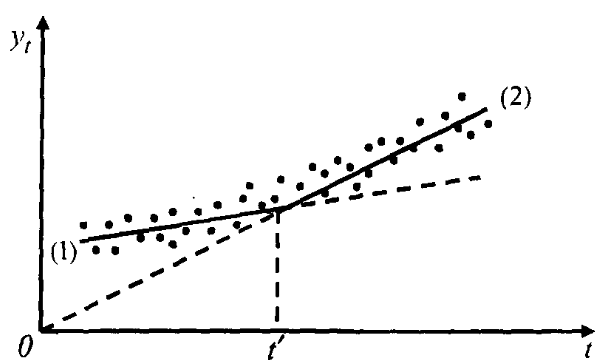


Рис. 8.8. Изменение тенденции временного ряда при статистически значимом различии между a_0^1 и a_0^2: a_1^1 и a_1^2

Геометрически эта ситуация означает, что изменение характера тенденции сопровождается изменением как начального уровня ряда, так и среднего за исследуемый период абсолютного прироста.

Статистический метод для тестирования перечисленных ситуаций изменения тенденции временного ряда был предложен американским экономистом Дамодаром Гуйарати. Этот метод основан на включении в модель регрессии фиктивной переменной Z_t, которая принимает значения, равные 1 для всех $t < t'$, принадлежащих промежутку времени до изменения характера тенденции (1) и равные 0 для всех $t > t'$, принадлежащих промежутку времени после изменения характера тенденции (2).

Гуйарати предлагает определять параметры следующего уравнения регрессии:

$$y_t = A_0 + A_1 Z_t + A_2 t + A_3 (Z_t t) + \Delta_t.$$

Таким образом, для каждого промежутка времени получим следующие уравнения:

$$(1)\, Z = 1 \qquad y_t = A_0 + A_1 + A_2 t + A_3 t + \Delta_t;$$
$$(2)\, Z = 2 \qquad y_t = A_0 + A_2 t + \Delta_t.$$

Сопоставив полученные уравнения с уравнениями (1) и (2), видим:

$$a_0^1 = A_0 + A_1; \qquad a_1^1 = A_2 + A_3;$$
$$a_0^2 = A_0; \qquad a_1^2 = A_2.$$

Таким образом, оценка статистической значимости различий a_0^1 и a_0^2, а также a_1^1 и a_1^2 эквивалентна оценке статистической значимости параметров A_1 и A_3 уравнения Гуйарати (так как $A_1 = a_0^1 - a_0^2$, а $A_3 = a_1^1 - a_1^2$). Эту оценку можно провести при помощи t-критерия Стьюдента.

Следовательно, если в уравнении A_1 является статистически значимым, а A_3 — незначимо, то изменение тенденции вызвано различиями параметров a_0^1 и a_0^2 (рис. 8.6); если наоборот, то изменение характера тенденции вызвано различиями параметров a_1^1 и a_1^2 (рис. 8.7); а если оба коэффициента A_1 и A_3 являются статистически значимыми, то на изменение характера тенденции влияют различия между a_0^1 и a_0^2; a_1^1 и a_1^2 (рис. 8.8).

Этот метод можно использовать не только в дополнение к тесту Чоу, но и самостоятельно для проверки структурной стабильности тенденции изучаемого ряда динамики. Преимущество уравнения Гуйарати в том, что не нужно строить три уравнения тренда, а достаточно одного.

8.6. Корреляционный анализ временных рядов данных

При изучении развития явления или процесса во времени зачастую возникает необходимость оценки взаимосвязи в изменениях уровней двух или более рядов динамики различного содержания, но связанных между собой.

Поставленная задача решается методами коррелирования:

1) уровней ряда динамики;
2) отклонений фактических уровней от тренда;
3) последовательных разностей (путем исчисления парного коэффициента корреляции).

Для изучения взаимосвязи рядов динамики могут использоваться:

1. Парный коэффициент корреляции, показывающий тесноту связи при отсутствии автокорреляции. В этом случае величина коэффициента корреляции рассчитывается по формуле

$$r = \frac{\overline{yx} - \overline{y} \cdot \overline{x}}{\sigma_x \cdot \sigma_y}, \qquad (8.17)$$

где x_i – уровни факторного ряда динамики;

y_i – уровни результативного ряда динамики.

Следовательно, прежде чем изучить тесноту связи между уровнями ряда динамики, требуется проверить каждый из них на наличие или отсутствие автокорреляции. Это достигается при помощи коэффициента автокорреляции. В случае подтверждения наличия автокорреляции между уровнями ряда динамики ее исключают.

2. Парный коэффициент корреляции по отклонениям фактических уровней от выравненных по уравнению регрессии. Этот способ заключается в коррелировании отклонений фактических и выравненных уровней, отражающих общую тенденцию, т.е. коррелируются остаточные величины. Для этого каждый ряд динамики выравнивают по характерной для него аналитической формуле, а затем из эмпирических уровней вычитают выравненные, вычисляя:

$$\Delta_y = y_t - \hat{y}_t \text{ и } \Delta_x = x_t - \hat{x}_t,$$

и определяют тесноту связи между рассчитанными отклонениями по формуле

$$r_{\Delta_y \Delta_x} = \frac{\sum \Delta_y \cdot \Delta_x}{\sqrt{\sum \Delta_y^2 \cdot \sum \Delta_x^2}}. \qquad (8.18)$$

3. Парный коэффициент корреляции по абсолютным отклонениям уровней ряда динамики. Влияние автокорреляции исключается путем вычитания из каждого уровня временного ряда предшествующего, находя последовательные разности уровней:

$$\Delta_{y_1} = y_t - y_{t-1}; \qquad \Delta_{x_1} = x_t - x_{t-1};$$
$$\Delta_{y_2} = y_{t-1} - y_{t-2}; \qquad \Delta_{x_2} = x_{t-1} - x_{t-2};$$
$$\ldots\ldots\ldots\ldots \qquad \ldots\ldots\ldots\ldots$$
$$\Delta_{y_n} = y_{t-n} - y_{t-n-1}; \qquad \Delta_{x_n} = x_{t-n} - x_{t-n-1}.$$

При переходе от уровней к их разностям исключается влияние тенденций на колеблемость. При этом при изменении уровней на прямой можно коррелировать первые разности (Δ_{y_1} и Δ_{x_1}) при изменении по параболе n-го порядка — n-е разности.

Измерение тесноты связи между исследуемыми временными рядами на основе коэффициента разностей осуществляется по формуле

$$r_{\Delta_{y_k} \Delta_{x_k}} = \frac{\sum \Delta_{y_k} \cdot \Delta_{x_k}}{\sqrt{\sum \Delta_{y_k}^2 \sum \Delta_{x_k}^2}}. \qquad (8.19)$$

Коэффициент корреляции, рассчитанный для измерения тесноты зависимости уровней временных рядов, является обобщающим показателем. Однако для продолжительного периода времени эта зависимость непостоянна и меняется во времени. Поэтому рекомендуется рассчитывать скользящие коэффициенты корреляции для определенного промежутка времени с целью выявления силы зависимости (слабая или высокая) между изменениями уровней временных рядов.

8.7. Прогнозирование тенденции временного ряда

Исследование динамики социально-экономических явлений и процессов, выявление и характеристика основного тренда развития и моделей взаимосвязи дают основание для прогнозирования, т.е. определения будущих размеров уровня явления.

Основа прогнозирования — статистические методы. Применение прогнозирования основывается на предположении, что закономерность развития, действовавшая в прошлом внутри ряда динамики, сохранится и в прогнозируемом будущем. Такой прогноз основан на перспективной экстраполяции.

Теоретической основой распространения тенденции является *инертность* социально-экономических явлений. Инертность позволяет выявить взаимосвязь между уровнями динамического ряда или между группой связных временных рядов. Надежные результаты прогнозирования временных рядов получают, если уровни ряда динамики сопоставимы и синтезированы на основе единого методологического подхода.

Применение перспективной экстраполяции в практическом прогнозировании основывается на следующих предпосылках:

1) развитие изучаемого явления графически описывается плавной линией (кривой);
2) общая тенденция развития явления как в прошлом, так и в будущем серьезно не изменяется.

Надежность прогноза зависит от того, как точны эти предположения в действительности, а также как точно охарактеризована выявленная закономерность.

Экстраполяция — это начальная стадия построения прогноза. Механическое, без учета условий, предпосылок и содержательного экономического анализа, применение экстраполяции может стать причиной неадекватных выводов.

Чем шире временной горизонт прогноза, тем очевиднее недостаточность простого метода экстраполяции в результате изменения тенденции, неизвестных точек, влияния новых факторов и т.д. В этом случае динамичность экономических процессов противоречит инертности их развития.

Так как исследуемые ряды динамики часто краткосрочны, то временной горизонт экстраполяции также ограничен. Поэтому чем короче срок экстраполяции (период упреждения), тем более надежны результаты прогноза. За относительно короткий период условия развития явлений не успевают измениться, что сохраняет характер его динамики.

В общем виде экстраполяции можно представить:

$$\hat{y}_{t+T} = f\left(y_t, T, a_i\right), \qquad (8.20)$$

где $\hat{y}_{t+T}$ — прогнозный уровень ряда динамики;

y_t — текущий уровень прогнозируемого ряда динамики;

T — период упреждения;

a_i — параметр уравнения регрессии временного ряда.

В зависимости от принципов и эмпирических данных ряда выделяют следующие методы экстраполяции:

1) по среднему абсолютному приросту;
2) по среднему темпу роста;
3) на основе аналитического выравнивания временного ряда.

Прогнозирование по среднему абсолютному приросту может быть выполнено в случае линейной тенденции развития. Этот метод основывается на предположении о стабильности (равномерности) изменения уровня.

Для экстраполирования по среднему абсолютному приросту необходимо определить средний абсолютный прирост и последовательно увеличивать конечный уровень ряда на его величину на требуемое число периодов экстраполяции:

$$\hat{y}_{t+T} = y_t + \bar{\Delta} \cdot T, \qquad (8.21)$$

где $\hat{y}_{t+T}$ — экстраполируемый уровень ряда;

T — срок прогноза, или период упреждения;

y_t — последний уровень ряда, за который рассчитан $\bar{\Delta}$;

$\bar{\Delta}$ — средний абсолютный прирост.

Следует иметь в виду, что использование среднего абсолютного прироста для прогноза возможно только при соблюдении условия

$$S^2_{\text{ост}} \le p^2, \tag{8.22}$$

где $p^2 = \dfrac{\sum \Delta^2_{\text{ц}}}{2n}$; $S^2_{\text{ост}} = \dfrac{\sum\left(y_t - \bar{y}_{\bar{\Delta}}\right)^2}{n}$.

Экстраполирование по среднему темпу роста осуществляется, когда общая тенденция ряда динамики характеризуется показательной (экспоненциальной) кривой. В этом случае для прогнозирования определяют средний коэффициент роста и возводят его в степень, соответствующую периоду экстраполяции:

$$\hat{y}_{t+T} = y_t \cdot \bar{K}_p^T, \tag{8.23}$$

где $\bar{K}_p$ — средний коэффициент роста.

Если временному ряду соответствует другая закономерность развития, то прогнозные значения, полученные по среднему темпу роста, будут отличаться от рассчитанных другими способами экстраполяции.

Рассмотренные два способа экстраполяции тренда являются самыми простыми и приближенными.

На практике наибольшее распространение получил метод экстраполирования на основе аналитического выравнивания временного ряда. При этом для получения прогнозного значения продолжают значения независимой переменной времени за границы исследуемого периода.

Этот подход прогнозирования предполагает, что уровень ряда динамики формируется под воздействием множества факторов, при этом отдельно влияние каждого из них не выделяется. Следовательно, тенденция развития связана не с каким-либо фактором, а с течением времени:

$$y_t = f(t).$$

На основе применения экстраполяции получают точечные значения прогноза, при этом полное совпадение фактических данных и прогнозных оценок маловероятно. Возникновение отклонений фактических уровней ряда динамики от выравненных по уравнению тренда связано со следующими причинами:

1) всегда существует кривая, которая дает более точные результаты по сравнению с избранной для описания тенденции;
2) кривая, избранная для экстраполяции, содержит случайную компоненту, так как каждый уровень ограниченных исходных данных обладает случайной компонентой;
3) выявленная тенденция характеризует движение среднего уровня ряда динамики, следовательно, возможны отклонения от него.

При использовании методов экстраполяции ввиду их приближенного характера рекомендуется определение доверительных интервалов прогноза:

$$\hat{y}_t \pm t_\alpha \cdot S_{\hat{y}_t}, \qquad (8.24)$$

где t_α – доверительная величина по распределению Стьюдента;

$S_{\hat{y}_t}$ — остаточная средняя квадратическая ошибка тренда, которая рассчитывается по формуле

$$S_{\hat{y}_t} = \sqrt{\frac{\sum(y - y_t)^2}{n-(m+1)}}, \qquad (8.25)$$

где n — число уровней базисного ряда динамики;

m — число параметров адекватной модели тренда.

Вместо t-критерия Стьюдента возможно применение коэффициента Четыркина (k), рассчитанного для соответствующего периода прогнозного интервала. Так, для первого периода во временном горизонте прогноза k_1=2,0153; для второго k_2=2,0621; третьего k_3=2,1131; четвертого k_4=2,1680. Следовательно, величина доверительного интервала определяется по формуле

$$\hat{y}_t \pm k \cdot S_{\hat{y}_t}.$$

Важно помнить, что экстраполяция носит условный характер, поэтому применение методов экстраполяции не является самоцелью. При построении прогнозов социально-экономических явлений следует привлекать дополнительную информацию, на основе которой корректируются количественные оценки, полученные методом экстраполяции.

Контрольные вопросы

1. Понятие и классификация временных рядов.
2. Компонентный анализ рядов динамики.
3. Методология регрессионного анализа тенденции.
4. Гармонический анализ динамического ряда.
5. Методы выявления периодической компоненты.
6. Методы измерения устойчивости тенденций динамики.
7. Моделирование тенденции ряда динамики при наличии структурных изменений.
8. Регрессионный анализ связных динамических рядов.
9. Автокорреляция временных рядов.
10. Критерий Дарбина-Уотсона.

ТЕСТОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ ПРОВЕРКИ ЗНАНИЙ

1. Ряд динамики характеризует:

а) факторы изменения показателя на определенную дату или за определенный период;

б) изменение значений признака во времени;

в) определенное значение варьирующего признака в совокупности;

г) структуру совокупности по какому-либо признаку.

2. Уровни ряда динамики — это:

а) значение варьирующего признака в совокупности;

б) показатели, числовые значения которого составляют динамический ряд.

3. Моментным рядом динамики является:
а) сумма вкладов населения в сберегательные кассы на конец каждого года последнего десятилетия;
б) производительность труда на промышленном предприятии за каждый месяц года;
в) состав населения по национальности на 9 октября 2002 г.;
г) остаток оборотных средств предприятия по состоянию на 1-е число каждого месяца.

4. Разность уровней ряда динамики называется:
а) темпом прироста;
б) темпом роста;
в) абсолютным приростом;
г) коэффициентом роста.

5. Базисный абсолютный прирост равен:
а) произведению цепных абсолютных приростов;
б) корню $n - 1$ степени из произведения цепных абсолютных приростов;
в) корню $n - 1$ степени из суммы абсолютных приростов;
г) сумме цепных абсолютных приростов.

6. Отношение текущего уровня ряда динамики к базисному называется:
а) цепным темпом роста;
б) базисным темпом роста;
в) цепным темпом прироста;
г) базисным темпом прироста;
д) абсолютным значением 1% прироста.

7. Для выявления основной тенденции развития явления используется:
а) метод укрупнения интервалов;
б) индексный метод;
в) метод скользящей средней;
г) расчет средней гармонической;
д) аналитическое выравнивание.

8. Средний уровень интервального ряда динамики с равными временными промежутками исчисляется по формуле средней:

а) хронологической простой;
б) гармонической взвешенной;
в) арифметической взвешенной;
г) гармонической простой;
д) хронологической взвешенной;
е) арифметической простой.

9. Средний уровень интервального ряда динамики с неравными временными промежутками исчисляется по формуле средней:

а) хронологической простой;
б) гармонической взвешенной;
в) арифметической взвешенной;
г) гармонической простой;
д) хронологической взвешенной;
е) арифметической простой.

10. Средний уровень моментного ряда динамики с равными временными промежутками исчисляется по формуле средней:

а) арифметической простой;
б) арифметической взвешенной;
в) гармонической простой;
г) гармонической взвешенной;
д) хронологической простой;
е) хронологической взвешенной.

11. Средний уровень моментного ряда динамики с неравными временными промежутками исчисляется по формуле средней:

а) хронологической простой;
б) гармонической взвешенной;
в) арифметической взвешенной;
г) гармонической простой;
д) хронологической взвешенной;
е) арифметической простой.

12. Уровень временного ряда может содержать:
а) любое сочетание тенденции, циклических, сезонных, случайных колебаний;
б) тенденцию, циклические, сезонные колебания, случайные колебания;
в) тенденцию и сезонные колебания;
г) сезонные и случайные колебания.

13. Аддитивная модель временного ряда имеет вид:
а) $Y_t = T_t + S_t + V_t + \varepsilon_t$;
б) $Y_t = T_t \cdot S_t \cdot V_t \cdot \varepsilon_t$;
в) $Y_t = T_t \cdot S_t \cdot V_t + \varepsilon_t$.

14. Автокорреляцией уравнений временного ряда называют:
а) автокорреляционную зависимость между последовательными уровнями временного ряда;
б) значение аналитической функции, характеризующей зависимость уровней ряда от времени;
в) значение перехода.

15. Автокорреляционная функция временного ряда — это:
а) последовательность коэффициентов автокорреляции уровней временного ряда;
б) коррелограмма;
в) последовательность уровней временного ряда.

ПРАКТИКУМ

Задача 8.1. По данным за 3 года о поголовье крупного рогатого скота в хозяйствах населения, представленным в разрезе кварталов (табл. 8.2), оценить внутригодовые сезонные колебания с помощью индексов сезонности и сделать прогноз исследуемого показателя на следующий год. В ходе решения задачи необходимо:

1) установить наличие тенденции во временном ряду;

2) в случае, если ряд является:

а) стационарным, то индекс сезонности определить как отношение средних уровней ряда в соответствующем периоде к общей средней;

б) нестационарным, то использовать альтернативный способ расчета индексов сезонности. При этом необходимые теоретические значения уровней определить на основании аналитического выравнивания. Сравнить теоретические значения, получаемые по уравнениям тренда различных форм.

Выбрать наилучший с точки зрения статистической корректности ряд, характеризующий основную тенденцию;

3) осуществить прогнозирование объема изучаемого показателя на следующий год с учетом сезонных колебаний по мультипликативной модели.

Таблица 8.2

Поголовье крупного рогатого скота в хозяйствах населения, тыс. гол.

Год	Квартал	Поголовье
2010	1	204,9
	2	198,0
	3	188,9
	4	185,9
2011	1	190,8
	2	198,8
	3	211,6
	4	218,5
2012	1	215,0
	2	201,8
	3	199,2
	4	214,6

Решение

В качестве показателя наличия тенденции ряда динамики используется коэффициент рангов Спирмена (8.6)

Для расчета коэффициента составим вспомогательную таблицу в среде *SPSSStatistics 20.* Вначале создадим количественную переменную *y*, содержащую уровни исследуемого динамического ряда:

- во вкладке «Данные» введем значения переменных;
- во вкладке «Переменные» укажем тип шкалы переменной: «Количественная»;
- в окне «Метка» укажем наименование переменной: «Поголовье».

Для вычисления рангов уровней динамического ряда выполним следующие действия:

- в главном меню выберем «Преобразовать» → «Ранжировать наблюдения»;
- в появившемся диалоговом окне в окне «Переменные» укажем «Поголовье [*y*]» (рис. 8.9), нажмем кнопку «ОК».

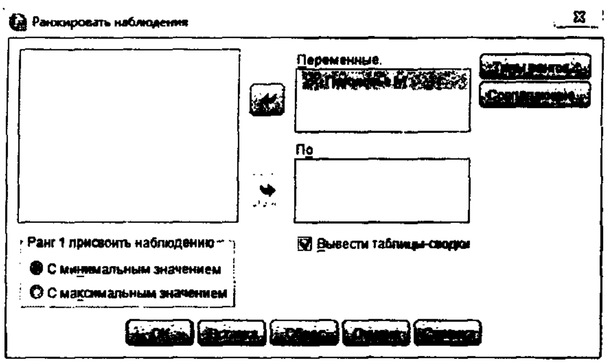


Рис. 8.9. Окно ранжирования наблюдений в *SPSSStatistics 20*

В результате в окне ввода данных появилась новая переменная *Ry*, указывающая ранги каждого наблюдения по порядку возрастания. Обозначим ее как *Py* и отметим количество десятичных знаков: «0».

Далее введем значения хронологических рангов уровней временного ряда и обозначим переменную как *Pt*.

На следующем этапе необходимо вычислить текущие значения d^2 для каждого наблюдения и рассчитать их сумму:

- в главном меню выберем «Преобразовать» → «Вычислить переменную»;
- в окне «Вычисляемая переменная» введем имя переменной: d^2;
- в окне «Числовое выражение» укажем формулу для расчета значений переменной: «$(Py-Pt)$**2» (в данном программном продукте знак «**» используется в качестве знака степени), нажмем кнопку «ОК» (рис. 8.10).

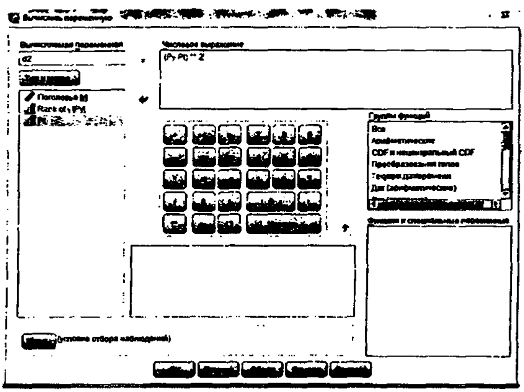


Рис. 8.10. Вычисление значений d^2 в *SPSSStatistics 20*

Полученная вспомогательная таблица изображена на рисунке 8.11.

	204,9	8	1	49
	198,0	4	2	4
	188,9	2	3	1
	185,9	1	4	9
	190,8	3	5	4
	198,8	5	6	1
	211,6	9	7	4
	218,5	12	8	16
	215,0	11	9	4
	201,8	7	10	9
	199,2	6	11	25
	214,6	10	12	4

Рис. 8.11. Вспомогательная таблица в *SPSSStatistics 20*

Далее рассчитаем значение $\sum d^2$. Для этого можно воспользоваться инструментом «Описательные статистики» из меню «Анализ»:

- в окно «Переменные» переместим переменную d^2;
- во вкладке «Параметры» отметим элемент «Сумма», нажмем «ОК».

В результате в окне «Вывод» получим искомое значение $\sum d^2 = 130$.

Рассчитаем коэффициент рангов Спирмена:

$$K_S = 1 - \frac{6 \cdot 130}{1728 - 12} = 0{,}55.$$

Таким образом, ряд является нестационарным с заметно выраженной тенденцией.

Соответственно, индекс сезонности необходимо определять вторым способом, для чего определим сначала теоретические значения y_t с помощью аналитического выравнивания по уравнению тренда.

Воспользуемся несколькими возможными моделями и выберем наилучшую из них с точки зрения статистической адекватности изучаемому процессу.

а) *линейная модель тренда* $y_t = a_0 + a_1 \cdot t$.

Для нахождения параметров воспользуемся *SPSS 20*:

- в окне ввода данных сформируем таблицу исходных данных, содержащую управляющую переменную t и управляемую переменную y (рис. 8.12);

1	204,9
2	198,0
3	188,9
4	185,9
5	190,8
6	198,8
7	211,6
8	218,5
9	215,0
10	201,8
11	199,2
12	214,6

Рис. 8.12. Исходные данные для моделирования линейного тренда в *SPSSStatistics 20*

- в меню «Анализ» выберем «Регрессия» → «Линейная»;
- в появившемся окне «Линейная регрессия» в качестве зависимой переменной укажем «Поголовье [y]», в качестве независимой – «*t*» (рис. 8.13).

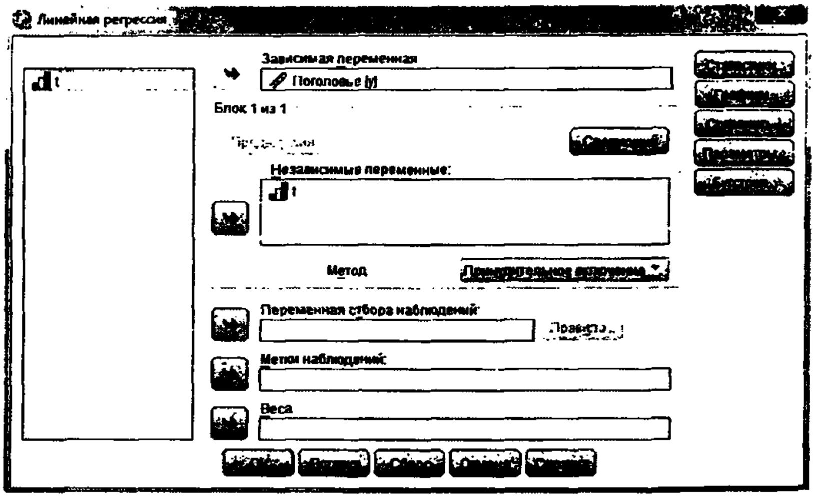


Рис. 8.13. Окно моделирования линейного тренда в *SPSSStatistics 20*

В окне вывода результатов получим значения коэффициентов уравнения линейного тренда (рис. 8.14):

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знч.
		B	Стд. Ошибка	Бета		
1	Константа	192,124	5,957		32,251	,000
	t	1,571	,809	,523	1,940	,081

а. Зависимая переменная: Поголовье

Рис. 8.14. Результаты моделирования линейного тренда в *SPSSStatistics 20*

Таким образом, уравнение линейного тренда имеет вид $y_t = 192{,}12 + 1{,}57t$. Таблица результатов моделирования также содержит значения t-статистик Стьюдента для параметров a_0 и a_1, а также соответствующие им уровни значимости p:

$$t_{a_0} = 32{,}25, p = 0{,}000;$$

$$t_{a_1} = 1{,}94, p = 0{,}081.$$

Указанные характеристики позволяют сделать вывод о статистической значимости найденных параметров, а именно:

- среднее значение поголовья крупного рогатого скота в хозяйствах населения находится на уровне 192,12 тыс. голов, при этом вероятность ошибочного заключения составляет 0,000%;
- средний ежеквартальный прирост численности животных составляет 1,571 тыс. голов, при этом вероятность ошибки составляет 8,1%.

Кроме того, в окне вывода имеется таблица «Сводка для модели», содержащая показатели статистической значимости построенного уравнения тренда в целом (рис. 8.15):

Сводка для модели

Модель	R	R-квадрат	Скорректированный R-квадрат	Стд. ошибка оценки
1	,523[a]	,274	,201	9,6792

а. Предикторы: (конст) t

Рис. 8.15. Оценка качества линейного тренда средствами *SPSSStatistics 20*

Коэффициент корреляции $R = 0{,}52$. Таким образом, теснота связи между значениями уровней ряда и фактором времени умеренная (в соответствии со шкалой Чеддока).

Коэффициент детерминации $R^2 = 0{,}52$, следовательно, вариация уровней ряда на 27% обусловлена влиянием фактора времени.

б) *альтернативные модели тренда.*

В соответствии с полученными результатами можно сделать вывод, что линейная модель тренда не вполне адекватна исследуемому процессу. Поэтому целесообразно провести оценку параметров других форм зависимости, например, логарифмической, степенной, квадратической, экспоненциальной, логистической. Для этого воспользуемся инструментом «Подгонка кривых», который находится в меню «Анализ» → «Регрессия»:

- в диалоговом окне «Подгонка кривых» укажем в качестве зависимой переменной «Поголовье [y]», в качестве независимой переменной отметим «Время», укажем виды искомых форм зависимости;
- отметим пункты «Включить в уравнение константу» и «Графики моделей», нажмем «ОК».

Итоговая таблица результатов имеет вид

Сводка модели и оценки параметров

Зависимая переменная: y

Уравнение	Сводка для модели					Оценки параметров			
	R-квадрат	F	ст.св.1	ст.св.2	Знч.	Константа	b1	b2	b3
Логарифмическая	.158	1.877	1	10	.201	192.849	5.694		
Квадратичный	.281	1.760	2	9	.226	194.732	.453	.086	
Степенная	.155	1.838	1	10	.205	192.883	.028		
Экспоненциальная	.273	3.762	1	10	.081	192.117	.008		
Логистическая	.273	3.762	1	10	.081	.005	.992		

Рис. 8.16. Оценка параметров моделей в *SPSSStatistics 20*

Анализируя полученные результаты, сделаем вывод, что наиболее оптимальной из всех исследованных моделей является квадратическая, поскольку коэффициент детерминации для этого вида регрессии максимальный: R^2=0,281. Соответственно, коэффициент корреляции для этой модели составит: $R=\sqrt{0{,}281}=0{,}53$, что свидетельствует о заметной тесноте связи между результативной переменной y и фактором времени t, а значит, о высоком качестве модели.

Окно вывода результатов моделирования содержит графическое изображение линий тренда, которое также свидетельствует о том, что линия кубического тренда расположена ближе других линий к эмпирическим уровням ряда динамики (рис. 8.17).

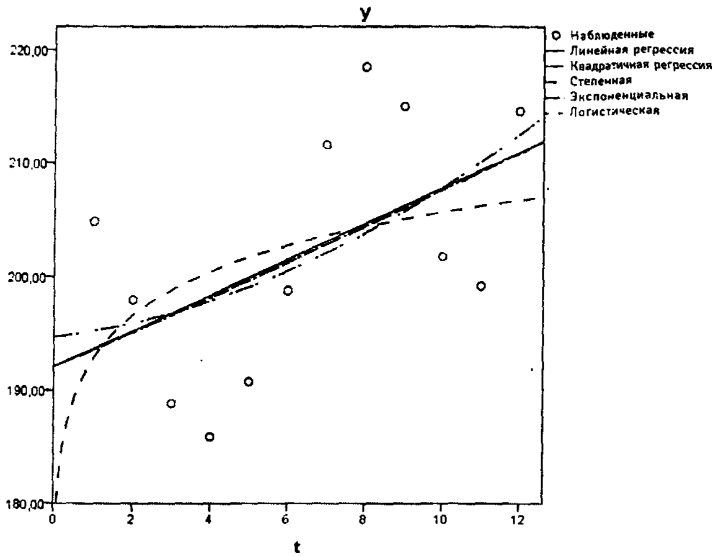


Рис. 8.17. Графическое изображение различных моделей тенденции в *SPSSStatistics 20*

Следовательно, для расчета прогнозных значений исследуемого процесса целесообразно использовать теоретические значения, полученные из кубической модели регрессии вида $y_t = 194{,}73 + 0{,}45 \cdot t + 0{,}86 \cdot t^2$.

Для вычисления теоретических значений вернемся к инструменту «Подгонка кривых», указав только квадратическую функцию в качестве вида модели и отметив пункт «Предсказанные значения» в меню «Сохранить» (рис. 8.18).

В результате в таблице исходных данных появится новая переменная «FIT_1», содержащая теоретические уровни ряда динамики, вычисленные по уравнению кубической модели тренда. Присвоим этой переменной имя «*yt*».

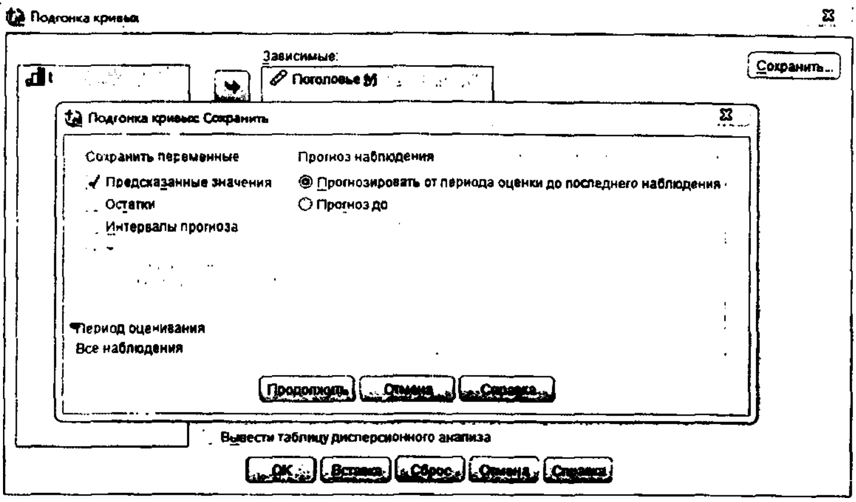


Рис. 8.18. Расчет теоретических значений по квадратическому уравнению тренда средствами *SPSSStatistics 20*

На следующем этапе решения задачи необходимо определить индексы сезонности по формуле

$$I_{S|i|} = 1/m \cdot \sum (y/y_t), \text{ где } m = 3.$$

Для этого вычислим значения новой переменной y/y_t для каждого наблюдения:

- в меню «Преобразовать» выберем пункт «Вычислить переменную»;
- в поле «Вычисляемая переменная» введем имя переменной, например «*yt_y*»;
- в поле «Числовое выражение» введем формулу для расчета переменной: y/yt, нажмем «ОК».

В результате получим необходимые значения для каждой пары эмпирических и теоретических значений уровней исследуемого ряда динамики (рис. 8.19):

[illegible]	1	204,9	195,3	1,05
[illegible]	2	198,0	196,0	1,01
[illegible]	3	188,9	196,9	,96
[illegible]	4	185,9	197,9	,94
[illegible]	5	190,8	199,1	,96
[illegible]	6	198,8	200,5	,99
[illegible]	7	211,6	202,1	1,05
[illegible]	8	218,5	203,9	1,07
[illegible]	9	215,0	205,8	1,04
[illegible]	10	201,8	207,9	,97
[illegible]	11	199,2	210,1	,95
[illegible]	12	214,6	212,5	1,01

Рис. 8.19. Расчет индексов сезонности средствами *SPSSStatistics 20*

Далее определим индексы сезонности для каждого квартала:

$$I_{S[1]} = \frac{1}{3} \cdot (1,05 + 0,96 + 1,04) = 1,02;$$

$$I_{S[2]} = \frac{1}{3} \cdot (1,01 + 0,99 + 0,97) = 0,99;$$

$$I_{S[3]} = \frac{1}{3} \cdot (0,96 + 1,05 + 0,95) = 0,99;$$

$$I_{S[4]} = \frac{1}{3} \cdot (0,94 + 1,07 + 1,01) = 1,007.$$

Определим прогнозное значение показателей объемов оказанных платных услуг населению по каждому кварталу на 2013 год по кубическому уравнению тренда: $y_t = 194,73 + 0,45 \cdot t + 0,86 \cdot t^2$, продолжив ряд времени t на следующие 4 квартала:

$$y^*_{пр2013[1]} = 194,73 + 0,45 \cdot 13 + 0,086 \cdot 13^2 = 215,11,;$$

$$y^*_{пр2013[2]} = 194,73 + 0,45 \cdot 14 + 0,086 \cdot 14^2 = 217,89,;$$

$$y^*_{пр2013[3]} = 194,73 + 0,45 \cdot 15 + 0,086 \cdot 15^2 = 220,83,;$$

$$y^*_{пр2013[4]} = 194,73 + 0,45 \cdot 16 + 0,086 \cdot 16^2 = 223,95.$$

В завершение осуществим корректировку прогнозных значений с учетом индексов сезонности по мультипликативной модели:

$$y_{np[i]} = y^*_{np[i]} \cdot I_{S[i]}.$$

$$y_{np[1]} = 215{,}11 \cdot 1{,}02 = 219{,}4,;$$

$$y_{np[2]} = 217{,}89 \cdot 0{,}99 = 215{,}7;$$

$$y_{np[3]} = 220{,}83 \cdot 0{,}99 = 218{,}6,;$$

$$y_{np[4]} = 223{,}95 \cdot 1{,}007 = 225{,}5.$$

Задача 8.2. По данным о внутригодовой динамике изменения стоимости производимой предприятием продукции построить уравнение Фурье по первой и второй гармоникам; оценить их статистическую значимость и сделать вывод о наиболее приемлемой форме модели для оценки сезонных колебаний анализируемого показателя.

Решение

Для решения поставленной задачи необходимо осуществить параметризацию каждой из моделей и осуществить их корреляционный анализ с помощью индекса корреляции, индекса детерминации, F-критерия Фишера и средней ошибки аппроксимации.

Для построения уравнения Фурье по первой гармонике и оценки его статистической корректности составим и рассчитаем таблицу 8.3.

Общий вид уравнения Фурье по первой гармонике имеет следующий вид:

$$Y_t = a_0 + a_1 \cdot \cos t + b_1 \cdot \sin t.$$

Для оценки параметров уравнения Фурье введем условное обозначение времени (t) в графе 3.

Определим величину прироста переменной t как $\frac{2\pi}{12} = \frac{2\pi}{6} = \frac{\pi}{3}$.

Для определения параметров a_0, a_1, b_1 будем использовать формулы:

$$a_0 = \frac{\sum Y_t}{n}; \quad a_1 = \frac{2}{n} \cdot \sum Y_t \cdot \cos t; \quad b_1 = \frac{2}{n} \cdot \sum Y_t \cdot \sin t.$$

Таблица 8.3

Месяцы	Y	t	ni st	$\sin t$	$Y_t \cdot \cos t$	$Y_t \cdot \sin t$	$Y_t^{(1)}$	$(Y_t - Y_t^1)^2$	$(Y_t - \bar{Y}_t)^2$	$\left\|\frac{Y - Y_t^{(1)}}{Y}\right\|$
1	578	0	1	0	578	0	558,62	375,67	466,56	0,034
2	539	$\frac{\pi}{6}$	0,866	0,5	466,774	269,5	553,64	213,31	302,76	0,027
3	545	$\frac{\pi}{3}$	0,5	0,866	272,5	471,97	549,4	19,006	129,96	0,008
4	552	$\frac{\pi}{2}$	0	1	0	552	547,05	25	19,36	0,0089
5	549	$\frac{2\pi}{3}$	–0,5	0,866	–274,5	475,434	547,2	3,387	54,76	0,0032
6	550	$\frac{5\pi}{6}$	-0,866	0,5	–476,3	275	549,83	0,042107	40,96	0,0003
7	554	π	–1	0	–554	0	554,22	0,04	5,76	0,0004
8	558	$\frac{7\pi}{6}$	-0,866	–0,5	–483,23	- 279	559,2	1,427547	2,56	0,0021
9	564	$\frac{4\pi}{3}$	-0,5	–0,866	- 282	–488,42	563,43	0,313152	57,76	0,001
10	566	$\frac{3\pi}{2}$	0	–1	0	–566	565,79	0,04	92,16	0,0004
11	562	$\frac{5\pi}{3}$	0,5	–0,866	281	–486,69	565,63	13,2525	31,36	0,0065
12	560	$\frac{11\pi}{6}$	0,866	–0,5	484,96	–280	563,01	9,031227	12,96	0,0054
Σ	6677	×	×	×	13,206	–56,212	6677	661,2116	1216,92	0,0971

Для определения параметров a_0 найдем сумму Y_t;

$$a_0 = \frac{6677}{12} = 556,4.$$

Расчет сумм произведений $\sum Y_t \cdot \sin t$, $\sum Y_t \cdot \cos t$ осуществим таблично.

После этого рассчитаем значения параметров a_1 и b_1:

$$a_1 = \frac{1}{6} \cdot 13,3 = 2,2;$$

$$b_1 = \frac{1}{6} \cdot (-56,2) = -9,4.$$

Таким образом, уравнение Фурье по первой гармонике имеет вид

$$Y_t^{(1)} = 556,4 + 2,2 \cdot \cos t - 9,4 \cdot \sin t.$$

В соответствии с ним определим теоретическое значение ряда динамики и отобразим в вышеприведенной таблице.

Оценку тесноты связи проводят с помощью индекса корреляции:

$$R = \sqrt{\frac{\sum (Y_t - Y_t^{(1)})^2}{\sum (Y_t - \bar{Y}_t)^2}} = \sqrt{1 - \frac{661,2116}{1216,92}} = \sqrt{1 - 0,543348} = 0,676.$$

В соответствии со шкалой Чеддока связь между факторами по модели можно охарактеризовать как заметную.

На основании полученного значения индекса корреляции рассчитаем значение коэффициента детерминации:

$$R^2 = (R)^2 = (0,67576)^2 = 0,45.$$

Таким образом, полученная модель уравнения Фурье позволила объяснить изменение стоимости производимой продукции предприятием от сезонных колебаний на 45%.

В заключение следует проверить статистическое значение полученной модели в целом. Для этого необходимо рассчитать значение F-критерия Фишера по формуле

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - (m+1)}{m}, \quad m = 2; \; F = 3,78.$$

Так как для полученной модели F-критерий меньше табличного, то можно сделать вывод, что она является статистически незначимой.

Расчет средней ошибки аппроксимации осуществляется по формуле

$$A=\frac{1}{n}\cdot\left|\sum\frac{Y-Y_t^{(1)}}{Y}\right|\cdot 100\,\%.\quad A=\frac{1}{12}\cdot 0{,}09708\cdot 100\,\%=0{,}8094.$$

Для построения уравнения Фурье по второй гармонике и оценки его статистической корректности составим и рассчитаем таблицу 8.4.

Месяцы	Y	t	$\cos 2t$	$\sin 2t$	$Y_t\cdot\cos 2t$	$Y_t\cdot\sin 2t$	$Y_t^{(2)}$	$(Y_t-Y_t^{2})^2$	$(Y_t-\bar{Y}_t)^2$	$\left\lvert\frac{Y-Y_t^{(2)}}{Y}\right\rvert$
1	578	0	1	0	578	0	559,87	328,78	466,56	0,03137
2	539	$\frac{\pi}{6}$	0,5	0,866	269,5	466,77	552,3	179,252	302,76	0,0248
3	545	$\frac{\pi}{3}$	–0,5	0,866	–272,5	471,97	546,90	3,62525	129,96	0,00349
4	552	$\frac{\pi}{2}$	–1	0	–552	0	545,79	38,464804	19,36	0,01124
5	549	$2\frac{\pi}{3}$	–0,5	–0,866	–274,5	–475,434	548,45	0,299437	54,76	0,00099
6	550	$5\frac{\pi}{6}$	0,5	–0,866	275	–476,3	552,32	5,411007	40,96	0,00422
7	554	π	1	0	554	0	555,46	2,148178	5,76	0,00264
8	558	$7\frac{\pi}{6}$	0,5	0,866	279	483,228	557,94	0,003020	2,56	0,08487
9	564	$4\frac{\pi}{3}$	–0,5	0,866	–282	488,424	560,92	9,427712	57,76	0,00544
10	566	$3\frac{\pi}{2}$	–1	0	–566	0	564,53	2,145248	92,16	0,00258
11	562	$5\frac{\pi}{3}$	–0,5	–0,866	–281	–486,692	566,88	23,817542	31,36	0,00868
12	560	$11\frac{\pi}{6}$	0,5	–0,866	280	–484,96	565,50	30,326564	12,96	0,00983
Σ	6677	×	×	×	7,5	–12,99	6677	623,70246	1216,92	0,10545

Общий вид уравнения Фурье по второй гармонике имеет следующий вид:

$$Y_t = a_0 + a_1 \cdot \cos t + b_1 \cdot \sin t + a_2 \cdot \cos t + b_2 \cdot \sin t.$$

Для определения параметров a_2 и b_2 будем использовать формулы:

$$a_2 = \frac{2}{n} \cdot \sum Y_t \cdot \cos 2t; \quad b_2 = \frac{2}{n} \cdot \sum Y_t \cdot \sin 2t.$$

Расчет сумм произведений $\sum Y_t \cdot \cos 2t$ и $\sum Y_t \cdot \sin 2t$ осуществим таблично.

После этого рассчитаем значения параметров a_2 и b_2:

$$a_2 = \frac{1}{6} \cdot 7{,}5 = 1{,}25; \quad b_2 = \frac{1}{6} \cdot (-12{,}9) = -2{,}165.$$

Таким образом, уравнение Фурье по второй гармонике имеет вид

$$Y_t^{(2)} = 556{,}4 + 2{,}2 \cdot \cos t - 9{,}4 \cdot \sin t + 1{,}25 \cdot \cos 2t - 2{,}165 \cdot \sin 2t.$$

В соответствии с ним определим теоретическое значение ряда динамики и отобразим в приведенной выше таблице.

Оценку тесноты связи проводят с помощью индекса корреляции:

$$R = \sqrt{1 - \frac{\sum (Y_t - Y_t^{(2)})^2}{\sum (Y_t - \overline{Y}_t)^2}} = \sqrt{1 - \frac{623{,}70247}{1216{,}92}} = \sqrt{1 - 0{,}512525} = 0{,}698192772.$$

В соответствии со шкалой Чеддока связь между факторами по модели можно охарактеризовать как заметную.

На основании полученного значения индекса корреляции рассчитаем значение коэффициента детерминации:

$$R^2 = (R)^2 = (0{,}698192772)^2 = 0{,}48.$$

Таким образом, полученная модель уравнения Фурье позволила объяснить изменение стоимости производимой продукции предприятием от сезонных колебаний на 48%.

В заключение необходимо проверить статистическое значение полученной модели в целом. Для этого необходимо рассчитать значение F-критерия Фишера по формуле:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-(m+1)}{m}, \quad m = 4;$$

$$F = 1{,}6644.$$

Так как полученная модель F-критерия меньше табличного, то можно сказать, что полученная модель статистически не значима.

Расчет средней ошибки аппроксимации осуществляется по формуле

$$A = \frac{1}{n} \cdot \left| \sum \frac{Y - Y_t^{(1)}}{Y} \right| \cdot 100\%;$$

$$A = \frac{1}{12} \cdot 0{,}1054 \cdot 100\% = 0{,}8788.$$

Составим обобщающую таблицу 8.5.

Модель	Вид уравнения регрессии	Проверка статистической значимости			Ошибка аппроксимации
		Фактическое значение	Табличное значение	Существенна или несущественна модель	
Уравнение Фурье по первой гармонике	$Y_t^{(1)} = 556{,}4 + 2{,}2 \cdot \cos t - 9{,}4 \cdot \sin t$	3,78	4,26	несущественна	0,8094
Уравнение Фурье по второй гармонике	$Y_t^{(2)} = 556{,}4 + 2{,}2 \cdot \cos t - 9{,}4 \cdot \sin t + 1{,}25 \cdot \cos 2t - 2{,}165 \cdot \sin 2t$	3, 3289	4,12	несущественна	0,8788

В соответствии с величиной ошибки аппроксимации можно сделать вывод о том, что ни одна из моделей не подходит для оценки сезонных колебаний анализируемого показателя.

Задача 8.3. По данным таблицы 8.6 о надое молока, полученном сельскохозяйственными товаропроизводителями (y), и расходе кормов на 1 голову крупного рогатого скота (x) в Ставропольском крае за 1996–2012 годы постройте уравнение регрессии и рассчитайте теоретические значения результативного признака, установите наличие автокорреляции остатков, используя критерий Дарбина–Уотсона, полученную величину сравните с табличной и сделайте вывод. При наличии автокорреляции устраните ее методом включения в уравнение регрессии лаговых переменных в качестве факторных признаков и постройте авторегрессионную модель.

Таблица 8.6

Данные о надое молока, полученном сельскохозяйственными организациями, и расходе кормов на 1 голову крупного рогатого скота в Ставропольском крае за 1996–2012 годы

Год	Надой молока, тыс. т	Расход кормов на одну голову КРС, ц корм. ед.
t	y	x
1996	1066,1	22,6
1997	1014,2	25,6
1998	856,0	22,0
1999	811,0	23,0
2000	786,5	25,1
2001	732,1	22,7
2002	654,1	22,7
2003	572,9	22,4
2004	524,9	20,9
2005	526,7	20,4
2006	542,8	18,6
2007	544,6	20,1
2008	553,4	19,5
2009	568,9	19,2
2010	544,2	23,4
2011	557,1	20,4
2012	574,4	20,2

Решение

Решим задачу с использованием программы *Statistica* 6.0.

На первом этапе построим уравнение линейной регрессии, характеризующее зависимость надоя молока (y) от расхода кормов на 1 голову крупного рогатого скота (x) вида $y_x = a_0 + a_1 x$. Для этого выполним последовательность действий:

- создадим таблицу исходных данных, содержащую данные о временном периоде (t), надое молока (y) и расходе кормов на одну голову КРС (x);
- откроем меню «Анализ» → «Множественная регрессия»;
- в качестве зависимой переменной укажем y, в качестве независимой — x, нажмем «ОК».
- в появившемся диалоговом окне «Результаты множественной регрессии» во вкладке «Быстрый» нажмем «Итоговая таблица регрессии».

Проанализируем полученные результаты моделирования. Итоговая таблица регрессии содержит в столбце «В» значения оцененных параметров и в столбце «t» – t-критерии Стьюдента, характеризующие их статистическую значимость (рис. 8.20).

	Итоги регрессии для зависимой переменной: y (Таблица R= ,67503934 R2= ,45567812 Скорректир. R2= ,41938999 F(1,15)=12,557 p<,00295 Станд. ошибка оценки: 133,60					
N=17	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(15)	p-уров.
Св.член			-608,804	362,9868	-1,67721	0,114211
x	0,675039	0,190494	59,055	16,6652	3,54362	0,002948

Рис. 8.20. Итоговая таблица регрессии в *Statistica 6.0*

Таким образом, уравнение регрессии имеет вид

$$y_x = -608{,}8 + 59{,}06x.$$

Значения t-критериев Стьюдента и их уровни значимости свидетельствуют о значимости коэффициента регрессии a_1 и незначимости a_0. Следовательно, можно утверждать, что увеличение расхода кормов на 1 голову КРС на 1 ц кормовых единиц приводит к увеличению надоя молока на 59,06 тыс. т.

Далее проанализируем таблицу «Итоговые статистики» (рис. 8.21).

Статистика	Итоговые статистики; Значение
Множест. R	0,6750
Множест. R2	0,4557
Скорр. R2	0,4194
F(1,15)	12,5572
p	0,0029
Стд. Ош. Оценки	133,6022

Рис. 8.21. Итоговая таблица регрессии в *Statistica 6.0*

Значение коэффициента корреляции $R = 0{,}68$ свидетельствует о заметной тесноте связи между исследуемыми признаками.

В соответствии с коэффициентом детерминации $R^2 = 0{,}46$ можно сделать вывод, что надой молока на 46% зависит от расхода кормов в расчете на 1 голову КРС, а его значение, скорректированное на число степеней свободы, снижает данную характеристику до 42%.

Для определения критерия Дарбина–Уотсона обратимся к окну «Результаты множественной регрессии» → «Остатки/предсказанные/наблюдаемые значения» → «Анализ остатков» (рис. 8.22).

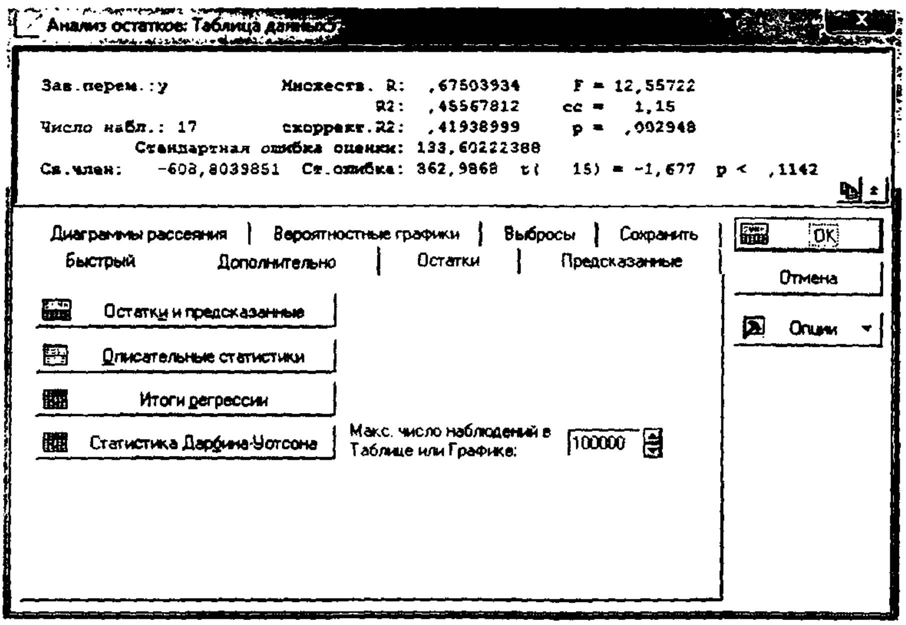


Рис. 8.22. Окно «Анализ остатков» в *Statistica 6.0*

В данном диалоговом окне нажмем «Статистика Дарбина–Уотсона». В результате в окне вывода данных появится таблица «Дабрина–Уотсона *d*» (рис. 8.23).

	Дарбина-Уотсона d и сериальная корреляция остатков	
	Дарбина-Уотсон.d	Сериал. Корр.
Оценка	0.917342	0.325064

Рис. 8.23. Окно «Анализ остатков» в *Statistica 6.0*

Итак, критерий Дарбина–Уотсона $d = 0{,}917$. Полученное значение необходимо сравнить с табличным при 5% уровне значимости (приложение 5). Для $n = 17$(число лет в периоде) и $m = 1$ (число факторов в модели) нижнее значение $d_l = 1{,}13$, а верхняя граница $d_u = 1{,}38$. По полученным значениям числовой промежуток от 0 до 4 поделим на пять отрезков:

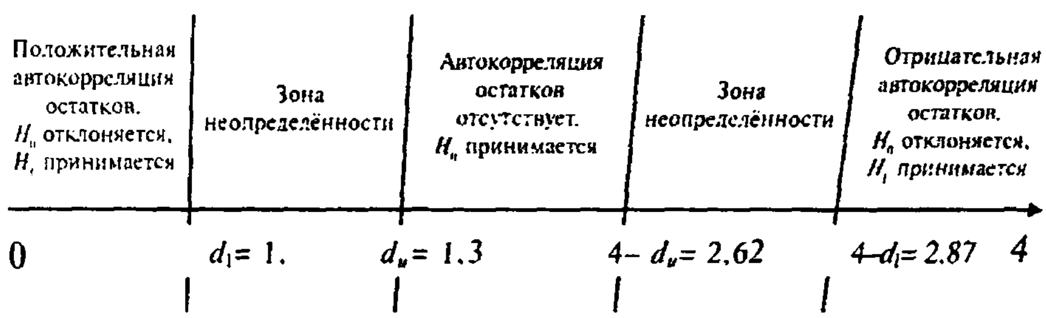


Рис. 8.24. Механизм проверки гипотезы о наличии автокорреляции остатков по критерию Дарбина–Уотсона

Так как фактическое значение критерия попадает в интервал для первого отрезка, то можно считать, что существует положительная автокорреляция остатков. Следовательно, полученное уравнение регрессии, описывающее изменение надоя молока в Ставропольском крае, не может быть использовано для прогноза, поскольку в нем не устранена автокорреляция остатков.

На практике автокорреляция остатков означает, что в уравнение не включен какой-либо существенный фактор. К таким факторам могут относиться, например, технологические или институциональные, воздействие которых на результативный признак характеризуется определенным запаздыванием или инертностью влияния. Следовательно, значение результативного признака в текущий момент времени t формируется под воздействием факторов, действовавших в прошлые моменты времени $t-1$, $t-2$, $t-3$, ..., $t-l$, где l – лаг, характеризующий запаздывание воздействия фактора на результат. В связи с этим предлагается расширить синтезированную модель по числу включенных переменных и по числу факторов, где в качестве лаговой переменной включить значения зависимой переменной y_{t-l}.

В результате моделирования построим авторегрессионную модель надоя молока. Для решения поставленной задачи, требующей расчета коэффициентов автокорреляции l-го порядка, воспользуемся вспомогательной таблицей 8.7.

Рассчитаем коэффициенты автокорреляции уровней ряда по формуле

$$r_a = \frac{\overline{y_t \cdot y_{t-l}} - \overline{y_t} \cdot \overline{y_{t-l}}}{\overline{y_t^2} - \overline{y}^2}.$$

При $l = 1$ коэффициент автокорреляции первого порядка равен:

$$r_1 = \frac{\overline{y_t \cdot y_{t-1}} - \overline{y_t} \cdot \overline{y_{t-1}}}{\overline{y_t^2} - \overline{y}^2} = \frac{\frac{8030886}{17} - \frac{11429{,}9}{17} \cdot \frac{11429{,}9}{17}}{\frac{8176744}{17} - \left(\frac{11429{,}9}{17}\right)^2} = 0{,}70347;$$

при $l = 2$ коэффициент автокорреляции второго порядка равен:

$$r_2 = \frac{\overline{y_t \cdot y_{t-2}} - \overline{y_t} \cdot \overline{y_{t-2}}}{\overline{y_t^2} - \overline{y}^2} = \frac{\frac{7870157}{17} - \frac{11429{,}9}{17} \cdot \frac{11429{,}9}{17}}{\frac{8176744}{17} - \left(\frac{11429{,}9}{17}\right)^2} = 0{,}37671;$$

при $l = 3$ коэффициент автокорреляции третьего порядка равен:

$$r_3 = \frac{\overline{y_t \cdot y_{t-3}} - \overline{y_t} \cdot \overline{y_{t-3}}}{\overline{y_t^2} - \overline{y}^2} = \frac{\frac{7764229}{17} - \frac{11429{,}9}{17} \cdot \frac{11429{,}9}{17}}{\frac{8176744}{17} - \left(\frac{11429{,}9}{17}\right)^2} = 0{,}16136.$$

Таблица 8.7

Данные для расчета значений коэффициентов автокорреляции *i*-го порядка

t	y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_t^2	yy_{t-1}	yy_{t-2}	yy_{t-3}
1996	1066,1	1014,2	856	811	1136569	1081239	912581,6	864607,1
1997	1014,2	856	811	786,5	1028602	868155,2	822516,2	797668,3
1998	856	811	786,5	732,1	732736	694216	673244	626677,6
1999	811	786,5	732,1	654,1	657721	637851,5	593733,1	530475,1
2000	786,5	732,1	654,1	572,9	618582,3	575796,7	514449,7	450585,9
2001	732,1	654,1	572,9	524,9	535970,4	478866,6	419420,1	384279,3
2002	654,1	572,9	524,9	526,7	427846,8	374733,9	343337,1	344514,5
2003	572,9	524,9	526,7	542,8	328214,4	300715,2	301746,4	310970,1
2004	524,9	526,7	542,8	544,6	275520	276464,8	284915,7	285860,5
2005	526,7	542,8	544,6	553,4	277412,9	285892,8	286840,8	291475,8
2006	542,8	544,6	553,4	568,9	294631,8	295608,9	300385,5	308798,9
2007	544,6	553,4	568,9	544,2	296589,2	301381,6	309822,9	296371,3
2008	553,4	568,9	544,2	557,1	306251,6	314829,3	301160,3	308299,1
2009	568,9	544,2	557,1	574,4	323647,2	309595,4	316934,2	326776,2
2010	544,2	557,1	574,4	1066,1	296153,6	303173,8	312588,5	580171,6
2011	557,1	574,4	1066,1	1014,2	310360,4	319998,2	593924,3	565010,8
2012	574,4	1066,1	1014,2	856	329935,4	612367,8	582556,5	491686,4
Итого	11429,9	11429,9	11429,9	11429,9	8176744	8030886	7870157	7764229

В соответствии со значениями коэффициентов автокорреляции связь между результативным признаком y_t и лаговой переменной y_{t-1} можно охарактеризовать как высокую; между y_t и y_{t-2} — как умеренную; между y_t и y_{t-3} – как слабую. Таким образом, в авторегрессионную модель следует включить лаговую переменную y_{t-1} в качестве дополнительного фактора. Тогда модель будет иметь вид $y_x = a_0 + a_1 \cdot x + a_2 \cdot y_{t-1}$.

Для оценки параметров и статистической адекватности уравнения авторегрессии выполним следующие шаги:

- составим таблицу исходных данных (рис. 8.25);

	1 t	2 y	3 x	4 y t 1
1	1996	1066,1	22,6	1014,2
2	1997	1014,2	25,6	856
3	1998	856	22	811
4	1999	811	23	786,5
5	2000	786,5	25,1	732,1
6	2001	732,1	22,7	654,1
7	2002	654,1	22,7	572,9
8	2003	572,9	22,4	524,9
9	2004	524,9	20,9	526,7
10	2005	526,7	20,4	542,8
11	2006	542,8	18,6	544,6
12	2007	544,6	20,1	553,4
13	2008	553,4	19,5	568,9
14	2009	568,9	19,2	544,2
15	2010	544,2	23,4	557,1
16	2011	557,1	20,4	574,4
17	2012	574,4	20,2	1066,1

Рис. 8.25. Таблица исходных данных для авторегрессионной модели в *Statistica 6.0*

- в диалоговом окне «Множественная регрессия» укажем в качестве зависимой переменной переменную «*y*», а в качестве независимых – переменные «*x*» и «*y_t_1*», нажмем кнопку «OK».

В появившемся окне вывода результатов во вкладке «Быстрый» в таблице «Коэффициенты» найдем пункт «Итоговая таблица регрессии», в которой указаны значения параметров атoрегрессионной модели (рис. 8.26), которая имеет вид

$$y_x = -615{,}23 + 42{,}83x + 0{,}53y_{t-1}.$$

Итоги регрессии для зависимой переменной y (Таблица данных
R= .83995377 R2= .70552233 Скорректир. R2= .66345409
F(2.14)=16.771 p<,00019 Станд ошибка оценки 101 72

N=17	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(14)	p-уров.
Св.член			[illegible]	[illegible]	[illegible]	[illegible]
x	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
y_t_1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Рис. 8.26. Коэффициенты и статистическая значимость авторегрессионной модели в *Statistica 6.0*

Кроме того, данная таблица содержит следующую информацию:

$$R = 0{,}84;$$

$$R^2 = 0{,}71;$$

$$S = 101{,}72;$$

$$F = 16{,}77.$$

Таким образом, можно сделать вывод об улучшении качества авторегрессионной функции, аппроксимированной при помощи включения в исходное уравнение парной линейной регрессии лаговой переменной y_{t-1}.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 8.4. По данным за 3 года об объеме платных услуг, оказанных населению, представленным в разрезе кварталов (табл. 8.8), оценить внутригодовые сезонные колебания с помощью индексов сезонности и сделать прогноз исследуемого показателя на следующий год. В ходе решения задачи необходимо:

1) установить (классифицировать) временной ряд на наличие тенденции, то есть выяснить, является ли он стационарным или нестационарным, с помощью коэффициента рангов Спирмена;

2) в случае если ряд является:

а) стационарным, то индекс сезонности определить как отношение средних уровней ряда в соответствующем периоде к общей средней;

б) нестационарным, то использовать альтернативный способ расчета индексов сезонности. При этом необходимые теоретические значения уровней определить на основании аналитического выравнивания. Сравнить теоретические значения, получаемые по линейному уравнению тренда $y_t = a_0 + a_1 \cdot t$ и по степенному $y_t = a_0 + t^{a_1}$.

Выбрать наилучший с точки зрения статистической корректности ряд, характеризующий основную тенденцию;

3) осуществить прогнозирование объема оказываемых платных услуг на следующий год с учетом сезонных колебаний по мультипликативной модели.

Таблица 8.8

Объем платных услуг, оказанных населению, млн руб.

Год	Квартал	Объем услуг, млн руб.
2010	1	2228
	2	1810
	3	2781
	4	2874
2011	1	2693
	2	2998
	3	3050
	4	3295
2012	1	3328
	2	3638
	3	3716
	4	3942
Σ	×	36353

Задача 8.5. По данным за 3 года о поголовье свиней в хозяйствах фермеров, представленным в разрезе кварталов (табл. 8.9), оценить внутригодовые сезонные колебания с помощью индексов сезонности и сделать прогноз исследуемого показателя на следующий год. В ходе решения задачи необходимо:

1) установить (классифицировать) временной ряд на наличие тенденции;

2) в случае если ряд является:

а) стационарным, то индекс сезонности определить как отношение средних уровней ряда в соответствующем периоде к общей средней;

б) нестационарным, то использовать альтернативный способ расчета индексов сезонности. При этом необходимые теоретические значения уровней определить на основании аналитического выравнивания. Сравнить теоретические значения, получаемые по линейному уравнению тренда $y_t = a_0 + a_1 \cdot t$, степенному $y_t = a_0 \cdot t^{a_1}$, показательному $y_t = a_0 \cdot a_1^t$.

Выбрать наилучший с точки зрения статистической корректности ряд, характеризующий основную тенденцию;

3) осуществить прогнозирование объема изучаемого показателя на следующий год с учетом сезонных колебаний по мультипликативной модели.

Таблица 8.9

Поголовье свиней в крестьянских (фермерских) хозяйствах, тыс. голов

Год	Квартал	Поголовье
2010	1	17,1
	2	16.5
	3	16,4
	4	18,0
2011	1	16,0
	2	17,1
	3	16,3
	4	16,6
2012	1	19,4
	2	18,0
	3	17,9
	4	18,8

Задача 8.6. По данным за 3 года об урожайности овощей закрытого грунта в сельскохозяйственных предприятиях региона, представленным в разрезе кварталов (табл. 8.10), оценить внутригодовые сезонные колебания с помощью индексов сезонности и сделать прогноз исследуемого показателя на следующий год. В ходе решения задачи необходимо:

1) установить (классифицировать) временной ряд на наличие тенденции;

2) в случае если ряд является:

а) стационарным, то индекс сезонности определить как отношение средних уровней ряда в соответствующем периоде к общей средней;

б) нестационарным, то использовать альтернативный способ расчета индексов сезонности. При этом необходимые теоретические значения уровней определить на основании аналитического выравнивания. Сравнить теоретические значения, получаемые по линейному уравнению тренда $y_t = a_0 + a_1 \cdot t$, степенному $y_t = a_0 \cdot t^{a_1}$, показательному $y_t = a_0 \cdot a_1^t$.

Выбрать наилучший с точки зрения статистической корректности ряд, характеризующий основную тенденцию;

3) осуществить прогнозирование объема изучаемого показателя на следующий год с учетом сезонных колебаний по мультипликативной модели.

Таблица 8.10

Урожайность овощей закрытого грунта в сельскохозяйственных предприятиях региона, ц/га

Год	Квартал	Урожайность
2010	1	89,4
	2	95,3
	3	99,4
	4	86,4
2011	1	95,5
	2	106,2
	3	122,8
	4	115,4

Окончание табл. 8.10

Год	Квартал	Урожайность
2012	1	103,5
	2	126,4
	3	157,2
	4	138,5

Задача 8.7. По данным о внутригодовой динамике изменения индексов физического объема продукции животноводства (табл. 8.11) построить уравнение Фурье по первой и второй гармоникам, оценить их статистическую значимость и сделать вывод о наиболее приемлемой форме модели для оценки сезонных колебаний анализируемого показателя.

Таблица 8.11

Индексы физического объема продукции животноводства в регионе в 2012 году, %

Месяцы года	Индексы(Y)	Месяцы года	Индексы (Y)
1	96,9	7	101,4
2	89,9	8	106,5
3	91,4	9	98,0
4	95,0	10	98,4
5	98,8	11	103,1
6	102,6	12	105,2

Задача 8.8. По данным о внутригодовой динамике изменения показателя производства яиц в хозяйствах населения региона (табл. 8.12) построить уравнения Фурье по первой и второй гармоникам, оценить их статистическую значимость и сделать вывод о наиболее приемлемой форме модели для оценки сезонных колебаний анализируемого показателя.

Таблица 8.12

Производство яиц в хозяйствах населения в 2012 г., тыс. штук

Месяцы года	Производство (Y)
1	636,0
2	600,0
3	605,8

Окончание табл. 8.12

Месяцы года	Производство (*Y*)
4	531,0
5	491,9
6	482,3
7	483,4
8	465,0
9	462,2
10	454,0
11	442,1
12	464,0

Задача 8.9. По данным таблицы 8.13 об изменении объема валового сбора овощей (y) и о внесении минеральных удобрений на 1 га удобренной площади (x) в Ставропольском крае за 1996–2012 годы постройте уравнение регрессии и рассчитайте теоретические значения результативного признака, определите автокорреляцию остатков, используя критерий Дарбина–Уотсона, полученную величину сравните с табличной и сделайте вывод.

Таблица 8.13

Данные об объеме валового сбора овощей и о внесении минеральных удобрений на 1 га удобренной площади в Ставропольском крае за 1996–2012 годы

Год	Валовой сбор овощей, тыс. т	Внесено минеральных удобрений на 1 га удобренной площади, кг
t	y	x
1996	4511,7	176
1997	3839,7	127
1998	3558,8	143
1999	3827,4	158
2000	2994,4	210
2001	2834,1	134
2002	2504,1	139
2003	2970,9	123
2004	2759,4	90

Окончание табл. 8.13

Год	Валовой сбор овощей, тыс. т	Внесено минеральных удобрений на 1 га удобренной площади, кг
t	y	x
2005	2346.5	144
2006	2740.5	123
2007	3541,8	136
2008	4670,6	104
2009	2978,6	94
2010	4866,8	158
2011	5748,6	134
2012	5108,3	127

Задача 8.10. По данным таблицы 8.14 о приплоде телят, полученном сельскохозяйственными товаропроизводителями, (y) и падеже крупного рогатого скота в процентах к обороту стада (x) в Ставропольском крае за 1996–2012 годы постройте уравнение регрессии и рассчитайте теоретические значения результативного признака, определите автокорреляцию остатков, используя критерий Дарбина – Уотсона, полученную величину сравните с табличной и сделайте вывод.

Таблица 8.14

Данные о приплоде телят, полученном сельскохозяйственными товаропроизводителями, и падеже крупного рогатого скота в процентах к обороту стада в Ставропольском крае за 1996–2012 годы

Год	Приплод телят, тыс. гол.	Падеж крупного рогатого скота в % к обороту стада, %
t	y	x
1996	207,7	1,8
1997	195,3	2,2
1998	164,0	3,4
1999	138,4	3,1
2000	113,9	3,5
2001	92,8	3,9

Окончание табл. 8.14

Год	Приплод телят, тыс. гол.	Падеж крупного рогатого скота в % к обороту стада, %
t	y	x
2002	89,2	4,1
2003	81,4	3,9
2004	73,4	2,9
2005	67,8	1,9
2006	63,3	1,8
2007	51,7	1,5
2008	43,9	1.6
2009	41,0	2,0
2010	33,2	2,0
2011	42,1	1,6
2012	43,9	1,3

Задача 8.10. По данным таблицы 8.15 о валовом сборе зерновых и зернобобовых культур в хозяйствах всех категорий Ставропольского края за 1913–2008 годы постройте авторегрессионную модель с лаговой переменной.

Таблица 8.15

Данные о валовом сборе зерновых и зернобобовых культур в хозяйствах всех категорий Ставропольского края за 1913–2008 годы, тыс. т

Год	Валовой сбор	Год	Валовой сбор
1913	2082,2	1972	2589,1
1928	929,8	1973	4530,9
1932	1138,2	1974	3163,1
1940	1928,9	1975	2270,2
1945	834,5	1976	2261,1
1950	1225,4	1977	4505,2
1955	1661,7	1978	5222,7
1960	3525,6	1979	2015,8
1965	2650,3	1980	4143,8
1970	4186,6	1981	4849,2
1971	4211,0	1982	3806,7

Окончание табл. 8.15

Год	Валовой сбор	Год	Валовой сбор
1983	3603,5	1996	3398,6
1984	3965,0	1997	3838,9
1985	2648,0	1998	3398,6
1986	4766,7	1999	3106,0
1987	3949,0	2000	3788,2
1988	4969,8	2001	4908,6
1989	5464,0	2002	6297,8
1990	6223,5	2003	3851,8
1991	5373,2	2004	6375,7
1992	5102,6	2005	6872,2
1993	5416,2	2006	6426,4
1994	4023,0	2007	6787,4
1995	3973,6	2008	7034,9

Задача 8.11. По данным таблицы 8.16 о валовом сборе сахарной свеклы в хозяйствах всех категорий Ставропольского края за 1960–2008 годы постройте авторегрессионную модель с лаговой переменной.

Таблица 8.16

Данные о валовом сборе сахарной свеклы в хозяйствах всех категорий Ставропольского края за 1960–2008 годы, тыс. т

Год	Валовой сбор	Год	Валовой сбор
1960	162,7	1982	867,9
1965	605,0	1983	648,1
1970	653,0	1984	764,3
1971	429,0	1986	777,0
1972	716,1	1987	630,3
1973	832,2	1988	760,6
1974	748,7	1989	997,6
1975	794,7	1990	910,6
1976	803,2	1991	807,7
1977	732,3	1992	672,1
1978	720,2	1993	798,9
1979	444,5	1994	521,6
1980	591,3	1995	609,0
1981	562,5	1996	647,1

Окончание табл. 8.16

Год	Валовой сбор	Год	Валовой сбор
1997	557,4	2003	484,0
1998	512,3	2004	929,1
1999	468,4	2005	721,7
2000	346,1	2006	1057,2
2001	456,7	2007	1078,2
2002	685,1	2008	1112,4

ГЛАВА

9 ЭКОНОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ СЛОЖНЫХ ДИНАМИЧЕСКИХ СИСТЕМ

9.1. Характеристика классов динамических эконометрических моделей

В эконометрике к числу динамических моделей относят не все модели, построенные по временным рядам данных. Динамические модели характеризуются каждым отдельным моментом времени t в отдельности, а не всем периодом.

Эконометрическую модель называют динамической, если в данный момент времени t она учитывает значения входящих в нее переменных, относящихся как к текущему, так и к предыдущим моментам времени, т.е. если эта модель отражает динамику последующих переменных в каждый момент времени.

Таким образом, при изучении зависимостей между показателями, для анализа развития во времени которых в качестве объясняющих переменных используются как текущие значения переменных, так и предыдущие во времени, а также само время T, используются динамические модели.

При исследовании социально-экономических процессов зачастую требуется моделировать ситуации, когда значение результативного признака в текущий момент времени t формируется под воздействием факторов, действовавших в прошлые моменты времени $t-1$, $t-2, t-3,..., t-l$. Величину l называют *лагом*, характеризующим запаздывание воздействия фактора на результат. В свою очередь переменные, влияние которых характеризуется определенным запаздыванием, называют *лаговыми переменными*.

Как правило, динамические модели подразделяются на два вида:

1. Модели с распределенным лагом – это модели, содержащие в качестве лаговых переменных независимые переменные. Эта модель имеет вид

$$y_t = a_0 + \alpha_0 \cdot x_t + \alpha_1 \cdot x_{t-1} + \ldots + \alpha_n \cdot x_{t-l} + \Delta_t. \qquad (9.1)$$

2. Авторегрессионные модели — это модели, уравнения которых в качестве лаговых переменных включают значения зависимых переменных. Модель авторегрессии имеет вид

$$y_t = a_0 + \alpha_0 \cdot x_t + \beta_1 \cdot y_{t-1} + \Delta_t. \tag{9.2}$$

В эконометрическом анализе динамические модели используются широко, так как воздействие ряда экономических факторов на другие осуществляется не мгновенно, а с запаздыванием. В качестве факторов этого запаздывания можно назвать:

1) психологические факторы, выражающиеся в инертности поведения людей, так, например, люди тратят свои доходы постепенно;
2) технологические факторы;
3) институциональные факторы;
4) механизмы формирования экономических показателей, например появление денег в банковской системе проявляет себя через определенный интервал времени.

Синтезирование динамических моделей имеет свои особенности:

1) оценка параметров авторегрессионных моделей, а иногда и моделей с распределенным лагом не может быть произведена с помощью МНК из-за нарушения его предпосылок, а требует применения специальных методов параметризации;

2) необходимо выбрать оптимальную величину лага и определить его структуру;

3) между двумя видами динамических моделей существует взаимосвязь, в результате которой требуется осуществлять переход от одной к другой.

9.2. Регрессионный анализ связных динамических рядов

Многомерные временные ряды, характеризующие зависимость результативного признака от одного или нескольких факторов, называют *связными рядами динамики*. Применение метода наименьших квадратов для обработки рядов динамики не требует выдвижения никаких предположений о законах распределения исходных данных. Однако при использовании метода наименьших квадратов для обработки связных рядов следует учитывать наличие автокорреляции (авторегрессии), которая не учитывалась при обработке одномерных рядов динамики, поскольку ее наличие способствовало более плотному и четкому выявлению тенденции развития изучаемого социально-экономического явления во времени.

В рядах динамики социально-экономических процессов между близко расположенными уровнями существует взаимосвязь. Это явление удобно представить в виде корреляционной зависимости между рядами:

$$y_1, y_2, y_3, ..., y_n$$

и этим же рядом, сдвинутым относительно первоначального положения на h моментов времени:

$$y_{1+h}, y_{2+h}, y_{3+h}, ..., y_{n+h}.$$

Временное смещение L называется *сдвигом*, а само явление взаимосвязи — *автокорреляцией*.

Автокорреляционная зависимость особенно существенна между последующими и предыдущими уровнями временного ряда. Так как классические методы математической статистики применяются лишь в случае независимости отдельных членов ряда между собой, то при анализе нескольких взаимосвязанных рядов динамики необходимо установить наличие и степень их автокорреляции.

Различают два вида автокорреляции:

- автокорреляция в наблюдениях за одной или более переменными;
- автокорреляция ошибок или автокорреляция в отклонениях от тренда.

Наличие автокорреляции остатка приводит к искажению величин средних квадратических ошибок коэффициентов регрессии, что затрудняет построение доверительных интервалов для коэффициентов регрессии, а также проверку их значимости.

Автокорреляцию измеряют при помощи нециклического коэффициента автокорреляции, который может рассчитываться не только между соседними уровнями, т. е. сдвинутыми на один период, но и между сдвинутыми на любое число единиц времени L.

Этот сдвиг (*временной лаг*) определяет также порядок коэффициентов автокорреляции:

$L = 1$ — коэффициент автокорреляции первого порядка;

$L = 2$ — коэффициент автокорреляции второго порядка и т. д.

На практике наибольшее искажение результатов анализа возникает при корреляции между исходными уровнями ряда y_t и теми же уровнями, сдвинутыми на единицу y_{t-1} или y_{t+1}. Поэтому наибольший интерес представляет вычисление нециклического коэффициента автокорреляции первого порядка.

Следовательно, коэффициент автокорреляции можно определить по формуле

$$r_a = \frac{\overline{y_t \cdot y_{t+1}} - \overline{y}_t \cdot \overline{y}_{t+1}}{\sigma y_t \cdot \sigma y_{t+1}}, \tag{9.3}$$

где σy_t — среднее квадратическое отклонение рядов y_t;

σy_{t+1} — среднее квадратическое отклонение рядов y_{t+1}.

Для суждения о наличии (отсутствии) автокорреляции в исследуемом ряду фактическое значение коэффициента автокорреляции сопоставляется с табличным (критическим) для соответствующего уровня значимости. Критическая область проверяемой гипотезы об отсутствии автокорреляции приведена в специальной таблице, составленной *Андерсеном*.

Если фактическое значение коэффициента автокорреляции меньше табличного, то гипотеза об отсутствии автокорреляции в ряду может быть принята, а если наоборот, то делают вывод о наличии автокорреляции во временном ряду.

Последовательность коэффициента автокорреляции уровней первого, второго и т. д. порядков называют *автокорреляционной функцией ряда динамики*, а график зависимости ее значений от величины лага называется *коррелограммой*.

Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором автокорреляция наиболее высокая, а соответственно, и лаг, при котором связь между текущим и предыдущим (последующем) уровнями ряда наиболее тесная. При помощи анализа автокорреляционной функции и коррелограммы выявляют структуру ряда.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый временной ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции n-го порядка, то ряд содержит периодические колебания в n моментов времени. Если же ни один из коэффициентов автокорреляции не является значимым, то делают предположения относительно структуры ряда:

– ряд не содержит тенденцию и периодические колебания;

– ряд содержит сильную нелинейную тенденцию, для выявления которой проводят дополнительный анализ.

Известен метод определения автокорреляции остатков, основанный на использовании критерия Дарбина–Уотсона, который рассчитывается по формуле

$$d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t+1})^2}{\sum_{t=1}^{n}\varepsilon_t^2}. \qquad (9.4)$$

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина–Уотсона следующий. Выдвигается гипотеза H_0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H_1 и H'_1 состоят соответственно в наличии положительной или отрицательной автокорреляции в остатках. Далее по специальным таблицам определяют критические значения критерия Дарбина–Уотсона d_l и d_u для заданного числа наблюдений n, числа независимых переменных модели и уровня значимости a. По этим значениям числовой промежуток от 0 до 4 разбивают на пять отрезков (рис. 9.1).

Рис. 9.1. Механизм проверки гипотезы о наличии автокорреляции остатков

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

Для исключения или уменьшения автокорреляции в рядах динамики могут использоваться две группы методов:

1. Методы, основанные на преобразовании уровней исходного ряда в новые переменные, не содержащие тенденции, которые используются в дальнейшем для анализа взаимосвязи изучаемых рядов динамики. Эта группа методов предполагает устранение трендовой компоненты из каждого уровня временного ряда. В этой группе наиболее широко распространены метод последовательных разностей и метод отклонения от трендов.
2. Методы, основанные на изучении взаимосвязи исходных уровней рядов динамики при элиминировании воздействия фактора времени на зависимую и независимые переменные модели. К этой группе методов в первую очередь относится метод включения времени в модель регрессии в качестве дополнительного фактора.

При изучении развития явления или процесса во времени зачастую возникает необходимость оценки взаимосвязи в изменениях уровней двух или более рядов динамики различного содержания, но связанных между собой.

Поставленная задача решается методами коррелирования:

- уровней ряда динамики;

- отклонений фактических уровней от тренда;
- последовательных разностей (путем исчисления парного коэффициента корреляции).

Для изучения взаимосвязи рядов динамики могут использоваться:

- парный коэффициент корреляции, показывающий тесноту связи при отсутствии автокорреляции. В этом случае величина коэффициента корреляции рассчитывается по формуле

$$r = \frac{\overline{yx} - \overline{y} \cdot \overline{x}}{\sigma_x \cdot \sigma_y}, \tag{9.5}$$

где x_i — уровни факторного ряда динамики;

y_i — уровни результативного ряда динамики.

Следовательно, прежде чем изучить тесноту связи между уровнями ряда динамики, требуется проверить каждый из них на наличие или отсутствие автокорреляции. Это достигается при помощи коэффициента автокорреляции. В случае подтверждения наличия автокорреляции между уровнями ряда динамики ее исключают;

- парный коэффициент корреляции по отклонениям фактических уровней от выравненных по уравнению регрессии. Этот способ заключается в коррелировании отклонений фактических и выравненных уровней, отражающих общую тенденцию, т. е. коррелируются остаточные величины. Для этого каждый ряд динамики выравнивают по характерной для него аналитической формуле, а затем из эмпирических уровней вычитают выровненные, вычисляя:

$$\Delta_y = y_t - \hat{y}_t \text{ и } \Delta_x = x_t - \hat{x}_t,$$

и определяют тесноту связи между рассчитанными отклонениями по формуле

$$r_{\Delta_y\Delta_x} = \frac{\sum \Delta_y \cdot \Delta_x}{\sqrt{\sum \Delta_y^2 \cdot \sum \Delta_x^2}}; \tag{9.6}$$

- парный коэффициент корреляции по абсолютным отклонениям уровней ряда динамики. Влияние автокорреляции исключается путем вычитания из каждого уровня временного ряда предшествующего, находя последовательные разности уровней:

$$\Delta_{y_1} = y_t - y_{t-1}; \qquad \Delta_{x_1} = x_t - x_{t-1};$$
$$\Delta_{y_2} = y_{t-1} - y_{t-2}; \qquad \Delta_{x_2} = x_{t-1} - x_{t-2};$$
$$\ldots\ldots\ldots\ldots\ldots\ldots \qquad \ldots\ldots\ldots\ldots\ldots\ldots$$
$$\Delta_{y_n} = y_{t-n} - y_{t-n-1}; \qquad \Delta_{x_n} = x_{t-n} - x_{t-n-1}.$$

При переходе от уровней к их разностям исключается влияние тенденций на колеблемость. При этом при изменении уровней на прямой можно коррелировать первые разности (Δ_{y_i} и Δ_{x_i}) при изменении по параболе n-го порядка n-е разности.

Измерение тесноты связи между исследуемыми временными рядами на основе коэффициента разностей осуществляется по формуле

$$r_{\Delta_{y_k}\Delta_{x_k}} = \frac{\sum \Delta_{y_k} \cdot \Delta_{x_k}}{\sqrt{\sum \Delta^2_{y_k} \sum \Delta^2_{x_k}}}. \tag{9.7}$$

Коэффициент корреляции, рассчитанный для измерения тесноты зависимости уровней временных рядов, является обобщающим показателем. Однако для продолжительного периода времени эта зависимость непостоянна и меняется во времени. Поэтому рекомендуется рассчитывать скользящие коэффициенты корреляции для определенного промежутка времени с целью выявления силы зависимости (слабая или высокая) между изменениями уровней временных рядов.

9.3. Теория коинтеграции временных рядов

Недостаток методов исключения или уменьшения автокорреляции в рядах динамики заключается в модификации исходной модели временного ряда в результате замены переменных либо включения в модель фактора времени. Однако большая часть соотношений сформулирована в уровнях временных рядов, а не их последовательных разностей или отклонений от трендов и предполагает измерение взаимосвязи переменных без включения в модель дополнительных факторов.

Иногда наличие тенденции в одном временном ряду является следствием наличия тенденции в другом ряду, включенном в модель, а не результатом случайных причин. Поэтому направленность (одинаковая либо противоположная) тенденций рядов может иметь устойчивый характер и наблюдаться на протяжении длительного периода времени. При этом коэффициент корреляции уровней временных рядов может характеризовать истинную причинно-следственную зависимость между ними и не содержать ложной корреляции.

Эти предположения легли в основу теории коинтеграции временных рядов. Коинтеграция – это причинно-следственная зависимость в уровнях двух или более временных рядов, выражающаяся в совпадении или обратной направленности их тенденции и случайной компоненты.

В соответствии с этой теорией между временными рядами существует коинтеграция в случае, если линейная комбинация этих временных рядов есть стационарный временной ряд (т.е. ряд, содержащий только случайную компоненту и имеющий постоянную дисперсию в длительном периоде времени).

Остатки исходного уравнения регрессии временного ряда представляют собой линейную комбинацию рядов y_t и x_t:

$$\Delta_t = y_t - a_0 - a_1 x_t.$$

Одним из методов тестирования гипотезы о коинтеграции временных рядов y_t и x_t является *критерий Энгеля–Грангера*. Алгоритм его применения следующий:

1. Формируется нулевая гипотеза H_0 об отсутствии коинтеграции между рядами y_t и x_t.
2. Рассчитывают параметры уравнения регрессии вида

$$\Delta'_t = a_0 + a_1 \cdot \Delta_{t-1},$$

где Δ'_t – первые разности остатков, полученных из соотношения $\Delta_t = y_t - a_0 - a_1 x_t$.

3. Вычисляют фактическое значение t-критерия Стьюдента для коэффициента регрессии a_0 в уравнении $\Delta'_t = a_0 + a_1 \cdot \Delta_{t-1}$.

4. Сравнивают полученное значение с критическим. Критические значения, рассчитанные Энгелем и Грангером для уровней значимости 1%, 5% и 10%, составляют соответственно 2,5899; 1,9439; 1,6177.

Если фактическое значение t-критерия больше критического для заданного уровня значимости α, нулевую гипотезу об отсутствии коинтеграции изучаемых временных рядов отклоняют и с вероятностью $(1 - \alpha)$ принимают альтернативную гипотезу H_1 о наличии коинтеграции временных рядов y_t и x_t. В противном случае гипотеза об отсутствии коинтеграции между исследуемыми временными рядами не отклоняется.

Другой метод тестирования нулевой гипотезы об отсутствии коинтеграции между двумя или более временными рядами связан с использованием величины критерия Дарбина–Уотсона. В основе этого метода лежит проверка гипотезы о том, что полученное значение критерия Дарбина–Уотсона в генеральной совокупности равно нулю.

Критические значения критерия Дарбина–Уотсона, полученные методом Монте-Карло, для уровней значимости 1%, 5% и 10% составляют соответственно 0,511; 0,386 и 0,322.

Если результаты тестирования показали, что фактическое значение критерия Дарбина–Уотсона нельзя признать равным нулю (следовательно, оно превышает критическое значение для заданного уровня значимости), гипотезу H_0 об отсутствии коинтеграции временных рядов отклоняют. И наоборот, если расчетное значение критерия Дарбина–Уотсона меньше критического значения при заданном уровне значимости, то гипотеза H_0 об отсутствии коинтеграции подтверждается.

Коинтеграция двух временных рядов упрощает методы и процедуры, используемые в целях их анализа, так как можно строить уравнения регрессии и определять показатели корреляции на основе непосредственно исследуемых исходных значений уровней временных рядов. Однако коинтеграция — это совпадение динамики временных рядов в течение продолжительного периода времени, поэтому теория коинтеграции применима только к временным рядам, охватывающим длительные промежутки времени. При наличии сравнительно коротких временных рядов моделирование взаимосвязей по уровням этих рядов ведет к неверным результатам вследствие нарушения предпосылок теории коинтеграции.

9.4. Интерпретация параметров моделей с распределенным лагом

Модель с распределенным лагом можно записать в виде

$$y_t = a_0 + \alpha_0 \cdot x_t + \alpha_1 \cdot x_{t-1} + \dots + \alpha_n \cdot x_{t-l} + \Delta_l.$$

Очевидно, что в момент времени t происходит изменение независимой переменной x и это изменение влияет на результативный показатель y в течение l моментов времени.

Оценка данной модели зависит от того, конечно или бесконечно число лагов в ней.

В модели с распределенными лагами коэффициент регрессии a_0 при переменной x_t характеризует среднее абсолютное изменение результата y_t при изменении фактора x_t на 1 единицу в некоторый фиксированный момент времени t без учета воздействия лаговых значений x. Этот коэффициент называется *краткосрочным мультипликатором.*

В момент $t+1$ совокупное воздействие факторной переменной x_t на результат y_t составит $\alpha_0 + \alpha_1$, а в момент $t+2$ соответственно $\alpha_0 + \alpha_1 + \alpha_2$ и т.д. Следовательно, любую сумму коэффициентов $\sum_{i=1}^{h} \alpha_i$ $(h < l)$ называют *промежуточным мультипликатором.*

Сумму всех α_i $(i = 0, 1, 2, \dots, h)$ называют *долгосрочным мультипликатором*, так как она характеризует изменение Y под воздействием единичного изменения X в каждом из рассматриваемых периодов времени:

$$M_д = \sum_{i=1}^{l} \alpha_i, \text{ где } i = 0, 1, 2, \dots, l.$$

Предположим, относительные коэффициенты модели с распределенным лагом равны:

$$A_i = \frac{\alpha_i}{\alpha},\ y = \frac{0}{l},\ \text{где}\ \alpha = \alpha_0 + \alpha_1 + \ldots \alpha_l.$$

Если все коэффициенты α_i имеют одинаковые знаки, то для любого i-го значения

$$0 < A_i < 1 \ \text{ и } \ \sum_{i=1}^{l} A_i = 1.$$

В этом случае относительные коэффициенты A_i являются весами для соответствующих коэффициентов α_i, каждый из них измеряет долю общего изменения результативного признака в момент времени $t + i$.

Зная величину A_i, можно определить:

1. Средний лаг по формуле средней арифметической взвешенной:

$$\bar{l} = \sum_{i=1}^{l} i \cdot A_i. \tag{9.8}$$

Средний лаг представляет собой средний период, в течение которого происходит изменение результата под воздействием изменения фактора в данный момент времени t.

2. Медианный лаг — это лаг, для которого

$$\sum_{i=1}^{l_{me}} A_i \approx 0{,}5.$$

9.5. Выбор формы модели с распределенным лагом

Сила воздействия текущих и лаговых значений факторного признака различна. Если построить график зависимости этих коэффициентов от величины лага, можно получить распределение во времени воздействия фактора на результат. Возможная структура лага представлена на рисунке 9.2.

Предположим, графический анализ показал, что в изучаемой модели имеет место полиноминальная структура лага, т.е. веса коэффициентов α_i аппроксимируются полиномами определенной степени от величины лага степени l. Частным случаем полиноминальной структурой лага является линейная модель (рис. 9.2, *а*).

Графиком полинома второй степени являются варианты рисунка 9.2, *г* и 9.2, *д*. Перевернутая V-образная структура лага аппроксимируется с помощью полинома второй степени. Примером модели лагов в форме полиномов третьей степени является

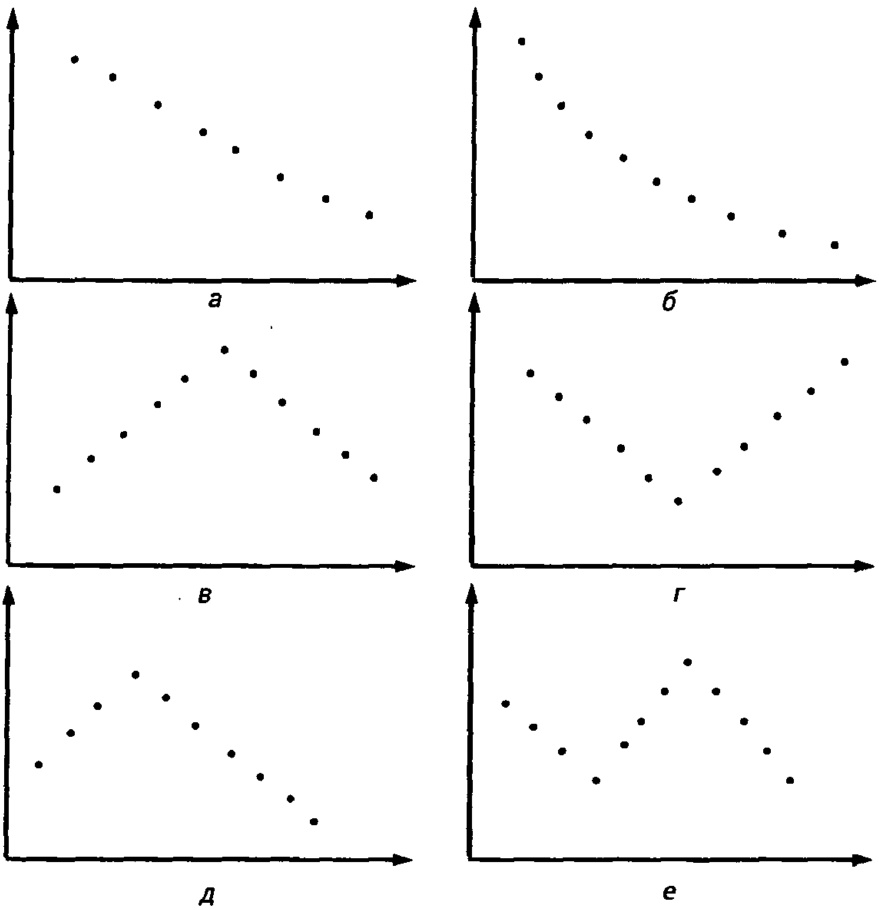


Рис. 9.2. Графическое изображение структуры лага:
а — линейная; *б* — геометрическая; *в* — перевернутая V-образом;
г, д, е — полиноминальные

рисунок 9.2, *е*. Лаги, структура которых может быть описана с помощью полиномов, называется лагами Алмон. Модель зависимости коэффициентов α_i, где $i = 0, 1, 2, ..., l$ от величины лага и в форме полинома имеет вид:

1) для полинома первой степени: $\alpha_i = a_0 + \alpha_1 \cdot i$;
2) для полинома второй степени: $\alpha_i = a_1 + \alpha_1 \cdot i + \alpha_2 \cdot i^2$;
3) для полинома третьей степени: $\alpha_i = a_0 + \alpha_1 \cdot i + \alpha_1 \cdot i^2 + \alpha_2 \cdot i^3$

и т.д.

В общем виде для полинома n-й степени модель имеет вид

$$\alpha_i = a_0 + \alpha_1 \cdot i + \alpha_1 \cdot i^2 + ... + \alpha_n \cdot i^n.$$

Следовательно, каждый из коэффициентов α_i модели можно записать:

$$\alpha_0 = \alpha_0;\ \alpha_1 = \alpha_0 + \alpha_1 + ... + \alpha_n;$$

$$\alpha_2 = \alpha_0 + 2\alpha_1 + 4\alpha_2 + ... + 2^n \cdot \alpha_n; \text{ и т.д.}$$

$$\alpha_l = \alpha_0 + l \cdot \alpha_1 + l^2 \cdot \alpha_2 + ... + l^n \cdot \alpha_n.$$

Тогда уравнение при $n = 2$ может быть записано в виде

$$y_t = a_0 + \sum_{i=0}^{n} (\alpha_0 + \alpha_{1i} + \alpha_2 i^2) \cdot x_{t-i} + \Delta_t =$$

$$= a_0 + \alpha_0 \cdot \sum_{i=0}^{n} x_{t-1} + \alpha_1 \cdot \sum_{i=0}^{n} i \cdot x_{t-1} + \alpha_2 \cdot \sum_{i=0}^{n} i^2 \cdot x_{t-1} + \Delta_r.$$

Обозначим $\quad z_0 = \sum_{i=0}^{n} x_{t-1};$

$$z_1 = \sum_{i=0}^{n} i \cdot x_{t-1};$$

$$z_2 = \sum_{i=0}^{n} i^2 \cdot x_{t-1}.$$

Имеем: $y_t = a_0 + \alpha_0 \cdot z_0 + \alpha_1 \cdot z_1 + \alpha_2 \cdot z_2 + \Delta_r$.

Процедура применения метода Алмон для целей параметризации моделей можно представить в виде алгоритма.

1. Определяют максимальную величину лага l.

2. Определяют степень полинома n, описывающего структуру лага.

3. Рассчитывают значение переменных z_i.

4. Определяют параметры преобразования уравнения регрессии в линейной форме.

5. Рассчитывают параметры исходной модели.

При применении метода Алмон необходимо решить ряд проблем.

1. Так как величина лага должна быть известна заранее, то исходят из максимально возможного лага. Выбор меньшего лага, чем реальное значение, ведет к тому, что в уравнении регрессии не учитывается фактор, оказывающий существенное влияние на результат. Влияние этого фактора в модели выражается в остатках, следовательно, не соблюдаются предпосылки МНК о случайности остатков, полученные оценки будут смещенными и неэффективными. В другом случае (если величина лага будет больше реального значения) оценки будут не смещенными, однако их эффективность снизится ввиду включения в модель статистически незначимого фактора. Оптимальную величину лага можно определить на основе априорной информации экономической теории или эмпирических исследований. Также простым способом определения величины лага является измерение тесноты связи между результатом и лаговыми значениями факторов.

2. Установить степень полинома n. На практике применяют следующее правило: выбранная степень полинома n должна быть на единицу больше числа экстремумов в структуре лага, если априорная информация о структуре лага недоступна, величину n определяют по наилучшей модели путем сравнения уравнений, построенных для различных значений n.

3. Переменные z_i, определяемые как линейная комбинация исходных факторов x_i, коррелируют между собой, когда существует высокая связь исходных переменных, поэтому оценку параметров преобразованной модели проводят в условиях мультиколлинеарности факторов.

Преимущество метода Алмон заключается в том, что он применим для моделирования процессов, характеризующихся разнообразными структурами лагов и при относительно небольшом количестве переменных в преобразованной модели. С помощью метода Алмон можно строить модели с распределенным лагом любой длины.

Модель с конечным числом лагов достаточно просто оценивать путем ее сведения к уравнению множественной регрессии (метод Алмон). Для оценки моделей с бесконечным числом лагов разработано несколько методов:

1) метод последовательного увеличения количества лагов;

2) метод преобразования Койка (метод геометрической прогрессии);

3) метод главных компонент.

По методу последовательного увеличения количества лагов уравнение $y_t = a_0 + \alpha_0 \cdot x_t + \alpha_1 \cdot x_{t-1} + \alpha_2 \cdot x_{t-2} + \ldots + \Delta_t$ рекомендуется оценивать с последовательно увеличивающимся количеством лагов. Завершение процедуры увеличения количества лагов может быть охарактеризовано несколькими признаками:

1) при добавлении нового лага коэффициент регрессии α_i при переменной x_{t-i} меняет знак, тогда в уравнении регрессии оставляют переменные x_t, $x_{t-1}, \ldots, x_{t-i-1}$, коэффициенты при которых знак не поменяли;

2) при добавлении нового лага коэффициент регрессии α_i при переменной x_{t-i} становится статистически незначимым. Следовательно, в уравнении используются переменные x_t, $x_{t-1}, \ldots,$ x_{t-i+1}, коэффициенты при которых статистически значимы.

Применение этого метода ограничено в результате изменяющегося в сторону уменьшения числа степеней свободы с одновременным увеличением стандартных ошибок и ухудшением качества оценок, а также возникновение мультиколлинеарности.

Для моделей с бесконечным числом лагов Койк предположил невозможность оценки параметров обычным МНК, так как число факторов бесконечно. При допущениях о наличии геометрической структуры лагов (рис. 9.2, *б*), то есть структуры, когда воздействие лаговых значений фактора на результат уменьшается с увеличением величины лагов в геометрической прогрессии, оценки параметров можно получить на основе подхода, предложенного Койком.

В *распределении Койка* предполагается, что коэффициенты α_i при лаговых значениях факторного признака убывают в геометрической прогрессии:

$$\alpha_i = \alpha_0 \cdot \lambda_i,\ i = 0, 1, \dots\ 0 < \lambda < 1.$$

Ограничение назначения $\lambda > 0$ обеспечивает одинаковые знаки для всех коэффициентов $\lambda_i > 0$, а ограничение $\lambda_i < 1$ означает, что с увеличением лага значение параметров модели убывает в геометрической прогрессии. Чем ближе λ к 0, тем выше темп снижения воздействия фактора на результат. В этом случае уравнение преобразуется:

$$y_t = a_0 + \alpha_0 \cdot x_t + \alpha_0 \cdot \lambda \cdot x_{t-1} + \alpha_0 \cdot \lambda^2 \cdot x_{t-2} + \dots + \Delta_t.$$

Параметры полученного уравнения можно определить двумя способами:

1) параметру λ присваиваются последовательно значения из интервала (0,1) с произвольным фиксированным шагом (0,01; 0,001; 0,0001). Для каждого λ рассчитывается:

$$z_t = x_t + \lambda \cdot x_{t-1} + \lambda^2 \cdot x_{t-2} + \lambda^3 \cdot x_{t-3} + \dots + \lambda^n \cdot x_{t-n}.$$

Значение n определяется из условия, что при дальнейшем добавлении лаговых значений x величина изменения z_t менее любого заданного ранее числа. Следовательно, необходимо оценить уравнение регрессии:

$$y_t = a_0 + \alpha_0 \cdot z_t + \Delta_t.$$

Из всех значений λ выбираются значения с наибольшим коэффициентом детерминации, а найденные параметры a_0, a_0, λ подставляются в уравнение:

$$y_t = a_0 + \alpha_0 \cdot x_t + \alpha_0 \cdot \lambda \cdot x_{t-1} + \alpha_0 \cdot \lambda^2 \cdot x_{t-2} + \dots + \Delta_t;$$

2) для периода $(t-1)$ модель можно записать следующим образом:

$$y_{t-1} = a_0 + \alpha_0 \cdot x_{t-1} + \alpha_0 \cdot \lambda \cdot x_{t-2} + \alpha_0 \cdot \lambda^2 \cdot x_{t-3} + \dots + \Delta_{t-1}.$$

Умножим обе части уравнения на λ:

$$\lambda \cdot y_{t-1} = \lambda \cdot a_0 + \lambda \cdot \alpha_0 \cdot x_{t-1} + \alpha_0 \cdot \lambda^2 \cdot x_{t-2} + \alpha_0 \cdot \lambda^3 \cdot x_{t-3} + \ldots + \lambda \cdot \Delta_{t-1}.$$

Проведем разностные преобразования моделей:

$$y_t - \lambda \cdot y_{t-1} = a_0 - \lambda \cdot a_0 + \alpha_0 \cdot x_t + \Delta_t - \lambda \cdot \Delta_{t-1}.$$

Произведенные преобразования дают возможность получить *модель Койка*:

$$y_t = a_0 \cdot (1-\lambda) + \alpha_0 \cdot x_t + (1-\lambda) \cdot y_{t-1} + v_t, \text{ где } v_t = \Delta_t \Delta_{t-1}.$$

С помощью преобразованного уравнения с бесконечным числом лагов получаем уравнение авторегрессии, для которого оценивают лишь три коэффициента: λ, α_0, a_0. В результате преобразования решается проблема мультиколлинеарности.

Модель Койка используется для анализа краткосрочных и долгосрочных свойств переменных. В краткосрочном периоде значение y_{t-1} можно рассматривать как фиксированное, а краткосрочный мультипликатор считать равным α_0. Долгосрочный мультипликатор вычисляется по формуле суммы бесконечно убывающей геометрической прогрессии.

Предположим, в долгосрочном периоде x_t стремится к равновесному значению x, а значение y_t и y_{t-1} — к своему равновесному значению y. Тогда без учета случайной ошибки модель примет вид

$$Y = a_0 \cdot (1-\lambda) + \alpha_0 \cdot X + \lambda \cdot Y.$$

Следовательно, $Y = a_0 + \dfrac{\alpha_0}{1-\lambda} \cdot X.$

В силу формулы суммы бесконечно убывающей геометрической прогрессии:

$$\frac{\alpha_0}{1-\lambda} = \alpha_0 + \alpha_0 \cdot \lambda + \alpha_0 \cdot \lambda^2 + \alpha_0 \cdot \lambda^3 + \ldots$$

Полученная дробь является долгосрочным мультипликатором, отражающим долгосрочное воздействие x на y. При $0 < \lambda < 1$ долгосрочное воздействие будет сильнее краткосрочного, так как $\dfrac{\alpha_0}{1-\lambda}$ будет больше α_0.

При применении преобразования Койка возможно возникновение трех проблем:

1) среди объясняющих переменных появляется переменная y_{t-1}, которая носит случайный характер, что нарушает условия Гаусса–Маркова, кроме того, она, возможно, коррелирует со случайным отклонением v_t;

2) если для случайных отклонений Δ_t и Δ_{t-1} исходной модели выполняется предпосылка № 3 МНК, то для случайных отклонений v_t имеет место автокорреляция. Для ее анализа используют h-статистику Дарбина, которую можно вычислить по формуле

$$h=(1-\frac{r_{dw}}{2})\cdot\sqrt{\frac{n}{1-n\cdot S^2_{y_{t-1}}}}, \qquad (9.9)$$

где n – объем выборки;

r_{dw} — коэффициент автокорреляции первого порядка, вычисленный на основе статистики Дарбина–Уотсона;

3) оценки, полученные МНК при наличии указанных проблем, являются смещенными и несостоятельными.

Для решения возникающей проблемы мультиколлинеарности используют *метод главных компонент*. Суть метода — в сокращении числа объясняющих переменных до более существенно влияющих факторов.

Метод главных компонент может применяться для исключения или уменьшения мультиколлинеарности объясняющих переменных регрессии. Основная идея заключается в сокращении числа объясняющих переменных путем линейного преобразования объясняющих переменных x_i в новые переменные — главные компоненты. Для выделения первой главной компоненты требуется соответствие максимума общей дисперсии всех объясняющих переменных. Второй компоненте – максимум оставшейся дисперсии после исключения влияния первой главной компоненты. Выполнение этих преобразований содействует уменьшению мультиколлинеарности новых переменных по сравнению с мультиколлинеарностью исходного набора переменных.

Процедура выделения главных компонент состоит из следующих шагов.

1. Строится матрица, элементами которой являются отклонения результатов наблюдения над n-ми переменными от существующих средних $x_{ij} - \bar{x}_i$, где $i = 0, ..., m; j = 0, ..., n$:

$$X^* = \begin{vmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \dots & x_{1n} - \bar{x}_n \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{2n} - \bar{x}_n \\ \dots & \dots & \dots & \dots \\ x_{m1} - \bar{x}_1 & x_{m2} - \bar{x}_2 & \dots & x_{mn} - \bar{x}_n \end{vmatrix}. \qquad (9.10)$$

2. Определяется матрица дисперсией и ковариацией объясняющих переменных:

$$D_{xx} = \frac{1}{n-1} X^{*\prime} X^*. \qquad (9.11)$$

Матрица дисперсий имеет размерность $n \times n$.

Главные компоненты z_j являются линейными комбинациями объясняющих переменных и могут быть записаны в виде

$$z_j = x_j^{*\prime} \cdot \alpha_j. \qquad (9.12)$$

Главные компоненты должны удовлетворять указанным выше требованиям, а векторы коэффициентов α_j должны быть нормированными и некоррелированными.

К недостаткам метода главных компонент относят следующие:

1) главным компонентам трудно подобрать экономические аналоги;

2) оценки параметров регрессии получают не по исходным объясняющим переменным, а по главным компонентам.

На практике метод главных компонент применяется в основном для оценки значений регрессии и для определения прогнозных значений зависимой переменной.

9.6. Авторегрессионные модели

Модель авторегрессии можно записать в виде

$$y_t = a_0 + \alpha_0 \cdot x_t + \beta_1 \cdot y_{t-1} + \Delta_t. \qquad (9.13)$$

В этой модели параметр a_0 характеризует краткосрочное применение y_t в результате изменения x_t на 1. К моменту времени $t+1$ результат y_t изменяется под воздействием изменения изучаемого фактора в момент времени t на a_0, а y_{t+1} — под воздействием своего изменения в непосредственно предшествующий момент времени на β_1. Таким образом, абсолютное изменение результата в момент времени $t+1$ составит $\alpha_0\beta_1$, в момент времени $t+2$ абсолютное изменение составит $\alpha_0\beta_1 l$ и т. д.

Таким образом, долгосрочный мультипликатор в модели авторегрессии рассчитывается как сумма краткосрочного и промежуточных мультипликаторов:

$$B = \alpha_0 + \alpha_0\beta_1 + \alpha_0\beta_1^2 + \alpha_0\beta_1^3 + \ldots + \alpha_0\beta_1^i.$$

Интерпретация коэффициентов модели авторегрессии и расчет долгосрочного мультипликатора основываются на предпосылке о бесконечном лаге, в воздействии текущего значения зависимой переменной на ее будущее значение.

В зависимости от лежащей в основе модели гипотезы о механизме формирования различают модели *адаптивных ожиданий*, *частичной корректировки* и *рациональных ожиданий*.

В модели *адаптивных ожиданий* происходит постоянная корректировка ожиданий на основе полученной информации о реализации исследуемого показателя. Если реальное значение показателя оказалось больше ожидаемого, то ожидаемое в следующем периоде значение корректируется в сторону увеличения, и наоборот. Величина корректировки должна быть пропорциональна разности между реальным и ожидаемым значениями.

Модель адаптивных ожиданий имеет вид

$$y_t = a_0 + \alpha_0 \cdot x'_{t+2} + \Delta_t,$$

где y_t — фактическое значение результативного признака;
x'_{t+1} – ожидаемое значение факторного признака.

В модели вместо текущего значения x'_{t+1} используются предыдущие x_t:

$$x'_{t+1}+x'_t=\varphi\cdot(x_t+x'_t) \text{ или } x'_{t+1}=\varphi\cdot x_t+(1-\varphi)\cdot x'_t,$$

где $0<\varphi<1$.

Ожидаемое значение x'_t является средней взвешенной арифметической фактического и ожидаемого значения в предыдущий период. Приведенное соотношение показывает, как в период времени $t+1$ корректируется ожидание на долю φ разности между фактическим значением факторного признака и его ожидаемым значением в предыдущий период. Параметр φ в модели называют *коэффициентом ожидания*. При $\alpha=0$ $x'_t=x'_{t+1}$, что показывает, что действующие факторы сохранят свое влияние на будущий период времени. Ожидаемые значения совпадут с текущими. Приближение величины φ к нулю свидетельствует об устойчивости тенденции. Чем ближе коэффициент ожидания к единице, тем в большей степени реализуются ожидания экономических агентов. Если $\varphi=0$, то $x'_t=x_t$.

Подставив полученное соотношение в модель, имеем:

$$y_t=a_0+\alpha_0\cdot(\varphi\cdot x_t+(1-\varphi)\cdot x'_t)+\Delta_t.$$

Вычитая аналогичное уравнение для y'_{t-1}, умноженное на $1-\varphi$, получим:

$$y_t-(1-\varphi)\cdot y_{t-1}=\varphi\cdot a_0+\varphi\cdot\alpha_0\cdot x_t+[\Delta_t\cdot(1-\varphi)\cdot\Delta_{t-1}] \text{ или}$$

$$y_t=a_0\cdot\varphi+\alpha_0\cdot\varphi\cdot x_t+(1-\varphi)\cdot y_{t-1}+V_t,$$

где $V_t=\Delta_t-(1-\varphi)\cdot\Delta_{t-1}$.

Коэффициент α_0 определяет величину изменения в среднем текущего значения y_t при изменении ожидаемого значения x'_t на единицу. При изменении текущего значения x_t на единицу значение y_t меняется в среднем на $\varphi\cdot\alpha_0$. На практике при оценке параметров авторегрессионного уравнения вначале оценивается параметр φ, а затем рассчитывают a_0 и α_0, используя значение свободного члена $a_0=\frac{a_0\cdot\varphi}{\varphi}$ и коэффициента регрессии $\alpha_0=\frac{\alpha_0\cdot\varphi}{\varphi}$.

Отметим также, что в модели адаптивных ожиданий включены ожидаемые значения факторной переменной, которые нельзя получить эмпирически. Следовательно, статистические методы оценки параметров модели неприемлемы. Преобразованная модель включает только фактические значения переменных. Поэтому ее параметры определяются с помощью стандартных статистических методов.

Преобразованная модель адаптивных ожиданий аналогична по форме модели Койка. Поэтому, как и в случае с преобразованием Койка, обобщенный метод наименьших квадратов приводит к смещенным оценкам вследствие наличия лагового значения результативного признака y_{t-1}.

Модель, характеризующая зависимость результативного признака от ожидаемых значений факторного, называется долгосрочной функцией модели адаптивных ожиданий. Модель, описывающая зависимость результата от фактических значений независимой переменной, называется краткосрочной функцией модели адаптивных ожиданий.

Модель адаптивных ожиданий используется при анализе зависимости потребления от дохода, спроса на деньги или инвестиций от процентной ставки и в других случаях, когда экономические показатели чувствительны к ожиданиям.

В отличие от модели адаптивных ожиданий *в модели неполной* (частичной) *корректировки* (модель акселератора) в уравнении регрессии в качестве зависимой переменной входит долгосрочное (желаемое) значение $y'_t : y'_t = a_0 + \alpha_0 \cdot x_t + \Delta_t$.

Формирование ожиданий относительно значений y'_t происходит на основе предпосылки частичной корректировки:

$$y_t - y_{t-1} = C \cdot (y'_t - y_{t-1}),$$

где $0 < C < 1$.

В модели предполагается, что абсолютные изменения фактических уровней результата — это доля его ожидаемого абсолютного изменения. Параметр C называют коэффициентом корректировки. Чем ближе величина C к 1, тем больше степень динамики показателя отвечает ожиданиям. Чем ближе C к 0, тем меньше реальное изменение показателя соответствует ожидаемому изменению. При $C = 0$ корректировка не происходит.

Модель, построенная на основе корректировки значений, основывается на гипотезе о формировании y'_t в виде

$$y_t = C \cdot y'_t + (1 - C) \cdot y_{t-1}. \tag{9.14}$$

Фактическое значение текущего результата есть средняя взвешенная арифметическая его ожидаемого значения y'_t и фактического значения предыдущего периода y'_{t-1}. В результате получим:

$$y_t = C \cdot a_0 + C \cdot \alpha_0 \cdot x_t + (1 - C) \cdot y_{t-1} + C \cdot \Delta_t. \tag{9.15}$$

Это соотношение – основное уравнение неполной корректировки. Общий вид модели неполной корректировки называют долгосрочной функцией, а преобразованное уравнение — краткосрочной функцией. Оценка параметров осуществляется аналогично модели адаптивных ожиданий, вначале C при y_{t-1}, затем

$$a_0 = \frac{C \cdot a_0}{C} \text{ и } \alpha_0 = \frac{C \cdot \alpha_0}{C}.$$

Модель неполной корректировки аналогична модели Койка, также включает случайно объясняющую переменную y_{t-1}, но в данной модели переменная y_{t-1} не коррелирует с текущим значением случайного отклонения, так как Δ_t рассчитывается после определения значения y_{t-1}. Применение метода наименьших квадратов позволяет получить несмещенные и эффективные оценки.

9.7. Оценка параметров моделей авторегрессии

Описанные выше авторегрессионные модели сводятся к уравнению вида

$$y_t = a_0 + \alpha_0 \cdot x_t + C_1 \cdot y_{t-1} + \Delta_t. \tag{9.16}$$

При построении модели авторегрессии возникают 3 проблемы:

1) наличие лаговых значений результативного признака приводит к нарушению предпосылки метода наименьших квадратов о делении переменных на результативную и факторную;

2) существует вероятность наличия автокорреляции между случайными отклонениями u_t;

3) так как существует зависимость между текущими значениями y_t и текущими остатками u_t, очевидно, что существует взаимозависимость между y_{t-1} и u_{t-1}. Следовательно, нарушается предпосылка об отсутствии связи между факторным признаком и остатками.

Поэтому применение классического метода наименьших квадратов приводит к получению смещенных и несостоятельных оценок параметров.

Наиболее распространенным методом оценивания авторегрессионных уравнений, позволяющих сгладить указанную корреляцию, является *метод инструментальных переменных*. Основная идея этого метода состоит в том, чтобы заменить переменную в правой части модели, для которой нарушаются предпосылки МНК, на новую, включение которой не приводит к нарушению его предпосылок.

В моделях авторегрессии удаляют из правой части переменную y_{t-1}. Новая переменная, с одной стороны, должна тесно коррелировать с y_{t-1}, а с другой — не коррелировать с остатками u_t.

Подбор инструментальной переменной зависит от практической ситуации, в частности, в качестве инструментальной переменной предлагают оценить y_{t-1} как результат регрессии переменной y на независимые переменные x_i, входящие в первоначальную авторегрессионную модель. Однако такая замена может привести к появлению мультиколлинеарности.

Существует несколько способов получения инструментальной переменной. В модели авторегрессии переменная y_t зависит не только от y_{t-1}, но и от x_t. Следовательно, может иметь место зависимость y_{t-1} от x_{t-1}:

$$y_{t-1} = \gamma_0 + \gamma_1 \cdot x_{t-1} + y_t,$$

следовательно:

$$y_{t-1} = y_{t-1}^{\text{рас}} + y_t. \qquad (9.17)$$

Оценка этого уравнения производится классическим методом наименьших квадратов. Оценка $y_{t-1}^{\text{рас}}$ может служить в качестве инструментальной переменной для фактора y_t. Такая инструментальная переменная имеет тесную связь с y_{t-1} и представляет собой линейную зависимость переменной x_{t-1}. При использовании этого метода не нарушается предпосылка об отсутствии зависимости между факторным признаком и остатками в модели. Следовательно, переменная $y_{t-1}^{\text{рас}}$ не коррелирует с ошибкой y_t.

Таким образом, оценки параметров уравнения можно найти из соотношения

$$y_t = a_0 + \alpha_0 \cdot x_t + C_1 \cdot y_{t-1}^{\text{рас}} + V_t. \qquad (9.18)$$

Однако практическая реализация метода инструментальных переменных приводит к появлению проблемы мультиколлинеарности. В некоторых случаях эту проблему решают включением в модель фактора времени в качестве независимой переменной.

9.8. Новые направления в анализе многомерных временных рядов

Современная система анализа многомерных временных рядов включает следующие модели:

1) векторная авторегрессия;
2) рациональных ожиданий;
3) коинтеграция временных рядов и механизм исправления ошибок.

Модели векторной авторегрессии

Временные ряды — основной источник данных для построения эконометрических моделей в форме систем одновременных уравнений. Однако методы построения структурных моделей (особенно моделей, содержащих значительное количество уравнений и переменных) сложны, поэтому был разработан и получил широкое распространение подход, основывающийся на построении *моделей векторной авторегрессии*. В разработку этого подхода большой вклад внесли Р. Лукас, Т. Сарджент, К. Симс и ряд других макроэкономистов.

Основной проблемой, которую приходится решать на этапе оценки параметров систем одновременных уравнений, является идентификация. Именно эта проблема (наряду с разделением переменных эконометрической модели на эндогенные и предопределенные) послужила основным поводом для критики стандартного подхода к системам одновременных уравнений.

Во-первых, в больших эконометрических моделях иногда трудно определить некоторые переменные как чисто экзогенные. Например, управляемые переменные экономической политики часто считают экзогенными. Однако в ответ на макроэкономические условия правительство может внести изменения в эти управляемые переменные и параметры. В этом смысле все переменные экономической политики являются эндогенными.

Во-вторых, проблемы другого характера существуют и в небольших моделях, построенных по принципу систем одновременных уравнений. Например, стандартная практика построения этих моделей — включение лаговой эндогенной переменной в правую часть некоторых уравнений с целью увеличения числа предопределенных переменных модели.

Однако иногда этот шаг не подтверждается веским теоретическим обоснованием. Таким образом, решая проблему идентификации, эконометристы вынуждены снижать обоснованность структурной формы исходной эконометрической модели.

Вместе с тем существует и другое понимание этой модели.

Одним из основных направлений использования эконометрических моделей является прогнозирование знаний эндогенных переменных на будущие периоды. Однако, как отмечал Р. Лукас, когда прогноз сделан и доведен до сведения экономических единиц, последние могут изменить свое поведение, если они считают этот прогноз достоверным. Это изменение поведения, в свою очередь, приводит к изменению значений ряда переменных, например, предельной склонности к потреблению, инфляционных ожиданий, которые использовались в структурной модели при составлении прогноза. Следовательно, все экономические переменные потенциально являются эндогенными, и эконометрическая идентификация теоретически невозможна.

Одним из результатов подобного рода дискуссий является разработка моделей векторной авторегрессии (*VAR*). В моделях *VAR* не делается попыток воссоздать реальную структуру экономики, в них не проводится различий между эндогенными и экзогенными переменными. Каждое уравнение модели *VAR* описывает зависимость одной из переменных модели от лаговых значений всех переменных модели. Следовательно, каждое уравнение модели есть комбинация модели с распределенным лагом и модели авторегрессии. Число уравнений модели *VAR* равно числу ее переменных.

Примером простейшей модели векторной авторегрессии является модель, предложенная в 1984 году *Уэббом* для прогнозирования процентных ставок (*R*) и процентного изменения денежной массы (*M*):

$$\begin{cases} R_t = a_1 + \alpha_1 \cdot M_{t-1} + C_1 \cdot R_{t-1} + \Delta_{t_1}; \\ M_t = a_2 + \alpha_2 \cdot M_{t-1} + C_2 \cdot R_{t-1} + \Delta_{t_2}. \end{cases} \tag{9.19}$$

Основная характеристика этой модели — симметрия переменных; обе переменные появляются по обе стороны каждого уравнения.

Реальные модели *VAR* имеют более длительные лаги и большее число переменных. Однако по сравнению со структурными моделями модели *VAR* имеют меньшее число параметров и менее строгие ограничения на их значения, что делает модели векторной авторегрессии чрезвычайно полезными при возникновении трудностей со сбором исходной информации.

Модели векторной авторегрессии имеют недостатки. Например, в ряде случаев трудно подвести теоретическое обоснование и дать экономическую интерпретацию параметрам модели *VAR*.

Однако определенная теоретическая база заложена в начальный выбор переменных, которые войдут в модель.

Другой недостаток моделей векторной авторегрессии – необходимость принятия решения относительно величины лага, адекватных методов оценки параметров модели, поскольку классический МНК чаще всего неприменим при оценке параметров моделей с распределенным лагом и тем более неприменим для оценки параметров моделей авторегрессии. Поэтому методы оценки параметров моделей *VAR* очень громоздки, и в настоящее время далеко не все статистические пакеты прикладных программ имеют эту функцию. Однако в целом модели *VAR* потенциально значительно проще структурных моделей.

Модели рациональных ожиданий

Теория рациональных ожиданий оказывает все большее воздействие на макроэкономическую теорию и прикладную экономику.

В наиболее упрощенной форме рациональные ожидания означают, что экономические агенты имеют доступ ко всей адекватной информации и наилучшим образом ее используют при формировании ожиданий относительно будущих значений экономических переменных. Адекватная информация предполагает знание целей экономической политики государства. Формально в моделях рациональных ожиданий предполагается следующее: экономические агенты на основе опыта и существующего уровня знаний убеждены, что для любого момента времени t каждая переменная X определяется следующим образом:

$$x_t = a_0 + \alpha_0 \cdot x_{t-1} + C_1 \cdot z_{t-1} + \Delta_t, \qquad (9.20)$$

где z — экзогенная переменная, которая, по убеждению экономических агентов, влияет на X;

Δ_t — случайная ошибка.

Предполагается, что в момент времени t экономические агенты ничего не знают о переменных X_t и Z_t, поэтому первая часть уравнения содержит только лаговые переменные. В момент времени $(t-1)$ экономические агенты формируют свои ожидания относительно X_t на базе уравнения

$$E_{t-1}(x_t) = a_0 + \alpha_0 \cdot x_{t-1} + C_1 \cdot Z_{t-1}. \qquad (9.21)$$

Эта величина и есть рациональные ожидания относительно переменной X_t. Таким образом, имеем:

$$x_t - E_{t-1}(x_t) = \Delta_t. \qquad (9.22)$$

В сущности, Δ_t является ошибкой прогноза экономической единицы. Относительно этой ошибки делаются две предпосылки. Во-первых, ее среднее значение должно быть равно нулю. Во-вторых, это должна быть действительно случайная, то есть непрогнозируемая, ошибка. Например, в этих ошибках не может наблюдаться автокорреляция. Прогнозируемость ошибки означала бы наличие информации, которая не использовалась экономическими агентами при формировании ожиданий, что противоречит гипотезе о рациональных ожиданиях. В случае, когда возможно получение прогноза, экономические агенты могли бы просто переформулировать исходное уравнение до тех пор, пока ошибка не станет непрогнозируемой.

В настоящее время рациональные ожидания часто используются в прикладных исследованиях как альтернатива адаптивным ожиданиям. Например, предполагается, что переменная y определяется, как и в модели адаптивных ожиданий. Однако вместо предпосылки об адаптивных ожиданиях относительно будущих значений переменной x^*_t используются рациональные ожидания.

$$y_t = a_0 + \alpha_0 \cdot E_{t-1}(x^*_t) + \Delta_t. \qquad (9.23)$$

Проблема получения оценок параметров a_0 и α_0 состоит в том, что данные о рациональных ожиданиях периода $(t-1)$ относительно величины x_t отсутствуют.

Процедура эмпирического оценивания этих параметров состоит из двух шагов.

ШАГ 1. Оценивают параметры уравнения

$$\hat{x}_t = C_0 + C_1 \cdot x_{t-1} + C_2 \cdot z_{t-1}, \qquad (9.24)$$

а также расчетные значения $\hat{x}_t$. Эти значения принимают за аппроксимацию $E_{t-1}(x^*_t)$.

ШАГ 2. В исходном уравнении ожидаемые значения $E_{t-1}(x_t^*)$ заменяют на $\hat{x}_t$ и определяют оценки параметров a_0 и α_0 обычным МНК.

Коинтеграция временных рядов и механизм исправления ошибок

Сама коинтеграция имеет место только для длительных промежутков времени или в долгосрочной перспективе. Рассмотрим модель регрессии по двум временным рядам y_t и x_t, между которыми существует коинтеграция:

$$y_t = a_0 + \alpha_0 \cdot x_t + \Delta_t.$$

Из этой модели мы получаем следующее выражение для ошибки:

$$\Delta_t = y_t - a_0 - \alpha_0 \cdot x_t.$$

Это соотношение описывает механизм расчета случайной ошибки в модели в долгосрочной перспективе. Д. Сарган предложил называть совпадение тенденций (коинтеграцию) рядов x_t и y_t в долгосрочной перспективе их равновесным состоянием, а ошибку Δ_t, описываемую соотношением, — равновесной ошибкой.

При рассмотрении коротких временных рядов данных, или краткосрочной перспективы, равновесное состояние может не достигаться, и коинтеграции между рядами x_t и y_t может не быть.

Поэтому каждый уровень временного ряда случайных ошибок Δ_t, или остатков, можно считать корректирующей компонентой модели, характеризующей степень достижения равновесного состояния динамики рядов x_t и y_t в долгосрочной перспективе. Основываясь на этих рассуждениях, Д. Сарган выдвинул предположение о том, что на формирование уровней ряда y_t (точнее — на их ценные абсолютные приросты) оказывают влияние два фактора – изменение, или цепные абсолютные приросты ряда x_t, который является причиной y_t, и величина ошибки предыдущего периода Δ_{t-1}. Формально это предположение можно описать с помощью следующей модели:

$$\Delta_{y_t} = a_0 + \alpha_0 \cdot \Delta_{x_t} + \alpha_1 \cdot \Delta_{t-1} + u_t, \qquad (9.25)$$

где $\Delta_{y_t}, \Delta_{x_t}$ — первые разности изучаемых рядов;

Δ_{t-1} — ошибка предыдущего периода в модели;

u_t — случайная ошибка уравнения.

Модель регрессии называется механизмом корректировки посредством ошибок. Если коэффициент α_1 статистически значим, то его величина характеризует долю неравновесного состояния временного ряда y_t, которая корректируется в каждом следующем периоде. Поскольку Δ_{y_t} в модели есть первые разности исходных уровней ряда, можно сказать, что коэффициент α_1 характеризует скорость корректировки ряда y_t во времени по направлению к достижению равновесного состояния.

Таким образом, механизм корректировки посредством ошибок позволяет количественно охарактеризовать взаимосвязь между краткосрочной и долгосрочной динамикой во временных рядах экономических показателей. Впервые этот механизм был описан Д. Сарганом, дальнейшая его разработка и эмпирическая проверка проводились также Энгелем и Грангером.

Контрольные вопросы

1. Общее представление о детерминированных и стохастических процессах.
2. Методы прогнозирования.
3. Экстраполяция в эконометрическом прогнозировании.
4. Интерполяционные методы прогнозирования.
5. Трендовое прогнозирование.
6. Общее понятие эконометрических моделей, их типы.
7. Модели временных рядов.
8. Регрессия модели с одним уравнением.
9. Комплексные эконометрические модели.
10. Применение эконометрических моделей.
11. Понятие и классификация временных рядов.
12. Компонентный анализ рядов динамики.
13. Методология регрессионного анализа тенденции.
14. Гармонический анализ.

15. Методы выявления периодической компоненты.
16. Методы измерения устойчивости тенденций динамики.
17. Моделирование тенденции ряда динамики при наличии структурных изменений.
18. Регрессионный анализ связных динамических рядов.
19. Автокорреляция временных рядов.
20. Критерий Дарбина-Уотсона.
21. Методы исключения автокорреляции.
22. Теория коинтеграции временных рядов.
23. Характеристика классов динамических эконометрических моделей.
24. Интерпретация параметров моделей с распределенным лагом.
25. Выбор формы модели с распределенным лагом.
26. Лаги Алмон.
27. Новые направления в анализе многомерных временных рядов.

☑ ТЕСТОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ ПРОВЕРКИ ЗНАНИЙ

1. Временным рядом является совокупность значений:
а) экономического показателя за несколько последовательных моментов (периодов) времени;
б) экономического показателя для однотипных объектов на определенный момент времени;
в) последовательных моментов (периодов) времени и соответствующих им значений экономического показателя;
г) экономических однотипных объектов по состоянию на определенный момент времени.

2. Компонентами временного ряда являются:
а) случайная;
б) прогнозная;
в) независимая;
г) трендовая.

3. Задачами построения эконометрической модели временного ряда являются:

а) определение доверительных интервалов для параметров модели;

б) выявление и придание количественного значения каждой из трех компонент;

в) расчет показателей существенности параметров;

г) изучение структуры временного ряда.

4. Доля объясненной дисперсии зависимой переменной в общей дисперсии этого признака составила 100%. Можно сделать вывод о том, что связь, описываемая построенным уравнением, является:

а) случайной;

б) функциональной;

в) несущественной;

г) стохастической.

5. Наиболее высокий коэффициент автокорреляции первого порядка свидетельствует о том, что:

а) исследуемый ряд содержит только тенденцию;

б) исследуемый ряд содержит циклические колебания;

в) ряд не содержит тенденции и циклических колебаний.

6. Если ни один из коэффициентов автокорреляции не является значимым, это свидетельствует о том, что:

а) ряд не содержит тенденции и циклических колебаний;

б) исследуемый ряд содержит только тенденцию;

в) исследуемый ряд содержит циклические колебания.

7. Кусочно-линейная модель регрессии применяется:

а) для моделирования тенденции временного ряда, испытывающего влияние структурных изменений;

б) для моделирования тенденции временного ряда за небольшой промежуток времени;

в) для моделирования тенденции временного ряда.

8. Тест Чоу применяется для:
а) выбора модели временного ряда;
б) определения наличия гетероскедастичности;
в) выбора метода оценки системы одновременных уравнений.

9. Коинтеграция временных рядов — это:
а) причинно-следственная зависимость в уровнях двух (или более) временных рядов;
б) корреляционная зависимость между последовательными уровнями временного ряда;
в) последовательность коэффициентов автокорреляции уровней временного ряда.

10. Авторегрессионные модели включают в качестве объясняющих переменных лаговые значения:
а) зависимых переменных;
б) независимых переменных;
в) зависимых и независимых переменных.

11. Модели с распределенными лагами включают в качестве объясняющих переменных лаговые значения:
а) независимых переменных;
б) зависимых переменных;
в) зависимых и независимых переменных.

12. Тенденция временного ряда характеризует совокупность факторов:
а) оказывающих долговременное влияние и формирующих общую динамику изучаемого показателя;
б) не оказывающих влияния на уровень ряда;
в) оказывающих единовременное влияние;
г) оказывающих долговременное влияние и формирующих общую динамику изучаемого показателя.

13. Хронологическая последовательность значений признака, характеризующего состояние данного объекта, называется:
а) автокорреляционной функцией;
б) корреляционным полем;
в) случайной выборкой;
г) временным рядом.

14. Отличительной особенностью аддитивных моделей следует считать:
а) уменьшающуюся амплитуду сезонных колебаний;
б) возрастающую амплитуду сезонных колебаний;
в) резкое затухание амплитуды колебаний;
г) неизменность амплитуды сезонных колебаний.

15. Укажите предположения, которые можно сделать, когда ни один из коэффициентов автокорреляции временного ряда не является значимым:
а) ряд содержит сильную нелинейную тенденцию;
б) ряд содержит только сезонные колебания;
в) ряд содержит только тенденцию;
г) ряд содержит только случайную компоненту.

ПРАКТИКУМ

Задача 9.1. По данным о ценах реализации зерновых культур в регионе (табл. 9.1) необходимо определить индивидуальные показатели динамики, сделать выводы.

Таблица 9.1

Средние цены реализации зерновых культур в Ставропольском крае за период 1995–2012 годов

Год	Средние цены реализации зерновых культур, руб./т	Год	Средние цены реализации зерновых культур, руб./т
1995	281	1997	597
1996	544	1998	483

Окончание табл. 9.1

Год	Средние цены реализации зерновых культур, руб./т	Год	Средние цены реализации зерновых культур, руб./т
1999	1183	2006	2945
2000	1718	2007	4570
2001	1790	2008	3518
2002	1490	2009	3616
2003	2488	2010	4349
2004	2619	2011	4933
2005	2187	2012	6864

Решение

Первоначально рассчитаем индивидуальные показатели динамики временного ряда с помощью программы *SPSSStatistics* 20.

Для осуществления этого вначале необходимо создать таблицу исходных данных в окне *SPSSStatistics 20*, обозначив исследуемый динамический ряд переменной *x*. На следующем этапе нужно создать ряд значений исследуемой переменной, смещенный на 1 период времени вперед, обозначив новую переменную как «x_1» (рис. 9.3).

год	x1	x_1
1995	281,00	
1996	544,00	281,00
1997	597,00	544,00
1998	483,00	597,00
1999	1183,00	483,00
2000	1718,00	1183,00
2001	1790,00	1718,00
2002	1490,00	1790,00
2003	2488,00	1490,00
2004	2619,00	2488,00
2005	2187,00	2619,00
2006	2945,00	2187,00
2007	4570,00	2945,00
2008	3518,00	4570,00
2009	3616,00	3518,00
2010	4349,00	3616,00
2011	4933,00	4349,00
2012	6864,00	4933,00

Рис. 9.3. Исходные данные для расчета показателей ряда динамики в *SPSSStatistics 20*

Для вычисления базисных абсолютных приростов в меню «Преобразовать» выберем инструмент «Вычислить переменную». В появившемся диалоговом окне выполним последовательность действий:

- в поле «Вычисляемая переменная» введем имя новой переменной, например «Пб»;
- в окне «Числовое выражение» введем выражение «x-281» (рис. 9.4), нажмем «ОК».

В результате получим значения абсолютных приростов исследуемого динамического ряда.

Для вычисления цепных абсолютных приростов в диалоговом окне «Вычислить переменную» выполним следующие действия:

- в поле «Вычисляемая переменная» введем имя новой переменной, например «Пц»;
- в окне «Числовое выражение» введем выражение «x-x_1», нажмем «ОК».

Для вычисления базисных темпов роста в диалоговом окне «Вычислить переменную» выполним следующие действия:

- в поле «Вычисляемая переменная» введем имя новой переменной: «ТРб»;

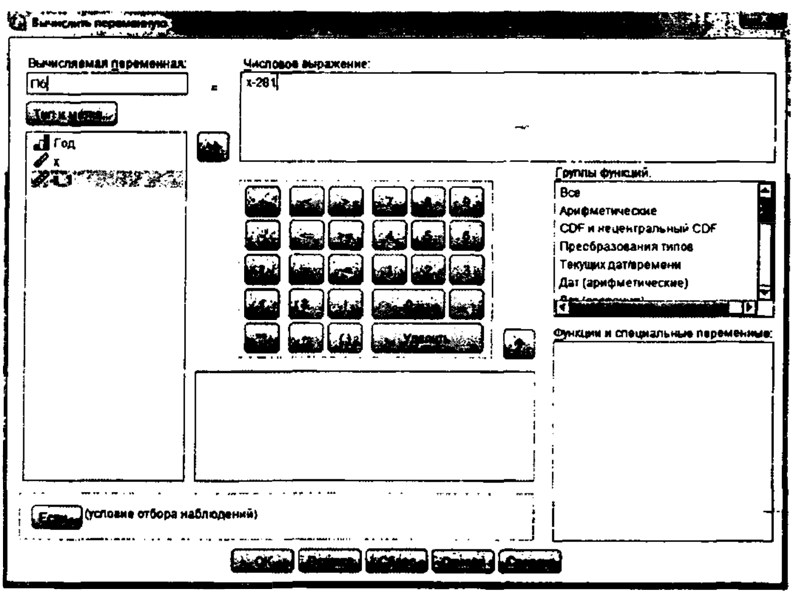


Рис. 9.4. Вычисление новой переменной в *SPSSStatistics 20*

- в окне «Числовое выражение» введем выражение «x/281*100», нажмем «ОК».

Для вычисления цепных темпов роста в диалоговом окне «Вычислить переменную» выполним следующие действия:

- в поле «Вычисляемая переменная» введем имя новой переменной: «Трц»;
- в окне «Числовое выражение» введем выражение «x/x_1*100», нажмем «ОК».

Для вычисления базисных темпов прироста в диалоговом окне «Вычислить переменную» выполним следующие действия:

- в поле «Вычисляемая переменная» введем имя новой переменной: «ТПб»;
- в окне «Числовое выражение» введем выражение «Пб/281*100», нажмем «ОК».

год	x	x_1	Пб	Пц	Трб	Трц	ТПб	ТПц
1995	281,00	.	.	.	.	.	.	.
1996	544,00	281,00	263,00	263,00	193,59	193,59	93,59	93,59
1997	597,00	544,00	316,00	53,00	212,46	109,74	112,46	9,74
1998	483,00	597,00	202,00	-114,00	171,89	80,90	71,89	-19,10
1999	1183,00	483,00	902,00	700,00	421,00	244,93	321,00	144,93
2000	1718,00	1183,00	1437,00	535,00	611,39	145,22	511,39	45,22
2001	1790,00	1718,00	1509,00	72,00	637,01	104,19	537,01	4,19
2002	1490,00	1790,00	1209,00	-300,00	530,25	83,24	430,25	-16,76
2003	2488,00	1490,00	2207,00	998,00	885,41	166,98	785,41	66,98
2004	2619,00	2488,00	2338,00	131,00	932,03	105,27	832,03	5,27
2005	2187,00	2619,00	1906,00	-432,00	778,29	83,51	678,29	-16,49
2006	2945,00	2187,00	2664,00	758,00	1048,04	134,66	948,04	34,66
2007	4570,00	2945,00	4289,00	1625,00	1626,33	155,18	1526,33	55,18
2008	3518,00	4570,00	3237,00	-1052,00	1251,96	76,98	1151,96	-23,02
2009	3616,00	3518,00	3335,00	98,00	1286,83	102,79	1186,83	2,79
2010	4349,00	3616,00	4068,00	733,00	1547,69	120,27	1447,69	20,27
2011	4933,00	4349,00	4652,00	584,00	1755,52	113,43	1655,52	13,43
2012	6864,00	4933,00	6583,00	1931,00	2442,70	139,14	2342,70	39,14

Рис. 9.5. Расчет показателей ряда динамики в *SPSSStatistics 20*

Для вычисления цепных темпов прироста в диалоговом окне «Вычислить переменную» выполним следующие действия:

- в поле «Вычисляемая переменная» введем имя новой переменной: «ТПц»;
- в окне «Числовое выражение» введем выражение «Пц/x_1*100», нажмем «ОК».

В итоге в окне ввода данных получим значения индивидуальных показателей динамики (рис. 9.5).

Задача 9.2. По данным таблицы 9.2 осуществите расчет прогнозных значений производства говядины в Ставропольском крае до 2017 г.

Таблица 9.2

Производство говядины во всех категориях хозяйств Ставропольского края в 2007–2013 годах

Период	Производство говядины во всех категориях хозяйств Ставропольского края, тонн (*У*)
2007	13288
2008	12874,4
2009	12506,7
2010	12700,2
2011	11551,8
2012	11288,6
2013	11849,9

Решение

Для расчета прогнозных значений объемов производства говядины во всех категориях хозяйств Ставропольского края можно воспользоваться инструментом «Подгонка кривых» пакета *SPSSStatistics* 20. Для этого сформируем таблицу исходных данных в редакторе данных *SPSSStatistics* 20. Затем в меню «Анализ» выберем пункт «Подгонка кривых».

В появившемся диалоговом окне необходимо указать ряд значений переменной в качестве зависимой. В пункте диалогового окна «Модели» можно отметить различные формы линии аналитического выравнивания (рис. 9.6). В качестве независимой переменной отметим пункт «Время».

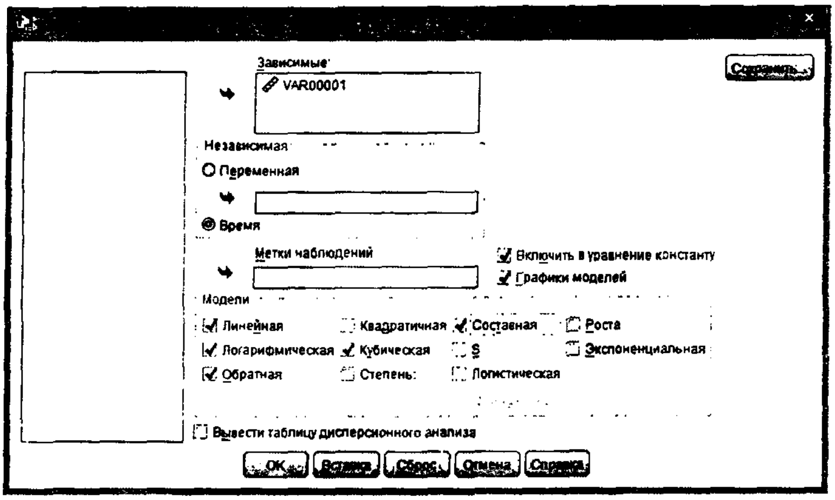


Рис. 9.6. Меню «Подгонка кривых» в *SPSSStatistics 20*

В результате в окне вывода получим таблицу «Сводка модели и оценки параметров» (таблица 9.3).

Из полученных результатов можно сделать вывод, что линейная функция представляется наиболее значимой для использования в качестве основного теоретического закона, по которому может быть произведено прогнозирование значений переменной до 2017 года, поскольку значение F-критерия Фишера для этой функции максимально из представленных (16,975).

В колонке «Оценки параметров» отображены рассчитанные параметры линии тренда: $y_t = 13500{,}01 - 301{,}46 \cdot t$.

Для расчета прогнозных значений снова откроем окно «Подгонка кривых», отметим во вкладке «Модель» пункт «Линейная». Во вкладке «Сохранить» отметим пункты «Сохранить предсказанные значения» и «Интервалы прогноза». В результате получим теоретические значения (FIT_1), а также значения нижнего доверительного интервала (LCL_1) и верхнего доверительного интервала (UCL_1) для каждого периода времени (рис. 9.7).

Таблица 9.3

Сводка модели и оценки параметров

Уравнение	Сводка для модели					Оценки параметров		
	R-квадрат	F	ст.св.1	ст. св. 2	Знч.	Константа	b1	b2
Линейный	,772	16,975	1	5	,009	13500,057	-301,457	
Логарифмическая	,765	16,284	1	5	,010	13446,818	-946,390	
Обратная	,638	8,818	1	5	,031	11570,983	1952,564	
Квадратичный	,799	7,954	2	4	,040	13887,657	-559,857	32,300
Составная	,763	16,117	1	5	,010	13538,732	,976	

	y	FIT_1	LCL_1	UCL_1
1	13288,00	13198,60000	11994,25899	14402,94101
2	12874,40	12897,14286	11768,62406	14025,66165
3	12506,70	12595,68571	11515,21123	13676,16020
4	12700,20	12294,22857	11230,25085	13358,20630
5	11551,80	11992,77143	10912,29694	13073,24592
6	11288,60	11691,31429	10562,79549	12819,83308
7	11849,90	11389,85714	10185,51613	12594,19815
8		11088,40000	9785,29874	12391,50126
9		10786,94286	9366,92118	12206,96454
10		10485,48571	8934,48498	12036,48645
11		10184,02857	8491,25038	11876,80677

Рис. 9.7. Расчет прогнозных значений результативного признака средствами *SPSSStatistics 20*

В заключение изобразим результаты графически средствами *MS Excel* (рис. 9.8).

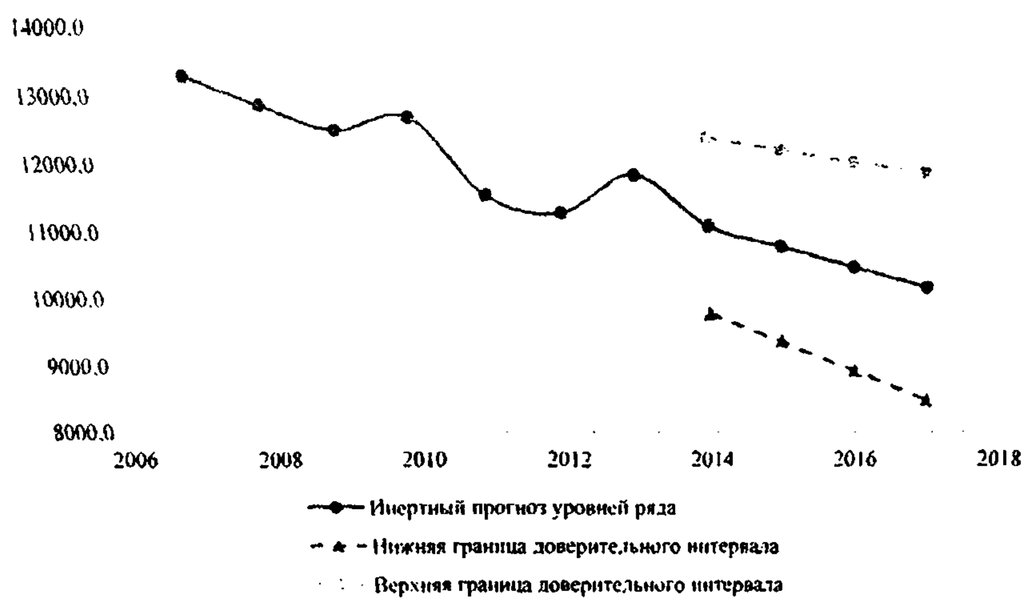


Рис. 9.8. Графическое изображение результатов прогнозирования средствами *MSExcel*

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 9.3. По данным временного ряда об объеме производства сельскохозяйственной продукции в Александровском муниципальном районе Ставропольского края (в фактически действовавших ценах) рассчитать индивидуальные и средние показатели динамики, сделать выводы.

Таблица 9.4

Объем производства сельскохозяйственной продукции в Александровском муниципальном районе Ставропольского края (в фактически действовавших ценах) за период 2009–2013 годов, тыс. руб.

Год	У, руб/т	Год	У, руб/т
2009	1078911	2012	1719063
2010	1499490	2013	2727583
2011	2443961		

Задача 9.4. По данным таблицы 9.5 осуществите расчет прогнозных значений инвестиций в основной капитал сельского хозяйства в Ставропольском крае до 2016 года.

Таблица 9.5

Инвестиции в основной капитал сельского хозяйства, млн руб.

Год	Уровень ряда	Год	Уровень ряда
2006	2863,4	2010	8552
2007	2237,2	2011	6875,3
2008	3871,8	2012	6861,8
2009	6332,9	2013	8429,2

Задача 9.5. По данным таблицы 9.6 осуществите прогноз среднемесячной начисленной заработной платы работников сельского хозяйства в Ставропольском крае до 2016 года.

Таблица 9.6

Среднемесячная начисленная заработная плата работников сельского хозяйства, руб.

Год	Уровень ряда	Год	Уровень ряда
2007	5623,62	2010	10412,15
2008	7814,65	2011	12374,31
2009	9001,17	2012	13464,82

Задача 9.6. По данным таблицы 9.7 осуществите расчет прогнозных значений себестоимости молока в Ставропольском крае до 2016 года.

Таблица 9.7

Производственная себестоимость молока, руб./т

Год	Уровень ряда	Год	Уровень ряда
2007	7630,67	2011	12609,96
2008	9840,76	2012	13189,57
2009	10506,54	2013	14068,5
2010	12327,73		

Задача 9.7. По данным таблицы 9.8 осуществите расчет прогнозных значений среднесложившихся цен на зерно в Ставропольском крае до 2016 года.

Таблица 9.8

Среднесложившиеся цены на зерно, реализованное по всем каналам, тыс. руб. за 1 тонну

Год	Уровень ряда	Год	Уровень ряда
2007	4636,39	2010	4348,97
2008	4369,24	2011	4965,77
2009	3632,45	2012	6892,15

ГЛАВА

10 СИСТЕМЫ ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ

10.1. Необходимость использования систем уравнений

Многие экономические взаимосвязи допускают моделирование одним уравнением. При формировании и построении эконометрических моделей в предыдущих главах предполагалось, что между независимыми переменными x_1, x_2, ..., x_m и зависимой переменной y в каждый момент времени существует только прямая связь: $x_i \to y$, $i = 1, 2, ..., m$. В такой ситуации зависимая переменная y не оказывает никакого влияния на переменные, входящие в правую часть модели. В большинстве случаев использование МНК для оценки параметров таких моделей является наиболее подходящей процедурой.

Применение отдельных уравнений регрессии, например для экономических расчетов, в большинстве случаев предполагает, что аргументы (факторы) можно изменять независимо друг от друга. Однако это предположение является очень грубым: практически изменение одной переменной, как правило, не может происходить при абсолютной неизменности других. Ее изменение повлечет изменения во всей системе взаимосвязанных признаков. Следовательно, отдельно взятое уравнение множественной регрессии не может характеризовать истинное влияние отдельных признаков на вариацию результирующей переменной.

Именно поэтому в последнее время при исследовании экономических и социальных явлений важное место заняла проблема описания структурных связей между переменными не одним, а несколькими уравнениями, содержащими как повторяющиеся, так и собственные переменные. В силу этого возникает необходимость использования так называемых ***систем эконометрических уравнений***.

Например, при оценке эффективности производства нельзя руководствоваться только моделью рентабельности. Она должна быть дополнена моделью производительности труда, а также моделью себестоимости единицы продукции.

В еще большей степени возрастает потребность в использовании системы одновременных уравнений при исследовании на макроэкономическом уровне. Так, модель национальной экономики включает в себя систему уравнений: функции потребления, инвестиций заработной платы, а также тождество доходов и т.д. Это связано с тем, что макроэкономические показатели, являясь обобщающими показателями состояния экономики, чаще всего взаимосвязаны и взаимозависимы. Так, расходы на конечное потребление в экономике зависят от валового национального дохода. Вместе с тем величина валового национального дохода рассматривается как функция инвестиций.

Одна из простейших систем одновременных уравнений используется при моделировании ***спроса и предложения в рыночной экономике***. В этом случае в предположении, что спрос Q^D и предложение Q^S являются линейными функциями от цены P, получим следующую систему уравнений:

$$\begin{array}{ll} \text{функция спроса:} & \left\{ \begin{array}{l} Q^D = \alpha_0 + \alpha_1 P + \Delta_1, \\ Q^S = \beta_0 + \beta_1 P + \Delta_2, \\ Q^D = Q^S. \end{array} \right. \\ \text{функция предложения:} & \\ \text{условие равновесия:} & \end{array}$$

Очевидно, что наличие случайных отклонений в данных моделях связано в первую очередь с отсутствием ряда важных объясняющих переменных (дохода, цен сопутствующих товаров, вкусов, ожиданий, цены ресурсов, налогов и т.д.). Изменение одного из этих факторов может отразиться на модели. Например, рост дохода потребителей может сдвинуть кривую спроса вверх (рис. 10.1). Это приведет к изменению равновесной цены и равновесного количества.

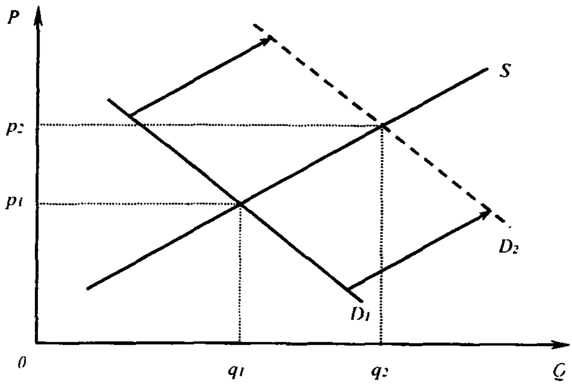


Рис. 10.1. Изменение равновесной цены и равновесного количества с ростом дохода потребителей

Модель спроса и предложения может быть усовершенствована. Если в функцию спроса добавить доход потребителей I, то система будет иметь вид:

функция спроса:
функция предложения:
условие равновесия:

$$\begin{cases} Q^D = \alpha_0 + \alpha_1 P + \alpha_2 I + \Delta_1, \\ Q^S = \beta_0 + \beta_1 P + \Delta_2, \\ Q^D = Q^S. \end{cases}$$

В качестве еще одного примера простейшей системы одновременных уравнений можно привести *кейнсианскую модель формирования доходов*. Данная модель описана в предположении, что рассматривается закрытая экономика без государственных расходов:

функция потребления:
макроэкономическое тождество:

$$\begin{cases} C = \beta_0 + \beta_1 Y + \Delta, \\ Y = C + I. \end{cases}$$

Здесь Y, C, I, представляют совокупный выпуск, объемы потребления и инвестиции соответственно.

Таким образом, системы одновременных (взаимозависимых) эконометрических моделей выражают через структурную форму взаимосвязей между явлениями их новое содержание, характеризующееся взаимным влиянием друг на друга некоторых зависимых и независимых переменных.

10.2. Составляющие и формы систем уравнений в эконометрических исследованиях

При рассмотрении систем эконометрических уравнений переменные делятся на два больших класса — *эндогенные* и *экзогенные* переменные. *Эндогенные переменные* — это переменные, значения которых определяются внутри модели и обозначаются обычно как y. *Экзогенные переменные* — это внешние по отношению к модели переменные, их значения определяются вне модели, и поэтому они считаются фиксированными, обозначаются обычно как x.

С математической точки зрения, главное отличие между экзогенными и эндогенными переменными заключается в том, что экзогенные переменные не коррелируют с ошибками регрессии, между тем как эндогенные могут коррелировать (и, как правило, коррелируют). Такая классификация переменных позволяет указать методы определения эндогенных переменных.

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

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

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \Delta_1, \\ y_2 = a_{20} + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \Delta_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = a_{n0} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \Delta_n. \end{cases} \tag{10.1}$$

Набор факторов x_i в каждом уравнении может варьировать. Отсутствие того или иного фактора в уравнении системы может быть следствием как экономической нецелесообразности его включения в модель, так и несущественности его воздействия на результативный признак (незначимо значение t-критерия или частного F-критерия для данного фактора).

В качестве примера такой модели может быть приведена модель экономической эффективности сельскохозяйственного производства вида:

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \Delta_1, \\ y_2 = a_{20} + a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \Delta_2, \end{cases} \tag{10.2}$$

где y_1 — продуктивность коров;

y_2 — себестоимость 1 ц молока;

x_1 — специализация хозяйства;

x_2 — количество голов на 100 га пашни;

x_3 — затраты труда.

Каждое уравнение системы независимых уравнений может рассматриваться самостоятельно. Для нахождения его параметров используется метод наименьших квадратов. Так как факторы, включенные в модель, полностью не объясняют изменение зависимых переменных, то в уравнениях присутствует свободный член a_0.

2. *Система рекурсивных уравнений,* когда зависимая переменная y включается в каждое последующее уравнение в качестве факторов:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \Delta_1, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \Delta_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 + \ldots + a_{3m}x_m + \Delta_3, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \Delta_n. \end{cases} \tag{10.3}$$

Примером такой системы может служить модель производительности труда и фондоотдачи вида:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \Delta_1, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \Delta_2, \end{cases} \tag{10.4}$$

где y_1 — производительность труда;

y_2 — фондоотдача;

x_1 — фондовооруженность труда;

x_2 — энерговооруженность труда;

x_3 — квалификация рабочих.

В таких моделях параметры уравнения оцениваются поэтапно ($y_1 \to y_2 \to y_3 \to \dots \to y_n$). Применение МНК для таких моделей позволяет получить несмещенные и состоятельные оценки. Однако модели данного типа встречаются достаточно редко.

3. *Система взаимозависимых*, совместных, одновременных *уравнений* (или *структурная форма модели*). В ней одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других уравнениях — в правую часть системы:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \dots + b_{1n}y_n + a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m + \Delta_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \dots + b_{2n}y_n + a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m + \Delta_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \dots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nm}x_m + \Delta_n. \end{cases} \tag{10.5}$$

В данной системе одни и те же переменные одновременно рассматриваются как зависимые в одних уравнениях и как независимые в других. В эконометрических исследованиях системы совместных уравнений получили наибольшее распространение.

Уравнения, составляющие систему взаимозависимых уравнений, называются ***структурными уравнениями модели***. Обычно их подразделяют на ***поведенческие уравнения*** и ***уравнения-тождества***. В первых из них описываются взаимодействия между переменными (регрессивные уравнения). Во вторых — соотношения, которые должны выполняться во всех случаях (при этом тождества не содержат подлежащие оценке параметры и случайные составляющие).

При этом эндогенные и экзогенные переменные принято, как уже отмечалось ранее, обозначать через y и x соответственно, а коэффициенты при них, которые называются ***структурными коэффициентами модели***, — через b_i и a_i, где b_i — коэффициент при эндогенной переменной, a_i — коэффициент при экзогенной переменной.

Структурная форма модели позволяет увидеть влияние изменений любой экзогенной переменной на значения эндогенной переменной. Целесообразно в качестве экзогенной переменной выбирать такие переменные, которые могут быть объектом регулирования. Меняя их и управляя ими, можно заранее иметь целевые значения эндогенных переменных.

В отличие от предыдущих систем (независимых, рекурсивных) каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим. С этой целью необходимо использовать специальные методы оценивания.

Как один из необходимых этапов определения структурных коэффициентов модели структурная форма модели преобразуется в ***приведенную форму модели***, которая представляет собой систему линейных функций эндогенных переменных от экзогенных:

$$\begin{cases} y_1 = \lambda_{11}x_1 + \lambda_{12}x_2 + \ldots + \lambda_{1m}x_m, \\ y_2 = \lambda_{21}x_1 + \lambda_{22}x_2 + \ldots + \lambda_{2m}x_m, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = \lambda_{n1}x_1 + \lambda_{n2}x_2 + \ldots + \lambda_{nm}x_m, \end{cases} \tag{10.6}$$

где λ_i — коэффициент приведенной формы модели.

По своему виду приведенная форма модели ничем не отличается от системы независимых уравнений, параметры которых оцениваются традиционным МНК. Применяя МНК, можно оценить λ, а затем оценить значение эндогенных переменных через экзогенные.

Рассмотрим пример преобразования структурной формы модели в приведенную на основе кейнсианской модели формирования доходов:

$$\begin{array}{ll} \text{функция потребления:} & \left\{\begin{array}{l} C = \beta_0 + \beta_1 Y + \Delta, \\ Y = C + I. \end{array}\right. \\ \text{макроэкономическое тождество:} & \end{array}$$

В данной модели C и Y являются эндогенными переменными, которые оцениваются внутри модели. Переменная I задается вне модели, следовательно, она является экзогенной переменной. Обозначив эндогенные и экзогенные переменные через y и x соответственно, выразим коэффициенты приведенной формы модели (λ) через коэффициенты структурной модели:

$$\begin{cases} y_1 = b_0 + b_1 y_2 + \Delta, \\ y_2 = y_1 + x, \end{cases}$$

где y_1 — объем потребления (C);

y_2 — совокупный выпуск (Y);

x — инвестиции (I);

b_0 и b_1 — структурные коэффициенты линейной зависимости y_1 от y_2.

Приведенная форма модели имеет вид:

$$\begin{cases} y_1 = \lambda_{10} + \lambda_{11} x + \varepsilon_i, \\ y_2 = \lambda_{20} + \lambda_{21} x + \varepsilon_i. \end{cases}$$

Первое уравнение приведенной формы модели получено путем подстановки в первое уравнение структурной модели вместо y_2 правой части второго уравнения системы, в результате чего было получено равенство

$$y_1 = b_0 + b_1 \cdot (y_1 + x) + \Delta_i,$$

отсюда $$y_1 - b_1 y_1 = b_0 + b_1 x + \Delta_i,$$

или $$y_1 \cdot (1 - b_1) = b_0 + b_1 x + \Delta_i.$$

Тогда $$y_1 = \frac{b_0}{(1-b_1)} + \frac{b_1}{(1-b_1)} \cdot x + \frac{\Delta_i}{(1-b_1)}.$$

Таким образом, мы представили первое уравнение структурной формы модели в виде *уравнения приведенной формы модели*:

$$y_1 = \lambda_{10} + \lambda_{11}x + \varepsilon_i.$$

Из уравнения следует, что коэффициенты приведенной формы модели представляют собой ***нелинейные соотношения коэффициентов структурной формы модели***, т.е.

$$\lambda_{10} = \frac{b_0}{1-b_1} \text{ и } \lambda_{11} = \frac{b_1}{1-b_1} \text{ и } \varepsilon_i = \frac{\Delta_i}{1-b_1}.$$

Аналогично можно показать, что коэффициенты приведенной формы модели второго уравнения системы (λ_{20} и λ_{21}) также нелинейно связаны с коэффициентами структурной модели. Для этого необходимо из второго уравнения структурной модели выразить y_1 и подставить в первое уравнение:

$$y_2 - x = b_0 + b_1 y_2 + \Delta_i,$$

отсюда
$$y_2 = \frac{b_0}{1-b_1} + \frac{1}{1-b_1}x + \frac{\Delta_i}{1-b_1},$$

что соответствует уравнению приведенной формулы модели:

$$y_2 = \lambda_{20} + \lambda_{21}x + \varepsilon_i,$$

т.е. $$\lambda_{20} = \frac{b_0}{1-b_1} \text{ и } \lambda_{21} = \frac{1}{1-b_1} \text{ и } \varepsilon_i = \frac{\Delta_i}{1-b_1}.$$

При этом заметим, что $\frac{1}{1-b_1}$ представляет собой денежный мультипликатор, определяющий, насколько увеличивается совокупный доход при увеличении объема инвестиций на единицу.

10.3. Смещенность и несостоятельность оценок МНК для систем одновременных уравнений

Непосредственное применение МНК для каждого из уравнений системы одновременных уравнений приводит к получению смещенных и несостоятельных оценок. Обычно это происходит вследствие коррелированности одной или нескольких объясняющих переменных со случайным отклонением. Для демонстрации данного вывода рассмотрим кейнсианскую модель. Для получения несмещенных и состоятельных оценок параметров уравнения

$$C = \beta_0 + \beta_1 Y + \Delta$$

по МНК необходимо выполнение ряда предпосылок.

1. $M(\Delta) = 0$ (первая предпосылка МНК): данное условие означает, что случайное отклонение не оказывает влияния на зависимую переменную. В каждом конкретном случае ошибка может быть либо положительной, либо отрицательной, но не должна иметь систематического смещения.

2. $S_\Delta^2 = \mathrm{const}$ (вторая предпосылка МНК): дисперсия случайных отклонений Δ_i постоянна. Данное условие подразумевает: несмотря на то что при каждом конкретном наблюдении случайное отклонение может быть либо большим, либо меньшим, не должно быть некой априорной причины, вызывающей большую ошибку (отклонение). Поскольку

$$S_\Delta^2 = M(\Delta_i - M(\Delta_i))^2 = M(\Delta_i^2),$$

то данную предпосылку можно перезаписать в форме

$$M(\Delta_i^2) = \sigma^2.$$

3. $\mathrm{cov}(\Delta_i, \Delta_j) = 0$ где $i \neq j$ (третья предпосылка МНК): случайные отклонения являются независимыми друг от друга, что означает отсутствие систематической связи между любыми случайными отклонениями, т.е. величина и определенный знак любого случайного отклонения не должны быть причинами величины и знака любого другого отклонения. Данное условие может быть записано в следующем виде:

$$M(\Delta_i \Delta_j) = 0.$$

4. $\mathrm{cov}(y_i, \Delta_i) = 0$ для любых отклонений, т.е. значения зависимых факторов модели и ошибки должны быть независимыми.

Однако из приведенных уравнений системы кейнсианской модели нетрудно заметить, что Y(совокупный выпуск) линейно зависит от случайного отклонения Δ_i:

$$y = \frac{b_0}{1-b_1} + \frac{1}{1-b_1} I + \frac{\Delta_i}{1-b_1}.$$

Следовательно, объясняющая переменная Y коррелирует со случайным отклонением Δ_i ($\mathrm{cov}(Y, \Delta_i) \neq 0$). Действительно:

$$M(Y) = \frac{b_0}{1-b_1} + \frac{1}{1-b_1} \cdot I.$$

Вычитая математическое ожидание $M(Y)$ из Y, имеем:

$$y - M(Y) = \frac{\Delta_i}{1-b_1}.$$

Следовательно,

$$\mathrm{cov}(Y, \Delta_i) = M((Y - M(Y))(\Delta_i - M(\Delta_i))) =$$
$$= M(\frac{\Delta_i}{1-b_1} \cdot \Delta_i) = \frac{1}{1-b_1} \cdot M(\Delta_i^2) = \frac{\sigma^2}{1-b_1} > 0.$$

Здесь используется предположение экономической теории о том, что предельная склонность к потреблению $0 < \beta_1 < 1$.

Если для нахождения неизвестных параметров уравнения парной линейной регрессии разделить оба уравнения системы на n, получим:

$$\begin{cases} Nb_0 + b_1 \sum x_i = \sum y_i; \\ b_0 \sum x_i + b_1 \sum x_i^2 = \sum x_i y_i; \end{cases} \Rightarrow \quad \begin{cases} b_0 + b_1 \bar{x} = \bar{y}; \\ b_0 \bar{x} + b_1 \overline{x^2} = \overline{xy}. \end{cases}$$

Отсюда

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum (x_i - \bar{x})^2},$$

$$b_0 = \bar{y} - b_1 \bar{x}.$$

Таким образом, оценка b_1 параметра β_1 по МНК для кейнсианской модели определяется по формуле

$$b_1 = \frac{\sum(C_i - \bar{C})\cdot(Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2} = \frac{\sum C_i \cdot (Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2} - \frac{\sum \bar{C}\cdot(Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2} = \frac{\sum C_i \cdot (Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2}.$$

Подставив вместо C_i правую часть функции потребления кейнсианской модели, получим

$$b_1 = \frac{\sum(\beta_0 + \beta_1 Y_i + \Delta_i)\cdot(Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2} = \frac{\beta_0 \cdot \sum(Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2} +$$

$$+ \frac{\beta_1 \cdot \sum Y_i \cdot (Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2} + \frac{\sum \Delta_i \cdot (Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2} = \beta_1 + \frac{\sum \Delta_i \cdot (Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2}.$$

Тогда

$$M(b_1) = \beta_1 + M\left[\frac{\sum \Delta_i \cdot (Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2}\right].$$

Математическое ожидание

$$M\left[\frac{\sum \Delta_i \cdot (Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2}\right]$$

нельзя вычислить непосредственно, так как числитель и знаменатель дроби не являются независимыми случайными величинами (оба зависят от Δ). Однако при больших объемах выборки можно сделать следующие выводы:

$$\left[\frac{\sum \Delta_i \cdot (Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2}\right] = \frac{\frac{1}{n}\sum \Delta_i \cdot (Y_i - \bar{Y})}{\frac{1}{n}\sum(Y_i - \bar{Y})} \xrightarrow[n\to\infty]{} \frac{\operatorname{cov}(\Delta_i, Y_i)}{S_{\hat{y}}^2}.$$

Таким образом, оценка b_1 является смещенной (скорее всего, завышенной при условии, что $0 < \beta_1 < 1$) оценкой параметра β_1. Причем эту смещенность нельзя преодолеть даже при бесконечном увеличении выборки, т.е. b_1 является еше и несостоятельной оценкой β_1.

10.4. Проблема идентификации

Изменение формы уравнения хотя и позволяет устранить проблему коррелированности объясняющей переменной и случайного отклонения, но может привести к другой, не менее серьезной проблеме — *проблеме идентификации*, под которой понимается возможность численной оценки параметров структурных уравнений по оценкам коэффициентов приведенных уравнений.

Исходную систему уравнений называют *идентифицируемой* (точно определенной), если по коэффициентам приведенных уравнений можно однозначно определить значения коэффициентов структурных уравнений. Обычно это удается сделать, когда количество уравнений для определения коэффициентов структурных уравнений в точности равно количеству этих коэффициентов.

Исходную систему уравнений называют *сверхидентифицируемой* (переопределенной), если по коэффициентам приведенных уравнений можно получить несколько вариантов значений коэффициентов структурных уравнений. Обычно в таких случаях число уравнений для оценки коэффициентов структурных уравнений больше числа определяемых коэффициентов.

Исходную систему уравнений называют *неидентифицируемой* (недоопределенной), если по коэффициентам приведенных уравнений невозможно определить значения коэффициентов структурных уравнений. В этом случае система, связывающая коэффициенты структурных уравнений с коэффициентами приведенных уравнений, является несовместимой. Обычно это происходит тогда, когда количество уравнений менее числа коэффициентов структурных уравнений.

Выполнение условия идентифицируемости модели проверяется для каждого уравнения системы. Чтобы уравнение 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 данном уравнении системы через H, а число экзогенных (предопределенных) переменных, которые содержатся в системе, но не входят в данное уравнение, — через D, то условие идентифицируемости модели может быть записано в виде следующего счетного правила:

$D + 1 = H$ — уравнение идентифицируемо;

$D + 1 < H$ — уравнение неидентифицируемо;

$D + 1 > H$ — уравнение сверхидентифицируемо.

Приведем примеры использования данных условий для определения идентифицируемости структурных уравнений.

1. В простой модели «спрос 3 предложение»:

$$\begin{cases} Q_i^D = \alpha_0 + \alpha_1 P_i + \Delta_{1i}, \\ Q_i^S = \beta_0 + \beta_1 P_i + \Delta_{2i}, \end{cases}$$

где Q_i^D, Q_i^S, P_i — эндогенные переменные, соответственно величины спроса, предложения и цены.

Для каждого из уравнений $H = 2$, $D = 0$. Следовательно, $D + 1 < H$, это означает, что оба они неидентифицируемы.

2. Добавим в исходную модель в первое уравнение экзогенную переменную I (доход потребителей):

$$\begin{cases} Q_i^D = \alpha_0 + \alpha_1 P_i + \alpha_2 I_i + \Delta_{1i}, \\ Q_i^S = \beta_0 + \beta_1 P_i + \Delta_{2i}. \end{cases}$$

Для первого уравнения системы $H = 2$ (Q_i^S и P_i) и $D = 0$, тогда выполняется неравенство $D + 1 < H$, т.е. уравнение неидентифицируемо. Для второго уравнения системы выполняется равенство $D + 1 = H$, так как $H = 2$, $D = 0$. Следовательно, функция предложения идентифицируема и может быть определена однозначно.

3. В модели

$$\begin{cases} Q_i^D = \alpha_0 + \alpha_1 P_i + \alpha_2 I_i + \Delta_{1i}, \\ Q_i^S = \beta_0 + \beta_1 P_i + \beta_2 P_{i-1} + \Delta_{2i}, \end{cases}$$

где P_{i-1} — экзогенная переменная, цены в предыдущий момент времени.

Для каждого уравнения системы $D = 1$, $H = 2$. В этом случае для любого из уравнений $D + 1 = H$, что означает: оба уравнения системы точно идентифицируемы.

4. Рассмотрим модель «спроса — предложение» с числом экзогенных переменных, превышающим количество структурных уравнений:

$$\begin{cases} Q_i^D = \alpha_0 + \alpha_1 P_i + \alpha_2 I_i + \alpha_3 S_i + \Delta_{1i}, \\ Q_i^S = \beta_0 + \beta_1 P_i + \beta_2 P_{i-1} + \Delta_{2i}, \end{cases}$$

где S_i — объем сбережений (экзогенная переменная).

Для первого уравнения $H = 2$ (Q_i^D и P_i) и $D = 1$ (P_{i-1}), тогда $D + 1 = H$, уравнение точно идентифицируемо. Во втором уравнении системы $H = 2$ (Q_i^D и P_i) и $D = 1$ (I_i и S_i), так как $2 + 1 > 2$. Данное уравнение системы является переопределенным (сверхидентифицируемым).

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

Рассмотренное счетное правило отражает *необходимое условие идентификации,* но недостаточное.

Более точно условия идентификации определяются, если накладывать ограничение 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 не равен нулю, а ранг матрицы не менее чем число эндогенных переменных без одного.

Достаточные условия идентифицируемости:

а) $\det A \neq 0$;

б) $\operatorname{rank}(A) \geq M - 1$, где M — число эндогенных переменных в системе.

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 системы выполнено счетное правило, а определитель матрицы названных коэффициентов равен нулю. В этом случае соблюдается необходимое, но недостаточное условие идентификации.

Когда достаточное условие идентификации не выполняется, то уравнение неидентифицируемо. Следовательно, рассматриваемая в целом структурная модель идентифицируема по счетному правилу, не может считаться идентифицируемой исходя из достаточного условия идентификации.

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

Рассмотрим в качестве примера эконометрическую модель страны:

$$\begin{cases} y_1 = a_{01} + b_{13}y_3 + b_{14}y_4 + \Delta_1, \\ y_2 = a_{02} + b_{23}y_3 + a_{21}x_1 + \Delta_2, \\ y_3 = a_{03} + b_{34}y_4 + a_{31}x_1 + \Delta_3, \\ y_4 = y_1 + y_2 + x_2, \end{cases}$$

где y_1 — расходы на конечное потребление данного года;
y_2 — валовые инвестиции в текущем году;
y_3 — расходы на заработную плату в текущем году;
y_4 — валовой доход за текущий год;
x_1 — валовой доход предыдущего года;
x_2 — государственные расходы текущего года;
a_0 — свободный член уравнения;
Δ — случайные ошибки.

В данной модели четыре эндогенные переменные, причем переменная y_4 задана тождеством. Поэтому практически статистическое решение необходимо только для первых трех уравнений системы, которые необходимо проверить на идентификацию. Модель также содержит две предопределенные (экзогенные) переменные — x_2 и x_1 (лаговая).

В рассматриваемой эконометрической модели первое уравнение системы точно идентифицируемо, так как $H = 3$, $D = 2$, выполняется необходимое условие идентификации ($D + 1 = H$). Кроме того, выполняется и достаточное условие идентификации, т.е. ранг матрицы равен 3, а определитель ее не равен нулю: $\det A = -a_{31}$, что видно из следующей таблицы:

Уравнение	y_2	x_1	x_2
2	–1	a_{21}	0
3	0	a_{31}	0
4	1	0	1

Второе уравнение системы также точно идентифицируемо: $H = 2$ и $D = 1$, т.е. счетное правило выполняется: $D + 1 = H$, также выполнено достаточное условие идентификации: $\operatorname{rank}(A) = 3$ и $\det A = -b_{34}$:

Уравнение	y_1	y_4	x_2
1	–1	b_{14}	0
3	0	b_{34}	0
4	1	–1	1

Аналогично третье уравнение системы также идентифицируемо: $H = 2$, $D = 1$, $D + 1 = 2$ и $\det A \neq 0$, а ранг матрицы $A = 3$ и $\det A = 1$:

Уравнение	y_1	y_2	x_2
1	–1	0	0
2	0	–1	0
4	1	1	1

Идентификация уравнений — достаточно сложная процедура. На структурные модели могут накладываться и другие ограничения, например, в производственной функции сумма эластичностей может быть равна по предположению 1. Могут накладываться ограничения на дисперсии и ковариации остаточных величин.

10.5. Методология оценивания параметров систем уравнений

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

- косвенный метод наименьших квадратов;
- двухшаговый метод наименьших квадратов;
- трехшаговый метод наименьших квадратов;
- метод максимального правдоподобия с полной информацией;
- метод максимального правдоподобия при ограниченной информации.

Косвенный и двухшаговый методы наименьших квадратов рассматриваются как традиционные методы оценки коэффициентов структурной модели. Эти методы достаточно легко реализуемы. Косвенный метод наименьших квадратов (КМНК) применяется для идентифицируемой системы одновременных уравнений, а двухшаговый метод наименьших квадратов (ДМНК) используется для оценки коэффициентов сверхидентифицируемой модели. Остальные перечисленные методы также используются для сверхидентифицируемых систем уравнений.

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 в качестве модификации применяется метод максимального правдоподобия при ограниченной информации (метод наименьшего дисперсионного отношения), разработанный в 1949 году Т. Андерсоном и Н. Рубиным. Математическое описание метода дано, например, в работе Дж. Джонстона. Несмотря на значительную популярность данного метода, к середине 60-х годов он был практически вытеснен двухшаговым методом наименьших квадратов (ДМНК) в связи с гораздо большей простотой последнего.

Дальнейшим развитием ДМНК является трехшаговый МНК (ТМНК), предложенный в 1962 году А. Зельнером и Г. Тейлом. Этот метод оценивания пригоден для всех видов уравнений структурной модели. Однако при некоторых ограничениях на параметры более эффективным оказывается ДМНК.

10.5.1. Косвенный метод наименьших квадратов (КМНК)

В силу невозможности получения на основе «обычного» МНК качественных оценок параметров системы одновременных уравнений необходимо использовать другие методы. Один из таких методов — *косвенный метод наименьших квадратов* (КМНК), основанный на использовании приведенных уравнений.

Как уже отмечалось, КМНК применяется в случае точно идентифицируемой структурной модели. Смысл названия данного метода очевиден в силу вычисления оценок параметров через оценки приведенных уравнений. Оценки, полученные по КМНК, являются состоятельными, а следовательно, при больших выборках велика вероятность, что они будут близки к истинным значениям параметров.

Таким образом, КМНК включает в себя следующие этапы:

1. Исходя из структурных уравнений строятся уравнения в приведенной форме.

2. Параметры уравнений в приведенной форме оцениваются по МНК.

3. На основе оценок, найденных на 2-м этапе, оцениваются параметры структурных уравнений.

Рассмотрим модель «спрос — предложение»:

$$\begin{array}{ll} \text{спрос:} & \begin{cases} q_t = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \Delta_{1t}, \\ q_t = \beta_0 + \beta_1 p_t + \Delta_{2t}, \end{cases} \\ \text{предложение:} & \end{array}$$

где q_t, p_t — эндогенные переменные: количество товара и цена в году t;

y_t — экзогенная переменная: доход потребителей;

Δ_{1t}, Δ_{2t} — случайные отклонения.

На основании следующих статистических данных необходимо оценить коэффициенты *функции предложения*, используя для этого МНК и КМНК. Сравнить результаты.

№ п/п	p_t	q_t	y_t	p_t^2	y_t^2	$p_t q_t$	$p_t y_t$	$q_t y_t$
1	1	8	2	1	4	8	2	16
2	2	10	4	4	16	20	8	40
3	3	7	3	9	9	21	9	21
4	4	5	5	16	25	20	20	25
5	5	1	2	25	4	5	10	2
Сумма	5	31	16	55	58	74	49	104
Среднее	3	6,2	3,2	11	11,6	14,8	9,8	20,8

Построим приведенные уравнения данной системы. Для этого вычтем из функции предложение функцию спрос. Получим:

$$(\beta_0-\alpha_0)+(\beta_1 p_t-\alpha_1 p_t)-\alpha_2 y_t+(\Delta_{2t}-\Delta_{1t})=0,$$

отсюда

$$p_t=\frac{\alpha_0-\beta_0}{\beta_1-\alpha_1}+\frac{\alpha_2}{\beta_1-\alpha_1}\cdot y_t+\frac{\Delta_{1t}-\Delta_{2t}}{\beta_1-\alpha_1}.$$

Следовательно, приведенное уравнение будет иметь вид

$$p_t=\lambda_{10}+\lambda_{11}\cdot y_t+v_{1t},$$

где $\lambda_{10}=\dfrac{\alpha_0-\beta_0}{\beta_1-\alpha_1}$, $\lambda_{11}=\dfrac{\alpha_2}{\beta_1-\alpha_1}$, $v_{1t}=\dfrac{\Delta_{1t}-\Delta_{2t}}{\beta_1-\alpha_1}$.

Подставим правую часть приведенного уравнения в функцию предложения, получим:

$$q_t=\beta_0+\beta_1\cdot(\lambda_{10}+\lambda_{11}\cdot y_t+v_{1t})+\Delta_{2t}.$$

Отсюда

$$q_t=(\beta_0+\beta_1\lambda_{10})+(\beta_1\cdot\lambda_{11})y_t+(\beta_1\cdot v_{1t}+\Delta_{2t}).$$

Следовательно, приведенное уравнение имеет вид

$$q_t=\lambda_{20}+\lambda_{21}\cdot y_t+v_{2t},$$

где $\lambda_{20} = \beta_0 + \beta_1\lambda_{10}$, $\lambda_{21} = \beta_1 \cdot \lambda_{11}$, $v_{2t} = \beta_1 \cdot v_{1t} + \Delta_{2t}$.

Приведенная форма модели составит:

$$\begin{cases} p_t = \lambda_{10} + \lambda_{11} \cdot y_t + v_{1t}, \\ q_t = \lambda_{20} + \lambda_{21} \cdot y_t + v_{2t}. \end{cases}$$

Для оценки неизвестных параметров парной линейной регрессии воспользуемся следующими формулами:

$$a = \overline{y} - b \cdot \overline{x} \text{ и } b = \frac{\overline{yx} - \overline{y} \cdot \overline{x}}{\overline{x^2} - \overline{x}^2}.$$

По имеющимся статистическим данным оценим коэффициенты приведенных уравнений:

$$\lambda_{11} = \frac{\overline{p_t y_t} - \overline{p_t} \cdot \overline{y_t}}{\overline{y_t^2} - \overline{y}_t^2} = \frac{9{,}8 - 3 \cdot 3{,}2}{11{,}6 - 3{,}2^2} = \frac{0{,}2}{1{,}36} = 0{,}1471.$$

$$\lambda_{10} = \overline{p_t} - \lambda_{11} \cdot \overline{y_t} = 3 - 0{,}1471 \cdot 3{,}2 = 2{,}5294.$$

$$\lambda_{21} = \frac{\overline{q_t y_t} - \overline{q_t} \cdot \overline{y_t}}{\overline{y_t^2} - \overline{y}_t^2} = \frac{20{,}8 - 6{,}2 \cdot 3{,}2}{11{,}6 - 3{,}2^2} = \frac{0{,}96}{1{,}36} = 0{,}7059.$$

$$\lambda_{20} = \overline{q_t} - \lambda_{21} \cdot \overline{y_t} = 6{,}2 - 0{,}7059 \cdot 3{,}2 = 3{,}9412.$$

Таким образом, приведенная форма модели имеет вид

$$\begin{cases} p_t = 2{,}5294 + 0{,}1471 \cdot y_t + v_{1t}, \\ q_t = 3{,}9412 + 0{,}7059 \cdot y_t + v_{2t}. \end{cases}$$

Нетрудно заметить, что в рассматриваемой модели в ее структурной форме только функция предположения является идентифицируемой, потому что для нее выполняется условие $D + 1 = H$. Оценки параметров β_1 и β_2 могут быть определены на основе оценок коэффициентов приведенных уравнений:

$$\lambda_{21} = \beta_1 \cdot \lambda_{11} \Rightarrow \beta_1 = \frac{\lambda_{21}}{\lambda_{11}},$$

$$\lambda_{20} = \beta_0 + \beta_1 \cdot \lambda_{10} \Rightarrow \beta_0 = \lambda_{20} - \beta_1 \cdot \lambda_{10}.$$

Следовательно, оценки коэффициентов функции предложения по КМНК будут равны:

$$\beta_1 = 0,7059 : 0,1471 = 4,7988;$$

$$\beta_0 = 3,9412 - 4,7988 \cdot 2,5294 = -8,1969.$$

Следовательно, функция предложения имеет вид

$$q_t = -8,1969 + 4,7988 \cdot p_t.$$

В то же время рассчитанные непосредственно по МНК оценки данного уравнения будут:

$$\beta_1 = \frac{\overline{q_t p_t} - \overline{q_t} \cdot \overline{p_t}}{\overline{p_t^2} - \overline{p}_t^2} = \frac{14,8 - 6,2 \cdot 3,0}{11,0 - 3,0^2} = \frac{-3,8}{2} = -1,9;$$

$$\beta_0 = \overline{q_t} - \beta_1 \cdot \overline{p_t} = 6,2 - (-1,9) \cdot 3 = 11,9.$$

Тогда функция предложения имеет вид

$$q_t = 11,9 - 1,9 \cdot p_t.$$

Полученные результаты позволяют сделать вывод о том, что применение МНК в несоответствующих ситуациях может существенно исказить картину зависимости. Нарушение предпосылки независимости факторов друг от друга при использовании МНК в системе одновременных уравнений приводит к несостоятельности оценок структурных коэффициентов; в ряде случаев они оказываются экономически бессмысленными. Опасность таких результатов возрастает при увеличении числа эндогенных переменных в правой части системы, ибо становится невозможным расщепить совместное влияние эндогенных переменных и видеть изолированные меры их воздействия в соответствии с предпосылками традиционного МНК.

10.5.2. Двухшаговый метод наименьших квадратов (ДМНК)

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

Основная идея ДМНК — 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 применить обычный МНК к структурной форме сверхидентифицируемого уравнения. Метод получил название двухшагового МНК, ибо дважды используется МНК: на первом шаге — при определении параметров приведенной формы модели и нахождении на их основе оценок теоретических знаний эндогенной переменной:

$$y_i = \lambda_{i1} \cdot x_1 + \lambda_{i2} \cdot x_2 + \ldots + \lambda_{ij} x_j$$

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.

Если все уравнения системы сверхидентифицируемые, то для оценки структурных коэффициентов каждого уравнения используется ДМНК. Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

Покажем использование ДМНК на примере модели IS-LM для закрытой экономики при фиксированной налоговой ставке (t):

$$\begin{cases} y = \alpha_0 + \alpha_1 r + \alpha_2 g + \alpha_3 t + \Delta_1, & (\alpha_1 < 0), \\ y = \beta_0 + \beta_1 r + \beta_2 m + \Delta_2 & (\beta_1 > 0), \end{cases}$$

где y, r — эндогенные переменные: национальный доход (Y), процентная ставка (r);

g, t, m — экзогенные переменные: государственные расходы (G), объем налогов (T), денежная масса (M).

Второе уравнение системы является переопределенным (сверхидентифицируемым) относительно эндогенной переменной r, так как $D+1>H$ ($H=2$, $D=2$). Для оценки коэффициентов данного уравнения рекомендуется использовать *метод инструментальных переменных*. Суть данного метода состоит в замене коррелирующей переменной на другую — *инструментальную переменную* (ИП), которая обладает следующими свойствами:

- она должна коррелировать (желательно сильно) с объясняющей переменной;
- она не должна коррелировать со случайным отклонением.

Итак, для того чтобы определить оценки сверхидентифицируемого уравнения, необходимо найти соответствующие ИП. Этот поиск позволяет осуществить *двухшаговый* МНК. Суть данного метода состоит в использовании в качестве ИП оценки переопределенной переменной, полученной на базе экзогенных (или предопределенных) переменных модели.

ШАГ 1

Во втором уравнении структурной формы модели переопределенной переменной является процентная ставка r. Ее можно оценить, опираясь на экзогенные переменные (например, вычитая из второго уравнения первое):

$$(\beta_0-\alpha_0)+(\beta_1 r-\alpha_1 r)+\beta_2 m-\alpha_2 g-\alpha_3 t+(\Delta_2-\Delta_1).$$

Приведенная форма уравнения будет иметь вид

$$r=\lambda_0+\lambda_1 m+\lambda_2 g+\lambda_3 t+v,$$

где $\lambda_0=\dfrac{\alpha_0-\beta_0}{\beta_1-\alpha_1}$; $\lambda_1=-\dfrac{\beta_2}{\beta_1-\alpha_1}$; $\lambda_2=\dfrac{\alpha_2}{\beta_1-\alpha_1}$;

$\lambda_3=\dfrac{\alpha_3}{\beta_1-\alpha_1}$; $v=\dfrac{\Delta_1-\Delta_2}{\beta_1-\alpha_1}$.

Применяя для приведенной формы уравнения МНК, получим оценку r переменной r:

$$r = \lambda_0 + \lambda_1 m + \lambda_2 g + \lambda_3 t,$$

где r — условная средняя при фиксированных значениях M, G, T.

ШАГ 2

Подставляя оценку $\hat{r}$ во второе уравнение исходной модели, имеем:

$$y = \beta_0 + \beta_1 r + \beta_2 m + \Delta_2.$$

Данная замена позволяет преодолеть такую существенную проблему сверхидентифицируемых моделей, как коррелированность объединяющей переменной со случайным членом (такая коррелированность приводит к получению смещенных и несостоятельных оценок). Действительно, оценка $\hat{r}$ выражается только через экзогенные переменные и, следовательно, не коррелирует со случайным отклонением. Фактически ее можно рассматривать как новую экзогенную переменную.

Заменив в исходной структурной модели второе уравнение на уравнение, полученное в результате *шага 2*, получим систему, которую можно оценить с помощью метода наименьших квадратов (МНК).

При наличии в модели более одной переопределенной переменной на первом этапе необходимо оценить все такие переменные.

ДМНК обладает определенными свойствами, делающими его весьма привлекательными для практического применения.

1. В данном методе первый этап (этап построения приведенных уравнений) применяется для конкретных уравнений, не затрагивая оставшиеся уравнения модели. Это позволяет минимизировать объем вычислений.

2. При наличии переопределенных (сверхидентифицируемых) уравнений ДМНК в отличие от МНК определяет единственные оценки параметров модели.

3. Применяя данный метод, достаточно использовать лишь экзогенные и переопределенные переменные модели.

4. Применение ДМНК будет эффективным лишь в том случае, когда коэффициент детерминации (R^2) приведенных уравнений, построенных на первом этапе, будет достаточно высоким. При этом ИП в очень малой степени коррелирует со случайным отклонением и будут близки к истинному значению (r) заменяемых переменных. При низком значении R^2 использование ДМНК малопродуктивно, так как в этом случае ИП весьма слабо соответствует истинному значению заменяемой переменной.

Использование метода инструментальных переменных как составной части ДМНК позволяет получать состоятельные оценки и оценки стандартных отклонений для выборок больших объемов. Для таких выборок выводы будут не столь конкретными.

ДМНК является наиболее обшим и широко распространенным методом решения системы одновременных уравнений. Для точно идентифицируемых уравнений ДМНК дает тот же результат, что и КМНК. Поэтому в ряде компьютерных программ для решения системы одновременных уравнений рассматривается лишь ДМНК.

10.5.3. Трехшаговый метод наименьших квадратов (ТМНК)

Как уже отмечалось, наличие корреляционных связей между ошибками различных эконометрических моделей, входящих во взаимозависимую систему, ведет к потере свойства эффективности оценок их коэффициентов. В такой ситуации рекомендуется для получения этих оценок вместо *двухшагового* использовать *трехшаговый* МНК, который включает в себя дополнительный этап, связанный с применением *обобщенного* МНК. В результате ТМНК применяется как метод оценивания коэффициентов структурной формы всей системы моделей, а не отдельных ее уравнений.

Таким образом, процедура оценки коэффициентов структурной формы всей системы взаимозависимых эконометрических моделей состоит из трех последовательных этапов, определяющих содержание ТМНК.

ШАГ 1

На данном этапе с использованием обычного МНК на основании приведенной формы определяются расчетные значения переменных $\hat{y}_i$, рассматриваемых в качестве независимых эндогенных переменных в каждом из уравнений системы.

ШАГ 2

Как и в ДМНК, на этом этапе с использованием значений $\bar{y}_i$ определяются оценки коэффициентов структурной формы каждого из уравнений системы.

ШАГ 3

С помощью обобщенного МНК определяются «окончательные» оценки коэффициентов структурной формы всей системы взаимозависимых эконометрических моделей, которые теоретически при наличии корреляции между ошибками различных уравнений являются «более эффективными» по сравнению с аналогичными оценками ДМНК.

Если ошибки системы не коррелируют между собой, то ТМНК не имеет преимуществ перед двухшаговым. При применении ТМНК необходимо соблюдать некоторые дополнительные правила, что делает его процедуру менее универсальной по сравнению с двухшаговой. Они состоят в следующем:

1. Процедура выполняется только для идентифицируемых и сверхидентифицируемых уравнений системы. Тождества и неидентифицируемые уравнения в ней не участвуют.

2. Процедуру желательно выполнять для групп идентифицируемых и сверхидентифицируемых уравнений раздельно. При этом если в соответствующую группу входит только одно сверхидентифицируемое уравнение, то трехшаговая процедура для него превращается в двухшаговую.

Практическое применение трехшагового метода ограничено из-за большого количества расчетов и чрезвычайной чувствительности результатов к ошибкам спецификации.

10.6. Применение систем эконометрических уравнений

Под системой эконометрических уравнений обычно понимается система одновременных, совместных уравнений. Ее применение имеет ряд сложностей, которые связаны с ошибками спецификации модели. Ввиду большого числа факторов, влияющих на экономические переменные, нет уверенности в точности предлагаемой модели для описания экономических процессов. Набор эндогенных и экзогенных переменных моделей соответствует теоретическому представлению о моделируемом объекте, которое сложилось на данный момент и может изменяться. Соответственно, может меняться и вид модели с точки зрения ее идентифицируемости.

Сверхидентифицируемую модель можно превратить в точно идентифицируемую путем добавления некоторых переменных или отбрасывания некоторых ограничений на параметры. При правильной спецификации модели она может оказаться идентифицируемой и поэтому переходит к сверхидентифицируемым и точно идентифицируемым моделям, несколько упрощающим характер взаимосвязей эконометрических явлений. Наличие множества прикладных моделей для решения одного и того же класса задач не случайно. Наиболее ярко это проявляется при построении макроэкономических моделей, когда одна и та же функция может включать в себя разный набор экономических переменных.

Рассмотрим основные направления практического использования эконометрических систем уравнений.

Наиболее широко системы одновременных уравнений используются при построении макроэкономических моделей функционирования экономики той или иной страны. Большинство из них представляют собой мультипликаторные модели кейнсианского типа с той или иной мерой сложности. В качестве примеров мы уже рассматривали кейнсианскую модель формирования доходов. В более поздних исследованиях статическая модель Кейнса включила уже не только функцию потребления, но и функцию сбережений:

$$\begin{cases} C = \beta_0 + \beta_1 y + \Delta_1, \\ r = T + K \cdot (C + I) + \Delta_2, \\ y = C + I - r, \end{cases}$$

где C, Y и I — те же по смыслу переменные, что и в предыдущей модели;

r — сбережение.

Данная модель содержит три эндогенные переменные — C, r, Y и одну экзогенную переменную I. Система идентифицируема: в первом уравнении $H = 2$ и $D = 1$, во втором $H = 1$ и $D = 0$; $C + I$ рассматривается как предопределенная переменная.

Наряду со статическими широкое распространение получили динамические модели экономики. В отличие от статических они содержат в правой части лаговые переменные, а также учитывают тенденцию (факторов времени). Примером могут служить модели А. Клейна, разработанные им для экономики США в 1950–1960-е годы. В упрощенном варианте модель Клейна рассматривается как конкретная модель:

$$\begin{cases} C_t = b_1 \cdot S_t + b_2 \cdot P_t + b_3 + \Delta_1, \\ I_t = b_4 \cdot P_t + b_5 \cdot P_{t-1} + b_6 + \Delta_2, \\ S_t = b_7 \cdot R_t + b_8 \cdot R_{t-1} + b_9 \cdot t + b_{10} + \Delta_3, \\ R_t = S_t + P_t + T_t, \\ R_t = C_t + I_t + G_t, \end{cases}$$

где C_t — функция потребления в период t;

S_t — заработная плата в период t;

P_t — прибыль в период t;

P_{t-1} — прибыль в период $t - 1$, т.е. в предыдущем году;

R_t — общий доход в период t;

R_{t-1} — общий доход в предыдущий год;

t — время;

T_t — чистые трансферты в пользу администрации в период t;
I_t — капиталовложения в период t;
G_t — спрос административного аппарата, правительственные расходы в период времени t.

Модель содержит пять эндогенных переменных — C_t, I_t, S_t, R_t (расположены в левой части системы) и P_t (зависимая переменная, определяемая по первому тождеству), три экзогенные переменные — T_t, G_t, t и две предопределенные, лаговые переменные — P_{t-1} и R_{t-1}. Как и большинство моделей такого типа, данная модель сверхидентифицируема и решаема ДМНК. Для прогнозных целей используется приведенная форма модели:

$$\begin{cases} C_t = d_1 \cdot T + d_2 \cdot G + d_3 \cdot t + d_4 \cdot P_{t-1} + d_5 \cdot R_{t-1} + v_1, \\ I_t = d_6 \cdot T + d_7 \cdot G + d_8 \cdot t + d_9 \cdot P_{t-1} + d_{10} \cdot R_{t-1} + v_2, \\ S_t = d_{11} \cdot T + d_{12} \cdot G + d_{13} \cdot t + d_{14} \cdot P_{t-1} + d_{15} \cdot R_{t-1} + v_3, \\ R_t = d_{16} \cdot T + d_{17} \cdot G + d_{18} \cdot t + d_{19} \cdot P_{t-1} + d_{20} \cdot R_{t-1} + v_4, \\ P_t = d_{21} \cdot T + d_{22} \cdot G + d_{23} \cdot t + d_{24} \cdot P_{t-1} + d_{25} \cdot R_{t-1} + V_5. \end{cases}$$

Динамическая модель может и не содержать учета тенденций, но лаговые переменные в ней обязательны. Динамическая модель Кейнса представлена следующими тремя уравнениями:

$$\begin{cases} C_t = a + b_1 \cdot Y_t + b_2 \cdot Y_{t-1} + \Delta_1, \\ Y_t = C_t + G_t + I_t + L_t, \\ P_t = Y_t + Z_t. \end{cases}$$

В этой системе три эндогенные переменные:
Y_t — имеющийся в распоряжении доход в период времени t;
C_t — частное потребление в период времени t;
P_t — валовой национальный продукт (ВНП) в период времени t;
Y_{t-1} — доход предыдущего года;
G_t — общественное потребление;
I_t — валовые капиталовложения;
L_t — изменение складских запасов;
Z_t — сальдо платежного баланса.

Если в модели Кейнса доход рассматривается как лаговая переменная, то в других исследованиях функции потребления в виде лаговой переменной используется потребление предыдущего года, т.е. считается, что потребление текущего года зависит не только от дохода, но и от достигнутого в предыдущий период уровня потребления.

Примером динамической модели экономики, учитывающей для каждой эндогенной переменной лаговые переменные соответствующего экономического содержания, может служить модель открытой экономики с экономической активностью со стороны государства:

$$\begin{cases} C_t = a_0 + a_1 Y_t + a_2 C_{t-1} + \Delta_1, \\ I_t = b_0 + b_1 Y_t + b_2 U_{t-1} + \Delta_2, \\ IM_t = k_0 + k_1 Y_t + k_2 IM_{t-1} + \Delta_3, \\ Y_t = C_t + I_t + G_t - IM_t. \end{cases}$$

В этой модели четыре эндогенные переменные:

C_t — личное потребление в период времени t;

I_t — частные чистые инвестиции в отрасли экономики в период времени t;

IM_t — импорт в период времени t;

Y_t — национальный доход за период времени t.

Все переменные приведены в постоянных ценах.

Предопределенными переменными в модели являются следующие три переменные:

C_{t-1} — личное потребление за предыдущий период;

U_{t-1} — доход личных домохозяйств от предпринимательской деятельности за предыдущий период и доход от имущества плюс нераспределенная прибыль предприятий до налогообложения;

IM_{t-1} — импорт за предыдущий период времени.

Экзогенная переменная:

G_t — общественное потребление плюс государственные капиталовложения в экономику страны, плюс изменение запасов, плюс косвенные налоги, плюс дотации, плюс экспорт.

Первые три уравнения системы являются сверхидентифицируемыми, а четвертое представляет собой балансовое тождество.

Широкий класс моделей в эконометрике представляют производственные функции $P = f(x_1, x_2, \ldots, x_n)$, где P — объем выпуска (уровень производства); $x_1, x_2, \ldots, x_n$ — факторы производства (труд, капитал и др.). Однако реализация такого рода моделей, как правило, не связана с системой одновременных уравнений. Производственная функция в упрощенном виде может быть включена в систему одновременных уравнений.

Не все эконометрические модели имеют вид системы одновременных уравнений. Так, широкий класс функций спроса на ряд потребительских товаров часто представляет собой рекурсивную систему, в которой с уравнениями можно работать последовательно и проблем одновременного оценивания не возникает.

Контрольные вопросы

1. Общие понятия о системах одновременных уравнений, необходимость их использования.
2. Составляющие системы одновременных уравнений.
3. Формы представления системы одновременных уравнений.
4. Состоятельность и несмещенность оценок системы одновременных уравнений.
5. Идентификация системы одновременных уравнений.
6. Методы оценки коэффициентов регрессии в структурной модели.
7. Косвенный метод наименьших квадратов.
8. Двухшаговый метод наименьших квадратов.
9. Трехшаговый метод наименьших квадратов.
10. Применение системы эконометрических уравнений.

☑ ТЕСТОВЫЕ ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ ПРОВЕРКИ ЗНАНИЙ

1. Принцип построения системы независимых уравнений состоит в том, что:
   а) каждая зависимая переменная рассматривается как функция одного и того же набора факторов;
   б) одни и те же зависимые переменные в одних уравнениях входят в левую часть, в других — в правую часть системы;
   в) модель содержит как в правой, так и в левой части эндогенные и экзогенные переменные.

2. Принцип построения системы взаимозависимых уравнений состоит в том, что:
   а) каждая зависимая переменная рассматривается как функция одного и того же набора факторов;
   б) одни и те же зависимые переменные в одних уравнениях входят в левую часть, в других – в правую часть системы;
   в) модель содержит как в правой, так и в левой части эндогенные и экзогенные переменные.

3. Система одновременных уравнений – это:
   а) система взаимозависимых уравнений;
   б) система независимых уравнений;
   в) приведенная форма модели;
   г) система взаимозависимых уравнений или структурная форма модели.

4. Идентификация модели – это:
   а) единственность соответствия между приведенной и структурной формами модели;
   б) преобладание эндогенных переменных над экзогенными;
   в) преобладание экзогенных переменных над эндогенными.

5. Модель идентифицируема, если:
а) число коэффициентов структурной модели равно числу коэффициентов приведенной формы модели;
б) число приведенных коэффициентов меньше числа структурных коэффициентов;
в) число приведенных коэффициентов больше числа структурных коэффициентов.

6. Модель неидентифицируема, если:
а) число коэффициентов структурной модели равно числу коэффициентов приведенной формы модели;
б) число приведенных коэффициентов меньше числа структурных коэффициентов;
в) число приведенных коэффициентов больше числа структурных коэффициентов.

7. Модель сверхидентифицируема, если:
а) число коэффициентов структурной модели равно числу коэффициентов приведенной формы модели;
б) число приведенных коэффициентов меньше числа структурных коэффициентов;
в) число приведенных коэффициентов больше числа структурных коэффициентов.

8. Модель считается идентифицируемой, если:
а) каждое уравнение системы идентифицируемо;
б) хотя бы два уравнения модели идентифицируемы;
в) большинство уравнений модели идентифицируемо.

9. Необходимое условие идентификации выполняется, если для уравнения модели соблюдается счетное правило:
а) $D + 1 = H$;
б) $D + 1 < H$;
в) $D + 1 > H$.

10. Структурные коэффициенты модели можно оценить тогда, когда модель:

а) идентифицируема;

б) сверхидентифицируема;

в) идентифицируема или сверхидентифицируема.

11. Методы оценивания коэффициентов структурной модели:

а) косвенный МНК;

б) двухшаговый и трехшаговый МНК;

в) метод максимального правдоподобия;

г) косвенный МНК, двухшаговый и трехшаговый МНК, метод максимального правдоподобия.

12. Под системой или моделью одновременных уравнений понимается:

а) случай, когда зависимая переменная в одном или нескольких уравнениях является объясняющей переменной в других уравнениях системы;

б) система из нескольких независимых уравнений, описывающих изучаемое явление;

в) система уравнений с одной и той же зависимой переменной, но с разным набором объясняющих переменных.

13. Эндогенные переменные — это:

а) зависимые переменные в системе одновременных уравнений, определяемые данной системой, даже если они появляются в качестве объясняющих переменных в других уравнениях системы;

б) переменные, определяемые внешними факторами;

в) переменные в каждом уравнении, не коррелированные с соответствующей ошибкой.

14. Предопределенные переменные включают в себя:

а) экзогенные переменные, определенные внешними для данной модели факторами;

б) экзогенные переменные и лаговые эндогенные переменные;
в) эндогенные переменные.

15. Уравнения приведенной формы получаются:
а) путем решения структурных уравнений, когда каждая эндогенная переменная в системе выражается как функция только экзогенных или предопределенных переменных системы;
б) при решении структурных уравнений обычным МНК;
в) при уменьшении количества независимых переменных.

ПРАКТИКУМ

Задача 10.1. На основании имеющихся данных построить следующую структурную форму модели:

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + b_{11}y_2, \\ y_2 = a_{20} + a_{21}x_2 + a_{22}x_3 + b_{21}y_1, \\ y_3 = a_{30} + a_{31}x_3 + b_{31}y_1, \end{cases}$$

где y_1 — заработная плата работников, руб.;
y_2 — темп прироста цен, %;
y_3 — доходы населения, млрд руб.;
x_1 — процент безработных;
x_2 — прирост цен на импорт, %;
x_3 — прирост экономически активного населения, %.

Таблица 10.1

Исходные данные к задаче 10.1

Год	y_1	y_2	y_3	x_1	x_2	x_3
2005	2500	21	10	1	2	1
2006	3100	18	12	2	3	2
2007	4200	16	11	3	2	3
2008	6775	16	15	2	4	4
2009	7690	14	14	3	3	5

Окончание табл. 10.1

Год	y_1	y_2	y_3	x_1	x_2	x_3
2010	8241	15	16	4	3	6
2011	10276	12	18	5	2	7
2012	11336	10	21	4	5	8

На основе табличных данных получить уравнение. Для каждого из полученных уравнений структурной формы модели необходимо определить показатели тесноты связи R, R^2 и проверить статистическую значимость моделей. Решение осуществите с помощью средств анализа данных *Excel*.

Решение

На первоначальном этапе необходимо произвести идентификацию уравнений системы, где через H обозначим количество эндогенных переменных в уравнении, а через D — экзогенных.

1) $D(1); H(2) \Rightarrow D(1) + 1 = H(2)$ — уравнение идентифицируемо;
2) $D(1); H(2) \Rightarrow D(1) + 1 = H(2)$ — уравнение идентифицируемо;
3) $D(2); H(2) \Rightarrow D(2) + 1 > H(2)$ — уравнение сверхидентифицируемо.

Таким образом, делаем вывод о том, что представленная система структурных уравнений является сверхидентифицируемой. Для определения ее параметров используют двухшаговый метод наименьших квадратов (ДМНК). На первоначальном этапе его применения необходимо:

а) построить систему приведенных уравнений;

б) определить параметры системы;

в) рассчитать теоретические значения эндогенных переменных (y):

$$\begin{cases} \hat{y}_1 = \lambda_{10} + \lambda_{11}x_1 + \lambda_{12}x_2 + \lambda_{13}x_3, \\ \hat{y}_2 = \lambda_{20} + \lambda_{21}x_2 + \lambda_{22}x_3 + \lambda_{23}x_3, \\ \hat{y}_3 = \lambda_{30} + \lambda_{31}x_1 + \lambda_{32}x_3 + \lambda_{33}x_3. \end{cases}$$

Для 1-го уравнения приведенной формы (рис. 10.2) составим систему уравнений для определения неизвестных параметров λ_{10}, λ_{11}, λ_{12}, λ_{13}:

$$\begin{cases} \sum y_1 = n\lambda_{10} + \lambda_{11}\sum x_1 + \lambda_{12}\sum x_2 + \lambda_{13}\sum x_3, \\ \sum y_1 x_1 = \lambda_{10}\sum x_1 + \lambda_{11}\sum x_1^2 + \lambda_{12}\sum x_1 x_2 + \lambda_{13}\sum x_1 x_3, \\ \sum y_1 x_2 = \lambda_{10}\sum x_2 + \lambda_{11}\sum x_1 x_2 + \lambda_{12}\sum x_2^2 + \lambda_{13}\sum x_2 x_3, \\ \sum y_1 x_3 = \lambda_{10}\sum x_3 + \lambda_{11}\sum x_1 x_3 + \lambda_{12}\sum x_2 x_3 + \lambda_{13}\sum x_3^2. \end{cases}$$

Определение параметров данной системы целесообразно осуществить в электронной таблице *MicrosoftExcel*, с помощью которой может быть решено большое количество эконометрических и статистических задач.

Для проведения регрессионного анализа и оценки качества синтезированной модели используется процедура *Регрессия* из пакета анализа. В *MSExcel* экспериментальные данные могут быть аппроксимированы линейным уравнением только до 16-го порядка.

Решение системы осуществляют в несколько шагов.

Шаг 1. На листе необходимо сформировать исходную матрицу, в состав которой войдут табличные значения заработной платы работников за указанный период и соответствующие значения факторов x_1, x_2, x_3.

Шаг 2. В соответствии с заданием определим параметры уравнения, получим R, для чего в пакете анализа *Excel* используется процедура *Регрессия*, которая позволяет получить все необходимые для анализа данные. Для выполнения данной процедуры необходимо:

- выполнить команду *Сервис → Анализ данных*;
- в появившемся списке *Инструменты анализа* выбрать строку *Регрессия* и нажать кнопку *ОК* (рис. 10.3);
- в появившемся диалоговом окне (рис. 10.4) в рабочем поле *Входной интервал Y* указать входной диапазон, т. е. ввести ссылку на ячейки, содержащие анализируемые данные.

	A	B	C	D	E	F	G
1	Год	y_1	y_2	y_3	x_1	x_2	x_3
2	2005	2500	21	10	1	2	1
3	2006	3100	18	12	2	3	2
4	2007	4200	16	11	3	2	3
5	2008	6775	16	15	2	4	4
6	2009	7690	14	14	3	3	5
7	2010	8241	15	16	4	3	6
8	2011	10276	12	18	5	2	7
9	2012	11336	10	21	4	5	8
10							
11							
12	Год	y_1	x_1	x_2	x_3		
13	2005	2500	1	2	1		
14	2006	3100	2	3	2		
15	2007	4200	3	2	3		
16	2008	6775	2	4	4		
17	2009	7690	3	3	5		
18	2010	8241	4	3	6		
19	2011	10276	5	2	7		
20	2012	11336	4	5	8		
21							

Рис. 10.2. Исходная матрица по первому уравнению приведенной формы

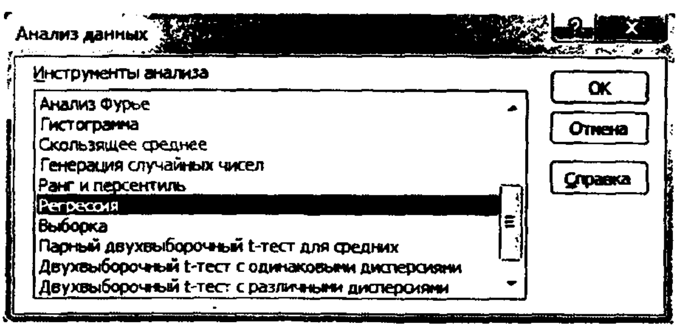


Рис. 10.3. Окно выбора метода обработки данных

Для этого следует навести указатель мыши на левую верхнюю ячейку данных, нажать левую кнопку мыши и, не отпуская ее, протянуть указатель мыши к правой нижней;

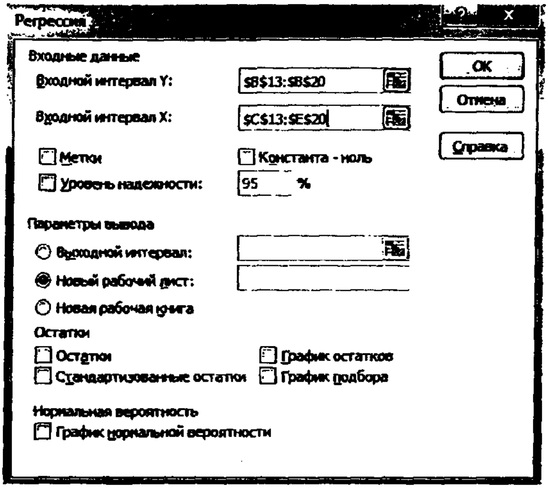


Рис. 10.4. Пример заполнения диалогового окна *Регрессия*

– указать *Входной интервал X*, то есть ввести ссылку на диапазон независимых данных, содержащий до 16 столбцов анализируемых данных;
– установить в *Параметрах вывода* переключатель в положение *Новый рабочий лист* и назвать его «Приведенная форма 1-го уравнения»;
– нажать кнопку *OK*.

Шаг 3. Получение результатов анализа (рис. 10.5): показателей дисперсионного анализа, значений коэффициентов корреляции и детерминации, стандартной погрешности вычисления Y, среднеквадратичного отклонения, коэффициентов регрессии и т. д.

Шаг 4. Интерпретация результатов. Мерой эффективности регрессионной модели являются множественный коэффициент детерминации (R^2) и корреляции (R). В соответствии с полученными результатами можно отметить, что 1-е уравнение приведенной формы характеризуется весьма высокой статистической

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,996910328
R-квадрат	0,993830203
Нормированный R-квадрат	0,989202854
Стандартная ошибка	338,8899012
Наблюдения	8

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	3	73997752,04	24665917,35	214,7731651	7,12275E-05
Остаток	4	459385,4607	114846,3652		
Итого	7	74457137,5			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	2855,505618	882,8401853	3,234453603	0,031840813	404,3483068	5306,662929	404,3483068	5306,662929
Переменная X 1	-1136,893258	405,6193982	-2,802857219	0,048669631	-2263,073251	-10,71326582	-2263,073251	-10,71326582
Переменная X 2	-515,9775281	259,039084	-1,991890645	0,11719715	-1235,185325	203,2302689	-1235,185325	203,2302689
Переменная X 3	1970,634831	247,5770756	7,959682157	0,001349577	1283,250672	2658,018991	1283,250672	2658,018991

Лист4 / Лист1 / Лист2 / Лист3

Рис. 10.5. Результаты анализа по регрессионной модели значимостью. Так, на основании величины показателя множественной корреляции (0,997) можно сделать вывод в соответствии со шкалой Чеддока о весьма высокой взаимосвязи между уровнем заработной платы работников и такими факторами, как процент безработных, прирост цен на импорт и прирост экономически активного населения. На основании значения коэффициента множественной детерминации можно утверждать, что факторные показатели модели позволяют на 99,3% объяснить вариацию результативного. Считается, что если $R^2 > 0{,}95$, то это говорит о высокой точности аппроксимации (модель хорошо описывает явление), как в данном уравнении.

Исследуется также значимость регрессионной модели с помощью F-критерия (Фишера). Если величина F-критерия значима (т. е. фактическое значение F-критерия, показывающего соотношение факторной и остаточной дисперсии в расчете на одну степень свободы, больше табличного, критического значения), то регрессионная модель признается значимой.

В таблице *Дисперсионный анализ* по столбцу *df* приводятся значения числа степеней свободы. На пересечении столбца *SS* и строки *Регрессия* находится значение факторной суммы квадратов отклонений, *SS* и *Остаток* — остаточной суммы квадратов отклонений, *SS* и *Итого* — общей суммы квадратов отклонений. В результате отношения соответствующей суммы квадратов отклонений и числа степеней свободы по столбцу *MS* получены значения факторной и остаточной дисперсии в расчете на одну степень свободы. Фактическое значение F-критерия показывает, что факторная дисперсия в 214,8 раза превышает остаточную.

В столбце *Значимость F* приведена вероятность ошибки, при которой допустимо отвергнуть нулевую гипотезу и принять альтернативную. Существует общепринятая терминология, которая относится к доверительным интервалам вероятности. Высказывания, имеющие вероятность ошибки $p \le 0{,}05$, называются значимыми; высказывания с вероятностью ошибки $p \le 0{,}01$ — очень значимыми, а высказывания с вероятностью ошибки $p \le 0{,}001$ — максимально значимыми.

Таким образом, для нашего примера дисперсионный анализ дает значимый результат ($p \leq 0{,}05$), и полученная модель в целом может быть признана значимой.

Значения коэффициентов регрессии полученной модели находятся в столбце *Коэффициенты* и соответствуют:

- Y-пересечению — λ_{10};
- переменной $X1$ — λ_{11};
- переменной $X2$ — λ_{12};
- переменной $X3$ — λ_{13}.

Таким образом, полученное уравнение множественной регрессии имеет следующий вид для нашего примера:

$$y = 2855{,}5056 - 1136{,}8933x_1 - 515{,}9775x_2 + 1970{,}6348x_3.$$

Полученные значения коэффициентов регрессии можно охарактеризовать следующим образом: для фактора $X1$ — увеличение доли безработных на 1% приводит в среднем к снижению оплаты работников на 1136,89 руб.; для фактора $X2$ — прирост цен на импорт на 1% приводит к снижению заработной платы работников на 515,98 руб.; для фактора $X3$ — прирост экономически активного населения на 1% влечет рост заработной платы на 1970,63 руб.

По столбцу *t*-статистика приведены фактические значения *t*-критерия Стьюдента для каждого из коэффициентов регрессии модели. В столбце *Р-Значение* приводится достоверность отличия соответствующих коэффициентов от нуля. В случаях, когда $p \leq 0{,}05$, коэффициент может считаться нулевым, это означает, что влияние соответствующей независимой переменной на зависимую переменную Y недостоверно, и эта независимая переменная может быть исключена из уравнения.

Для нашего примера незначимым следует считать полученное значение коэффициента регрессии при факторной переменной $X3 = 7{,}95$. Это означает, что для улучшения качества модели в дальнейшем следовало бы исключить данную переменную из исходной модели.

В столбцах *Нижние 95%* и *Верхние 95%* итоговой таблицы приводятся значения верхних и нижних границ коэффициентов уравнения регрессии с 95-процентной вероятностью. Если в соответствии с ними рассчитать два ряда *Y* по нижним и верхним границам коэффициентов регрессии, то это позволит получить доверительные интервалы вариации выравненных по полученному уравнению регрессии значений результативного показателя модели.

Аналогично произвести определение неизвестных параметров для 2-го и 3-го уравнений приведенной формы.

Шаг 5. Аналогично произведем определение неизвестных параметров для 2-го и 3-го уравнений приведенной формы.

Полученные данные отобразим в таблице 10.2.

Таблица 10.2

Идентификация приведенной формы модели

Уравнение приведенной формы	Вид уравнения	R	R^2	F	Значимость (p)
1	$y = 2855{,}5 - 1136{,}9x_1 - 515{,}9x_2 + 1970{,}6x_3$	0,99	0,99	214,8	0,001
2	$y = 24{,}03 - 1{,}25x_1 - 0{,}84x_2 - 0{,}55x_3$	0,96	0,93	18,16	0,008
3	$y = 6{,}69 - 0{,}12x_1 + 0{,}76x_2 + 1{,}33x_3$	0,97	0,95	26,82	0,004

На основании данных, представленных в таблице, можно заключить, что наибольшую эффективность модели выражают показатели 3-го уравнения приведенной формы.

Найдем теоретические (выравненные) значения для использования их в качестве исходных в правой части уравнения. На следующем этапе необходимо определить параметры структурной формы модели МНК. При этом в качестве исходных данных в правой части уравнения будут использоваться фактические данные по факторным показателями (x) и теоретические значения эндогенных переменных, рассчитанные из приведенной формы уравнения.

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + b_{11}\hat{y}_2, \\ y_2 = a_{20} + a_{21}x_2 + a_{22}x_3 + b_{21}\hat{y}_1, \\ y_3 = a_{30} + a_{31}x_3 + b_{31}\hat{y}_1. \end{cases}$$

Для каждого уравнения структурной формы составим соответственно системы уравнений. Определение параметров систем целесообразно осуществить в электронной таблице *Microsoft Excel* описанными ранее пошаговыми действиями.

Полученные результаты корреляционно-регрессионного анализа обнародуем в итоговой таблице.

Таблица 10.3

Идентификация структурной формы модели

Уравнение структурной формы	Вид уравнения	R	R^2	F	Значимость (p)
1	$y=19362{,}78-1102.02x_1-582.12x_2-798.38y_2$	0,95	0,90	12,79	0,016
2	$y=21{,}94-1{,}95x_2-0{,}02x_3-0{,}012y_1$	0,96	0,93	18,16	0,008
3	$y=6{,}68-0{,}82x_3+0{,}005y_1$	0,97	0,95	48,73	0,0005

Задача 10.2. На основании имеющихся данных построить следующую структурную форму модели:

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2, \\ y_2 = a_{20} + a_{21}x_3 + b_{21}y_1, \end{cases}$$

где y_1 — урожайность зерновых культур, ц/га;

y_2 — уровень рентабельности зерновых культур, %;

x_1 — приходится зерноуборочных комбайнов на 1000 га посевов зерновых, шт.;

x_2 — внесено минеральных удобрений по зерновым в расчете на 1 га удобренной площади, ц;

x_3 — коэффициент обновления основных фондов сельскохозяйственных предприятий.

Таблица 10.4

Показатели деятельности отрасли растениеводства в разрезе муниципальных районов Ставропольского края

№ района	y_1	y_2	x_1	x_2	x_3
1	27,8	27,2	2,5	0,8	5,3
2	19,3	−7,8	1	1,3	8,2
3	29,8	65,7	3,7	0,5	7,1
4	25,2	57,8	2,3	0,3	9,5
5	32,9	66,6	2,7	0,3	11,2
6	32,1	35,1	2,9	0,3	8,4
7	36,7	25,3	2,4	0,9	11,3
8	25,9	35,2	0,7	0,4	9
9	35,5	30,1	3,6	0,9	2,1
10	30,5	45,3	1,9	0,7	10
11	32,0	36,1	2,6	1	4,4
12	38,0	48,4	3,5	0,9	12,4
13	39,0	48,3	2,5	1,2	23,9
14	28,4	47,2	2,9	0,6	10,1
15	26,3	35,5	3,2	0,3	12,6
16	26,4	21,9	2,3	1,5	8
17	28,5	32	2,6	0,6	14,9
18	42,5	43,3	3,5	1,3	10,8
19	41,3	75	3,4	0,7	15,5
20	31,7	34,4	3,5	0,7	8,9
21	32,1	33,5	2,1	1,1	15,9
22	45,2	60,1	3,1	0,8	7,5
23	32,6	49,5	2,9	0,5	9,9
24	37,7	44,2	3,2	1,1	11,9
25	31,1	42,8	2,5	1,4	4,6
26	27,2	29,6	2,2	1	9,2

Решение

На первоначальном этапе осуществим проверку необходимого условия идентифицируемости уравнений системы, где через H обозначим количество эндогенных переменных в уравнении, а через D — количество экзогенных переменных, которые содержатся в системе, но не входят в состав данного уравнения:

1) $D=1$, $H=1$; $\Rightarrow D+1>H$ — уравнение сверхидентифицируемо;

2) $D=1$, $H=2$; $\Rightarrow D+1=H$ — уравнение идентифицируемо.

Таким образом, делаем вывод о том, что представленная система структурных уравнений является сверхидентифицируемой. Для определения ее параметров используют двухшаговый метод наименьших квадратов (ДМНК). На первоначальном этапе его применения необходимо:

а) построить систему приведенных уравнений;

б) определить параметры системы;

в) рассчитать теоретические значения эндогенных переменных (y).

Система приведенных уравнений для данного примера содержит два уравнения:

$$\begin{cases} \hat{y}_1 = \lambda_{10} + \lambda_{11}x_1 + \lambda_{12}x_2 + \lambda_{13}x_3, \\ \hat{y}_2 = \lambda_{20} + \lambda_{21}x_1 + \lambda_{22}x_2 + \lambda_{23}x_3. \end{cases}$$

Поиск решения осуществим с помощью программного продукта *Statistica* 6.0. Для этого вначале составим таблицу исходных данных для первого уравнения из приведенной формы системы (рис. 10.6).

Для оценки параметров и идентификации первого уравнения выполним следующие действия:

- в меню «Анализ» выберем инструмент «Множественная регрессия». В появившемся диалоговом окне следует указать в качестве зависимой переменной переменную y_1, а в качестве независимых — переменные x_1, x_2, x_3; нажмем «ОК»;
- в окне «Результаты множественной регрессии» выберем пункт «Итоговая таблица регрессии».

Проанализируем полученные результаты.

В таблице «Итоги регрессии» в столбце «В» находятся значения параметров (рис. 10.7):

$\lambda_{10} = 11{,}5$; $\lambda_{11} = 5{,}08$; $\lambda_{12} = 3{,}84$; $\lambda_{13} = 0{,}39$.

	1 y1	2 x1	3 x2	[illegible]
1	27,8	2,5	0,8	5,3
2	19,3	1,0	1,3	8,2
3	29,8	3,7	0,5	7,1
4	25,2	2,3	0,3	9,5
5	32,9	2,7	0,3	11,2
6	32,1	2,9	0,3	8,4
7	36,7	2,4	0,9	11,3
8	25,9	0,7	0,4	9,0
9	35,5	3,6	0,9	2,1
10	30,5	1,9	0,7	10,0
11	32,0	2,6	1,0	4,4
12	38,0	3,5	0,9	12,4
13	39,0	2,5	1,2	23,9
14	28,4	2,9	0,6	10,1
15	26,3	3,2	0,3	12,6
16	26,4	2,3	1,5	8,0
17	28,5	2,6	0,6	14,9
18	42,5	3,5	1,3	10,8
19	41,3	3,4	0,7	15,5
20	31,7	3,5	0,7	8,9
21	32,1	2,1	1,1	15,9
22	45,2	3,1	0,8	7,5
23	32,6	2,9	0,5	9,9
24	37,7	3,2	1,1	11,9
25	31,1	2,5	1,4	4,6
26	27,2	2,2	1,0	9,2

Рис. 10.6. Исходные данные для первого уравнения приведенной формы в *Statistica 6.0*

	Итоги регрессии для зависимой переменной: y1 (Таблица данных в Workbook1) R= ,69421282 R2= ,48193144 Скорректир. R2= ,41128573 F(3,22)=6,8218 p<,00202 Станд. ошибка оценки: 4,6188					
N=26	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(22)	p-уров.
Св.член			11,50996	4,796005	2,399905	0,025301
x1	0,624274	0,154428	5,08015	1,256689	4,042488	0,000544
x2	0,229601	0,154395	3,83543	2,579139	1,487096	0,151182
x3	0,279535	0,153488	0,38624	0,212079	1,821217	0,082202

Рис. 10.7. Окно «Итоги регрессии» в *Statistica 6.0*

Таким образом, первое уравнение исследуемой системы имеет вид

$$y_1 = 11{,}5 + 5{,}08x_1 + 3{,}84x_2 + 0{,}39x_3$$

Кроме того, данная таблица содержит значения t-критериев Стьюдента для каждого из коэффициентов регрессии и соответствующие им уровни значимости p, характеризующие вероятность ошибочного заключения. Считается, что при вероятности ошибки $p \leq 0{,}05$ высказывания называются значимыми, при $p \leq 0{,}01$ — очень значимыми, а при $p \leq 0{,}001$ — максимально значимыми.

В соответствии с полученными результатами $t_{\lambda_{10}} = 2{,}4$; $p = 0{,}025$, следовательно, значение параметра $\lambda_{10} = 11{,}5$ следует признать статистически значимым при вероятности ошибки в 2,5%. Таким образом, среднее значение урожайности зерновых по Ставропольскому краю (y_i) в отсутствие воздействия таких факторов, как число зерноуборочных комбайнов на 1000 га посевов зерновых (x_1), количество внесенных минеральных удобрений по зерновым в расчете на 1 га удобренной площади (x_2) и коэффициент обновления основных фондов сельскохозяйственных предприятий (x_3), составит 11,5 ц/га.

$t_{\lambda_{11}} = 4{,}04$; $p = 0{,}0005$, следовательно, значение параметра $\lambda_{11} = 5{,}08$ можно считать статистически адекватным с вероятностью ошибки 0,05%, что позволяет говорить о максимальной значимости сделанного вывода. Интерпретируя полученный вывод к данной модели, можно сказать, что увеличение количества зерноуборочных комбайнов на 1000 га посевов приводит к росту урожайности на 5,08 ц/га.

$t_{\lambda_{12}} = 1{,}49$; $p = 0{,}15$. Таким образом, значение параметра $\lambda_{12} = 3{,}84$ следует признать статистически незначимым, поскольку уровень вероятности ошибки слишком высок и составляет 15%.

$t_{\lambda_{13}} = 1{,}82$; $p = 0{,}08$. Следовательно, коэффициент регрессии $\lambda_{13} = 0{,}39$ при переменной x_3 статистически неадекватен.

Полученные результаты позволяют сделать вывод, что коэффициенты регрессии при факторных переменных x_2, x_3 следует признать незначимыми. Это означает, что для улучшения качества модели в дальнейшем следовало бы исключить данные переменные из исходной модели.

Проанализируем показатели статистической значимости построенного регрессионного уравнения.

В соответствии с таблицей «Итоги регрессии» (рис. 10.7) значение индекса множественной корреляции $R = 0{,}69$. Таким образом, в соответствии со шкалой Чеддока, можно сделать вывод о заметной взаимосвязи между значением урожайности зерновых и совокупностью факторов, включенных в модель: числом зерноуборочных комбайнов на 1000 га посевов зерновых (x_1), количеством внесенных минеральных удобрений по зерновым в расчете на 1 га удобренной площади (x_2) и коэффициентом обновления основных фондов сельскохозяйственных предприятий (x_3). Значение коэффициента множественной детерминации $R^2 = 0{,}48$ позволяет утверждать, что факторные показатели модели позволяют на 48% объяснить вариацию результативного признака. В данном случае можно говорить об удовлетворительной точности аппроксимации и необходимости включения в модель дополнительных неучтенных факторов, воздействие которых достаточно значимо и оценивается в 52%.

Исследуется также значимость регрессионной модели с помощью F-критерия Фишера. Если величина F-критерия значима (т. е. фактическое значение F-критерия, показывающего соотношение факторной и остаточной дисперсии в расчете на одну степень свободы, больше табличного, критического значения), то регрессионная модель признается значимой.

В таблице «Итоги регрессии» имеются значение F-критерия Фишера $F = 6{,}82$ и его табличное значение (3,22). Таким образом, для нашего примера фактическое значение F-критерия больше критического (табличного): $F_{\text{факт}} > F_{\text{кр.}}$, следовательно, полученная модель в целом может быть признана значимой.

Для расчета теоретических значений результативного признака y_i воспользуемся пунктом меню «Множественная регрессия» «Анализ остатков» → «Остатки и предсказанные». Результаты отражены на рисунке 10.8.

Аналогично произведем оценку параметров для второго уравнения приведенной формы.

Для этого составим таблицу исходных данных для второго уравнения из приведенной формы системы (рис. 10.9).

Результаты применения инструмента «Множественная регрессия» для данного примера отражены на рисунке 10.10.

Набл. No.	Предсказанные значения и остатки (Таблица данных в Workbook1) Зависимая перемен.: y1		
	Наблюд. Значение	Предск. Значение	Остатки
1	27,80000	29,32576	-1,52576
2	19,30000	24,74335	-5,44335
3	29,80000	34,96655	-5,16655
4	25,20000	28,01423	-2,81423
5	32,90000	30,70290	2,19710
6	32,10000	30,63745	1,46255
7	36,70000	31,51874	5,18126
8	25,90000	20,07641	5,82359
9	35,50000	34,06149	1,43851
10	30,50000	27,70946	2,79054
11	32,00000	30,25324	1,74676
12	38,00000	37,53177	0,46823
13	39,00000	38,04403	0,95597
14	28,40000	32,44469	-4,04469
15	26,30000	33,78371	-7,48372
16	26,40000	32,03738	-5,63738
17	28,50000	32,77461	-4,27461
18	42,50000	38,44795	4,05205
19	41,30000	37,45402	3,84598
20	31,70000	35,41283	-3,71283
21	32,10000	32,53849	-0,43849
22	45,20000	33,22358	11,97642
23	32,60000	31,98390	0,61610
24	37,70000	36,58169	1,11831
25	31,10000	31,35664	-0,25664
26	27,20000	30,07514	-2,87514
Минимум	19,30000	20,07641	-7,48372
Максим.	45,20000	38,44795	11,97642
Среднее	32,14231	32,14231	-0,00000
Медиана	31,85000	32,24104	0,54216

Рис. 10.8. Окно «Предсказанные значения и остатки» в *Statistica 6.0*

Тогда второе уравнение приведенной формы имеет вид $y_2 = 16{,}9 + 10677x_1 - 17{,}54x_2 + 0{,}93x_3$.

Анализ статистической значимости параметров уравнения на основе t-критериев Стьюдента позволяет сделать вывод о статистической значимости параметра λ_{21} и λ_{22} и незначимости λ_{20} и λ_{23}.

	1 y2	2 x1	3 x2	4 x3
1	27,2	2,5	0,8	5,3
2	-7,8	1	1,3	8,2
3	65,7	3,7	0,5	7,1
4	57,8	2,3	0,3	9,5
5	66,6	2,7	0,3	11,2
6	35,1	2,9	0,3	8,4
7	25,3	2,4	0,9	11,3
8	35,2	0,7	0,4	9
9	30,1	3,6	0,9	2,1
10	45,3	1,9	0,7	10
11	36,1	2,6	1	4,4
12	48,4	3,5	0,9	12,4
13	48,3	2,5	1,2	23,9
14	47,2	2,9	0,6	10,1
15	35,5	3,2	0,3	12,6
16	21,9	2,3	1,5	8
17	32	2,6	0,6	14,9
18	43,3	3,5	1,3	10,8
19	75	3,4	0,7	15,5
20	34,4	3,5	0,7	8,9
21	33,5	2,1	1,1	15,9
22	60,1	3,1	0,8	7,5
23	49,5	2,9	0,5	9,9
24	44,2	3,2	1,1	11,9
25	42,8	2,5	1,4	4,6
26	29,6	2,2	1	9,2

Рис. 10.9. Таблица исходных данных для второго уравнения приведенной формы в *Statistica 6.0*

	Итоги регрессии для зависимой переменной: y2 (Таблица данных в Workbook1) R= ,67675932 R2= ,45800318 Скорректир. R2= ,38409452 F(3,22)=6,1969 p<,00325 Станд. ошибка оценки: 13,188					
N=26	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(22)	p-уров.
Св.член			16,8798	13,69365	1,23268	0,230707
x1	0,473957	0,157954	10,7665	3,58812	3,00060	0,006586
x2	-0,376045	0,157921	-17,5354	7,36401	-2,38123	0,026337
x3	0,239908	0,156993	0,9253	0,60553	1,52815	0,140727

Рис. 10.10. Итоги моделирования для второго уравнения приведенной формы в *Statistica 6.0*

Значение индекса множественной корреляции $R = 0{,}68$ позволяет сделать вывод о заметной тесноте связи между уровнем рентабельности зерновых культур в крае (y_2) и совокупностью факторных признаков: числом зерноуборочных комбайнов на 1000 га посевов зерновых (x_1), количеством внесенных минеральных удобрений по зерновым в расчете на 1 га удобренной площади (x_2) и коэффициентом обновления основных фондов сельскохозяйственных предприятий (x_3).

Значение индекса множественной детерминации $R^2 = 0{,}46$ свидетельствует о том, что совокупность перечисленных факторов на 46% объясняет вариацию результативного признака.

Фактическое значение F-критерия Фишера $F = 6{,}2$ больше критического (3,22), следовательно, модель следует признать в целом статистически значимой.

В заключение рассчитаем теоретические уровни результативного признака (рис. 10.11).

Итак, согласно произведенным расчетам, приведенная форма системы уравнений имеет вид

$$\begin{cases} \hat{y}_1 = 11{,}5 + 5{,}08\,x_1 + 3{,}84\,x_2 + 0{,}39\,x_3 \\ \hat{y}_2 = 16{,}9 + 10{,}77\,x_1 - 17{,}54\,x_2 + 0{,}93\,x_3 \end{cases}$$

На следующем этапе необходимо определить параметры структурной формы модели, но при этом в качестве исходных данных в правой части уравнения будут использоваться фактические значения экзогенных переменных (x_1, x_2, x_3) и теоретические значения эндогенных переменных, рассчитанных по уравнениям регрессий приведенной формы:

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2, \\ y_2 = a_{20} + a_{21}x_3 + b_{21}\hat{y}_1. \end{cases}$$

Составим таблицы исходных данных для оценки параметров уравнений системы (рис. 10.11).

Для оценки параметров и идентификации первого уравнения необходимо осуществить последовательность действий:

Набл. No.	Предсказанные значения и остатки (Таблица данных в Workbook1) Зависимая перемен.: y2 Наблюд. Значение	Предск. Значение	Остатки
1	27,20000	34,67212	-7,4721
2	-7,80000	12,43818	-20,2382
3	65,70000	54,51815	11,1818
4	57,80000	45,17294	12,6271
5	66,60000	51,05263	15,5474
6	35,10000	50,61497	-15,5150
7	25,30000	37,39400	-12,0940
8	35,20000	25,73033	9,4697
9	30,10000	41,80064	-11,7006
10	45,30000	34,31487	10,9851
11	36,10000	31,40889	4,6911
12	48,40000	50,25503	-1,8550
13	48,30000	44,86936	3,4306
14	47,20000	46,92744	0,2726
15	35,50000	57,73136	-22,2314
16	21,90000	22,74250	-0,8425
17	32,00000	48,13914	-16,1391
18	43,30000	41,76033	1,5397
19	75,00000	55,55401	19,4460
20	34,40000	50,52340	-16,1234
21	33,50000	34,91355	-1,4136
22	60,10000	43,16778	16,9322
23	49,50000	48,49591	1,0041
24	44,20000	43,05533	1,1447
25	42,80000	23,50317	19,2968
26	29,60000	31,54394	-1,9439
Минимум	-7,80000	12,43818	-22,2314
Максим.	75,00000	57,73136	19,4460
Среднее	40,85769	40,85769	-0,0000
Медиана	39,45000	43,11156	0,6383

Рис. 10.11. Расчет теоретических уровней результативного признака y_2 в *Statistica 6.0*

- в меню «Анализ» выберем инструмент «Множественная регрессия». В появившемся диалоговом окне следует указать в качестве зависимой переменной переменную y_1, а в качестве независимых — переменные x_1, x_2; нажмем «ОК»;
- в окне «Результаты множественной регрессии» выберем пункт «Итоговая таблица регрессии».

	1-е уравнение структурной формы				2-е уравнение структурной формы		
	1 y1	2 x1	3 x2		1 y2	2 x3	3 y1(x)
1	27,8	2,5	0,8	1	27,2	5,3	29,3
2	19,3	1,0	1,3	2	-7,8	8,2	24,7
3	29,8	3,7	0,5	3	65,7	7,1	35,0
4	25,2	2,3	0,3	4	57,8	9,5	28,0
5	32,9	2,7	0,3	5	66,6	11,2	30,7
6	32,1	2,9	0,3	6	35,1	8,4	30,6
7	36,7	2,4	0,9	7	25,3	11,3	31,5
8	25,9	0,7	0,4	8	35,2	9,0	20,1
9	35,5	3,6	0,9	9	30,1	2,1	34,1
10	30,5	1,9	0,7	10	45,3	10,0	27,7
11	32,0	2,6	1,0	11	36,1	4,4	30,3
12	38,0	3,5	0,9	12	48,4	12,4	37,5
13	39,0	2,5	1,2	13	48,3	23,9	38,0
14	28,4	2,9	0,6	14	47,2	10,1	32,4
15	26,3	3,2	0,3	15	35,5	12,6	33,8
16	26,4	2,3	1,5	16	21,9	8,0	32,0
17	28,5	2,6	0,6	17	32,0	14,9	32,8
18	42,5	3,5	1,3	18	43,3	10,8	38,4
19	41,3	3,4	0,7	19	75,0	15,5	37,5
20	31,7	3,5	0,7	20	34,4	8,9	35,4
21	32,1	2,1	1,1	21	33,5	15,9	32,5
22	45,2	3,1	0,8	22	60,1	7,5	33,2
23	32,6	2,9	0,5	23	49,5	9,9	32,0
24	37,7	3,2	1,1	24	44,2	11,9	36,6
25	31,1	2,5	1,4	25	42,8	4,6	31,4
26	27,2	2,2	1,0	26	29,6	9,2	30,1

Рис. 10.12. Набор переменных для структурной формы системы уравнений в *Statistica 6.0*

Результаты оценки параметров и статистической значимости уравнения отражены на рисунке 10.13.

	Итоги регрессии для зависимой переменной: y1 (сф_y1 в Workbook1) R= ,63547215 R2= ,40382485 Скорректир. R2= ,35198353 F(2,23)=7,7896 p<,00261 Станд. ошибка оценки: 4,8459					
N=26	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(23)	p-уров.
Св.член			15,54229	4,463482	3,482100	0,002014
x1	0,618476	0,161985	5,03297	1,318185	3,818103	0,000883
x2	0,229247	0,161985	3,82951	2,705921	1,415235	0,170398

Рис. 10.13. Результаты моделирования для первого уравнения структурной формы системы уравнений в *Statistica 6.0*

В столбце «В» указаны искомые коэффициенты регрессии. Таким образом, первое уравнение структурной формы имеет вид $y_1 = 15{,}54 + 5{,}03x_1 + 3{,}83x_2$.

Проанализируем показатели статистической значимости полученной модели в целом.

Значение индекса множественной корреляции $R = 0{,}64$ свидетельствует о заметной тесноте связи между урожайностью зерновых и совокупностью таких факторов, как число зерноуборочных комбайнов на 1000 га посевов зерновых и количество внесенных минеральных удобрений по зерновым в расчете на 1 га удобренной площади.

Значение индекса множественной детерминации $R^2 = 0{,}4$ позволяет утверждать, что факторные показатели модели позволяют на 40% объяснить вариацию результативного признака. Таким образом, можно сделать вывод об удовлетворительной точности аппроксимации.

На основе полученного значения F-критерия Фишера можно сделать заключение о статистической адекватности уравнения регрессии в целом. В нашем случае $F_{\text{факт}} = 7{,}79$; $F_{\text{кр.}} = 2{,}23$. Сравнивая эти значения, можно говорить, что уравнение статистически значимо: $F_{\text{факт}} > F_{\text{кр.}}$.

Полученное значение t-критерия Стьюдента для параметра a_{10} $t_{a_{10}} = 3{,}48$; $p = 0{,}002$; следовательно, доверять полученному коэффициенту можно с вероятностью ошибки 0,2%. Таким образом, среднее значение урожайности зерновых по Ставропольскому краю (y_1) в отсутствие воздействия таких факторов, как число зерноуборочных комбайнов на 1000 га посевов зерновых (x_1), количество внесенных минеральных удобрений по зерновым в расчете на 1 га удобренной площади (x_2) составит 11,5 ц/га.

$t_{a_{11}} = 3{,}82$; $p = 0{,}0009$. Таким образом, значение параметра $a_{11} = 5{,}03$ следует признать статистически значимым. Следовательно, увеличение количества зерноуборочных комбайнов на 1000 га посевов приводит к росту урожайности на 5,03 ц/га.

$t_{a_{12}} = 1{,}42$; $p = 0{,}17$. Полученные результаты позволяют сделать вывод о незначимости параметра a_{12}, а вероятность ошибочного вывода об изменении урожайности зерновых под воздействием количества внесенных минеральных удобрений составляет 17%, что является довольно значительной погрешностью. Кроме того, в качестве резерва повышения качества уравнения регрессии в целом можно использовать исключение переменной x_2 из модели.

Аналогично осуществим параметризацию и идентификацию второго уравнения структурной формы модели:

- в меню «Анализ» выберем инструмент «Множественная регрессия». В появившемся диалоговом окне следует указать в качестве зависимой переменной переменную y_2, а в качестве независимых — переменные x_3, $y_1(x)$; нажмем «ОК»;
- в окне «Результаты множественной регрессии выберем пункт «Итоговая таблица регрессии».

Результаты оценки параметров и статистической значимости уравнения отражены на рисунке 10.14.

	Итоги регрессии для зависимой переменной: y2 (сф_y2 в Workbook1) R= ,41735475 R2= ,17418499 Скорректир. R2= ,10237499 F(2,23)=2,4256 p<,11070 Станд. ошибка оценки: 15,921					
N=26	БЕТА	Стд.Ош. БЕТА	B	Стд.Ош. B	t(23)	p-уров.
Св.член			-11,2213	24,78682	-0,452714	0,654996
x3	0,084648	0,205264	0,3265	0,79172	0,412387	0,683875
y1(x)	0,377429	0,205264	1,5177	0,82538	1,838754	0,078899

Рис. 10.14. Результаты моделирования для второго уравнения структурной формы системы уравнений в *Statistica 6.0*

В столбце «В» указаны коэффициенты второго уравнения регрессии: $y_2 = -11,22 + 0,33x_3 + 1,52y_1$.

Анализ статистической значимости параметров уравнения на основе t-критериев Стьюдента позволяет сделать вывод о статистической незначимости коэффициентов регрессии a_{21} и b_{21}, а также свободного члена уравнения a_{20}.

Проанализируем показатели статистической значимости полученной модели в целом.

Значение индекса множественной корреляции $R = 0{,}42$ свидетельствует об умеренной тесноте связи между рентабельностью зерновых и совокупностью таких факторов, как коэффициент обновления основных фондов сельскохозяйственных предприятий и урожайность зерновых.

Значение индекса множественной детерминации $R^2 = 0{,}17$ позволяет утверждать, что рентабельность производства зерновых только на 17% обусловлена урожайностью данного направления растениеводства и обновлением основных фондов сельхозтоваропроизводителей.

Таким образом, можно сделать вывод о низкой точности аппроксимации.

На основе полученного значения F-критерия Фишера можно сделать заключение о статистической адекватности уравнения регрессии в целом. В нашем случае $F_{факт} = 2{,}43$; $F_{кр.} = 2{,}23$. Сравнивая эти значения, можно говорить, что уравнение статистически значимо: $F_{факт} > F_{кр.}$.

Интерпретация полученных результатов позволяет сделать вывод, что рентабельность производства зерновых в целом по Ставропольскому краю лишь на 17% зависит от урожайности данной культуры и обновления основных фондов. Это означает, что во второе уравнение следует включить дополнительные факторы, оказывающие более значимое влияние на результативный показатель.

В целом же полученная структурная форма модели имеет вид

$$\begin{cases} y_1 = 15{,}54 + 5{,}03x_1 + 3{,}83x_2, \\ y_2 = -11{,}22 + 0{,}33x_3 + 1{,}52y_1 \end{cases}$$

Задача 10.3. На основании имеющихся данных построить следующую форму модели:

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + b_{11}x_3, \\ y_2 = a_{20} + a_{21}x_1 + a_{22}x_2 + a_{23}x_3, \\ y_3 = a_{30} + a_{31}x_1 + a_{32}x_2 + a_{33}x_3, \end{cases}$$

где y_1 — заработная плата работников, руб.;

y_2 — темп прироста цен, %;
y_3 — доходы населения, млрд руб.;
x_1 — процент безработных;
x_2 — прирост цен на импорт, %;
x_3 — прирост экономически активного населения, %.

Таблица 10.5

Исходные данные к задаче 10.3

Год	y_1	y_2	y_3	x_1	x_2	x_3
2003	1100	21	10	1	2	1
2004	1250	18	12	2	3	2
2005	1280	16	11	3	2	3
2006	1700	16	15	2	4	4
2007	1950	14	14	3	3	5
2008	2500	15	16	4	3	6
2009	3100	12	18	5	2	7
2010	4200	10	21	4	5	8

Решение

Представленная в задаче эконометрическая модель является системой независимых уравнений. Следовательно, каждое из уравнений системы может быть рассмотрено самостоятельно, а поиск решения целесообразно осуществить с помощью классического метода наименьших квадратов.

Для оценки параметров первого уравнения системы осуществим следующие шаги:

- сформируем в *SPSSStatistics* 20 таблицу исходных данных;
- в меню «Анализ» выберем пункт «Регрессия» → «Линейная»;
- в диалоговом окне «Линейная регрессия» укажем зависимую переменную (y_1) и независимые переменные (x_1, x_2, x_3);
- в качестве метода выбора значимых факторов для модели укажем метод исключения. Это позволит в итоге улучшить качество модели путем исключения не значимых для результативного признака факторов (рис. 10.15);
- нажмем «ОК».

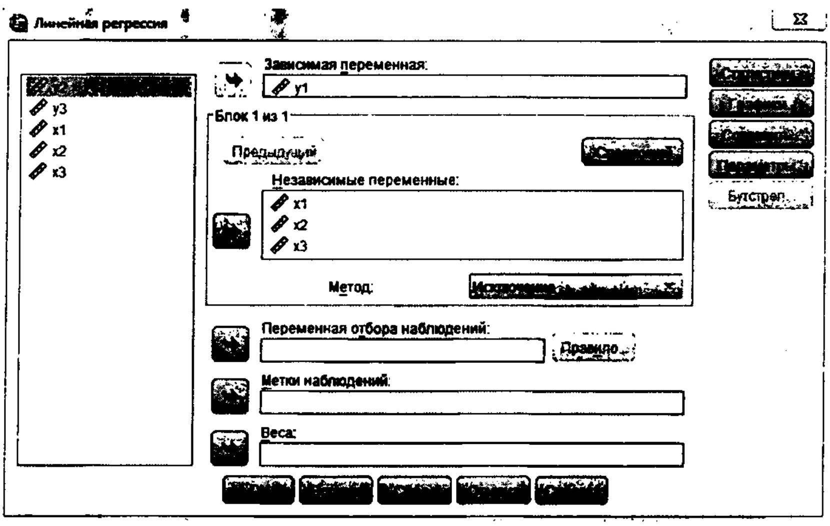


Рис. 10.15. Ввод условий поиска решения первого уравнения системы в *SPSSStatistics 20*

В окне результатов построения модели найдем таблицу «Коэффициенты» (рис. 10.16). Проанализируем полученные результаты.

Коэффициенты[a]

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знч.
		B	Стд. Ошибка	Бета		
1	(Константа)	170.169	1141,432		.149	,889
	x1	-101.798	524.429	-,124	-,194	,856
	x2	90,674	334,914	,090	,271	,800
	x3	444,045	320,095	1,009	1,387	,238
2	(Константа)	-29,235	447,149		-,065	,950
	x2	145,765	159,795	,145	,912	,404
	x3	383,765	69,740	,872	5,503	,003
3	(Константа)	267,500	302,475		,884	,411
	x3	415,000	59,899	,943	6,928	,000

a. Зависимая переменная: y1

Рис. 10.16. Результаты подбора модели для первого уравнения системы в SPSSStatistics *20*

Первоначально в уравнение были включены все указанные экзогенные переменные, а оценка параметров уравнения позволила установить его первоначальный вид:

$$y_1 = 170,17 - 101,8x_1 + 90,67x_2 + 444,05x_3.$$

Однако значения t-статистик Стьюдента показали, что ни один из полученных коэффициентов не является статистически значимым:

$t_{a_{10}} = 0,149;\ p = 0,889;$

$t_{a_{11}} = -0,194;\ p = 0,856;$

$t_{a_{12}} = 0,271;\ p = 0,8;$

$t_{a_{13}} = 1,387;\ p = 0,238.$

Поскольку незначимые экзогенные переменные необходимо исключать последовательно, вначале из уравнения была исключена наименее значимая с точки зрения величины вероятности ошибки p переменная x_1. Оценка параметров нового уравнения указана во второй строке исследуемой таблицы, а само уравнение имеет вид $y_1 = -29,24 + 145,77x_2 + 383,77x_3$. Анализируя значения t-критериев Стьюдента для коэффициентов регрессии, можно сделать вывод, что статистически адекватным является только коэффициент $a_{23} = 383,77$:

$t_{a_{10}} = -0,065;\ p = 0,95;$

$t_{a_{12}} = 0,912;\ p = 0,404;$

$t_{a_{13}} = 5,503;\ p = 0,003.$

Следовательно, на следующем этапе необходимо исключить переменную x_2 из уравнения. В третьей строке таблицы «Коэффициенты» окна вывода результатов отражены результаты оценки параметров первого уравнения системы, в котором исключен фактор x_2. Новое уравнение регрессии имеет вид $y_1 = 267,5 + 415x_3$.

Рассмотрим значения t-статистик Стьюдента для этого уравнения:

$t_{a_{10}} = 0,884;\ p = 0,411;$

$t_{a_{13}} = 6,928;\ p = 0,000.$

Следовательно, параметр a_{10} следует признать статистически незначимым, а параметр a_{13} отличается высокой степенью надежности. В рамках исследуемого значения это значит, что с увеличением прироста экономически активного населения на 1% заработная плата работников увеличивается на 415 рублей. Данное утверждение можно считать достаточно обоснованным, поскольку вероятность ошибки такого заключения в соответствии с полученными результатами находится на уровне 0,0%. С другой стороны, утверждение о том, что средний уровень заработной платы работников в рамках исследуемой системы без учета фактора прироста экономически активного населения находится на уровне 267,5 руб., можно считать статистически неадекватным, поскольку вероятность ошибочного вывода в этом случае составляет 41,1%.

Следует также обратить внимание на таблицу «Сводка для модели» окна вывода результатов моделирования (рис. 10.17). По данной таблице можно сделать вывод, что уравнение вида $y_1 = a_{10} + a_{13}x_3$ является наиболее адекватным исследуемому процессу, поскольку скорректированный индекс детерминации максимален: $\bar{R}^2 = 0,87$.

Сводка для модели[d]

Модель	R	R-квадрат	Скорректированный R-квадрат	Стд. ошибка оценки
1	,952[a]	,906	,835	438,154
2	,951[b]	,905	,867	393,738
3	,943[c]	,889	,870	388,190

a. Предикторы: (конст) x3, x2, x1

b. Предикторы: (конст) x3, x2

c. Предикторы: (конст) x3

d. Зависимая переменная: y1

Рис. 10.17. Оценка качества первого уравнения регрессии в *SPSSStatistics 20*

Для оценки параметров второго уравнения системы выполним следующие операции:

- в меню «Анализ» выберем пункт «Регрессия» → «Линейная»;
- в диалоговом окне «Линейная регрессия» укажем зависимую переменную (y_2) и независимые переменные (x_1, x_2, x_3);
- в качестве метода выбора значимых факторов для модели укажем метод исключения, нажмем «OK».

В окне результатов построения модели найдем таблицу «Коэффициенты» (рис. 10.18). Проанализируем полученные результаты.

Коэффициенты[a]

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знч.
		B	Стд. Ошибка	Бета		
1	Константа	24,039	3,075		7,818	,001
	x1	-1,253	1,413	-,481	-,887	,425
	x2	-,843	,902	-,264	-,934	,403
	x3	-,556	,862	-,399	-,645	,554
2	Константа	25,734	1,501		17,148	,000
	x1	-2,137	,322	-,820	-6,638	,001
	x2	-1,358	,394	-,425	-3,444	,018

a. Зависимая переменная: y2

Рис. 10.18. Результаты подбора модели для второго уравнения системы в *SPSSStatistics 20*

Первоначально в уравнение были включены все указанные экзогенные переменные, а оценка параметров уравнения позволила установить его первоначальный вид:

$$y_2 = 24{,}039 - 1{,}253x_1 - 0{,}843x_2 - 0{,}556x_3.$$

Однако значения t-статистик Стьюдента показали, что лишь коэффициент $a_{20} = 24{,}039$ можно считать статистически значимым:

$t_{a_{20}} = 7,818;\ p = 0,001;$

$t_{a_{21}} = -0,887;\ p = 0,425;$

$t_{a_{22}} = -0,934;\ p = 0,403;$

$t_{a_{23}} = -0,645;\ p = 0,554.$

На следующем этапе из уравнения регрессии была исключена экзогенная переменная x_3, поскольку коэффициент регрессии при ней наименее значим с точки зрения уровня значимости p, полученного для соответствующего значения t-критерия Стьюдента. Второе уравнение системы после выполнения этой операции приняло вид: $y_2 = 25,734 - 2,137x_1 - 1,358x_2$. Анализ t-статистик Стьюдента свидетельствует об адекватности полученных параметров:

$t_{a_{20}} = 17,148;\ p = 0,000;$

$t_{a_{21}} = -6,638;\ p = 0,001;$

$t_{a_{22}} = -3,444;\ p = 0,18.$

Таким образом, на основании второго уравнения системы можно сделать вывод, что среднее значение темпа прироста цен при элиминировании таких факторов, как процент безработных и прирост цен на импорт, находится на уровне 25,734% в течение исследуемого периода. С увеличением безработных на 1% темп прироста цен падает на 2,1%, а с увеличением прироста цен на импорт на 1% — снижается на 1,36%.

Анализируя таблицу «Сводка для модели» окна вывода результатов (рис. 10.19), можно сделать вывод, что уравнение вида $y_1 = a_{20} + a_{21}x_1 + a_{22}x_2$ является наиболее адекватным исследуемому процессу, поскольку скорректированный индекс детерминации максимален: $\bar{R}^2 = 0,894$. Таким образом, темп прироста цен можно объяснить воздействием процента безработных и приростом цен на импорт на 89,4%. Следует отметить, что предпочтение данному показателю при оценке статистической значимости модели в целом отдается по причине того, что скорректированный индекс детерминации учитывает поправку на число

Сводка для модели[c]

Модель	R	R-квадрат	Скорректированный R-квадрат	Стд. ошибка оценки
1	.965[a]	.932	.880	1,180
2	.962[b]	.925	.894	1.109

a. Предикторы: (конст) x3, x2, x1

b. Предикторы: (конст) x2, x1

c. Зависимая переменная: y2

Рис. 10.19. Оценка качества первого уравнения регрессии в *SPSSStatistics 20*

степеней свободы, а значит, дает более объективную оценку, чем индексы корреляции и детерминации.

Для оценки параметров третьего уравнения системы выполним последовательность действий:

- в меню «Анализ» выберем пункт «Регрессия» → «Линейная»;
- в диалоговом окне «Линейная регрессия» укажем зависимую переменную (y_3) и независимые переменные (x_1, x_2, x_3);
- в качестве метода выбора значимых факторов для модели укажем метод исключения, нажмем «ОК».

В окне результатов построения модели найдем таблицу «Коэффициенты» (рис. 10.20). Проанализируем полученные результаты.

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знч.
		B	Стд. Ошибка	Бета		
1	(Константа)	6.691	2.775		2.411	.073
	x1	-.121	1.275	-.043	-.095	.929
	x2	.764	.814	.221	.938	.401
	x3	1.334	.778	.883	1.714	.162
2	(Константа)	6.454	1.083		5.958	.002
	x2	.829	.387	.240	2.142	.085
	x3	1.263	.169	.836	7.474	.001

а. Зависимая переменная: y3

Рис. 10.20. Результаты оценки третьего уравнения системы в *SPSSStatistics 20*

Первоначально в уравнение были включены все указанные экзогенные переменные, а оценка параметров уравнения позволила установить его первоначальный вид:

$$y_3 = 6{,}691 - 0{,}121x_1 + 0{,}764x_2 + 1{,}334x_3.$$

Однако значения t-статистик Стьюдента показали, что ни один из полученных коэффициентов не является статистически значимым:

$t_{a_{30}} = 2{,}411$; $p = 0{,}073$;

$t_{a_{31}} = -0{,}095$; $p = 0{,}929$;

$t_{a_{32}} = 0{,}938$; $p = 0{,}401$;

$t_{a_{33}} = 1{,}714$; $p = 0{,}162$.

На следующем этапе из уравнения была исключена переменная x_1 как наименее значимая.

Оценка параметров нового уравнения указана во второй строке таблицы, а само уравнение имеет вид

$$y_3 = 6{,}454 + 0{,}829x_2 + 1{,}263x_3.$$

Анализируя значения t-критериев Стьюдента для коэффициентов регрессии, можно сделать вывод, что статистически адекватными можно признать все полученные значения параметров:

$t_{a_{30}} = 5{,}958$; $p = 0{,}002$;

$t_{a_{32}} = 2{,}142$; $p = 0{,}085$;

$t_{a_{33}} = 7{,}474$; $p = 0{,}001$.

Интерпретируя третье уравнение системы, можно сделать следующее заключение: значение доходов населения в рамках изучаемой системы в среднем за исследуемый период составляет 6,454 млрд руб. в случае отсутствия воздействия совокупности факторов — прирост цен на импорт и прирост экономически активного населения. В результате увеличения прироста цен на импорт на 1% доходы населения возрастают на 0,829 млрд руб., а в случае прироста экономически активного населения на 1% происходит увеличение доходов на 1,263 млрд руб.

В таблице «Сводка для модели» окна вывода результатов моделирования отражены показатели тесноты связи и статистической значимости полученного уравнения (рис. 10.21). Согласно данной таблице совокупность факторных признаков: прирост цен на импорт и прирост экономически активного населения — позволяет объяснить вариацию доходов населения на 93,4%.

Сводка для модели[c]

Модель	R	R-квадрат	Скорректированный R-квадрат	Стд. ошибка оценки
1	,976[a]	,953	,917	1,065
2	,976[b]	,953	,934	,954

a. Предикторы: (конст) x3, x2, x1

b. Предикторы: (конст) x3, x2

c. Зависимая переменная: y3

Рис. 10.21. Оценка качества третьего уравнения регрессии в *SPSSStatistics 20*

Отразим результаты моделирования в сводной таблице 10.6.

Таблица 10.6

Полученные уравнения системы и показатели статистической значимости

№	Уравнение	R	R^2	R^2
1	$y_1 = 267{,}5 + 415x_3$	0,943	0,889	0,87
2	$y_2 = 25{,}734 - 2{,}137x_1 - 1{,}358x_2$	0,962	0,925	0,894
3	$y_3 = 6{,}454 + 0{,}829x_2 + 1{,}263x_3$	0,976	0,953	0,934

Таким образом, все полученные уравнения системы следует признать статистически адекватными. Система уравнений имеет следующий вид:

$$\begin{cases} y_1 = 267{,}5 + 415x_3, \\ y_2 = 25{,}734 - 2{,}137\,x_1 - 1{,}358\,x_2 \\ y_3 = 6{,}454 + 0{,}829x_2 + 1{,}263x_3 \end{cases}$$

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

Задача 10.4. На основании имеющихся данных построить следующую структурную форму модели:

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2, \\ y_2 = a_{20} + a_{21}x_2 + a_{22}x_3 + b_{21}y_1, \\ y_3 = a_{30} + a_{31}x_3 + b_{31}y_1, \end{cases}$$

где y_1 — произведено мяса, тыс. т;

y_2 — коэффициент обеспеченности мясными продуктами населения;

y_3 — уровень рентабельности привеса КРС, %;

x_1 — получено приплода телят в СХП в расчете на 100 коров, гол.;

x_2 — заготовлено кормов в СХП в расчете на одну условную голову скота, ц корм. ед.;

x_3 — реализовано скота и птицы сельскохозяйственными предприятиями, тыс. т.

Таблица 10.7

Показатели деятельности отрасли животноводства в разрезе муниципальных районов Ставропольского края

№ района	y_1	y_2	y_3	x_1	x_2	x_3
1	4,3	0,654	–45,8	77,0	16,5	0,6
2	3,9	0,660	–34,3	53,0	22,0	0,3
3	7,2	0,662	–19,6	84,0	12,8	3,4
4	4,5	0,658	–50,9	66,0	14,8	1,5
5	3,9	0,663	–48,3	90,0	14,4	0,9
6	5,0	0,664	–35,8	88,0	25,4	0,4

Окончание табл. 10.7

№ района	y_1	y_2	y_3	x_1	x_2	x_3
7	6,5	0,715	–42,5	73,0	22,8	2,5
8	3,8	0,664	–34,7	84,0	29,1	1,3
9	11,0	0,665	–15,4	85,0	24,5	1,2
10	8,0	0,661	–20	82,0	16,1	2,6
11	5,4	0,670	–9,6	79,0	23,4	1,8
12	12,8	0,667	–29,1	44,0	23,9	4,9
13	7,9	0,665	–53,8	58,0	32,7	1,2
14	4,6	0,674	–34,5	83,0	15,6	1,1
15	8,9	0,665	–14,3	67,0	7,4	2,2
16	2,8	0,664	–30,1	80,0	37,1	0,1
18	9,3	0,660	–37,9	77,0	24,5	3,4
19	6,5	0,670	–42,6	33,0	18,0	3,1
20	8,2	0,663	–45,7	90,0	27,4	1,5
21	5,1	0,672	–31,9	74,0	45,0	2,6
22	10,6	0,663	–20,5	54,0	24,2	3,9
23	6,2	0,666	–15,1	58,0	22,0	1,1
24	12,2	0,658	–29,5	96,0	22,8	4,4
25	3,7	0,662	–31,8	82,0	13,5	0,9
26	54,5	0,687	–53,4	58,0	21,3	51,9

Задача 10.5. На основании имеющихся данных построить следующую структурную форму модели:

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + b_{11}y_2, \\ y_2 = a_{20} + a_{21}x_2 + a_{22}x_3 + a_{23}x_4 + b_{21}y_1, \\ y_3 = a_{30} + a_{31}x_3 + a_{32}x_4 + b_{31}y_1, \end{cases}$$

где y_1 — прибыль (убыток) сельхозпредприятий, тыс. руб.;

y_2 — кредиторская задолженность предприятий, млн руб.;

y_3 — инвестиции в основной капитал, млн руб.;

x_1 — среднее значение температуры воздуха, °С;

x_2 — доля зерновых в общей посевной площади, %;

x_3 — доля растениеводства в общем объеме произведенной сельскохозяйственной продукции, %;

x_4 — производственная себестоимость 1 ц зерна, руб.

Таблица 10.8

Данные к задаче 10.5 в разрезе муниципальных районов Ставропольского края

y_1	y_2	y_3	x_1	x_2	x_3	x_4
55310	201,1	123,0	11,7	81,2	69,13	141
50496	217,2	181,7	11,2	74,6	67,25	155
119755	376,8	255,4	11,8	89,2	72,07	187
93097	548,5	305,4	11,2	69,3	66,48	205
−29654	394,2	147,2	10,3	67,7	63,44	183
80739	1353,8	549,2	10,5	60,6	68,42	202
80420	269,8	255,7	11,7	67,4	67,78	150
−8616	277,5	187,4	11,2	64,9	63,53	190
139676	415,7	439,8	10,8	57,8	58,84	185
146581	299	319,5	11,8	71,6	79,15	176
99619	123,2	78,9	12	78,1	68,20	152
104558	316,3	182,9	11,7	80,5	58,73	152
13677	1575,1	1213,9	11,4	65,4	69,52	231
28778	1197,9	829,1	12	61,4	65,92	159
182380	738,1	456,8	11,7	65,0	77,35	186
192586	74	69,2	11,2	78,2	68,34	155

КОНТРОЛЬНОЕ ТЕСТИРОВАНИЕ ЗНАНИЙ

ВАРИАНТ 1

1. Случайным называется такое событие, которое:

а) не происходит никогда в условиях данного эксперимента;
б) может произойти или не произойти в условиях данного эксперимента;
в) происходит всегда в условиях данного эксперимента.

2. Спецификация модели — это:

а) формулировка вида модели;
б) пояснение к модели;
в) формулировка цели моделирования;
г) перечисление входящих в нее переменных;
д) оптимизация модели.

3. Корреляционно-регрессионный анализ относится к методам оценки взаимосвязи между переменными:

а) непараметрическим;
б) оптимизационным;
в) статистическим;
г) функциональным.

4. Общая сумма квадратов отклонений в парной регрессии имеет число степеней свободы, равное:

а) 1;
б) $n - 1$;
в) $n - 2$.

5. Значение коэффициента детерминации составило 0,64. Определите долю случайных факторов в общей дисперсии зависимой переменной:

а) 0,64%; б) 0,36%; в) 0,8%; г) 64%.

6. Множественный коэффициент корреляции $R_{yx_1 \cdot x_2} = 0{,}8$. Объясненная часть дисперсии зависимой переменной y влиянием факторов x_1 и x_2 составит:

а) 80%; б) 28%; в) 64%; г) 32%.

7. Под идентификацией понимается:

а) возможность или невозможность получения структурных параметров системы одновременных уравнений через приведенные формы уравнений;

б) определение количества эндогенных переменных в системе уравнений;

в) получение оценок параметров приведенных уравнений.

8. Общий недостаток всех методов расчета коэффициентов структурной формы системы взаимосвязанных уравнений в том, что не учитывается:

а) мультиколлинеарность между предопределенными переменными;

б) автокорреляция экзогенных переменных;

в) автокорреляция эндогенных переменных;

г) мультиколлинеарность случайных возмущений.

9. По результатам наблюдений получен парный коэффициент корреляции $r_{yx_1} = 0{,}6$. Известно, что x_2 занижает связь между y и x_1. Частный коэффициент корреляции принимает значение:

а) 0,8;

б) –0,5;

в) 0,5;

г) –0,6.

10. Число степеней свободы остаточной суммы квадратов отклонений при n наблюдениях для множественной линейной регрессии $y = a_0 + a_1 x_1 + \ldots + a_k x_k$ равно:

а) 1;

б) $n - k - 1$;

в) k;

г) $n - k$.

11. В модели вида $y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon$ количество объясняющих переменных равно:

а) 2; б) 4; в) 1; г) 3.

12. Для регрессионной модели вида $y = a + b \cdot x + \varepsilon$ рассчитано значение коэффициента парной корреляции r_{xy}. Если $r_{xy} < 0$, то связь между y и x:

а) обратная;
б) функциональная;
в) отсутствует;
г) прямая.

13. Одной из предпосылок метода наименьших квадратов является утверждение, что остатки регрессионной модели должны подчиняться закону распределения:

а) нормальному;
б) равномерному;
в) экспоненциальному;
г) геометрическому.

14. По результатам 50 статистических наблюдений построено уравнение множестенной регрессии $y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon$. Число степеней свободы остаточной суммы квадратов отклонений для этого уравнения равно:

а) 46; б) 48; в) 49; г) 47.

15. При моделировании линейного уравнения множественой регресии вида $y = a + b_1x_1 + b_2x_2 + \varepsilon$ необходимо, чтобы выполнялось требование отсутствия взаимосвязи между:

а) y и $\{x_1; x_2\}$;
б) a и $\{b_1; b_2\}$;
в) x_1 и x_2;
г) b_1 и b_2.

ВАРИАНТ 2

1. Достоверным называется такое событие, которое:
а) может произойти или не произойти в условиях данного эксперимента;
б) происходит всегда в условиях данного эксперимента;
в) не происходит никогда в условиях данного эксперимента.

2. Коэффициент регрессии показывает:
а) среднее изменение результата с изменением фактора на одну единицу;
б) уровень значимости уравнения регрессии;
в) степень разброса значений фактора;
г) степень разброса значений результата;
д) тесноту связи между результатом и фактором.

3. Эконометрика синтезирует в себе науки:
а) микроэкономику, математику и информатику;
б) экономическую теорию, математическую статистику и экономическую статистику;
в) макроэкономику, теорию вероятностей и линейную алгебру;
г) экономический анализ, статистику и информатику.

4. Построена эконометрическая модель зависимости валового регионального продукта (млн руб.) от стоимости основных средств в регионе (млн руб.): $y = 1100 + 0{,}11x + \varepsilon$. Если стоимость основных фондов в регионе составит 1 млн руб., то стоимость валового регионального продукта составит:
а) 1111,0 млн руб.;
б) 1100,11 млн руб.;
в) 0,11 млн руб.;
г) 1000 млн руб.

5. Доля объясненной дисперсии зависимой переменной в общей дисперсии этого признака составила 100%. Следовательно, доля влияния случайных факторов составляет:

а) 100%; б) 1; в) 0,1; г) 0%.

6. Множественный индекс корреляции не принимает значение:

а) –1; б) –0,5; в) 1,2; г) 0.

7. Косвенный МНК используется для определения состоятельных структурных параметров в системе одновременных уравнений, если уравнения:

а) точно идентифицированы;
б) неидентифицированы;
в) сверхидентифицированы.

8. Выберите возможные формы системы одновременных уравнений:

а) рекурсивная;
б) структурная;
в) приведенная;
г) стандартная.

9. Разность уровней ряда динамики называется:

а) темпом прироста;
б) темпом роста;
в) абсолютным приростом;
г) коэффициентом роста.

10. Хронологическая последовательность значений признака, характеризующего состояние данного объекта, называется:

а) автокорреляционной функцией;
б) корреляционным полем;
в) случайной выборкой;
г) временным рядом.

11. Степенной моделью *не является* регрессионная модель:
а) $y = x^a \cdot \varepsilon$;
б) $y = a \cdot b^x \cdot \varepsilon$;
в) $y = a \cdot x^b \cdot \varepsilon$;
г) $y = \ln a \cdot x^b \cdot \varepsilon$.

12. Для нелинейного уравнения регрессии рассчитано значение индекса детерминации, которое составило $R^2 = 0,7$. Следовательно, доля остаточной дисперсии в общей дисперсии зависимой переменной для данного уравнения составляет:
а) 0,7%;
б) 0,7%;
в) 0,3%;
г) 0,3%.

13. Левая часть системы эконометрических уравнений представлена совокупностью переменных:
а) экзогенных;
б) случайных;
в) независимых;
г) зависимых.

14. Для оценки параметров эконометрической модели линейного уравнения регрессии вида $y = a_0 + a_1x_1 + a_2x_2 + \dots a_ix_i + \dots + a_nx_n + \varepsilon$ используется метод наименьших квадратов. В системе нормальных уравнений МНК неизвестными величинами являются:
а) ε; б) a_i; в) y; г) x_n.

15. Для регрессионной модели известны следующие величины дисперсий:

$$\sum(y - y_x)^2 \,;\ \sum(y - \bar{y})^2 \,;\ \sum(y_x - \bar{y})^2 \,,$$

где y — значение зависимой переменной по исходным данным; y_x — значение зависимой переменной, вычисленное по регрессионной модели; $\bar{y}$ — среднее значение зависимой переменной, определенное по исходным статистическим данным. Для указанных дисперсий справедливо равенство:

а) $\sum(y-\bar{y})^2=\sum(y_x-\bar{y})^2+\sum(y-y_x)^2$;

б) $\sum(y_x-\bar{y})^2=\sum(y-\bar{y})^2+\sum(y-y_x)^2$;

в) $\sum(y-\bar{y})^2+\sum(y_x-\bar{y})^2=\sum(y-y_x)^2$;

г) $\sum(y-\bar{y})^2=\sum(y_x-\bar{y})^2-\sum(y-y_x)^2$.

ВАРИАНТ 3

1. Случайная величина — это:
а) исход или совокупность исходов вероятностного эксперимента;
б) количественная мера для сравнения событий по степени возможности их появления;
в) заранее неизвестное численное значение, зависящее от случайных обстоятельств.

2. В парной регрессии спецификация модели связана:
а) с анализом качества уравнения регрессии;
б) определением параметров регрессии;
в) переходом к стандартизации переменных;
г) выбором вида функциональной зависимости.

3. Оценка расстояния может быть определена по формуле:
а) метрика Евклидово расстояние: $d_e=(\sum\left(x_{ik}-x_{jk}\right)^2)^{\frac{1}{2}}$;
б) метрика Евклидово расстояние: $d_e=(\sum\left(x_{ik}-x_{jk}\right)^2)^{\frac{1}{2}}$;;
в) метрика Евклидово расстояние: $d_e=(\sum\left(\overline{x_{ik}-x_{jk}}\right)^2)^{\frac{1}{2}}$;;

г) метрика Евклидово расстояние: $d_e = (\sum \left(\overline{x_{ik}} - x_{jk}\right)^2)^{\frac{1}{2}}$.

4. При оценке параметров линейных уравнений регрессии используется метод:

а) наименьших квадратов;
б) максимальных квадратов;
в) наибольших квадратов;
г) нулевых квадратов.

5. Критерий Фишера (F-критерий) в эконометрических моделях служит:

а) для проверки статистической значимости уравнения регрессии;
б) для оценки параметров регрессии;
в) показателем линейной связи между переменными;
г) показателем преимущества выбранной модели перед другими.

6. Статистическая надежность оценок коэффициентов регрессии:

а) увеличивается с уменьшением числа степеней свободы;
б) не зависит от числа степеней свободы;
в) увеличивается с увеличением числа степеней свободы.

7. Для точно идентифицированных уравнений двухшаговый метод наименьших квадратов дает оценки:

а) одинаковые с косвенным МНК;
б) лучше, чем косвенный МНК;
в) хуже, чем косвенный МНК.

8. Выберите верные утверждения по поводу системы одновременных уравнений:

а) в каждом последующем уравнении системы зависимая переменная является функцией от всех зависимых и предопределенных переменных предшествующих уравнений;

б) каждое уравнение системы не может рассматриваться в качестве отдельного уравнения регрессии зависимости одной переменной от группы факторов;
в) зависимые переменные в одних уравнениях могут выступать в роли независимых переменных в других уравнениях системы;
г) оценки параметров уравнений определяются только методом наименьших квадратов.

9. Базисный абсолютный прирост равен:
а) произведению цепных абсолютных приростов;
б) корню $n-1$ степени из произведения цепных абсолютных приростов;
в) корню $n-1$ степени из суммы абсолютных приростов;
г) сумме цепных абсолютных приростов.

10. Методами выравнивания уровней временного ряда могут служить:
а) графическое представление временного ряда;
б) метод наименьших квадратов;
в) построение уравнения регрессии, характеризующего зависимость уровней ряда от времени;
г) метод скользящей средней.

11. Для регрессионной модели $y=f(x)+\varepsilon$, где $f(x)$ — нелинейная функция, $y_x=f(x)$ — рассчитанное по модели значение переменной y, получены значения дисперсий: $S_y^2=3{,}12$; $S_{y_x}^2=2{,}82$; $S_\Delta^2=0{,}3$. Не объяснена моделью часть дисперсии переменной y, равная:
а) 10,4;
б) 0,904;
в) 0,096;
г) 0,106.

12. Для системы одновременных уравнений

$$\begin{cases} R_t = a_{11}Y_t + a_{12}M_t + \varepsilon, \\ Y_t = a_{21}R_t + a_{22}I_t + a_{23}G_t, \\ I_t = a_{31}R_t + a_{32}I_t, \end{cases}$$

где R_t — процентная ставка;

Y_t — реальный ВВП;

M_t — объем денежной массы;

I_t — внутренние инвестиции;

G_t — реальные государственные расходы.

Эндогенными являются переменные:

а) M_t;

б) R_t;

в) G_t;

г) ε.

13. Для структурной формы модели системы одновременных уравнений

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

y_i являются:

а) эндогенными переменными;

б) структурными коэфициентами;

в) экзогенными переменными;

г) приведенными коэффициентами.

14. Значение коэффициента автокорреляции второго порядка равно (–0,6), следовательно, ряд содержит:

а) тенденцию;

б) полиномиальную тенденцию с точкой минимума;

в) затухающую сезонную волну периодичностью 2 момента времени;

г) убывающую тенденцию.

14. Известно, что доля остаточной дисперсии зависимой переменной в ее общей дисперсии равна 0,2. Тогда значение коэффициента детерминации составляет:
а) 0,8;
б) $\sqrt{0,8}$;
в) 0,64;
г) $\sqrt{0,2}$.

15. Система эконометрических уравнений вида:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \varepsilon, \\ y_2 = a_{21}x_1 + a_{22}x_2 + \varepsilon, \\ y_1 = a_{31}x_1 + a_{33}x_3 + \varepsilon \end{cases}$$

является системой эконометрических уравнений:
а) нормальных;
б) независимых;
в) одновременных;
г) рекурсивных.

ВАРИАНТ 4

1. Выбор формы связи между переменными называется:
а) идентификацией;
б) идентифицируемостью;
в) верификацией;
г) спецификацией.

2. Неправильный выбор вида эконометрической модели называют ошибкой:
а) измерения переменных;
б) агрегирования переменных;
в) спецификации модели;
г) параметризации модели.

3. Расчет расстояния между всеми парами объектов d_{ij} осуществляют по метрике:

а) взвешенное Евклидово расстояние:

$$d_e = (\sum w_k \left(x_{ik} + x_{jk}\right)^2)^{\frac{1}{2}};$$

б) взвешенное Евклидово расстояние:

$$d_e = (\sum w_k \left(x_{ik} - x_{jk}\right)^2)^{\frac{1}{2}};$$

в) взвешенное Евклидово расстояние:

$$d_e = (\sum w_k \left(\overline{x_{ik} - x_{jk}}\right)^2)^{\frac{1}{2}};$$

г) взвешенное Евклидово расстояние:

$$d_e = (\sum w_k \left(\overline{x_{ik}} - x_{jk}\right)^2)^{\frac{1}{2}}.$$

4. Нарушением предпосылки МНК является случай остатков:

а) гомоскедастичности;
б) нормального распределения;
в) наличия автокорреляции;
г) случайного характера.

5. Укажите верное утверждение о скорректированном коэффициенте детерминации:

а) скорректированный коэффициент детерминации больше обычного коэффициента детерминации для $m > 1$;
б) скорректированный коэффициент детерминации меньше или равен обычному коэффициенту детерминации для $m > 1$;
д) скорректированный коэффициент детерминации меньше обычного коэффициента детерминации для $m > 1$.

6. Известно, что x_2 усиливает связь между y и x_1. По результатам наблюдений получен частный коэффициент корреляции $R_{yx_1 \cdot x_2} = -0{,}45$. Парный коэффициент корреляции r_{yx_1} принимает значение:

а) 0,4; б) 0,2; в) –0,3; г) –1,2.

7. Дана следующая модель:

$$\begin{cases} Y_{1t} = a_0 + a_1 X_{1t} + u_{1t}, \\ Y_{2t} = b_0 + b_1 Y_{1t} + b_2 X_{2t} + u_{2t}, \\ Y_{3t} = c_0 + c_1 Y_{1t} + c_2 Y_{2t} + c_3 X_{2t} + u_{3t}. \end{cases}$$

Данная модель является:

а) системой рекурсивных уравнений;
б) системой независимых уравнений;
в) системой взаимосвязанных моделей.

8. Применение традиционного метода наименьших квадратов к системе одновременных уравнений нецелесообразно, так как получаемые в результате оценки параметров являются:

а) несмещенными и несостоятельными;
б) несмещенными и состоятельными;
в) смещенными и несостоятельными;
г) смещенными и состоятельными.

9. Отношение текущего уровня ряда динамики к базисному называется:

а) цепным темпом роста;
б) базисным темпом роста;
в) цепным темпом прироста;
г) базисным темпом прироста;
д) абсолютным значением 1% прироста.

10. Законы распределения случайной величины необходимы для:

а) определения интервальных оценок и проверки статистических гипотез;

б) определения интервальных оценок;
в) проверки статистических гипотез.

11. Уравнением нелинейной регрессии, являющейся нелинейной относительно параметров, является:

а) $y = a_0 + a_1 x + a_2 x^2 + \varepsilon$;

б) $y = a_0 \cdot x^{a_1} \cdot \varepsilon$;

в) $y = a_0 + a_1 \cdot \frac{1}{x^2} + \varepsilon$;

г) $y = a_0 + a_1 \cdot \frac{1}{x} + a_2 \cdot \frac{1}{x^2} + \varepsilon$.

12. Фиктивные переменные эконометрической модели:
а) отражают качественные признаки исследуемого объекта наблюдения;
б) отражают количественые признаки исследуемого объекта наблюдения;
в) используются в случае однородных совокупностей данных;
г) используются в случае неоднородных совокупностей данных.

13. В эконометрической модели линейного уравнения регресии $y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_i x_i + \ldots + a_k x_k + \varepsilon$ коэффициентом регрессии, характеризующим среднее изменение зависимой переменной при изменении независимой переменной на 1 единицу измерения, является:
а) x_i;
б) y;
в) a_0;
г) a_i.

14. Значение критерия Дарбина–Уотсона можно приблизительно рассчитать по формуле $d \approx 2 \cdot (1 - r_\varepsilon)$, где r_ε — значение коэффициента автокорреляции остатков модели. Максимальная величина значения d будет наблюдаться при автокорреляции остатков:

а) отрицательной;
б) положительной;
в) бесконечно малой;
г) нулевой.

15. Известно, что общая сумма квадратов отклонений $\sum(y-\bar{y})^2 = 150$, а остаточная сумма квадратов отклонений — $\sum(y-y_x)^2 = 30$. Тогда значение коэффициента детерминации равно:

а) 0,8; б) $\sqrt{0{,}2}$; в) 0,2; г) $\sqrt{0{,}8}$.

ВАРИАНТ 5

1. Модели в эконометрике — это:

а) словесное описание экономического процесса;
б) математическая формула;
в) графическое представление поведения экономических показателей.

2. В качестве экзогенных переменных могут рассматриваться следующие переменные:

а) эндогенные переменные за предыдущий период времени;
б) переменные, которые могут быть объектом регулирования;
в) зависимые переменные;
г) эндогенные переменные за текущий период времени.

3. Для преодоления проблемы автокорреляции служит:

а) косвенный метод наименьших квадратов;
б) двухшаговый метод наименьших квадратов;

в) метод наименьших квадратов;
г) обобщенный метод наименьших квадратов.

4. Важное значение коэффициента детерминации:
а) –0,5;
б) –0,2;
в) 0,4;
г) 1,2.

5. С увеличением числа объясняющих переменных скорректированный коэффициент детерминации:
а) не изменяется;
б) растет быстрее значения обычного коэффициента детерминации;
д) растет медленнее, чем обычный коэффициент детерминации.

6. Условие гомоскедастичности регрессионной модели:
а) $M\left(\varepsilon_i^{\,2}\right) \neq M\left(\varepsilon_j^{\,2}\right)$;
б) $M(\varepsilon_i) \cdot \varepsilon_j \neq 0$;;
в) $M\left(\varepsilon_i^{\,2}\right) = M\left(\varepsilon_j^{\,2}\right)$;
г) $M\left(\varepsilon_i\right) \cdot \varepsilon_j = 0$.

7. Преимуществом двухшагового МНК, по сравнению с косвенным МНК, является то, что он может быть использован для получения состоятельных оценок структурных параметров:
а) как для сверхидентифицированных, так и для точно идентифицированных уравнений в системе одновременных уравнений;
б) для неидентифицированных уравнений в системе одновременных уравнений;
в) как для неидентифицированных, так и для точно идентифицированных уравнений в системе уравнений.

8. В модели спроса — предложения:

$$\begin{cases} Q_t^d = a_0 + a_1 \cdot P_t + a_2 \cdot Y_t + \varepsilon_1, \\ Q_t^s = b_0 + b_1 \cdot P_t + \varepsilon_2, \\ Q_t^d = Q_t^s, \end{cases}$$

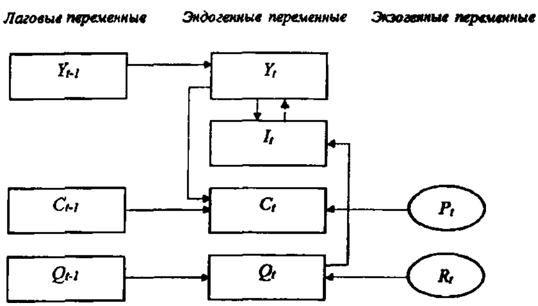
где Q_t^d и Q_t^s — количество покупаемых и продаваемых товаров; P_t — равновесная цена товара (определяется исходя из равновесия спроса и предложения), Y_t — доход покупателей, эндогенными являются переменные:

а) Q_t^d, Q_t^s;
б) P_t;
в) Y_t;
г) Q_t^d, Q_t^s, Q_t^s.

9. Для выявления основной тенденции развития явления используются:

а) метод укрупнения интервалов;
б) индексный метод;
в) метод скользящей средней;
г) расчет средней гармонической;
д) аналитическое выравнивание.

10. Схематическое изображение взаимосвязи различных переменных в системе эконометрических уравнений имеет вид:



Тогда соответствующая система эконометрических уравнений будет иметь вид:

а) $$\begin{cases} Y_t = a_1 + a_2 \cdot Y_{t-1} + a_3 \cdot I_t, \\ I_t = b_1 + b_2 \cdot Y_t + b_3 \cdot Q_t, \\ C_t = d_1 + d_2 \cdot C_{t-1} + d_3 \cdot Y_t + d_4 \cdot P_t, \\ Q_t = f_1 + f_2 \cdot Q_{t-1} + f_3 \cdot R_t. \end{cases}$$

б) $$\begin{cases} Y_t = a_1 \cdot Y_{t-1} + a_2 \cdot I_t + \varepsilon_1, \\ I_t = b_1 \cdot Y_t + \varepsilon_2, \\ C_t = d_1 \cdot C_{t-1} + d_2 \cdot Y_t + d_3 \cdot P_t + \varepsilon_3, \\ Q_t = f_1 \cdot Q_{t-1} + f_2 \cdot R_t + \varepsilon_4. \end{cases}$$

в) $$\begin{cases} Y_t = a_1 \cdot Y_{t-1} + a_2 \cdot I_t, \\ I_t = b_1 \cdot Y_t, \\ C_t = d_1 \cdot C_{t-1} + d_2 \cdot Y_t + d_3 \cdot P_t, \\ Q_t = f_1 \cdot Q_{t-1} + f_2 \cdot R_t. \end{cases}$$

г) $$\begin{cases} Y_t = a_1 + a_2 \cdot Y_{t-1} + a_3 \cdot I_t + \varepsilon_1, \\ I_t = b_1 + b_2 \cdot Y_t + b_3 \cdot Q_t + \varepsilon_2, \\ C_t = d_1 + d_2 \cdot C_{t-1} + d_3 \cdot Y_t + d_4 \cdot P_t + \varepsilon_3, \\ Q_t = f_1 + f_2 \cdot Q_{t-1} + f_3 \cdot R_t + \varepsilon_4. \end{cases}$$

11. Среди представленных нелинейных зависимостей нелинейной по параметрам является:

а) $y = a_0 + \dfrac{a_1}{x} + \varepsilon;$

б) $y = a_0 \cdot x^{a} \cdot \varepsilon;;$

в) $y = a_0 + a_1 \cdot \ln x + \varepsilon;$

г) $y = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \varepsilon.$

12. Для нелинейного уравнения регрессии рассчитано значение индекса детерминации $R^2 = 0,9$. Следовательно, доля остаточной дисперсии в общей дисперсии зависимой переменной для данного уравнения составляет:

а) 90%; б) 10%; в) 10%; г) 90%.

13. Совокупность значений экономического показателя за несколько последовательных моментов (периодов) времени называется:

а) автокорреляционной функцией;
б) тенденцией;
в) временным рядом;
г) коррелограммой.

14. Для аддитивной модели временного ряда $Y = T + S + \Delta$ лаг модели равен 4 и ивестны значения трех скорректированных сезонных компонент: $S_1 = 2$, $S_2 = -1$, $S_3 = -2$. S_4 равна:

а) 1; б) 2; в) 0; г) 4.

15. Система эконометрических уравнений содержит совокупность переменных:

а) стационарных и нестационарных;
б) гомоскедастичных и гетероскедастичных;
в) экзогенных и эндогенных;
г) постояных и изменяющихся.

ВАРИАНТ 6

1. Эконометрическая модель — это:

а) графическое представление экспериментальных данных;
б) совокупность числовых характеристик, характеризующих экономический объект;
в) линейная функциональная зависимость между экономическими показателями;
г) экономическая модель, представленная в математической форме.

2. В линейном уравнении парной регрессии $y = a + bx + \varepsilon$ переменными являются:

а) b; б) y; в) x; г) a.

3. Обобщенный метод наименьших квадратов подразумевает:

а) введение в выражение для дисперсии остатков коэффициента пропорциональности;
б) линеаризацию уравнения регрессии;
в) двухэтапное применение метода наименьших квадратов;
г) переход от множественной регрессии к парной.

4. Согласно методу наименьших квадратов минимизируется сумма:

а) $\sum\left|y_i - \hat{y}_i\right|$;
б) $\sum(y_i - \hat{y}_i)$;
в) $\sum(y_i - \hat{y}_i)^2$;
г) $\sum(y_i - y)^2$.

5. Скорректированный коэффициент детерминации увеличивается при добавлении новой объясняющей переменной только тогда:

а) когда t-статистика для параметра по этой переменной по модулю больше своего критического значения;
б) когда t-статистика параметра по модулю больше трех;
в) когда t-статистика параметра по модулю больше единицы.

6. Условие гетероскедастичности регрессионной модели:

а) $M(\varepsilon_i)\cdot\varepsilon_j \neq 0$;
б) $M(\varepsilon_i^2) = M(\varepsilon_j^2)$;
в) $M(\varepsilon_i^2) \neq M(\varepsilon_j^2)$;
г) $M(\varepsilon_i)\cdot\varepsilon_j = 0$.

7. В правой части структурной формы взаимозависимой системы могут стоять:

а) только экзогенные лаговые переменные;
б) только экзогенные переменные (как лаговые, так и нелаговые);
в) только эндогенные лаговые переменные;
г) только эндогенные переменные (как лаговые, так и нелаговые);
д) любые экзогенные и эндогенные переменные.

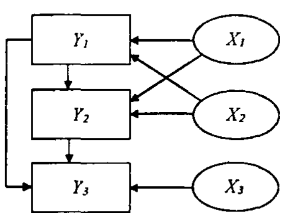
8. Число приведенных коэффициентов системы одновременных уравнений равно числу структурных коэффициентов, тогда модель:

а) не существует;
б) идентифицируема;
в) неидентифицируема;
г) сверхидентифицируема.

9. Средний уровень интервального ряда динамики с равными временными промежутками исчисляется по формуле средней:

а) хронологической простой;
б) гармонической взвешенной;
в) арифметической взвешенной;
г) гармонической простой;
д) хронологической взвешенной;
е) арифметической простой.

10. Для указанной схемы взаимосвязей между переменными справедливы утверждения:



а) может быть описана с помощью системы рекурсивных уравнений;
б) включает 6 уравнений;
в) может быть описана с помощью системы независимых уравнений;
г) включает 3 уравнения.

11. Система эконометрических уравнений вида:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \varepsilon_1, \\ y_2 = a_{21}x_1 + a_{22}x_2 + b_{21}y_1 + \varepsilon_2, \\ y_3 = a_{31}x_1 + a_{32}x_2 + b_{31}y_1 + b_{32}y_2 + \varepsilon_3 \end{cases}$$

является системой эконометрических уравнений:
а) рекурсивных;
б) нормальных;
в) независимых;
г) одновременных.

12. Для структурной формы модели системы одновременных уравнений

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

b_i являются:
а) эндогенными переменными;
б) структурными коэффициентами;
в) экзогенными переменными;
г) приведенными коэффициентами.

13. При исследовании зависимости потребления мяса от уровня дохода и пола потребителя можно рекомендовать:
а) исключить из рассмотрения пол потребителя, так как даный фактор нельзя измерить количественным образом;
б) использовать фиктивную переменную — уровень дохода;

в) использовать фиктивную переменную — пол потребителя;
г) разделить совокупность на две: для потребителей женского пола и для потребителей мужского пола.

14. Для нелинейного уравнения парной регрессии рассчитано значение индекса корреляции $R = 0,6$. Следовательно:
а) доля объясненной дисперсии в общей дисперсии зависимой переменой для данного уравнения составляет 36%;
б) доля объясненной дисперсии в общей дисперсии зависимой переменой для данного уравнения составляет 60%;
в) доля остаточной дисперсии в общей дисперсии зависимой переменой для данного уравнения составляет 64%;
г) доля остаточной дисперсии в общей дисперсии зависимой переменой для данного уравнения составляет 40%.

15. По типу функциональной зависимости между переменными эконометрической модели различают уравнения регрессии:
а) линейные и парные;
б) линейные и нелинейные;
в) множественные и парные;
г) стохастические и вероятностные.

ВАРИАНТ 7

1. Учет в эконометрической модели временного фактора тем или иным способом является одним из принципов:
а) верификации;
б) спецификации;
в) параметризации;
г) линеаризации.

2. В линейном уравнении парной регрессии $y = a + bx + \varepsilon$ параметрами не являются:
а) x; б) a; в) y; г) b.

3. Общая сумма квадратов отклонений может интерпретироваться как мера разброса:

а) реальных значений зависимой переменной относительно ее среднего значения;

б) реальных значений независимой переменной относительно ее среднего значения;

в) случайных отклонений;

г) расчетных значений зависимой переменной относительно ее среднего значения.

4. На практике гетероскедастичность имеет место, если есть основания считать, что:

а) вероятностные распределения случайных отклонений при различных наблюдениях будут одинаковы;

б) дисперсии случайных отклонений постоянны;

в) вероятностные распределения случайных отклонений при различных наблюдениях будут различны.

5. При добавлении существенной объясняющей переменной X в линейную модель множественной регрессии скорректированный коэффициент детерминации:

а) уменьшается;

б) не изменяется;

в) увеличивается.

6. На практике гетероскедастичность имеет место, если есть основания считать, что:

а) вероятностные распределения случайных отклонений при различных наблюдениях будут одинаковы;

б) дисперсии случайных отклонений постоянны;

в) дисперсии случайных отклонений непостоянны;

г) вероятностные распределения случайных отклонений при различных наблюдениях будут различны.

6. Имеется следующая модель:

$$\begin{cases} R_t = a_1 + b_{12}Y_t + b_{14}M_t + \varepsilon_1, \\ Y_t = a_2 + b_{21}R_t + b_{23}I_t + b_{25}G_t + \varepsilon_2, \\ I_t = a_3 + b_{31}R_t + \varepsilon_3. \end{cases}$$

Она имеет следующие характеристики:

а) 3 эндогенные и 2 экзогенные переменные, модель неидентифицируема;

б) 3 эндогенные и 2 экзогенные переменные, модель сверхидентифицируема;

в) 3 эндогенные и 3 экзогенные переменные, модель идентифицируема;

г) 3 эндогенные и 2 экзогенные переменные, все уравнения неидентифицируемы.

8. Предопределенные переменные включают:

а) все экзогенные и эндогенные переменные;

б) только экзогенные переменные;

в) лаговые экзогенные и эндогенные переменные;

г) все экзогенные переменные и лаговые эндогенные переменные.

9. Средний уровень интервального ряда динамики с неравными временными промежутками исчисляется по формуле средней:

а) хронологической простой;

б) гармонической взвешенной;

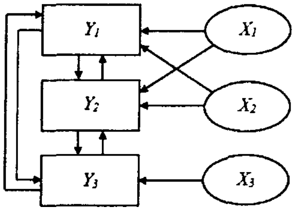
в) арифметической взвешенной;

г) гармонической простой;

д) хронологической взвешенной;

е) арифметической простой.

10. Для указанной схемы взаимосвязей между переменными справедливы утверждения:



а) включает 3 уравнения;
б) включает 6 уравнений;
в) может быть описана с помощью системы одновременных уравнений;
г) может быть описана с помощью системы рекурсивных уравнений.

11. При моделировании уравнения множественной регрессии проверку тесноты связи между независимыми переменными модели осуществляют на основе:
а) коэффициента множественной корреляции;
б) системы нормальных уравнений МНК;
в) матрицы парных коэффициентов линейной корреляции;
г) показателей существенности параметров модели.

12. В эконометрической модели линейного уравнения регрессии $y = a_0 + a_1x_1 + a_2x_3 + \ldots + a_ix_i + \ldots + a_nx_n + \varepsilon$ ошибкой модели является:
а) ε; б) a_i; в) y; г) x_n.

13. Для эконометрической модели уравнения регрессии ошибки модели определяется как ... между фактическим значением зависимой переменной и ее расчетным значением.
а) квадрат разности;
б) разность;
в) сумма квадратов разности;
г) сумма разности квадратов.

14. Для эконометрической модели вида $y = a_0 + a_1x_1 + \varepsilon$ показателем тесноты связи между переменными y и x является парный коэффициент линейной:

а) эластичности;
б) детерминации;
в) корреляции;
г) регрессии.

15. Уравнением нелинейной регрессии, линейной по параметрам, является:

а) $y = \dfrac{1}{a_0 + a_1x + \varepsilon}$;
б) $y = a_0 + a_1x + \varepsilon$;
в) $y = e^{a_0 + a_1x} \cdot \varepsilon$;
г) $y = a_0 + a_1x + a_2x^2 + a_3x^3 + \varepsilon$.

ВАРИАНТ 8

1. По своей природе результирующая переменная всегда:

а) функциональна;
б) стохастична;
в) предопределена;
г) постоянна.

2. Выберите верные утверждения по поводу экзогенных переменных:

а) предопределенные переменные;
б) число экзогенных переменных системы равно числу эндогенных переменных системы;
в) значения экзогенных переменных определяются вне модели;
г) зависимые переменные.

3. Объясненная сумма квадратов может интерпретироваться как мера:

а) влияния величины суммы квадратов отклонений на число степеней свободы;

б) разброса величины y, объясненного с помощью регрессии, относительно $\bar{y}$;

в) остаточного разброса, не объясненного уравнением регрессии;

г) общего разброса величины y относительно $\bar{y}$.

4. Для линейного уравнения множественной регрессии в качестве показателя тесноты связи результативной переменной с набором факторов используется коэффициент множественной:

а) регрессии;

б) корреляции;

в) эластичности;

г) детерминации.

5. Число степеней свободы для остаточной суммы квадратов отклонений в линейной модели множественной регрессии равно:

а) $n - 1$;

б) m;

в) $n - m - 1$.

6. При гетероскедастичности случайных отклонений оценки коэффициентов регрессии становятся:

а) смещенными;

б) нелинейными;

в) неэффективными.

7. Система эконометрических уравнений предполагает наличие:

а) нескольких зависимых и нескольких независимых признаков;

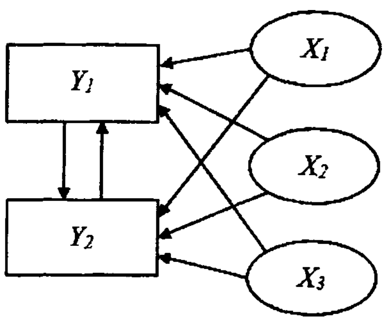
б) одного зависимого и нескольких независимых признаков;

в) нескольких зависимых и одного независимого признаков;
г) одного зависимого и совокупности независимых признаков.

8. Уравнения, в которых эндогенные переменные выражены только через экзогенные или предопределенные, называются:
а) приведенными;
б) структурными;
в) идентифицируемыми;
г) оцененными.

9. Средний уровень интервального ряда динамики с неравными временными промежутками исчисляется по формуле средней:
а) хронологической простой;
б) гармонической взвешенной;
в) арифметической взвешенной;
г) гармонической простой;
д) хронологической взвешенной;
е) арифметической простой.

10. Для указанной схемы взаимосвязей между переменными справедливы утверждения:



а) включает 2 уравнения;
б) может быть описана с помощью системы независимых уравнений;

в) включает 5 уравнений;
г) может быть описана с помощью системы одновременных уравнений.

11. Система эконометрических уравнений может быть использована для:
а) линеаризации моделируемого экономического процесса или явления;
б) упрощения вида моделируемой связи;
в) описания взаимосвязей между совокупностью зависимых и независимых переменных.

12. Для учета влияния на результативную переменную признаков качественного характера используются фиктивные переменные, при этом фиктивной переменной может присваиваться значение:
а) -1; б) 0; в) 1; г) 0,1.

13. Автокорреляционной функцией временного ряда называется последовательность коэффициентов автокорреляции:
а) первого, второго, третьего и последующих порядков;
б) между несколькими временными рядами;
в) между трендовой, сезонной и случайной компонентами;
г) факторов, формирующих уровень ряда.

14. При идентификации модели множественной регрессии $y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + \varepsilon$ количество оцениваемых параметров равно:
а) 6; б) 3; в) 4; г) 5.

15. В эконометрической модели линейного уравнения множественной регрессии $y = a_0 + a_1x_1 + a_2x_2 + ... + a_ix_i + ... + a_nx_n + \varepsilon$ величина параметра a характеризует среднее по совокупности значение зависимой переменной, при значениях ..., равных 0.
а) ε; б) a_i; в) y; г) x_j.

ВАРИАНТ 9

1. Укажите уравнения регрессии, в которых фиктивная переменная D используется только в мультипликативной форме:

а) $y_x = a_0 + a_1 D + a_2 Dx$;

б) $y_x = a_0 + a_1 x^2 + a_2 D$;

в) $y_x = a_0 + a_1 Dx$;

г) $y_x = a_0 + a_1 x + a_2 Dx$.

2. Выбор формы зависимости экономических показателей и определение количества факторов в модели называется ... эконометрической модели.

а) линеаризацией;

б) апробацией;

в) идентификацией;

г) спецификацией.

3. Общая сумма квадратов отклонений подсчитывается на основе отклонений:

а) расчетных значений результирующего признака, найденных по уравнению регрессии, от среднего значения результирующего признака;

б) индивидуальных значений результирующего признака от его среднего значения;

в) индивидуальных значений результирующего признака от расчетных значений результирующего признака, найденных по уравнению регрессии;

г) расчетных значений результирующего признака, найденных по уравнению регрессии, от нуля.

4. Коэффициент множественной корреляции используется для исследования силы связи между:

а) одной зависимой переменной и несколькими независимыми факторами;

б) несколькими зависимыми переменными и одной независимой переменной;

в) несколькими зависимыми переменными и несколькими независимыми факторами;
г) одной зависимой переменной и одной независимой переменной.

5. Число степеней свободы для регрессионной суммы квадратов отклонений в линейной модели множественной регрессии равно:
а) $n - m - 1$;
б) $n - 1$;
в) m.

6. При гетероскедастичности вероятнее всего, что t-статистики коэффициентов регрессии и F-статистика будут:
а) заниженные;
б) точные;
в) завышенные.

7. $\bar{Y}$ — вектор эндогенных переменных, B — матрица коэффициентов при эндогенных переменных, R — матрица коэффициентов при предопределенных переменных; XY_L — вектор предопределенных переменных; E — вектор случайных отклонений. Общий вид системы одновременных уравнений представляется в форме:
а) $B \cdot \bar{Y} + R \cdot XY_L = E$;
б) $R \cdot XY_L = E$;
в) $B \cdot \bar{Y} = E$;
г) $B \cdot \bar{Y} + R \cdot XY_L = 0$.

8. Приведена последовательность операций:
1) заданная система одновременных уравнений из структурной формы преобразуется в приведенную форму;
2) оценки параметров приведенной формы находятся традиционным методом наименьших квадратов;
3) по оценкам параметров приведенной формы вычисляются оценки структурных параметров.

Этот алгоритм соответствует методу наименьших квадратов:

а) трехшаговому;
б) двухшаговому;
в) косвенному;
г) обобщенному.

9. Средний уровень моментного ряда динамики с равными временными промежутками исчисляется по формуле средней:

а) арифметической простой;
б) арифметической взвешенной;
в) гармонической простой;
г) гармонической взвешенной;
д) хронологической простой;
е) хронологической взвешенной.

10. Автокорреляцией уровней ряда называется корреляционная зависимость между:

а) последовательными уровнями ряда;
б) последовательными коэффициентами корреляции;
в) уровнями ряда и соответствующими им значениями ошибок.

11. Из несмещенности оценки параметра следует, что среднее значение остатков равно:

а) 0; б) –1; в) 8; г) 1.

12. Если параметр эконометрической модели *не является* статистически значимым, то соответствующая независимая переменная:

а) не оказывает влияния на результативную переменную;
б) тесно связана с зависимой переменной;
в) оказывает статистически значимое влияние на результативную переменную;
г) оказывает доминирующее влияние на зависимую переменную.

13. Переменная x является нелинейной в уравнении:

а) $y=a_0+a_1\cdot x+\varepsilon;$

б) $y=a_0+a_1\cdot x+a_2^2\cdot x+\varepsilon;$

d) $y=a_0+a_1\cdot x+a_2\cdot x^2+\varepsilon;$

г) $y=a_0+a_1\cdot x+a_2\cdot x+\varepsilon.$

14. Для приведенной формы модели системы одновременных уравнений вида

$$\begin{cases} y_1=\lambda_{11}x_1+\lambda_{12}x_2+\lambda_{13}x_3+\lambda_{14}x_4+\varepsilon_1, \\ y_2=\lambda_{21}x_1+\lambda_{22}x_2+\lambda_{23}x_3+\lambda_{24}x_4+\varepsilon_2, \\ y_3=\lambda_{31}x_1+\lambda_{32}x_2+\lambda_{33}x_3++\lambda_{34}x_4+\varepsilon_3 \end{cases}$$

λ_{ij} являются:

а) экзогенными переменными;

б) приведенными коэффициентами;

в) эндогенными переменными;

г) структурными переменными.

15. Примерами фиктивных переменных в эконометрической модели зависимости стоимости 1 м^2 жилья *не является*:

а) величина прожиточного минимума в регионе;

б) принадлежность тому или иному региону;

в) прощадь жилья (м^2);

г) категория жилья: первичное жилье / вторичное жилье.

ВАРИАНТ 10

1. Укажите правильный вариант ответа относительно числа зависимых переменных, включаемых в уравнение регрессии:

а) несколько переменных;

б) только одна переменная;

в) количество зависимых переменных равно количеству независимых;

г) в парной регрессии одна зависимая переменная, во множественной — несколько зависимых переменных.

2. Остаток регрессионной модели представляет собой оценку:
   а) свободного члена;
   б) факторной переменной;
   в) коэффициента регрессии;
   г) случайной ошибки.

3. Остаточная сумма квадратов отклонений в парной регрессии имеет число степеней свободы, равное:
   а) $n - 1$;
   б) $n - 2$;
   в) 1.

4. Квадрат множественного коэффициента линейной корреляции зависимости результативной переменной y от набора факторов x_1, x_2, x_3 оценивает:
   а) долю дисперсии x_1, объясненной линейной регрессией y по x_1, x_2, x_3;
   б) долю дисперсии y, объясненной линейной регрессией y по x_2 при фиксированных значениях x_1, x_3;
   в) долю дисперсии y, объясненной линейной регрессией y по x_1 при фиксированных значениях x_2, x_3;
   г) долю дисперсии y, объясненной линейной регрессией y по x_1, x_2, x_3.

5. Известно, что при фиксированном значении переменной x_2 между переменными y и x_1 существует положительная связь. Значение частного коэффициента корреляции $R_{yx_1 \cdot x_2}$:
   а) 0;
   б) –0,8;
   в) 0,4;
   г) 1,3.

6. Дана следующая система из трех уравнений:

$$\begin{cases} Y_{1t} = a_0 + a_1 X_t + u_{1t}, \\ Y_{2t} = b_0 + b_1 Y_{1t} + b_2 X_t + u_{2t}, \\ Y_{3t} = c_0 + c_1 Y_{2t} + c_{2t} X_t + u_{3t}. \end{cases}$$

Обычный МНК может быть использован для оценки уравнений:

а) только первого уравнения;
б) только для второго и третьего;
в) для каждого уравнения;
г) не может вообще.

7. Система эконометрических уравнений в приведенной форме является системой уравнений:

а) рекурсивных;
б) нормальных;
в) независимых;
г) одновременных.

8. Приведена последовательность операций:

1) заданная система одновременных уравнений из структурной формы преобразуется в приведенную форму;
2) оценки параметров приведенной формы находятся традиционным методом наименьших квадратов;
3) определение расчетных значений эндогенных переменных, которые выступают в качестве факторов в структурной форме модели;
4) определение структурных параметров каждого уравнения в отдельности традиционным методом наименьших квадратов, используя в качестве факторов входящие в это уравнение предопределенные переменные и расчетные значения эндогенных переменных, полученные на первом шаге.

Этот алгоритм соответствует методу наименьших квадратов:

а) трехшаговому;
б) двухшаговому;
в) косвенному;
г) обобщенному.

9. Средний уровень моментного ряда динамики с неравными временными промежутками исчисляется по формуле средней:

а) хронологической простой;
б) гармонической взвешенной;
в) арифметической взвешенной;
г) гармонической простой;
д) хронологической взвешенной;
е) арифметической простой.

10. Автокорреляционной функцией временного ряда называется последовательность:

а) коэффициентов автокорреляции первого, второго, третьего и последующих порядков;
б) уровней динамического ряда;
в) значений результативной переменной, изображенных на поле корреляции.

11. Значение коэффициента автокорреляции первого порядка характеризует:

а) значимость тренда;
б) тесноту нелинейной связи;
в) качество модели временного ряда;
г) тесноту линейной связи.

12. Для регрессионной модели состоятельность оценки параметра означает, что при увеличении выборки значение оценки параметра стремится к:

а) свободному члену уравнения регрессии;

б) коэффициенту парной корреляции между зависимой переменной и соответствующей независимой переменной;
в) истинному значению параметра, вычисленному для генеральной совокупности;
г) оцениваемому параметру, рассчитанному по другой выборке, объем которой значительно меньше исходной совокупности данных.

13. Для регрессионной модели вида $y = a_0 + a_1x_1 + a_2x_2 + ... + a_ix_i + ... + a_nx_n + \varepsilon$ показателем тесноты связи является:
а) коэффициент множественной корреляции;
б) F-критерий Фишера;
в) коэффициент автокорреляции;
г) парный коэффициент корреляции.

14. Если известно уравнение множественной регрессии $y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \varepsilon$, построенное по результатам 50 наблюдений, для которой общая сумма квадратов отклонений равна 153, и остаточная сумма квадратов отклонений равна 3, то значение F-статистики равно:
а) 766,67;
б) 877,45;
в) 46;
г) 50.

15. Для мультипликативной модели временного ряда $Y = T \cdot S \cdot E$ сумма скорректированных сезонных компонент равна:
а) лагу;
б) 1;
в) половине лага;
г) 0.

ГЛОССАРИЙ

β-коэффициент линий регрессии — характеризует наиболее крупные резервы улучшения изучаемого признака.

Гармонический анализ — нахождение конечной суммы уровней с использованием функций косинусов и синусов времени.

Гомоскедастичность — дисперсия каждого отклонения ε одинакова для всех значений x.

Задачей эконометрики является оценка направленных действий на достижение и повышение экономической эффективности, кроме того, задачей эконометрики является прогнозирование путей развития макро- и микроэкономических фактов.

Индекс множественной корреляции — характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком, оценивает тесноту совместного влияния факторов на результат.

Индекс сезонности — это процентное соотношение фактических внутригодовых уровней и постоянной или переменной средней.

Индекс частной корреляции — характеризует тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии.

Интерполяция — применяется на этапе предварительной обработки данных и предполагает определение значений уровней ряда внутри заданного интервала.

Коллинеарными называются две переменные, которые находятся между собой в линейной зависимости.

Коэффициент детерминации — характеризует долю дисперсии результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака.

Коэффициент эластичности — показывает, на сколько процентов изменится в среднем результат, если фактор изменится на 1%.

Метод наименьших квадратов — согласно сумма отклонений фактических значений результативных показателей от теоретических, найденных по уравнению связи, должна быть минимальной.

Мультиколлинеарность — когда более чем два фактора связаны между собой линейной зависимостью.

Нелинейная регрессия внутренне линейная, т. е. она с помощью соответствующих преобразований может быть приведена к линейному виду.

Нелинейная регрессия внутренне нелинейная, т. е. она не может быть сведена к линейной функции.

Несмещенность оценки — означает, что математическое ожидание остатков равно нулю.

Перспективная экстраполяция — предполагает продолжение ряда динамики на будущее на основе выявления закономерности изменений уровней ряда в изучаемый период времени.

Предмет эконометрики — факты, формирующие развитие экономических процессов и явлений.

Производственная функция — характеризует связь между производственными факторами и величиной продукта.

Ретроспективная экстраполяция — продолжение уровней временного ряда в прошлое.

Ряды динамики (временные, хронологические ряды) — упорядоченные статистические данные по времени их получения.

Сезонные колебания — это колебания, периодически повторяющиеся в некоторое определенное время каждого года, месяца, дня или его часа.

Состоятельность оценок — характеризует увеличение их точности с увеличением объема выборки.

Стохастические модели — допускают наличие случайных воздействий на исследуемые показатели. Такой вид зависимости называется ***корреляционным***.

Тенденция автокорреляции — характеризует изменения связи между отдельными уровнями ряда динамики.

Тенденция дисперсии — тенденция изменения отклонений между эмпирическими уровнями и детерминированной компонентой ряда.

Тренд — это длительная тенденция изменения случайного процесса, определяющая основную тенденцию изменения экономических показателей.

Циклические (или периодические) ***колебания*** состоят в том, что значение изучаемого признака в течение какого-то времени возрастает, достигая определенного максимума, затем понижается, достигая определенного минимума, вновь возрастает до прежнего значения и т.д.

Цифровые метки, т. е. качественные переменные, преобразованные в количественные. Такого рода переменные в эконометрике принято называть ***фиктивными переменными***.

Частные уравнения регрессии — уравнения регрессии, которые связывают результативный признак с соответствующими факторами x при закреплении других учитываемых во множественной регрессии факторов на среднем уровне.

Экзогенные переменные — это внешние по отношению к модели переменные, их значения определяются вне модели и поэтому они считаются фиксированными, обозначаются обычно как x.

Эконометрика — это наука, которая дает количественное выражение взаимосвязей и взаимозависимостей социально-экономических явлений и процессов.

Эконометрической моделью называется совокупность уравнений, описывающих связи между некоторыми экономическими показателями. Соотношения могут быть **стохастическими** (случайными) и **детерминированными** (зависящими от чего-либо).

Экстраполяция — метод научного исследования, заключающийся в распространении выводов, полученных из наблюдений над одной частью явления, на другую ее часть.

Эндогенные переменные — это переменные, значения которых определяются внутри модели и обозначаются обычно как y.

Эффективными считаются оценки, если они характеризуются наименьшей дисперсией.

СТАТИСТИКО-МАТЕМАТИЧЕСКИЕ ТАБЛИЦЫ

1. Значение *F*-критерия Фишера при уровне значимости 0,05

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	∞
1	161	200	216	225	230	234	237	239	241	242	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,50
3	10,13	9,55	9,28	9,19	9,01	8,94	8,88	8,84	8,81	8,78	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	1,73
25	4,24	3,88	2,99	2,76	2,60	2,49	2,4.1	2,34	2,26	2,24	1,71

k_1 / k_2	1	2	3	4	5	6	7	8	9	10	∞
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,20	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,18	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,99	1,39
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,28
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,83	1,00

2. Значение *F*-критерия Фишера при уровне значимости 0,01

k_1 / k_2	1	2	3	4	5	6	7	8	9	10	∞
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6366
2	98,5	99,0	99,16	99,25	99,3	99,33	99,36	99,38	99,39	99.4	99.5
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	26,13
4	21,2	18,0	16,69	15,98	15,52	15,21	14,98	14,8	14,66	14,55	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,1	7,98	7,87	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	4,86
9	10,56	8,02	6,99	6,42	6,06	5,8	5,61	5,47	5,35	5,26	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,2	5,06	4,94	4,85	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	3,6
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,5	4,39	4,3	3,36
13	9,07	6,7	5,74	5,21	4,86	4,62	4,44	4,3	4,19	4,1	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,0
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,0	3,89	3,8	2,87
16	8,53	6,23	5,29	4,77	4,44	4,2	4,03	3,89	3,78	3,69	2,75
17	8,4	6,11	5,19	4,67	4,34	4,1	3,93	3,79	3,68	3,59	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,6	3,51	2,57
19	8,18	5,93	5,01	4,5	4,17	3,94	3,77	3,63	3,52	3,43	2,49
20	8,1	5,85	4,94	4,43	4,1	3,87	3,7	3,56	3,46	3,37	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,4	3,31	2,36

5. Критические значения коэффициентов корреляции для уровней значимости 0,05, 0,01

d.f.	α=0,05	α=0,01	*d.f.*	α=0,05	α=0,01
1	0,996917	0,9998766	17	0,4555	0,5751
2	0,995000	0,990000	18	0,4438	0,5614
3	0,8783	0,95873	19	0,4329	0,5487
4	0,8114	0,91720	20	0,4227	0,5368
5	0,7545	0,8745	25	0,3809	0,4869
6	0,07067	0,8343	30	0,3494	0,4487
7	0,6664	0,7977	35	0,3246	0,4182
8	0,6319	0,7646	40	0,3044	0,3932
9	0,6021	0,7348	45	0,2875	0,3721
10	0,5760	0,7079	50	0,2732	0,3541
11	0,5529	0,6835	60	0,2500	0,3248
12	0,5324	0,6614	70	0,2919	0,3017
13	0,5139	0,6411	80	0,2172	0,2830
14	0,4973	0,6226	90	0,2050	0,2673
15	0,4821	0,6055	100	0,1946	0,2540
16	0,4683	0,5897			

6. Значения статистик Дарбина–Уотсона *dLdU* при 5%-м уровне значимости

n	$k'=1$		$k'=2$		$k'=3$		$K'=4$		$k'=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40	–	–	–	–	–	–	–	–
7	0,70	1,36	0,47	1,90	–	–	–	–	–	–
8	0,76	1,33	0,56	1,78	0,37	2,29	–	–	–	–
9	0,82	1,32	0,63	1,70	0,46	2,13	–	–	–	–
10	0,88	1,32	0,70	1,64	0,53	2,02	–	–	–	–
11	0,93	1,32	0,66	1,60	0,60	1,93	–	–	–	–
12	0,97	1,33	0,81	1,58	0,66	1,86	–	–	–	–
13	1,01	1,34	0,86	1,56	0,72	1,82	–	–	–	–
14	1,05	1,35	0,91	1,55	0,77	1,78	–	–	–	–
15	1,08	1,36	0,95	1,52	0,82	1,75	0,69	1,97	0,59	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99

n	$k^1=1$		$k^1=2$		$k^1=3$		$K^1=4$		$k^1=5$	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83

СПИСОК ЛИТЕРАТУРЫ

1. Введение в эконометрическое моделирование / А. Клас [и др.]; пер. со словац. Л.А. Клименко; под ред. Е.М. Четыркина. — М.: Статистика, 1978. — 152 с.

2. *Герасимов А. Н.* Эконометрика: теория и практика [Электронный ресурс] : электрон. учеб. [для студентов вузов по эконом. специальностям] / А. Н. Герасимов, А. В. Гладилин, Е. И. Громов. — Электрон. дан. и прогр. (695 МБ). — М. : КНОРУС, 2011.

3. *Елисеева И.И.* Практикум по эконометрике : учебное пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко; под ред. И.И. Елисеева. — М.: Финансы и статистика, 2006. — 345 с.

4. *Емельянов А.С.* Эконометрика и прогнозирование / А.С. Емельянов — М.: Экономика, 1985. — 207 с.

5. *Маленво Э.* Статистические методы эконометрии / Э. Маленво — М.: Статистика, 1975, вып. 1; 1976, вып. 2

6. *Мордэкэй Елекиэл.* Методы анализа корреляции и регрессии / Елекиэл Мордэкэй, Карл А. Фокс; пер. с англ. А.С. Кучаева; под ред. Н.К. Дружинина. — М.: Статистика, 1966. — 560 с.

7. *Новиков А.И.* Эконометрика: учеб. пособие для бакалавров / А.И. Новиков. — М.: Дашков и К°, 2013. — 224 с.

8. *Путко Б.А.* Эконометрика: учебник / Б.А. Путко, Н.Ш. Кремер; под ред. Н.Ш. Кремер. — 3-е изд., перераб. и доп. — М.: Юнити-Дана, 2012. — 329 с.

9. *Тихомиров Н.П.* Эконометрика: учебник / Н.П. Тихомиров, Е.Ю. Дорохина. — М.: Экзамен, 2003. — 512 с.

10. *Уотшем Т. Дж.* Количественные методы в финансах: учебное пособие для студентов. / Т. Дж. Уотшем, К. Паррамоу. — М.: Финансы, Изд. объединен. ЮНИТИ, 1999. — 528 с.

11. Эконометрика: учебник / К.В. Балдин, В.Н. Башлыков, В.А. Брызгалов и др.; под ред. В.Б. Уткина. — 2-е изд. — М.: Дашков и К°, 2013. — 562 с.

12. Эконометрика: учебник / И. И. Елисеева, С. В. Курышева, Т. В. Костеева и др.; под ред. И. И. Елисеевой. — 2-е изд., перераб. и доп. — М.: Финансы и статистика, 2007. — 576 с.

Учебное издание

Алексей Николаевич ГЕРАСИМОВ,
Евгений Иванович ГРОМОВ,
Юрий Сергеевич СКРИПНИЧЕНКО

ЭКОНОМЕТРИКА

Ответственный за выпуск *Д. Хабихужин*
Выпускающий редактор *Г. Логвинова*
Технический редактор *Ю. Давыдова*
Верстка: *А. Патулова*

Формат 84 x 108 1/32. Бумага офсетная.
Тираж 1500 экз. Заказ № 220.

ООО «Феникс»
344011, Россия, Ростовская область,
г. Ростов-на-Дону, ул. Варфоломеева, 150.
Тел. (863) 261-89-59, тел./факс 261-89-50
Сайт издательства: www.phoenixrostov.ru
Интернет-магазин: www.phoenixbooks.ru

Изготовлено в России. Дата изготовления: 04.2017.

Изготовитель: АО «Книга»
344019, Россия, Ростовская обл.,
г. Ростов-на-Дону, ул. Советская, 57/1.