**МИНИСТЕРСТВО ВЫСШЕГО ОБРАЗОВАНИЯ, НАУКИ
И ИННОВАЦИЙ РЕСПУБЛИКИ УЗБЕКИСТАН**

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
“ТАШКЕНТСКИЙ ИНСТИТУТ ИНЖЕНЕРОВ ИРРИГАЦИИ
И МЕХАНИЗАЦИИ СЕЛЬСКОГО ХОЗЯЙСТВА“**

Г. Шодманова, С.С. Мирзаев

ПРИКЛАДНАЯ ЭКОНОМЕТРИКА

/ Учебное пособие /

**Ташкент
2023**

Учебное пособие утверждено и рекомендовано к публикации в соответствии с приказом ректора НИУ “ТИИИМСХ“ за № 418 a/f от 15 ноября 2022 года.

Регистрационный номер: 418 a/f-080

***

Учебное пособие «Прикладная эконометрика» предназначено для подготовки бакалавров и магистров, переподготовки специалистов в области экономики, финансов и менеджмента. Изучаются различные методы построения и настройки экономико-математических моделей на основе статистических данных с использованием компьютеров. Рассмотрены линейные и нелинейные модели парной и множественной регрессии. Учебное пособие предназначено для студентов следующих направлений бакалавриата: 60310100 – Экономика (в водном хозяйстве), 60411200 – Менеджмент (в водном хозяйстве), 60410100 – Бухгалтерский учет и аудит (в водном хозяйстве) и 60412300 – Организация и управление водным хозяйством.

Авторы: к.э.н., профессор **Г. Шодманова,**
к.т.н., доцент **С.С. Мирзаев**

Рецензенты: к.ф.-м.н., доц. **Ш. Насретдинова**, зав. каф. Национального университета Узбекистана

д.т.н., проф. **Б.А. Худояров**, НИУ «Ташкентский институт инженеров ирригации и механизации сельского хозяйства»

***

Ушбу «Амалий эконометрика» номли ўқув қўлланма иқтисодиёт, молия ва менежмент соҳасида таълим олаётган бакалавр, магистр ҳамда мутахассисларни қайта тайерлаш учун мўлжалланган. Компьютердан фойдаланган ҳолда статистик маълумотлар асосида иқтисодий-математик моделларни тузиш ва мослаштиришнинг турли усуллари ўрганилади. Бир ўзгарувчили ва кўп ўзгарувчили чизиқли ва чизиқли бўлмаган регрессия моделлари қаралган. Ўқув қўлланма бакалавриатнинг 60310100 – Иқтисодиёт (сув хўжалигида) ва 60411200 – Менежмент (сув хўжалигида), 60410100 – Бухгалтерия ҳисоби ва аудит (сув хўжалигида) ва 60412300 – Сув хўжалигини ташкил этиш ва бошқариш таълим йўналишлари талабалари учун мўлжалланган.

***

The manual "Applied Econometrics" is intended for the training of bachelors and masters, retraining of experts in of economics, finance and management. Various methods of construction and estimation of economic-mathematical models using statistical data with use of computers are studied. Considered linear and nonlinear models of pair and multiple regressions. The manual is intended for students of the following directions of a bachelor degree: 60310100 – Economics (in a water management), 60411200 – Management (in a water management), 60410100 – Accounting and audit (in a water management) ва 60412300 – Water management and organization.

Г. Шодманова, С.С. Мирзаев
/ ПРИКЛАДНАЯ ЭКОНОМЕТРИКА /
Учебное пособие, Ташкент – 2023 г., 177 стр.

Введение

В современном мире качество образования уже стало одним из важнейших факторов конкурентоспособности экономики. Именно поэтому в принятой в начале нового этапа реформ в Узбекистане в январе 2022 года Стратегии развития Нового Узбекистана на 2022 — 2026 годы одним из приоритетов является «продолжение курса дальнейшего совершенствования системы непрерывного образования, повышения доступности качественных образовательных услуг, подготовки высококвалифицированных кадров в соответствии с современными потребностями рынка труда». Вопросы развития повышения конкурентоспособности образования страны на национальном и международном рынках труда также вошли в Концепцию комплексного социально-экономического развития Республики Узбекистан до 2030 года.

Реформы в профессиональном образовании проводятся на основании Указа Президента «О дополнительных мерах усовершенствования системы профессионального образования», принятого 6 сентября 2019 года[12].

Эконометрика – динамично развивающаяся дисциплина, направленная на количественное определение экономических отношений. Эконометрика как наука развивается более 100 лет, входит в список обязательных дисциплин для подготовки специалистов экономического и финансового профиля.

Квалифицированные специалисты экономического и финансового профиля должны быть способны построить эконометрическую модель процесса, получить решение и анализировать решения модели. Для того, чтобы принимать компетентные решения по вопросам экономики и управления капиталами, необходимо анализировать большое количество информации. Благодаря современным информационным технологиям можно собирать, хранить и получать постоянный доступ к информации. Но как обработать, проанализировать необходимую информацию и принять решение? С этой целью были разработаны и усовершенствованы методики и программные продукты, обеспечивающие:

- *Представление данных в наглядном виде.* Графики и диаграммы являются наглядной и удобной для отображения большого количества данных. К тому же, они дают возможность применять анимацию. Человеческий глаз наделен аналитическими способностями, он может улавливать закономерности и отклонения в диаграммах и иллюстрациях, и это используется при мониторинге на электростанциях и в промышленности, в космосе, в военном деле и т.д.

- *Статистическая обработка данных:* расчет средних значений, дисперсий, корреляций, законов распределения, то есть представление данных в обобщенном виде.

- *Построение экономико-математических моделей и их настройка,* оценивание параметров с помощью методов эконометрики.

- *Построение на основе модели оптимальных планов и управленческих решений.*

Основная цель изучения эконометрики - увидеть за набором цифр модель, построить ее, настроить и оценить ее применимость для принятия управленческих решений.

При построении модели не обязательно описывать многочисленные связи, присущих моделируемой системе, так как могут быть погрешности в количественной специфике всех связей и зависимостей в изучаемой системе; кроме того, это может так усложнить и перегрузить модель, что решение окажется невозможным. Поэтому математическое моделирование предполагает абстрагирование, отвлечение от несущественных сторон моделируемого объекта и, следовательно, описание наиболее характерных закономерных его черт [2].

Задачи решаются на компьютерах, но построение модели и анализ выдаваемых компьютером результатов остается за человеком.

Учебное пособие имеет практическую направленность и рассчитан на будущих специалистов экономическо-финансового направления, для обучения бакалавров и магистров, а также на переподготовку специалистов.

За основу взят курс эконометрики профессора К. Доугерти, Лондонская школа экономики [1], а также использованы учебники профессора Н.В.Катаргина [2], профессора В.А. Бывшева [3], профессора В.Е.Гмурмана [4], профессора Л.О. Бабешко [5] и член-корр. РАН И.И. Елисеевой [6], профессора И.А. Кацко [7], профессора О.О.Замкова и др. [8], профессора Г. Шадмановой, доцента С.С. Мирзаева [9], А.И.Новикова [10].

Во многих разделах дается общее представление о построении экономико-математической модели, требующее знания и навыки хорошей математической подготовки и много времени. На достаточном уровне рассмотрены теоретические основы, которые обязательно надо знать, чтобы грамотно формировать и оценивать модели. Это учебное пособие способствует решению практических задач с использованием программных продуктов, которые, не требуют доказывания теорем и образования сложных формул, и эти все процессы отражены в программном обеспечении.

В качестве сервисов и функций электронных таблиц MS Excel, программирование сложных алгоритмов и других прикладных программ (в эконометрике – Stata, EViews и др.) позволило их использовать широкому кругу научных сотрудников и практиков, от которых требуется: грамотная постановка задачи; подбор и оценка исходных данных; построение структурной модели: системы уравнений, тождеств и неравенств, и алгоритма решения задачи, оформление результатов работы.

Глава 1. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ В ЭКОНОМИКЕ

1.1 Моделирование в экономике

Эконометрика и эконометрические модели широко внедряются во всех отраслях знаний, в частности, в такой сложной области знаний как экономика. Создание моделей экономического процесса и исследования, проводимые на построенной эконометрической модели представляет собой сущность эконометрического моделирования экономического процесса. Таким образом, эконометрическое моделирование представляет собой мощный инструмент не только научного познания, но и практического применения полученных результатов в хозяйственной деятельности. С другой стороны широкое использование информационных технологий позволяет упростить процесс эконометрического моделирования в решении задач экономики.

Математические модели использовались с иллюстративными и исследовательскими целями еще Ф.Кенэ (1758г. «Экономическая таблица»), А.Смитом (классическая макроэкономическая модель), Д.Рикардо (модель международной торговли). В XIX веке большой вклад в моделирование рыночной экономики внесла математическая школа (Л.Вальрас, О.Курно, В.Парето, Ф.Эджворт и др.). В XX веке математические методы моделирования применялись очень широко, с их использованием связаны практически все работы, удостоенные Нобелевской премии по экономике (Д.Хикс, Р.Солоу, В.Леонтьев, П.Самуэльсон и др.).

Развитие микроэкономики, макроэкономики, прикладных дисциплин связано с все более высоким уровнем их формализации. Основу для этого заложил прогресс в области прикладной математики – теории игр, математического программирования, математической статистики. В России в начале XX века большой вклад в математическое моделирование экономики внесли В.К.Дмитриев и Е.Е.Слуцкий. В 1960-е – 80-е годы экономико-математическое направление возродилось (В.С.Немчинов, В.В.Новожилов, Л.В.Канторович), но было связано в основном с попытками формально описать «систему оптимального функционирования социалистической экономики» (СОФЭ, Н.П.Федоренко, С.С.Шаталин и др.). Строились многоуровневые системы моделей народнохозяйственного планирования, оптимизационные модели отраслей и предприятий. Сейчас важной задачей является моделирование прогрессов переходного периода [8].

Любое экономическое исследование всегда предполагает объединение теории и практики. Мы используем теоретические модели для описания и объяснения наблюдаемых процессов и собираем статистические данные с целью эмпирического построения и обоснования моделей.

1.2. Моделирование как метод научного познания

Моделирование в научных исследованиях стало применяться еще в глубокой древности и постепенно захватывало все новые области научных знаний: техническое конструирование, строительство и архитектуру, астрономию, физику, химию, биологию и, наконец, общественные науки. Большие успехи и признание практически во всех отраслях современной науки принес методу моделирования XX в. Однако методология моделирования долгое время развивалась независимо отдельными науками. Отсутствовала единая система понятий, единая терминология. Лишь постепенно стала осознаваться роль моделирования как универсального метода научного познания.

Термин «модель» широко используется в различных сферах человеческой деятельности и имеет множество смысловых значений. Рассмотрим только такие «модели», которые являются инструментами получения знаний.

Модель – это такой материальный или мысленно представляемый объект, который в процессе исследования замещает объект-оригинал так, что его непосредственное изучение дает новые знания об объекте-оригинале. Под *моделированием* понимается процесс построения, изучения и применения моделей. Оно тесно связано с такими категориями, как абстракция, аналогия, гипотеза и др. Процесс моделирования обязательно включает и построение абстракций, и умозаключения по аналогии, и конструирование научных гипотез. Главная особенность моделирования в том, что это метод опосредованного познания с помощью объектов-заместителей.

Моделирование – циклический процесс. При этом знания об исследуемом объекте расширяются и уточняются, а исходная модель постепенно совершенствуется. Недостатки, обнаруженные после первого цикла моделирования, обусловленные малым знанием объекта и/или ошибками в построении модели, можно исправить в последующих циклах [7].

1.3. Особенности применения метода математического моделирования (эконометрики) в экономике

Проникновение математики в экономическую науку связано с преодолением значительных трудностей. В этом отчасти была «повинна» математика, развивающаяся на протяжении нескольких веков в основном в связи с потребностями физики и техники. Но главные причины лежат все же в природе экономических процессов, в специфике экономической науки. Большинство объектов, изучаемых экономической наукой, может быть охарактеризовано кибернетическим понятием *сложная система.*

Наиболее распространено понимание системы как совокупности элементов, находящихся во взаимодействии и образующих некоторую целостность, единство. Одна из трудностей экономических исследований заключается в том, что почти не существует экономических объектов, которые можно было бы рассматривать как отдельные элементы.

Сложность системы определяется количеством входящих в нее элементов, связями между этими элементами, а также взаимоотношениями между системой и средой. В народном хозяйстве взаимодействуют природные, технологические, социальные процессы, объективные и субъективные факторы.

Сложность экономики иногда рассматривалась как обоснование невозможности ее моделирования, изучения средствами математики. Но такая точка зрения в принципе неверна. Моделировать можно объект любой природы и любой сложности. И как раз сложные объекты представляют наибольший интерес для моделирования; именно здесь моделирование может дать результаты, которые нельзя получить другими способами исследования.

1.4. Особенности экономических наблюдений и измерений

Уже длительное время главным тормозом практического применения математического моделирования в экономике является наполнение разработанных моделей конкретной и качественной информацией. Точность и полнота первичной информации, реальные возможности ее сбора и обработки во многом определяют выбор типов прикладных моделей. В зависимости от моделируемых объектов и назначения моделей используемая в них исходная информация имеет существенно различный характер и происхождение. Она может быть разделена на две категории: о прошлом развитии и современном состоянии объектов (экономические наблюдения и их обработка) и о будущем развитии объектов, включающую данные об ожидаемых изменениях их внутренних параметров и внешних условий (прогнозы). Вторая категория информации является результатом самостоятельных исследований, которые также могут выполняться посредством моделирования.

Методы экономических наблюдений и использования результатов этих наблюдений разрабатываются экономической статистикой. Поэтому стоит отметить только специфические проблемы экономических наблюдений, связанные с моделированием экономических процессов.

Познание количественных отношений экономических процессов и явлений опирается на экономические измерения. Точность измерений в значительной степени предопределяет и точность конечных результатов количественного анализа посредством моделирования. Поэтому необходимым условием эффективного использования математического моделирования является совершенствование экономических измерителей. В процессе моделирования возникает взаимодействие «первичных» и «вторичных» экономических измерителей.

Любая модель народного хозяйства опирается на определенную систему экономических измерителей (продукции, ресурсов, элементов и т.д.). С точки зрения «интересов» моделирования экономики в настоящее время наиболее актуальными проблемами совершенствования экономических измерителей являются: оценка результатов интеллектуальной деятельности (особенно в сфере научно-технических разработок, индустрии информатики), построение обобщающих показателей социально-экономического развития, измерение эффектов обратных связей.

1.5. Случайность и неопределенность в экономическом развитии

Для методологии планирования экономики важное значение имеет понятие неопределенности экономического развития. В исследованиях по экономическому прогнозированию и планированию различают два типа неопределенности: «истинную», обусловленную свойствами экономических процессов, и «информационную», связанную с неполнотой и неточностью имеющейся информации об этих процессах.

В развитии экономики неопределенность вызывается двумя основными причинами. Во-первых, ход планируемых и управляемых процессов, а также внешние воздействия на эти процессы не могут быть точно предсказуемы из-за действия случайных факторов и ограниченности человеческого познания в каждый момент.

Особенно характерно это для прогнозирования научно-технического прогресса, потребностей общества, экономического поведения. Во-вторых, общегосударственное планирование и управление не только не всеобъемлющи, но и не всесильны, а наличие множества самостоятельных экономических субъектов с особыми интересами не позволяет точно предвидеть результаты их взаимодействий. Неполнота и неточность информации об объективных процессах и экономическом поведении усиливают истинную неопределенность.

На первых этапах исследований по моделированию экономики применялись в основном детерминированные модели. В этих моделях все параметры предполагаются точно известными. Однако детерминированные модели неправильно понимать в механическом духе и отождествлять их с моделями, которые лишены всех «степеней выбора» (возможностей выбора) и имеют единственное допустимое решение. Классическим представителем жестко детерминированных моделей является оптимизационная модель народного хозяйства, применяемая для определения наилучшего варианта экономического развития среди множества допустимых вариантов.

В результате накопления опыта использования жестко детерминированных моделей были созданы реальные возможности успешного применения более совершенной методологии моделирования экономических процессов, учитывающих стохастику и неопределенность. Здесь можно выделить два основных направления исследований. Во-первых, усовершенствуется методика использования детерминированных моделей: проведение многовариантных расчетов и модельных экспериментов с вариацией конструкции модели и ее исходных данных; изучение устойчивости и надежности получаемых решений, выделение зоны неопределенности; включение в модель резервов, применение приемов, повышающих приспособляемость экономических решений к вероятным и непредвидимым ситуациям.

Во-вторых, получают распространение модели, непосредственно отражающее стохастику и неопределенность экономических процессов и использующее соответствующий математический аппарат: теорию вероятностей и математическую статистику, эконометрику, теорию игр и статистических решений, теорию массового обслуживания, стохастическое программирование, теорию случайных процессов и др. [7].

1.6. Последовательность разработки математических моделей

В данном разделе рассмотрены общие принципы выполнения любого проекта с учетом особенностей разработки и использования экономико-математических, в частности эконометрических моделей. Если раньше предполагалось, что перечисленные этапы выполняются последовательно, то в настоящее время считается нормальным возврат к предыдущим этапам. Например, при разработке модели может возникнуть потребность в новых данных, и вообще этот этап может предшествовать сбору данных.

1. *Постановка задачи*. Необходимо понять потребности, сформулировать цель работы, предполагаемые результаты, имеющиеся ресурсы (денежные, технические, кадровые, юридические), объем работ, который предполагается выполнить; оценить имеющиеся разработки и программное обеспечение, стоимость закупки или разработки недостающего; решить вопрос о целесообразности разработки; разработать техническое задание, календарный план, соглашение о цене.

2. *Обследование предметной области, сбор и оценка качества информации*. От качества исходной информации об объекте моделирования зависят как адекватность модели, так и достоверность результатов моделирования. Возможно, исследуемые объекты придется разбивать на группы, и в каждой будут свои закономерности. Возможно наличие ошибочных или аномальных данных (например, о продажах перед праздниками), которые надо уметь выделить и рассмотреть отдельно. Обычно они видны на графиках.

Теоретические предпосылки также являются информацией об объекте и способствуют целенаправленному сбору данных. В эконометрических исследованиях данные разбиваются на три группы: *cross-sectional data*, в русскоязычных учебниках обозначаются как “*пространственные*”; *временные ряды (time series); panel data,* в русскоязычных учебниках *“панельные”*, содержащие набор временных рядов.

3. *Построение концептуальной модели*: В зависимости от характера изучаемых процессов в системе все виды моделирования могут быть разделены на детерминированные и стохастические, статические и динамические, дискретные, непрерывные и дискретно- непрерывные.

Детерминированное моделирование отображает детерминированные процессы, т.е. процессы, в которых предполагается отсутствие всяких случайных воздействий; *стохастическое моделирование* отображает вероятностные процессы и события; эконометрическое моделирование относится к этому виду. В этом случае анализируется ряд реализаций случайного процесса и оцениваются средние характеристики, т.е. набор однородных реализаций. *Статическое моделирование* служит для описания поведения объекта в какой-либо момент времени, в эконометрике такие модели называют пространственными. *Динамическое моделирование* отражает поведение объекта во времени. В эконометрике изучают временные ряды и их наборы (панельные данные). *Дискретное моделирование* служит для описания процессов, которые предполагаются дискретными, соответственно непрерывное моделирование позволяет отразить непрерывные процессы в системах, а *дискретно-непрерывное моделирование* используется для тех случаев, когда хотят выделить наличие как дискретных, так и непрерывных процессов. Эконометрика базируется на дискретных данных, но результатом является непрерывная функция. На этапе концептуальной модели можно использовать неформальное описание реального объекта (тексты, рисунки, схемы, диаграммы и т.д.), но используются и формализованные технологии: системный анализ, Универсальный Язык Моделирования UML и др.

Основные этапы построения концептуальной модели:

– выдвижение гипотез и предложений;

– определение параметров и переменных модели;

– обоснование выбора показателей и критериев эффективности системы;

– составление содержательного описания модели.

Выбранные показатели и критерии эффективности системы должны отражать цель функционирования системы и представлять собой функции переменных и параметров системы. Основные виды переменных в эконометрике:

- *эндогенные*, или зависимые переменные, прогнозирование которых является одной из основных задач эконометрики;
- *экзогенные,* или влияющие переменные; могут быть внешними по отношению к системе (курс доллара, учетная ставка, время), или мы можем ими управлять: расходы на разные цели;

- *лаговые:* переменные прошедших временных интервалов; вчера мы пытались их прогнозировать, а сегодня знаем.

Экзогенные и лаговые объединяют термином *предопределенные*. Кроме того, существуют фиктивные, замещающие, инструментальные переменные, которые мы обсудим далее.

Разработка концептуальной модели завершается составлением содержательного описания, которое может содержать тексты, формулы, графики, рисунки, таблицы, описывающие систему или процесс, и используется как основной документ для дальнейших действий.

4. *Формальное описание* задач, построение *структурной модели*: системы уравнений, тождеств, ограничений-равенств и ограничений-неравенств.

5. *Разработка алгоритма решения задачи*. Алгоритм – это конечная последовательность точно определенных действий, однозначно определяющая процесс преобразования исходных и промежуточных данных, приводящий к решению задачи. В эконометрике – это преобразование структурной модели к приведенной форме: уравнению или системе равенств, в которых эндогенные (прогнозируемые) переменные будут в левой части, а экзогенные и лаговые – в правой. Этот этап требует большого количества вычислений. В настоящее время многие программы для решения таких задач оформлены в виде сервисов различных прикладных пакетов: MS Excel,

MathCad, MatLab, Stata, EViews и др., решение задачи обычно заменяется *выбором пакета, сервиса, его настройкой и стыковкой с используемыми данными*, обычно в интерактивном графическом режиме: ввод формул, установка ограничений и т.д. Правильный выбор метода решения, программного обеспечения (и программиста в реальном проекте) могут уменьшить время и стоимость проекта в десятки раз! Личный опыт автора это подтверждает, и предлагаемые далее технологии решения задач позволяют это сделать.

6. *Тестирование*. Программа, не имеющая синтаксических ошибок, может иметь логические ошибки и выдавать неверные результаты. Поэтому как отдельные блоки, так и программа в целом должны быть проверены с помощью тестовых задач с известными решениями. 90% эконометрики – это методы оценки надежности модели в целом и ее параметров.

7. *Показатели качества эконометрической модели: коэффициент детерминации R^2, статистика Фишера F, t-статистика Стьюдента для коэффициентов уравнений, тест Дарбина-Уотсона на автокорреляцию DW, тест Голдфелда-Квандта на гетероскедастичность GQ, выявление мультиколлинеарности по матрице корреляции экзогенных переменных, а также оценка погрешности прогноза и проверка адекватности модели*.

Оформление и интерпретация результатов моделирования имеет целью переход от информации, полученной в результате машинного эксперимента с моделью, к выводам, касающимся процесса функционирования объекта-оригинала. Результаты моделирования могут быть представлены в виде таблиц, графиков, диаграмм, схем и т.п. В большинстве случаев наиболее простой формой считаются таблицы, хотя *графики более наглядно иллюстрируют результаты моделирования системы*.

Опорные слова

Модель, моделирование, построение модели, объект, система, объект-оригинал, модели экономики, информация, переменные, экзогенные, эндогенные, коэффициенты, параметры, статическая, динамическая модель, детерминированные математические модели, экономико-математические модели, оптимизация, случайность, стохастические модели, этапы построения модели, экономические процессы, математический анализ, адекватность модели, классификация модели.

Контрольные вопросы

1. Почему необходимо использование математики в экономике?
2. Что такое математическая модель?
3. Последовательность разработки математических моделей и решения задач.
4. Назначение эконометрических моделей. Принципы их спецификации.
5. Типы данных в эконометрике.
6. Виды экономико-математических моделей.
7. Построение концептуальной модели.
8. Как строится математическая модель экономических явлений или объекта? Приведите пример построения модели.
9. Какова связь между математической структурой модели и ее содержательной интерпретацией?
10. Какие переменные модели называются экзогенными, а какие эндогенными?
11. Чем отличается равновесные модели от оптимизационных?
12. В чем отличие статических моделей от динамических? К какому типу относится приведенная в главе модель?
13. В чем отличие математической экономики от эконометрики?
14. Предположим, перед вами стоит задача обоснования линейной зависимости объема потребления от дохода на основе эмпирических данных. Относится ли решение подобной задачи к математической экономике или к эконометрике?
15. Пусть фирма выпускает один вид продукции, причем объем выпуска зависит от затрат только двух факторов: труда и капитала. Заданы цена единицы труда (ставка заработной платы), цена единицы капитала (ставка процента) и издержки фирмы. Задача состоит в определении объемов затрат труда и капитала, максимизирую их объем выпуска. Постройте соответствующую экономико-математическую модель.

16. В модели, сформулированной в вопросе 15, укажите, какие переменные являются экзогенными, а какие эндогенными. К какому типу относится эта модель: макроэкономических или микроэкономических?

ГЛАВА 2. ОСНОВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Для успешного освоения эконометрики необходимо владение основами теории вероятностей и математической статистики. По этой причине большинство учебников по эконометрике в качестве приложения содержит изложение основ теории вероятностей и математической статистики. В этой главе приводятся некоторые основополагающие понятия этой теории.

2.1. Случайная величина

Величина X называется случайной, если в результате испытания она принимает только одно возможное значение, которое заранее неизвестно. Различают дискретную и непрерывную случайную величину.

Примеры.

1. Количество ясных дней в месяце марте в Ташкенте есть случайная величина, возможными значениями которой являются числа 0, 1, 2, …, 29, 30, 31. Это дискретная случайная величина, потому что может принимать отдельные изолированные значения.
2. Объем воды к началу оросительного сезона в водохранилище Чарвак есть случайная величина, возможными значениями которой являются все числа от 0 до 2 км^3. Это непрерывная случайная величина.

На первый взгляд может показаться, что возможные значения полностью характеризуют случайную величину. Однако, возможные значения не могут полностью охарактеризовать случайную величину, поскольку эти значения случайная величина принимает с разными вероятностями.

Наиболее полно случайную величину характеризует закон ее распределения, т.е. соответствие между возможными значениями и их вероятностями. Закон распределения можно задавать в виде таблицы. Пусть X – количество ясных дней в месяце марте в Ташкенте. Тогда закон распределения случайной величины X имеет следующий вид:

X	0	1	2	3	…	29	30	31
P	0,01	0,01	0,02	0,03	…	0,02	0,01	0,01

Отметим, что сумма вероятностей всех возможных значений равна единице:

$p_1 + p_2 + \ldots + p_n = 1$

Очевидно, что наибольшую вероятность имеют близкие к среднегодовому количеству ясных дней в этом месяце. И наоборот, наименьшую вероятность имеют крайние значения 0, 1, 2, 29, 30, 31.

2.2. Математическое ожидание, дисперсия и среднее квадратическое отклонение дискретной случайной величины

При статистическом анализе экспериментальных данных важное значение имеют числовые характеристики случайной величины, такие как, математическое ожидание, дисперсия и среднее квадратическое отклонение.

Математическое ожидание – это сумма произведений всех возможных значений случайной величины на их вероятности.

X	x_1	x_2	…	x_n
P	p_1	p_2	…	p_n

$$M(X) = x_1p_1 + x_2p_2 + \ldots + x_np_n \quad (2.1)$$

Пример 3. Найти математическое ожидание случайной величины X:

X	2	5	8
P	0,2	0,5	0,3

M(X) = 2*0,2 + 5*0,5 + 8*0,3 = 5,3.

Математическое ожидание характеризует среднее арифметическое всех возможных значений случайной величины.

Свойства математического ожидания:

1. Математическое ожидание постоянной равно самой постоянной.
2. Постоянный множитель можно выносить за знак математического ожидания.
3. Математическое ожидание произведения нескольких независимых случайных величин равно произведению их математических ожиданий.

4. Математическое ожидание суммы нескольких случайных величин равно сумме их математических ожиданий.

Рассмотрим еще два примера:

Пример 4. Найти математическое ожидание случайной величины X:

X	-0,001	0,001
P	0,5	0,5

M(X) = -0,001*0,5 + 0,001*0,5 = 0.

Пример 5. Найти математическое ожидание случайной величины X:

Y	-1000	1000
P	0,5	0,5

M(Y) = -1000*0,5 + 1000*0,5 = 0.

Математические ожидания случайных величин на примерах 4 и 5 равны M(X) = M(Y) = 0, однако их возможные значения совершенно различные. Возможные значения случайной величины X близки к математическому ожиданию, а случайной величины Y достаточно далеки. Т.е., кроме определения характеристики среднего значения, необходимо найти характеристику разброса (рассеяния) возможных значений случайной величины от ее среднего значения. Такой характеристикой является дисперсия.

Поскольку дисперсия характеризует среднее отклонение, то ее можно было бы определить как математическое ожидание отклонений от среднего значения. Часть отклонений положительны, а другая часть отрицательны, и при вычислении математического ожидания они будут взаимно уничтожаться. По этой причине определим математическое ожидание квадрата отклонений.

Таким образом, дисперсия – это математическое ожидание квадрата отклонения случайной переменной х от ее среднего значения:

$$D(X) = M[X - M(X)]^2 \quad (2.2)$$

Поскольку при вычислении дисперсии отклонения возводятся в квадрат, то размерность дисперсии тоже будет в квадрате размерности случайной величины. В связи с этим рассмотрим еще одну характеристику случайной величины, т.е. среднее квадратическое отклонение.

Среднее квадратическое отклонение (или стандартное отклонение) – это квадратный корень от дисперсии:

$$\sigma(X) = \sqrt{D(X)} \qquad (2.3)$$

Как и дисперсия, среднее квадратическое отклонение также характеризует отклонения значений случайной величины от ее среднего значения, но имеет ту же размерность, что и сама случайная величина.

Можно выделить три уровня параметров случайной величины:

1. Результаты замеров реально существующей константы. Примеры: масса протона, период полураспада (или вероятность распада) радиоактивного изотопа, вероятность падения монеты гербом кверху. Эти константы объективно существуют, и, проводя эксперименты, мы можем приближаться к ним, достигая заданной точности. Увеличивая число бросков монеты, мы можем сделать оценку вероятности выпадения герба сколь угодно близкой к 1/2. В экономике и социологии абсолютных констант не существует, нет абсолютно точных взаимозависимостей величин, как в физике. Существуют константы, устанавливаемые правительством, например, ставка налога, но они не являются фундаментальными, могут меняться, и их не оценивают с использованием статистики и эконометрики.

2. Роль абсолютных констант, характеризующих экономику и социальную сферу страны и региона, играют параметры генеральных совокупностей – всех доступных значений по стране или региону. Примеры: средний доход домохозяйств, процент заболевших гриппом. В принципе, эти параметры можно измерить во время переписей населения или тотальных проверок (при условии достоверной информации), но такие технологии дороги, а исследуемые параметры непрерывно меняются. Поэтому для оценки параметров природных и социально-экономических объектов служат случайные выборки.

3. Случайные выборки. Было доказано, что если замеры x независимы, то наилучшая оценка математического ожидания E(x) – среднее значение по выборке, а наилучшая оценка дисперсии σ2

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (2.4)$$

Почему *n-1*, а не *n*? Дело в том, что в формулах (2.2) и (2.3) используется не математическое ожидание *E(x)*, которое мы не знаем, а его оценка – среднее значение *x*, вычисляемое по выборке X$\{x_1, x_2, ..., x_n\}$, поэтому смещенное относительно *E(x)* и расположенное ближе к центру значений множества $\{x_1, x_2, ..., x_n\}$. Если делить на *n*, получим заниженную оценку дисперсии. *n* в формулах (2.2), (2.3) и *n-1* в формуле (2.4) – это число степеней свободы, независимых суммируемых переменных. Поскольку *x* вычислено по $\{x_1, x_2, ..., x_n\}$, одно из выражений в скобках в формуле (2.4) мы можем вычислить, зная *n-1* значений *x*.

Что такое наилучшая оценка, или наилучшая технология оценки (estimator) математического ожидания случайной величины? Каковы ее критерии?

1. *Несмещенность*. Применяя правильную технологию расчета, мы не получим в результате обработки серии замеров статистически значимого отклонения от реального значения оцениваемого параметра.

2. *Эффективность*. Если в формуле (2.4) мы используем вместо *x* другую величину, полученную по другой формуле, то оценка дисперсии *S* будет больше. Значит, среднее значение обеспечивает наиболее эффективную оценку математического ожидания *E(x)*. Эффективность может вступить в противоречие с несмещенностью. Например, исключение переменных из эконометрических моделей может привести к уменьшению дисперсий оцениваемых параметров и к их смещению относительно истинных значений.

3. *Consistency*. В русскоязычной литературе это слово переводят как “состоятельность”, но правильнее говорить о *сходимости*. Это значит, что увеличивая количество замеров в серии *n*, мы можем получить разность оценок исследуемого параметра меньше любого ε, то есть наши оценки сходятся к какому-то пределу.

2.3. Законы распределения случайной величины

Пример. Отдел проверки качества некоторого предприятия исследует соответствие размеров изготовленных предприятием товаров стандарту. Очевидно, размеры некоторых товаров могут быть больше, некоторых меньше стандарта. Отклонения могут иметь различные значения, поэтому сотрудники отдела распределяют товары по значению отклонений от стандарта на несколько интервалов и строят гистограммы частотных распределений, то есть считают количество товаров в каждом интервале размеров.

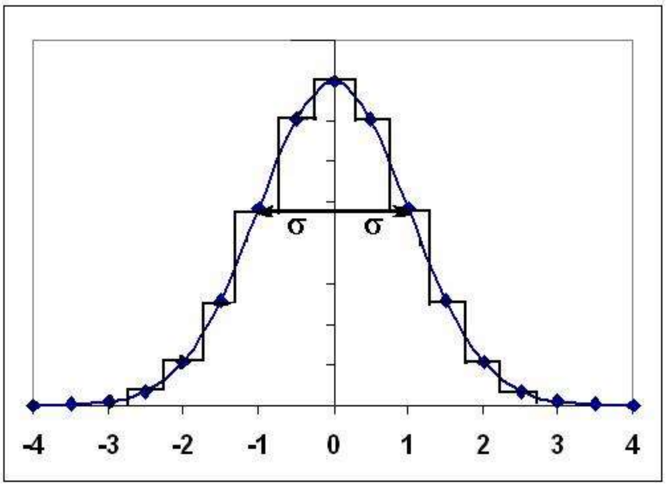


Рис.2.1. Гистограмма частотного распределения и кривая Гаусса с параметрами *E(x)* = 0 и σ = 1.

Инженеры считают, что размеры товаров подчиняются *закону нормального распределения (ЗНР)*, выведенного К.Гауссом

$$P(x_i) = \frac{1}{\sigma\sqrt{2\pi}}\exp(\frac{(x_{i-}\mu_x)^2}{2\sigma^2}$$

Как видите, в *функции Гаусса* всего два параметра: математическое ожидание μ_x и стандартное отклонение σ, которые сравнительно легко оценить по выборке, используя формулы (2.3) и (2.4). Эти формулы реализованы в MS Excel в функциях соответственно СРЗНАЧ, ДИСП и СТАНДОТКЛОН, категория «Статистические». Зная параметры гауссианы, можно вычислить процент деталей в различных диапазонах *x (квантили)*, используя таблицы или функцию НОРМРАСП MS Excel. Поэтому закон нормального распределения широко применяется при проектировании машин и механизмов. Например, можно вычислить количество событий (деталей) в диапазоне {*E(x)* -2σ, *E(x)* +2σ}. Это примерно 95%, то есть в "хвостах" останется по 2,5%.

В данном случае *p* = 0,95 – *доверительная вероятность*, а {*E(x)* -2σ, *E(x)* +2σ} - соответствующий *доверительный интервал*.

На рисунке 2.2 показано применение функции НОРМРАСП. Площадь левого хвоста гауссианы (Рисунок 2.1) от $-\infty$ до -1,96 (почти 2) равна 0,024997895, то есть 2,5%.

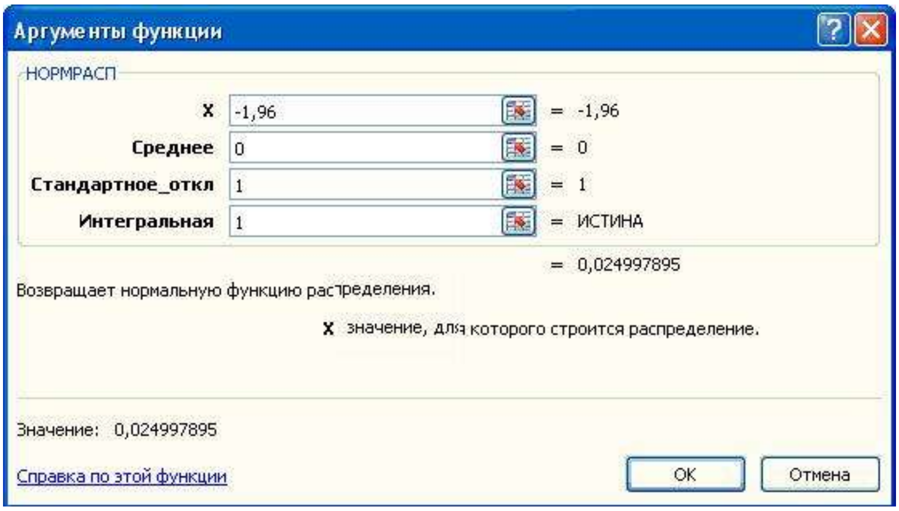


Рис.2.2.

В общем виде это утверждение выглядит следующим образом: *для уровня значимости* $\alpha = 1 - p$ *доверительный интервал равен* $\{E(x) - t_{крит}\sigma,\ E(x) + t_{крит}\sigma\}$, *где* $t_{крит}$ *–критические значения статистики Стьюдента* $t = E(x)/\sigma$. В нашем примере α – доля деталей в одном или двух “хвостах”. При уменьшении числа замеров надежность оценки $E(x)$ и дисперсии падают, и доверительный интервал надо расширять. Поэтому критические значения статистики Стьюдента зависят от уровня значимости (доверительной вероятности) и количества замеров (степеней свободы). Распределение Стьюдента $t_{крит}(\alpha,\ n)$ приведено во всех учебниках и практикумах по математической статистике и эконометрике. В MS Excel имеется функция СТЬЮДРАСП ($t_{крит}$, n, число хвостов (1 или 2)), которая возвращает долю событий в одном или двух “хвостах”. Для практических целей достаточно запомнить, что при числе замеров больше 30 и p=95% $t_{крит}$ примерно равно 2 (при “бесконечном” числе замеров – 1,96).

Инженеры используют правило, опирающееся на распределение Гаусса: “за тремя сигмами ничего нет”, то есть количество деталей с размерами, отклоняющимися от среднего более чем на 3σ, ничтожно мало, меньше 0,135% в каждом “хвосте” (сейчас переходят на шестисигмовый уровень надежности). Разница экономики и техники состоит в том, что 5%

невыгодных сделок – не страшно, а 5% или 2,5% (один хвост) заклиненных деталей – это много.

На рисунке 2.3 представлено окно функции СТЬЮДРАСП. Функция вычисляет площадь одного “хвоста” от - ∞ до -2 (или от 2 до ∞) : в данном случае 0,027312522, то есть 2,7%. В окне функции СТЬЮДРАСПОБР на рисунке 2.4 представлено значение *t*-статистики (отклонение от среднего значения в сигмах), равное 2,042; то есть площадь двух “хвостов” с границами ±2,042σ равна 5%.

В метеорологии, геохимии, биологии и экономике закон нормального распределения не работает, что связано с когерентностью, то есть взаимной зависимостью событий. Например, изъятие вкладов из банка может многократно превысить средний уровень из-за негативных публикаций или слухов. Для природы и экономики характерны распределения “с толстыми хвостами”, то есть количество аномальных замеров достаточно велико. Известно, что количество природных катастроф в зависимости от количества жертв подчиняется экспоненциальному закону.

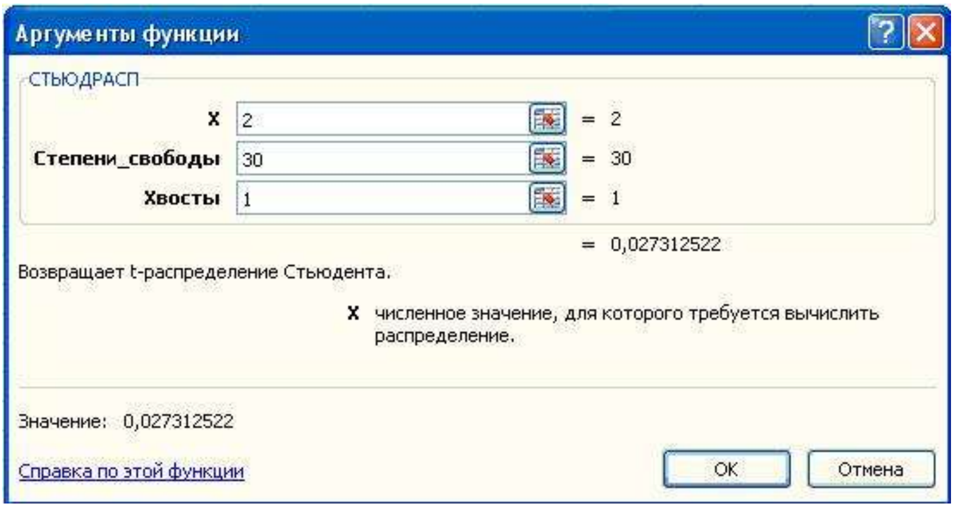


Рис.2.3.

Успешно используется логнормальное распределение, сводимое к нормальному заменой x_i *на* $log(x_i)$. Логнормальному распределению подчиняются, по данным автора, микроэлементы и чернобыльские радионуклиды в пробах, количество покупок магазине в зависимости от их стоимости.

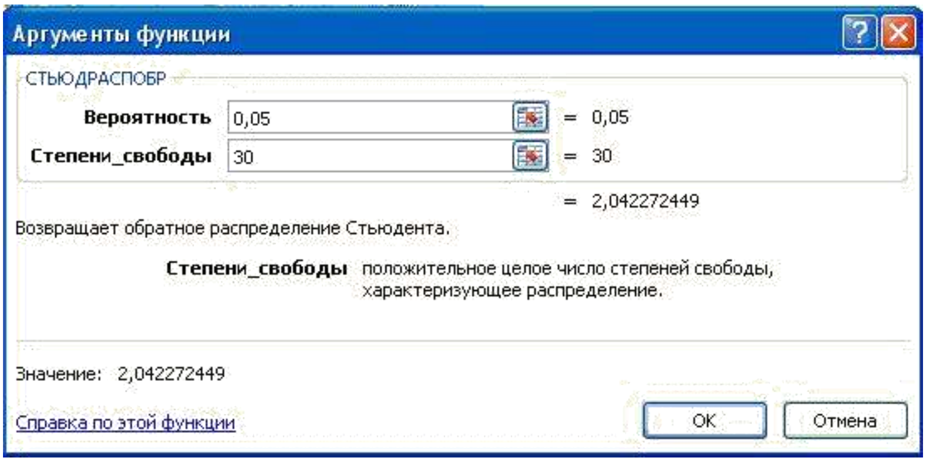


Рис.2.4.

Авторы не располагают данными о количестве льготников – пассажиров на городском и пригородном транспорте, но предполагают, что именно незнание законов частотных распределений в социальной сфере привело к бунтам и блокированию трасс при монетизации льгот. Предположим, что количество льготников ***N*** в зависимости от стоимости проезда распределено по логнормальному закону (Рис.2.5). По оси абсцисс указано количество поездок на городском транспорте в день.

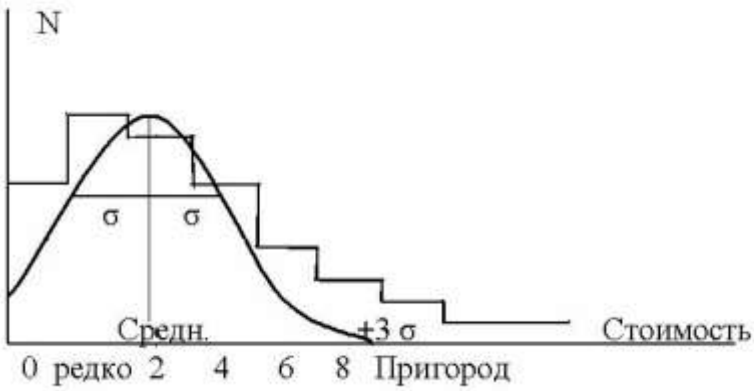


Рис.2.5. Количество льготников ***N*** в зависимости от стоимости проезда.

Видимо, при расчетах компенсаций был использован закон нормального распределения (плавная кривая), компенсировали средние затраты, но больше половины льготников были недовольны. Даже когда добавили σ, потом 2σ, может быть 3σ, то осталось много недовольных: бывшие военные, полярники, милиционеры, которые ездят из пригородов в Ташкент на заработки. В результате – огромные траты из казны, а льготный проезд из пригородов пришлось оставить.

В математической статистике используются также распределения Пирсона (хи-квадрат), Фишера, Пуассона.

2.4. Взаимосвязь случайных величин

Одна из основных задач эконометрики – выявление взаимосвязи переменных. Количественными оценками взаимосвязи служат ковариация и коэффициент корреляции. Ковариация переменных ***x*** и ***y*** – это ожидаемое значение произведения их отклонений от ожидаемых значений:

$$cov(x, y) = E((\mathrm{x} - E(\mathrm{x})) * (y - E(y))$$

Для оценки ковариации по выборке используется формула, аналогичная формуле дисперсии

$$cov(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Cov(x,x) – это дисперсия *x*. Коэффициент корреляции – это ковариация, нормированная на стандартные отклонения *x* и *y*:

$$r = \frac{cov(x, y)}{\sigma_x * \sigma_y}$$

Коэффициент корреляции – безразмерная величина, изменяется от –1 до +1; близость к нулю означает отсутствие связи переменных.

Опорные слова

Статистика, математическая статистика, дискретная переменная, непрерывная случайная величина, вероятность, закон распределения, распределение частот, плотность вероятности, математическое ожидание, дисперсия, квадратическое отклонение, несмещенность, эффективность, стандартное отклонение, распределение Стьюдента, ковариация переменных.

Контрольные вопросы

1. Дифференциальный и интегральный закон распределения случайной величины, виды функций распределения. Что такое “толстые хвосты”?
2. Параметры случайной величины: ожидаемое значение, дисперсия и среднее квадратическое отклонение, коэффициенты ковариации и корреляции.
3. Проверка статистических гипотез, t-статистика Стьюдента, доверительная вероятность и доверительный интервал, критические значения статистики Стьюдента.

ГЛАВА 3. РЕГРЕССИОННЫЙ И КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

3.1. Корреляционный анализ

В экономических исследованиях одной из основных задач является анализ зависимостей между переменными. Зависимость может быть строгой (функциональной) либо статистической. Алгебра и математический анализ занимаются изучением функциональных зависимостей, то есть зависимостей, заданных в виде точных формул. Но любая такая зависимость в определенной степени является абстракцией, поскольку в окружающем мире, частью которого является экономика, значение конкретной величины не определяется неизменной формулой ее зависимости от некоторого набора других величин. Всегда есть несколько величин, которые определяют главные тенденции изменения рассматриваемой величины, и в экономической теории и практике ограничиваются тем или иным кругом таких величин (объясняющих переменных). Однако всегда существует и воздействие большого числа других, менее важных или трудно идентифицируемых факторов, переменной от конкретной формулы ее связи с объясняющих переменных, построение формул зависимости и оценка их параметров являют не только одним из важнейших разделов математической статистики. Это своего рода искусство, учитывающее в каждой конкретной области знаний (в частности, в экономике, о которой идет речь), ее внутренние законы и потребности. Но это также и наука, поскольку выбираемый и оцениваемый вид формулы должен быть объяснен в терминах данной области знаний.

Пусть требуется оценить связь между переменными (например, связь показателей безработицы и инфляции в данной стране за определенный период времени). В частности, может стоять вопрос, связаны ли между собой эти показатели, и при положительном ответе на него, естественно, встает задача нахождения формулы этой связи. Основой для ответа на этот вопрос являются статистические данные о динамике этих показателей (годовые, квартальные, месячные и т.п.). Эти данные представляют собой некоторую, предположительно – случайную, выборку из генеральной совокупности, то есть из совокупности всех возможных сочетаний показателей инфляции и безработицы в сложившихся условиях. Таким образом, вывод о наличии связи для всей генеральной совокупности нужно делать по выборочным данным, что само по себе уже делает ответ на поставленный вопрос небезусловным. Более того, по данным выборки ответить на вопрос в приведенной постановке, то есть о наличии связи «вообще», невозможно. Действительно, через любые n точек на плоскости всегда можно провести полином степени m и объявить, что найдена точная формула связи. Однако опыт подсказывает, что если бы мы получили еще одну точку-наблюдение, то она наверняка не удовлетворяла бы найденной формуле. Поэтому вопрос о наличии связи между переменными (в частности – экономическими) следует ответить на вопрос о наличии конкретной формулы (спецификации) такой связи, устойчивой к изменению числа наблюдений. При этом нужно понимать, что ответ на этот вопрос по данным выборки не может быть однозначным и категоричным.

Простейшей формой зависимости между переменными является линейная зависимость, и проверка наличия такой зависимости, оценивание ее индикаторов и параметров является одним из важнейших направлений приложения математической статистики.

Рассмотрим вначале вопрос о линейной связи двух переменных:

1) Связаны ли между собой линейно переменные?
2) Какова формула связи переменных?

В первом случае переменные выступают как равноправные, здесь нет независимой и зависимости одной переменной от другой, например об оценивании формулы $y = a + bx$ (a и b - неизвестные коэффициенты такой зависимости). В этом случае переменная X является независимой (объясняющей), а переменная Y - зависимой (объясняемой). Вопрос о нахождении формулы зависимости можно ставить после положительного ответа на вопрос о существовании такой зависимости, но эти два вопроса можно решать и одновременно. Для ответа на поставленные вопросы существуют специальные статистические методы и, соответственно, показатели, значения которых определенным образом (и с определенной вероятностью) свидетельствуют о наличии или отсутствии линейной связи между переменными. В первом случае это коэффициент корреляции величин, во втором случае – коэффициенты линейной регрессии и их стандартные ошибки – статистики, по значениям которых проверяется гипотеза об отсутствии связи величин.

Вначале объясним логику появления такого показателя, как коэффициент корреляции. Предположим, что между переменными существует линейная связь. Наличие такой связи можно интерпретировать следующим образом. Если переменная принимает значения большие, чем ее среднее значение, и связь положительна (на языке формул это означает, что коэффициент положителен), то значение переменной также должно быть больше ее среднего значения и соотношение отклонений от их средних значений должно быть постоянным. Если связь переменных отрицательна, то положительное отклонение от среднего значения должно сочетаться с отрицательным отклонением от ее средней, а отрицательное отклонение от среднего значения – с положительным отклонением от ее средней – при постоянном соотношении этих отклонений. Если линейной связи между переменными нет, то положительные отклонения переменной от ее среднего значения могут (хотя и не обязательно будут) сочетаться как с положительными, так и с отрицательными отклонениями от ее среднего, то же можно сказать и про отрицательные отклонения от среднего.

В качестве меры для степени линейной связи двух переменных используется коэффициент их корреляции:

$$r(x,y) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (3.1)$$

По формуле коэффициента корреляции видно, что он будет положителен, если отклонения переменных от своих средних значений имеют, как правило, одинаковый знак, и отрицательным – если разные знаки.

Коэффициент корреляции является безразмерной величиной (так как размерности числителя и знаменателя есть размерности произведения); его величина не зависит от выбора единиц измерения обеих переменных. Величина коэффициента корреляции меняется от – 1 в случае строгой линейной отрицательной связи до +1 в случае строгой линейной положительной связи. Близкая к нулю величина коэффициента корреляции говорит об отсутствии линейной связи переменных, но не об отсутствии связи между ними вообще. Последнее вытекает из того, что каждой паре одинаковых отклонений переменной от ее среднего значения соответствуют равные по абсолютной величине положительное и отрицательное отклонения переменной от ее среднего. Соответственно, произведения этих отклонений «гасят» друг друга в числителе формулы коэффициента корреляции, и он оказывается близким к нулю. Заметим, что в числителе формулы для выборочного коэффициента корреляции величин стоит их показатель ковариации: $cov(x,y)=\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})$.

В отличие от коэффициента корреляции, показатель ковариации не нормирован – он имеет размерности, и его величина зависит от единиц измерения величин. В статистическом анализе показатель ковариации сам по себе используется редко; он фигурирует обычно как промежуточный элемент расчета коэффициента корреляции.

Далее, в анализе коэффициента корреляции возникает следующий вопрос. Если он равен нулю для генеральной совокупности, это вовсе не значит, что он в точности будет равен нулю для выборки. Наоборот, он обязательно будет отклоняться от истинного значения, но чем больше такое отклонение, тем менее оно вероятно при данном объеме выборки. Таким образом, при каждом конкретном значении коэффициента корреляции величин для генеральной совокупности выборочный коэффициент корреляции является случайной величиной. Следовательно, случайной величиной является также любая его функция, и требуется указать такую функцию, которая имела бы одно из известных распределений, удобное для табличного анализа. Для выборочного коэффициента корреляции такой функцией является статистика, рассчитываемая по $t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$.

Число степеней свободы меньше числа наблюдений на 2, поскольку в формулу выборочного коэффициента корреляции входят средние выборочные значения, для расчета которых используются две линейные формулы их зависимости от наблюдений случайных величин. Сразу уточним, что для коэффициента корреляции будет проверяться нулевая гипотеза, то есть гипотеза о равенстве его нулю в генеральной совокупности.

Эта гипотеза отвергается, если выборочный коэффициент корреляции слишком далеко отклонился от нулевого значения, то есть произошло событие, которое было бы маловероятным в случае $\rho_{xy}=0$. Здесь, конечно, очень важно понять, что конкретно значат слова «слишком далеко» и «маловероятное событие». В последнем случае нужно задать вероятность такого события, которая называется в статистике «уровень значимости».

Чаще всего задается уровень значимости 1% или 5%. Если для некоторого показателя проверяется гипотеза, то она отвергается в том случае, если оценка показателя по данным выборки такова, что вероятность получения такого или большего (по модулю) ее значения меньше, чем 1% или 5% соответственно [9].

3.2. Парный регрессионный анализ

Коэффициент корреляции показывает, что две переменные связаны друг с другом, однако он не дает представления о том, каким образом они связаны. Рассмотрим более подробно те случаи, что одна переменная зависит от другой.

$$Y = \alpha + \beta x + \varepsilon$$

Здесь $\alpha + \beta x$ - неслучайная составляющая, где x выступает как объясняющая переменная, ε – случайный член. Точки P – это единственные точки, отражающие реальные значения переменных. Фактические значения α и β и положения точки Q неизвестны, так же как и фактические значения случайного члена.

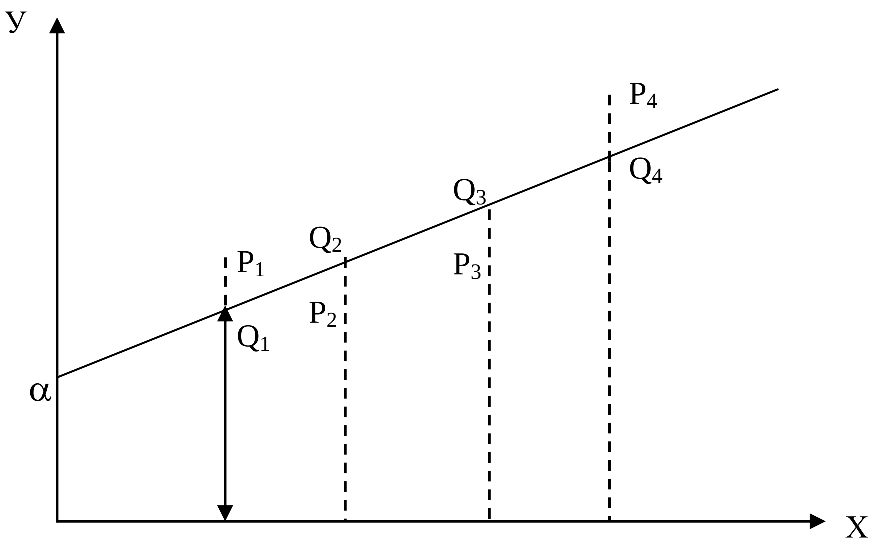


Рис.3.2.

Задача регрессионного анализа состоит в получении оценок α и β и следовательно, в определении положения прямой по точкам P. Очевидно, что чем меньше значения ε, тем легче эта задача.

Оценка параметров по методу наименьших квадратов.

Допустим, мы имеем четыре наблюдения для X и Y. Здесь $\alpha+\beta x$ - неслучайная составляющая, где x выступает как объясняющая переменная, ε – случайный член.

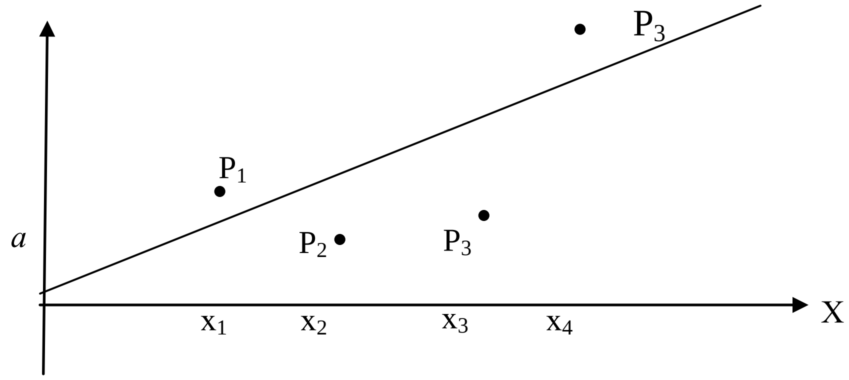
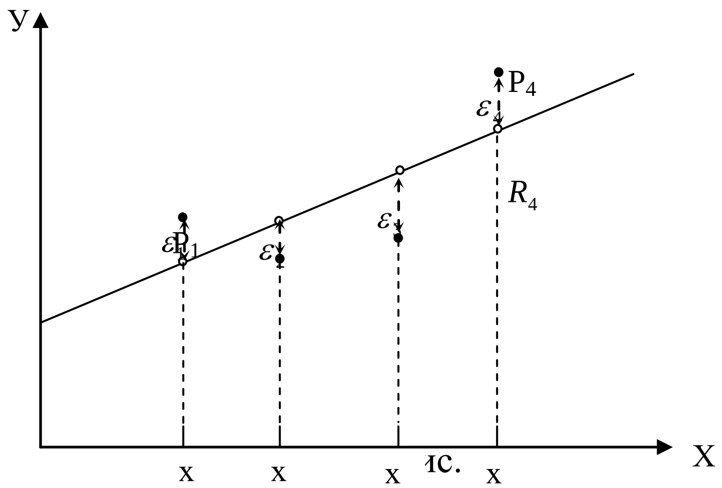


Рис. 3.3.

Отрезок, отсекаемый прямой на оси Y представляет собой оценку α и обозначен a, а угловой коэффициент прямой представляет собой оценку β, и обозначен b. Первым шагом является определение остатков для каждого наблюдения.

$$\varepsilon_1 = y_1 - \hat{y}_1, \quad \varepsilon_2 = y_2 - \hat{y}_2, \quad \varepsilon_3 = y_3 - \hat{y}_3, \quad \varepsilon_4 = y_4 - \hat{y}_4.$$

Мы хотим построить линии регрессии таким образом, чтобы эти остатки были минимальными. Одним из способов решения поставленной проблемы состоит в минимизации суммы квадратов остатков S.



$$S = \varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \varepsilon_4^2 \to min$$

$$S(ab) = \sum_i \varepsilon_i^2 \to min$$

$$S(ab) = \sum_{i} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \to min$$

Находим

$$\frac{\partial s}{\partial a} = 0. \quad \frac{\partial s}{\partial b} = 0$$

Поэтому

$$\begin{cases} \dfrac{\partial S}{\partial a} = -2\sum_{i=2}^{n} (y_i - a - bx_i) = 0 \\ \dfrac{\partial S}{\partial b} = -2\sum_{i=1}^{n} (y_i - a - bx_i)x_i = 0 \end{cases}$$

Отсюда $\begin{cases} \sum y_i = na + b\sum x_i \\ \sum x_i y_i = a\sum x_i + b\sum x_i^2 \end{cases}$ (3.2.)

Эта система называется *системой нормальных уравнений*. В ней два уравнения и два неизвестных *a* и *b*, а коэффициенты получаются суммированием ***x, y*** и т.д.

Решать ее можно разными способами. В данном случае использован сервис Excel *Поиск решения* для настройки линейной модели по данным ***X*** и ***Y***, представленным в Таблице 3.1. Коэффициенты системы нормальных уравнений расположены в виде матрицы (верхние строки таблицы 3.2), неизвестные ***a*** и ***b*** задаются произвольно и умножаются на коэффициенты (нижние строки). В окне *Поиска решения* задаются: *Целевая ячейка* – первая сумма, *Значение* равно 247 (*Σy), Изменяя ячейки* – ***a*** и ***b***, *Ограничения*: вторая сумма равна 3901 (*Σxy)*. Исходные данные *X* и *Y* приведены в Таблице 3.1. результаты расчета в

Таблице 3.1.

X	*Y*	x**2**	*XY*
10	12	100	120
11	15	121	165
12	18	144	216
13	16	169	208
14	24	196	336
15	22	225	330
16	27	256	432
17	28	289	476
18	25	324	450
19	32	361	608
20	28	400	560
Суммы 165	247	2585	3901

После этого можно построить функцию регрессии $\hat{\boldsymbol{Y}}$, сравнить ее с $\boldsymbol{Y}$ и использовать для прогноза. В принципе, МНК с *Поиском решения* можно использовать непосредственно [9].

Таблица 3.2.

11	165	247
165	2585	3901
a	*b*	
-4,27	1,78	
		Суммы по строкам
-47,00	294,00	246,9999
-705,00	4606,00	3901

Для этого надо задать произвольные коэффициенты a и b, построить по ним функцию $\hat{Y} = a + bX$, вычислить остатки $e = Y - \hat{Y}$ и их квадраты, сумму e^2. В окне *Поиска решения* установить *Целевая ячейка* Σe^2 *минимум*, *Изменяя ячейки* a и b, ограничений нет.

Таблица 3.3

X	***Y***	$\hat{Y}$	***Остатки е***	***е2***
10	12	13,545	-1,545	2,388
11	15	15,327	-0,327	0,107
12	18	17,109	0,890	0,793
13	16	18,890	-2,890	8,357
14	24	20,672	3,327	11,070
15	22	22,454	-0,454	0,206
16	27	24,236	2,763	7,637
17	28	26,018	1,981	3,927
18	25	27,8	-2,8	7,840
19	32	29,581	2,418	5,847
20	28	31,363	-3,363	11,314
		Суммы	1E-06	59,490
Дисперсии	40,872	34,923	5,949	
R^2	0,854		***A***	***B***
F	52,833		-4,27	1,78

.

Пример 3.1. Найти зависимость между стоимостью валовой продукции и баллом бонитета земли и количеством вносимых удобрений в фермерских хозяйствах. Данные по этим показателям приводятся в таблице 3.4:

Таблица 3.4.

Хозяйства №	Стоимость валовой продукции (тыс. сум/га) Y	Балл бонитета земли, X_1	Вносимые минеральные удобрения, ц.д.в. X_2
1	375	60	3,4
2	350	53	3,1
3	360	54	3,2
4	600	61	4,0
5	420	55	3,5
6	280	46	2,5
7	390	58	3,7
8	410	52	3,6
9	350	51	3,3

Сначала рассмотрим парную зависимость между стоимостью валовой продукции и баллом бонитета земли. Результат зависимости описывается следующим выражением: $y = \alpha + \beta x + \varepsilon$ и оценена регрессия $\hat{y} = -340{,}05 + 13{,}46x$.

Полученный результат можно истолковать следующим образом. Коэффициент при *x* (коэффициент наклона) показывает, что если *x* увеличивается на одну единицу, то *y* возрастает на 13,4 единиц. Как *x* измеряется в баллах, *y* измеряется тыс. сум/га; таким образом, коэффициент наклона показывает, что если балл бонитета земли увеличивается на 1 балл, то стоимость валовой продукции возрастает на 13,46 тыс. сум.

Что можно сказать о постоянных членах в уравнении? Формально говоря, она показывает прогнозируемый уровень *y*, когда *x* = 0. Иногда это имеет ясный смысл, иногда – нет. В данном случае константа выполняет единственную функцию: она позволяет определить положение линии регрессии на графике.

При интерпретации уравнения регрессии чрезвычайно важно помнить о трех вещах. Во-первых, *a* является лишь оценкой α и *b* – оценкой β. Поэтому вся интерпретация в действительности представляет собой лишь оценку. Во-вторых, уравнение регрессии отражает только общую тенденцию для выборки.

При этом каждое отдельное наблюдение подвержено воздействию случайностей. В-третьих, верность интерпретации зависит от правильности спецификации уравнения.

Оценка параметров конкретного уравнения является лишь отдельным этапом длительного и сложного процесса построения эконометрической модели. Первое же оцененное уравнение очень редко является удовлетворительным во всех отношениях. Обычно приходится постепенно подбирать формулу связи и состав объясняющих переменных, анализируя на каждом этапе качество оцененной зависимости. Этот анализ качества включает статистическую и содержательную составляющую. Проверка статистического качества оцененного уравнения состоит из следующих элементов:

- Проверка статистической значимости каждого коэффициента уравнения регрессии;
- Проверка общего качества уравнения регрессии;
- Проверка свойств данных, выполнение которых предполагалось при оценивании уравнения.

Под содержательной составляющей анализа качества понимается рассмотрение экономического смысла оцененного уравнения регрессии: действительно ли значимыми оказались объясняющие факторы, важные с точки зрения теории; положительные или отрицательные коэффициенты, показывающие направления воздействия этих факторов; попали ли оценки коэффициентов регрессии предполагаемым из теоретических соображении интервалам.

Методика проверки статистической значимости каждого отдельного коэффициента уравнения линейной регрессии была рассмотрена в предыдущей главе. Перейдем теперь к другим проверки качества уравнения.

3.3. Проверка общего качества уравнения регрессии. Коэффициент детерминации R^2

Для проверки общего качества оцененной линейной регрессии используют обычно коэффициент детерминации R^2, называемый также квадратом коэффициента множественной корреляции переменных X и Y. Коэффициент детерминации рассчитывается по формуле

$$D = R^2 = 1 - \frac{\sum_i e_i^2}{\sum_i (y_i - \bar{y})^2} \quad (3.2.)$$

Коэффициент детерминации характеризует доля вариации (разброса) зависимой переменной, объясненной с помощью данного уравнения. В качестве меры разброса зависимой переменной обычно используется ее дисперсия, а остаточная вариация может быть измерена как дисперсия отклонения вокруг линии регрессии. Если числитель и знаменатель вычитаемой из единицы дроби разделить на число наблюдений n, то получим, соответственно, выборочные оценки остаточной дисперсии и дисперсии зависимой переменной y. Отношение остаточной и общей дисперсий представляет собой долю необъясненной дисперсии. Если же эту долю вычесть из единицы, то получим долю дисперсии зависимой переменной, объясненной с помощью регрессии.

Иногда при расчете коэффициент детерминации для получения не смещенных оценок дисперсии в числители и знаменатели вычитаемой из единицы дроби делается поправка на число степеней свободы; тогда

$$D = R^2 = 1 - \left[\frac{\sum_i e_i^2}{n - m - 1}\right] / \left[\frac{\sum_i (y_i - \bar{y})^2}{n - 1}\right] \quad (3.3)$$

Или, для парной регрессии, где число независимых переменных m равно 1,

$$R^2 = 1 - \left[\frac{\sum_i e_i^3}{n - 2}\right] / \left[\frac{\sum_i (y_i - \bar{y})^2}{n - 1}\right] \quad (3.4)$$

В числители дроби, который вычитается из единицы, стоит сумма квадратов отклонений наблюдений y_i от линии регрессии, в знаменателе, от среднего значения переменной y. Таким образом, дробь эта мала (а коэффициент R^2, очевидно, близок к единице), если разброс точек вокруг линии регрессии значительно меньше, чем вокруг среднего значения. МНК позволяет найти прямую, для которой сумма e_i^2 минимальны, а $y = \bar{y}$ представляет собой одну из возможных линий, для которых выполняется условие $\bar{y} = a + b\bar{x}(b = 0, a = \bar{y})$.

Поэтому величина в числителе вычитаемой из единицы дроби меньше, чем величина в ее знаменателе, - иначе выбираемой по МНК линией регрессии была бы прямая $y = \bar{y}$. Таким образом, коэффициент детерминации R^2 является мерой, позволяющей определить, в какой степени найденная регрессионная прямая дает лучший результат для объяснения поведения зависимой переменной y, чем просто горизонтальная прямая $y = \bar{y}$.

Смысл коэффициента детерминации может быть пояснен и немного иначе. Можно показать, что $i \sum(y_i - \bar{y})^2 = \sum_i k_i^2 + \sum_i e_i^2$, где k_i - отклонение i – й точки на линии регрессии от $\bar{y}$. В данной формуле величина в левой части может интерпретироваться как мера общего разброса (вариации) переменной y, первое слагаемое в правой части $\sum_i k_i^2$ - как мера разброса, объясненного с помощью регрессии и второе слагаемое $\sum_i e_i^2$- как мера остаточного, необъясненного (разброса точек вокруг линий регрессии). Если разделить эту формулу на ее левую часть и перегруппировать члены, то $R^2 = \frac{\sum_i k_i^2}{\sum_i (y_i - \bar{y})^2}$, то есть коэффициент детерминации R^2- есть доля объясненной части разброса зависимой переменной (или доля объясненной дисперсии, если разделить числитель или знаменатель на n или n-1). Часто коэффициент детерминации R^2 иллюстрирует рис. 3.5.

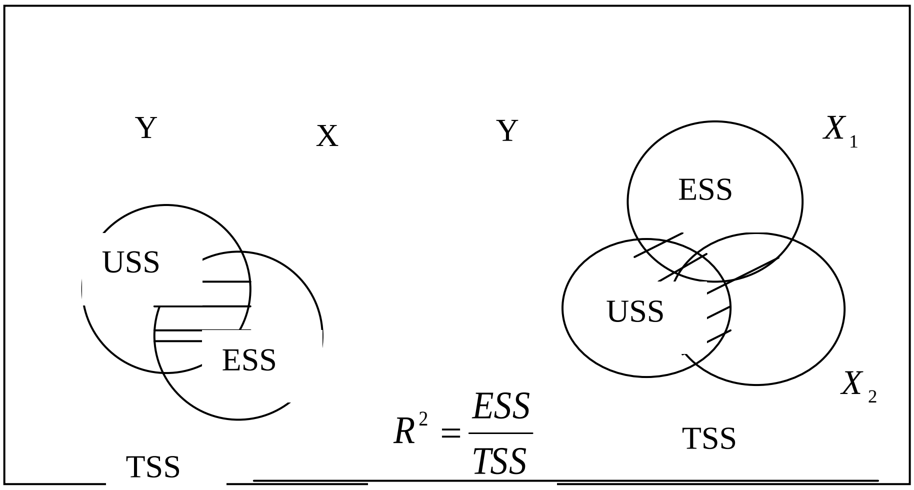


Рис. 3.5.

Здесь TSS (Total Sum of Squares) – общий разброс переменной y , ESS (Explained Sum of Squares) – разброс, объясненный с помощью регрессии, USS (Unexplained Sum of Squares) – разброс, необъясненный с помощью регрессии. Из рисунка видно, что с увеличением объясненной долей разброса коэффициент $\boldsymbol{R}^2$ приближается к 1. Кроме того, из рисунка видно, что с добавлением еще одной переменной R^2 обычно увеличивается, однако если объясняющие переменные x_1 и x_2 сильно коррелируют между собой, то они объясняют одну и ту же часть разброса переменной y, и в этом случае трудно идентифицировать вклад каждой из переменных в объяснении поведения y. Если существует статистически значимая линейная связь величин x и y, то коэффициент R^2 близок к единице. Однако он может быть близким к единице просто в силу того, что обе эти величины имеют выраженный временной тренд, не связанный с их причинно-следственной взаимозависимостью. В экономике обычно объемные показатели (доход, потребление, инвестиции) имеют такой тренд, а темповые и относительные (производительности, темпы роста, доли, отношения) – не всегда. Поэтому при оценивании линейных регрессиях по временным рядам объемных показателей (например, в зависимости выпуска от затрат ресурсов или объема потребления, от величины дохода) величина R^2 обычно очень близка к единице. Это говорит о том, что зависимую переменную нельзя описать просто как равная своему среднему значению, но это и заранее очевидно, раз она имеет временной тренд.

Если имеются не временные ряды, а перекрестная выборка, то есть данные однотипных объектов в один и тот же момент времени, то для оцененного по ним уравнения линейной регрессии величина R^2 не превышает обычно уровня 0,6-0,7. то же самое обычно имеет место и для регрессии по временным рядам, если они не имеют выраженного тренда. В макроэкономике примерами таких зависимостей являются связи относительных, удельных, темповых показателей: зависимость темпов инфляции от уровня безработицы, норма накоплений от величины процентной ставки, темпов прироста выпуска от темпов прироста затрат ресурсов. Таким образом, при построении макроэкономических моделей особенно - по временным рядам данных, нужно учитывать, являются входящие в них переменные объемными или относительными, имеют ли они временной тренд.

Точную границу приемлемости показателя R^2 указать сразу для всех случаев невозможно. Нужно принимать во внимание и число степеней свободы уравнения, и наличие трендов переменных, и содержательную интерпретацию уравнения. Показатель R^2 может оказаться даже отрицательным. Как правило, это случается в уравнении без свободного члена $y = \sum_i a_i x_i$. Оценивание такого уравнения производится, как и в общем случае, по методу наименьших квадратов. Однако множество выбора при этом существенно сужается: рассматриваются не все возможные прямые или гиперплоскости, а только проходящие через начало координат. Величина R^2 получится отрицательной в том случае, если разброс значений зависимой переменной вокруг прямой $y = \bar{y}$ меньше, чем вокруг даже наилучшей прямой из проходящих через начало координат. Отрицательная величина R^2 в уравнении $y = \sum_i a_i x_i$ говорит о целесообразности введения в него свободного члена. Эта ситуация проиллюстрирована на рис. 3.2.

Линия 1 на нем – график уравнения регрессии без свободного члена (он проходит через начало координат), линия 2 – со свободным членом (он равен a_0), линия 3 - $y = \bar{y}$. Горизонтальная линия 3 дает гораздо меньшую сумму квадратов отклонений e_i, чем линия 1, и поэтому для последней коэффициент детерминации R^2 будет отрицательным. Поправка на число степеней свободы всегда уменьшает значение R^2, поскольку $(n-1)\rangle(n-m-1)$. В результате величина R^2 также может стать отрицательной, но это означает, что она была близкой к нулю до такой поправки, и объясненная с помощью уравнения регрессии доля дисперсии зависимой переменной очень мала.

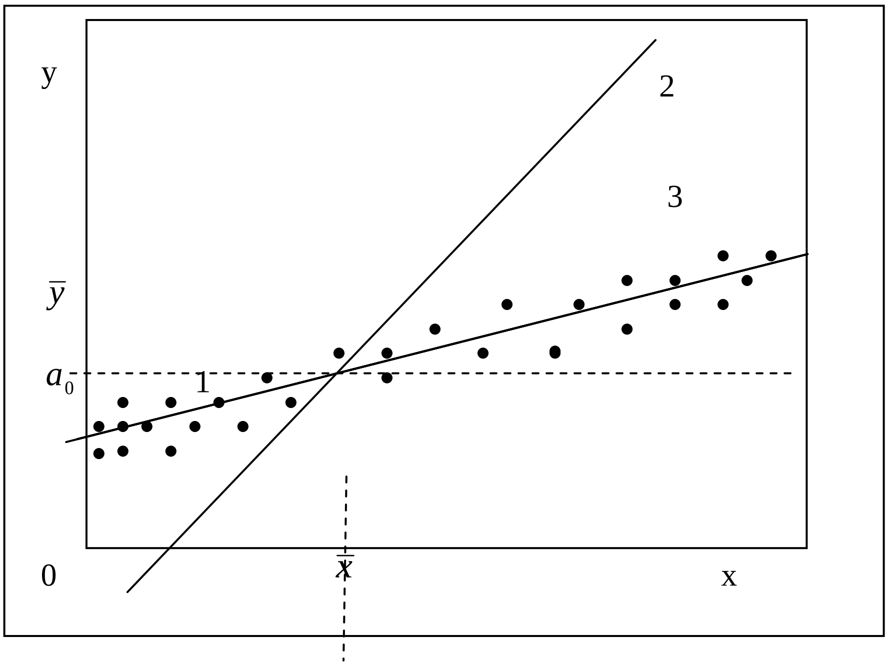


Рис.3.2. Линии уравнений линейной регрессии $y = f(x)$ без свободного члена (1) и со свободным членом (2).

3.4. F-статистика. Распределение Фишера в регрессионном анализе

Для статистической значимости коэффициента детерминации R^2 проверяется нулевая гипотеза для F-статистики, рассчитываемой по формуле:

$$F = \frac{R}{1 - R^2} * \frac{n - m - 1}{m} \qquad (3.5)$$

Соответственно, для парной регрессии

$$F = \frac{R^2(n-2)}{1 - R^2}$$

Смысл проверяемой гипотезы заключается в том, что все коэффициенты линейной регрессии должны иметь вид y=$\bar{y}$, а коэффициент детерминации R^2 и F-статистика Фишера также равны нулю. При этом их оценки для случайной выборки, конечно, отличаются от нуля, но чем больше такое отличие, тем менее оно вероятно. Логика проверки нулевой гипотезы заключается в том, что если произошло событие, которое было бы слишком маловероятным в том случае, если данная гипотеза действительно была бы верна, то эта гипотеза отвергается.

Величина F, если предположить, что выполнены предпосылки относительно отклонений e_i, имеет распределение Фишера с (m; n-m-1) степенями свободы, где m-число объясняющих переменных, n-число наблюдений. Распределение Фишера – двухпараметрическое распределение неотрицательной случайной величины, являющейся в частном случае, при m=1, квадратом случайной величины, распределенной по Стьюденту. Для распределения Фишера имеется таблицы критических значений, зависящих от чисел степеней свободы m и n-m-1 при различных уровнях значимости. В приложении дана некоторая вспомогательная информация о распределении Фишера и показано, как пользоваться таблицами этого распределения.

Показатели F и R^2 равны или не равны нулю одновременно, поэтому F=0 равнозначно тому, что линия регрессии y=$\bar{y}$ является наилучшей по МНК и, следовательно величина y статистически независима от x. Поэтому проверяется нулевая гипотеза для показателя F, который имеет хорошо известное, табулированное распределение Фишера. Для проверки этой гипотезы при заданном уровне значимости по таблицам находится критическое значение $F_{крит}$, и нулевая гипотеза отвергается, если $F \geq F_{крит}$.

Пусть, например, при оценке парной регрессии по 15 наблюдениям R^2= 0,7. В этом случае F=0,7* $\frac{13}{0,3} \approx 30,3$. По таблицам для распределения Фишера с (1,13) степенями свободы найдем, что при 5%- ном уровне значимости (доверительная вероятность 95%) критическое значение F=4,67, при 1% - ном- 9,07. Поскольку F=30,3>$F_{крит}$, нулевая гипотеза в обоих случаях отвергается. Если в той же ситуации R^2= 0,5 то F=13, и предположение о незначимости связи отвергается и здесь. Таким образом, для того, чтобы отвергнуть гипотезу о равенстве нулю одновременно всех коэффициентов линейной регрессии, коэффициент детерминации не должен быть очень близким к единице; его критическое значение для данного числа степеней свободы уменьшается при росте числа наблюдений и может стать сколько угодно малым. В то же время величина коэффициента R^2(точнее, рассчитанной по нему F- статистики, поскольку последняя учитывает число наблюдений и число объясняющих переменных) может служить отражением общего качества регрессионной модели.

Отметим, что в случае парной регрессии проверка нулевой гипотезы для t-статистики коэффициента регрессии равносильна проверке нулевой гипотезы для F- статистики (и, соответственно, показателя R^2). В этом случае F-статистика= квадрату t-статистики. В случае парной регрессии статистическая значимость величин R^2и t- статистики коэффициента регрессии определяется коррелированностью переменных x и y. Самостоятельную важность показатель R^2 приобретает в случае множественной линейной регрессии.

Распределение Фишера может быть использовано не только для проверки гипотезы об одновременном равенстве нулю всех коэффициентов линейной регрессии, но и гипотезы о равенстве нулю части этих коэффициентов. Это особенно важно при развитии линейной регрессионной модели, так как позволяет оценить обоснованность исключения отдельных переменных или их групп из числа объясняющих переменных, или же, наоборот, включение их в это число.

Пусть, например, в начале была оценена множественная линейная регрессия $y = a_0 + a_1 x_1 + a_2 x_2 + ... + a_m x_m$ по n наблюдениям с m объясняющими переменными, и коэффициент детерминации равен R_1^2. Затем последние k переменных исключены из числа объясняющих , и по тем же данным оценено уравнение $y = b_0 + b_1 x_1 + b_2 x_2 + ... + b_{m-k} x_{m-k}$, для которого коэффициент детерминации равен R_2^2(обязательно уменьшился, поскольку каждая дополнительная переменная объясняет часть, пусть небольшую, вариации зависимой переменной). Для того, чтобы проверить гипотезу об одновременном равенстве нулю всех коэффициентов регрессии при исключенных переменных, рассчитывается величина

F=$\frac{R_1^2 - R_2^2}{1 - R_1^2} * \frac{n-m-k-1}{k}$, имеющая распределение Фишера с (k, n-m-1) степенями свободы. По таблицам, при заданном уровне значимости, находится критическое значение F-статистики, и если ее рассчитанное значение превосходит критическое, то нулевая гипотеза отвергается. В таком случае исключать сразу из числа объясняющих все k переменных некорректно. F-статистика оказывается относительно большой, если велика разность (R_1^2 -R_2^2). В этом случае исключение данного набора k объясняющих переменных приводит к слишком большому сокращению доли объясненной дисперсии зависимой переменной, и поэтому недопустимо. Если, наоборот, эта доля сокращается незначительно, то F- статистика невелика, нулевая гипотеза не отвергается, и указанные k переменных могут быть проведены и по поводу обоснованности включения в уравнение регрессии одной или нескольких (k) новых объясняющих переменных. В этом случае рассчитывается F-статистика

$$F = \frac{R_2^2 - R_1^2}{1 - R_2^2} * \frac{n - m - k - 1}{k}$$

имеющая распределение F(k, n-m-k-1), и если она превышает критический уровень, то включение новых переменных объясняет существенную часть необъясненной ранее дисперсии зависимой переменной y. Отметим лишь, что добавлять новые переменные целесообразно, как правило, по одной.

В вопросе о добавлении объясняющих переменных в уравнение регрессии полезным может оказаться рассмотрение R^2 с поправкой на число степеней свободы:

$$\left[R^2 = 1 - \frac{\sum_i e_i^2}{n - m - 1}\right] / \left[\frac{\sum_i (y_i - y)^2}{n - 1}\right]$$

Обычный R^2 (без поправки) всегда растет при добавлении новой переменной; в R^2 с поправкой растет величина m, уменьшающая его. Если уменьшение доли объясненной дисперсии при добавлении новой переменной мало, то R^2 с поправкой может уменьшиться. Если это так, то добавлять переменную нецелесообразно.

F-статистика Фишера используется также для проверки гипотезы о совпадении уравнений регрессии для отдельных групп наблюдений. Пусть имеются две выборки, содержащие, соответственно, n_1 и n_2 наблюдений. Для каждой из этих выборок оценено уравнение регрессии вида $y = a_0 + a_1x_1 + a_2x_2 + ... + a_mx_m$. Пусть суммы квадратов отклонений y_1 от линии регрессии равны для них, соответственно, S_1 и S_2. Проверяется нулевая гипотеза, заключающаяся в том, что все соответствующие коэффициенты этих уравнений равны друг другу, то есть что уравнение для этих выборок одно и то же. Пусть оценено уравнение регрессии того же вида сразу для всех $(n_1 + n_2)$ наблюдений, и сумма квадратов отклонений y_i от линии регрессии равна для него S_0. Тогда рассчитывается F-статистика по формуле

$$F = \frac{(S_0 - S_1 - S_2)}{(S_1 + S_2)} * \frac{(n_1 + n_2 - 2m - 2)}{(m + 1)}$$

Она имеет распределение Фишера с $(m + 1, n_1 + n_2 - 2m - 2)$ степенями свободы. F-статистика будет близкой к нулю, если уравнение регрессии для обеих выборок одинаково, поскольку в этом случае $S_0 = S_1 + S_2$. Если же ее расчетное значение велико (то есть больше критического значения при данном уровне значимости), то нулевая гипотеза отвергается. Описанная процедура важна для ответа на вопрос, можно ли за весь рассматриваемый в модели период времени построить единое уравнение регрессии, или же нужно разбить его на части и на каждой из частей строить свое уравнение регрессии.

3.5. Проверка условий, выполнение которых предполагалось при оценивании уравнения регрессии. Автокорреляция остатков. Статистика Дарбина-Уотсона

Близкое к единице значение коэффициента детерминации R^2еще не свидетельство высокого качества уравнения регрессии. Рассмотрим рисунок 3.2. На нем показана зависимость реального объема потребления от численности населения в США за 1931-1990 гг., а также линия оцененного по этим данным уравнения парной линейной регрессии.

Формула этого уравнения, следующая:

$$CONS=-1817{,}3+16{,}7\ POP. \qquad (3.6)$$

Стандартные ошибки свободного члена и коэффициента регрессии равны, соответственно, 84,7 и 0,46; их t-статистики- (-21,4 и 36,8). По абсолютной величине t-статистики намного превышают 3, и это свидетельствует о высокой надежности оцененных коэффициентов. Коэффициент детерминации R^2 уравнения равен 0,96, то есть объяснено 96% дисперсии объема потребления. И в то же время уже по рисунку видно, что регрессия не очень хороша: зависимость величин POP и CONS явно не линейна.

Если использовать проведенную прямую, скажем, для прогнозирования дальнейшей динамики потребления, результат будет неудовлетворительным. Существо вопроса здесь понятно – в течении рассматриваемого периода значительно вырос объем потребления на душу населения. Численность населения США росла во времени почти линейно (то есть с примерно постоянным темпом). Это ясно и без уравнения линейной регрессии, но мы специально оценили его для иллюстрации.

Как же можно выразить формально неудовлетворенность полученного уравнения регрессии? Можно видеть, что не выполнены необходимые предпосылки об отклонении от линии регрессии e_1. Эти величины явно не являются взаимно независимыми, и дисперсия их непостоянна. Нарушение исходных предпосылок не только свидетельствует о неточной спецификации уравнения регрессии, но и делают неточным полученные оценки коэффициентов регрессии и их стандартных ошибок. Поэтому следующий этап проверки качества уравнения регрессии – проверка некоторых важных свойств, выполнение которых предполагалось при оценивании уравнения регрессии.

Приступая к оценке линейного уравнения регрессии, мы предполагали, что реальная взаимосвязь переменных линейна, а отклонения регрессионной прямой случайны, независимы между собой и имеют нулевое среднее постоянную дисперсию. Так ли это на самом деле? Если нет, то наш анализ статистической значимости коэффициента регрессии не точен и оценки этих коэффициентов не обладают такими желательными свойствами, как несмещенность, состоятельность и эффективность.

Попытаемся ответить на вопрос, в каких случаях отклонения не обладают предполагавшимися свойствами. Во-первых, если в действительности исследуемая взаимосвязь не линейна. Мы видим, например, на рис.3.2 что в этом случае отклонения от линии регрессии не случайно распределены вокруг нее, а обладают определенной закономерностью. Эта закономерность, в частности, выражается в одинаковом, как правило, знаке каждых двух соседних отклонений. Это может являться следствием нелинейного характера связи переменных, либо воздействия какого-то фактора, не включенного в уравнение регрессии. Величина такого неучтенного фактора может менять свою динамику в рассматриваемый период, отклоняясь в достаточно длительные промежутки времени в ту или иную сторону от своего среднего значения. Это, очевидно, может служить причиной длительных устойчивых отклонений зависимой переменной от линии регрессии. Обе указанные причины свидетельствуют о том, что существует возможность улучшить уравнение регрессии путем оценивания какой-то новой не линейной формулой или включения некоторой новой объясняющей переменной.

Зависимость, показанная на рис.3.2, очевидно, нелинейно. Но это-крайний случай. Далеко не всегда бывает столь же очевидно, что отклонения от регрессионной прямой имеют неслучайный, закономерный характер. Для оценки степени такой неслучайности необходимо ввести количественную меру.

Итак, одним из основных предполагаемых свойств отклонений ε_i значений y_i от регрессионной формулы $y = \alpha + \beta x$ является их статистическая независимость между собой. Поскольку значения ε_i остаются неизвестными ввиду неизвестности истинных значений, то проверяется статистическая независимость их аналогов – отклонений e_i. При этом проверяется обычно их некоррелированность (являющаяся необходимым, но не достаточным атрибутом независимости), при чем некоррелированность не любых, а соседних величин e_i. Соседними можно считать соседние во времени (в случае временных рядов) или по возрастанию переменной x (в случае перекрестных выборок) значения e_i. Для этих величин можно рассчитать, например, коэффициент корреляции (называемый *коэффициентом автокорреляции первого порядка)*:

$$r_{i,i-1} = \frac{\sum_i e_i e_{i-1}}{\sqrt{\sum_i e_i^2 \sum_i e_{i-1}^2}}$$

(считаем, что $M[e_i]$=0). Практически, однако, используют тесно связанную с $r_{i,i-1}$ статистику Дарбина-Уотсона DW, рассчитываемую по формуле:

$$DW = \frac{\sum_i (e_i - e_{i-1})^2}{\sum_i e_i^2} \qquad (3.7.)$$

Очевидно, $DW = \frac{\sum_i e_i^2 - 2\sum_i e_i e_{i-1} + \sum_i e_{i-1}^2}{\sum_i e_i^2}$,

и поскольку при больших n $\sum_i e_i^2 \approx \sum_i e_{i-1}^2$, получаем $DW \approx 2(1 - r_{i,i-1})$. Если e_i в точности равно e_{i-1}, то $DW = 0$; если $e_i = -e_{i-1}$, то $DW = 4$ во всех других случаях 0$< DW < 4$.

В случае, когда каждое отклонение e_i примерно совпадает с предыдущим отклонением e_{i-1}, каждое слагаемое в числителе величины DW близко к нулю. Сумма квадратов разностей отклонений в числителе будет намного меньше суммы квадратов отклонений в знаменателе, и поэтому статистика Дарбина-Уотсона окажется близкой к нулю. Рис.3.2 представляет такой случай – это случай положительной автокорреляции остатков первого порядка. Значение статистики Дарбина-Уотсона здесь равно 0,045, что очень мало и подтверждает статистическую зависимость отклонений e_i без всяких таблиц. Другой крайний случай возникает, когда точки наблюдений поочередно отклоняются в разные стороны от линии регрессии, и каждое следующее отклонение e_i имеет, как правило, противоположный знак, чем предыдущее отклонение e_{i-1}. В этом случае $(e_i - e_{i-1}) \approx 2e_i$, и

$$DW = \frac{\sum_i (2e_i)^2}{\sum_i e_i^2} = 4\frac{\sum_i e_i^2}{\sum_i e_i^2} = 4$$

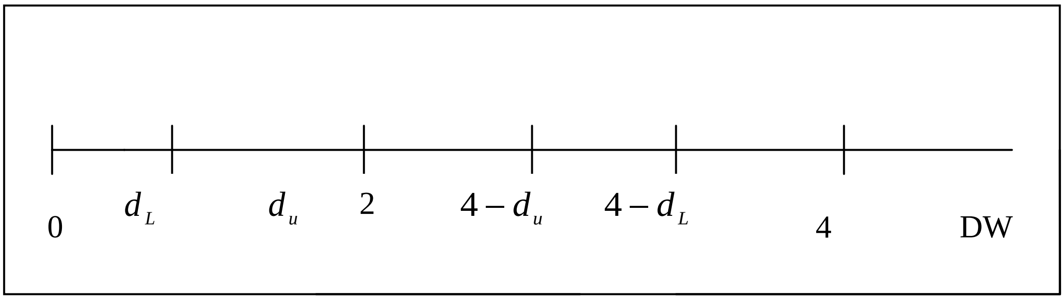
Это – случай отрицательной автокорреляции остатков первого порядка. Последняя достаточно редко встречается в экономическом анализе. Если рассматриваются временные ряды с годовыми данными, то подобную закономерность поведения последовательных отклонений довольно трудно проинтерпретировать. Однако она может встретиться при работе, например, с полугодовыми данными показателей с сезонным характером изменения. Наконец, если характер поведения отклонения случаен, можно предположить, что в половине случаев знак последовательных отклонений совпадает, а в половине – различен. Поскольку абсолютная величина их в среднем предполагается одинаковой, можно считать, что здесь в половине случаев e_i равно e_{i-1} , а в оставшейся половине равно $-e_{i-1}$.

Итак, при этом $DW \approx \frac{\sum_i 0.5*(2e_i)^2}{\sum_i e_i^2} = 0,5 \times 4 \times \frac{\sum_i e_i^2}{\sum_i e_i^2} = 2.$

Это показывает, что близость статистики Дарбина-Уотсона к двум является необходимым условием случайного характера отклонения от линии регрессии. Нужно, однако, иметь в виду, что в показателе DW сравниваются только соседние отклонения от регрессии, в то же время циклы изменения экономических переменных могут быть более или менее длительными, чем одна единица времени. Например, если рассматриваются поквартальные данные сельскохозяйственного производства (имеющего годовой цикл) и оценивается их линейная регрессии от времени, статистика Дарбина-Уотсона может быть близка к двум при выраженной регулярности отклонений зависимой переменной от линии регрессии.

Если статистика Дарбина-Уотсона близка к двум, мы считаем отклонения от регрессии случайными (хотя в действительности они могут и не быть таковыми). Это означает, что линейная функция вероятно, отражает реальную взаимосвязь; скорее всего, не осталось существенных не учтенных факторов, влияющих на зависимую переменную, и какая-либо другая нелинейная формула не превосходит по статистическим характеристикам данную линейную. Даже если доля дисперсии зависимой переменной, объясненной с помощью регрессии, при этом мала, можно ожидать, что другая часть этой дисперсии, оставшаяся необъясненной, порождена действием множества различных малых факторов и может быть описана как случайная нормальная ошибка. Но как определить, достаточно ли близка величина статистики DW к двум? Для этого имеются специальные таблицы, позволяющие при данном числе наблюдений и объясняющих переменных, для заданного уровня значимости, найти критическое значение статистики Дарбина-Уотсона.

Итак, статистика Дарбина-Уотсона применяется для проверки гипотезы об отсутствии автокорреляции остатков e_i первого порядка (нулевой гипотезы). Для этого по таблицам находятся (при данном уровне значимости, числе наблюдений и независимых переменных) доверительные интервалы в пределах, которых нулевая гипотеза принимается, отвергается или не может быть принята или отвергнута. Важно, что для статистики Дарбина-Уотсона существует два критических значения, меньше двух: нижнее d_1 как граница для признания положительной автокорреляции остатков и верхнее d_u как граница признания ее отсутствия. Для проверки гипотезы от отрицательной автокорреляции остатков этих критических значений отражается симметрично относительно числа 2:



Например, пусть оценена парная линейная регрессия по 15 наблюдениям, и DW =1,1. Зададим уровень значимости 15% и найдем по таблицам $d_i = 0{,}95$;

$d_u = 1{,}23$. нулевая гипотеза была принята при $d_u = 1{,}23, < DW < 2{,}77 = 4 - d_u$ и отвергнута при $DW < 0{,}95 = d_i$ или $DW > 3{,}05 = 4 - d_i$. Поскольку в данном случае DW лежит между d_u и d_i, нулевая гипотеза не может быть ни принята ни отвергнута. Если альтернативная гипотеза является гипотезой о положительной автокорреляции остатков (отрицательная из содержательных соображений отбрасывается), то критические значения $d_u = 1{,}23$ и $d_i = 0{,}95$ и соответствуют 2,5%-ному уровню значимости.

Как в общем случае выглядят примерно критические величины статистики DW? Если статистика DW находится приблизительно между 1,2-1,3 и 2,7-2,8, мы можем считать, что статистически значимая автокорреляция остатков отсутствует. В промежуточном случае достаточно надежный вывод быть не может. Если число наблюдений растет, то критические значения статистики Дарбина-Уотсона d_i и d_u приближаются к двум: для 60-70 наблюдений ее нижнее критическое значение d_i составляет примерно 1,4-1,5. Это верно для прежнего относительно малого числа объясняющих переменных; если число растет, то критическое значение DW становится меньше.

Итак, обобщая, если статистика Дарбина-Уотсона составляет 1,5-2,0-2,5, мы хотя и не можем быть абсолютно уверены, что отклонения от линии регрессии взаимно независимы, но обычно удовлетворяемся этим в проверке их независимости.

Статистика DW позволяет проверить некоррелированность отклонений от линии регрессии. Некоторые другие свойства этих отклонений (например, постоянство их дисперсии) могут быть также проверены с помощью специальных статистик. Мы не будем останавливаться на этом подробно, упомянув лишь о существовании самой проблемы. Рассуждения при этом могут быть подобными прежним: если значения тестовых статистик «плохие», то можно попытаться уточнить формулу связи, набор объясняющих переменных или процедуру оценивания.

Качество оценки.

Качество оценки уравнения регрессии проводится с определением показателя коэффициента детерминации

$$D = r^2 \quad \text{или} \quad R^2 = 1 - \frac{var(e)}{var(y)}$$

Максимальное значение коэффициента R^2 равно единице. Это происходит в том случае, когда линия регрессии точно соответствует всем наблюдениям, т.к. что

$$\hat{y}_i = y_i \text{ для всех } i \text{ и все остатки равны нулю.}$$

$$\text{Тогда } var(\hat{y}) = var(y), \rightleftarrows\rightleftarrows\ var(e) = 0 \ \text{ и } \ R^2 = 1.$$

Если в выборке отсутствует видимая связь между y и x, то коэффициент R^2 будет близок к нулю.

Пример вычисления R^2.

Уравнение регрессии $\hat{y} = 1{,}6667 + 1{,}5000x$ построено по наблюдениям x и y.

Набл.	X	У	$\hat{y}$	Е	$y-\bar{y}$	$y-\hat{y}$	$(y-\bar{y})^2$	$(\hat{y}-\bar{\hat{y}})^2$	e^2
1	1	3	3,1667	-0,1667	-1,6667	-1,5	2,7778	2,25	0,0278
2	2	5	4,6667	0,3333	0,3333	0,0	0,1111	0,00	0,1111
3	3	6	6,1667	-0,1667	1,3333	1,5	1,7778	2,25	0,0278
Сумма	6	14	14	0			4,6667	4,5	0,1667
Среднее	2	4,6667	4,6667	0			1,5556	1,5	0,0556

$$R^2 = 1 - \frac{var(e)}{var(y)} = 1 - \frac{0.0556}{0.5556} = 0.96\,.$$

3.6. Проверка гипотез, относящихся к коэффициентам регрессии $[a, b]$

Проверка гипотезы $H_0 : \beta = \beta_0$. Пусть в теоретической зависимости

$$Y = \alpha + \beta X + \varepsilon$$

случайный член ε распределен нормально с неизвестной дисперсией σ^2.

Хотя величина β и неизвестна, имеется основание предполагать, что она равна заданной величине β_0.

Выдвигаются гипотезы

$$\begin{cases} H_0 : \beta = \beta_0 \\ H_1 : \beta \neq \beta_0 \end{cases}$$

Задача заключается в проверке гипотезы H_0 на основании выборочных данных.

Пусть по выборочным данным получена оценка b.

В качестве критерия проверки гипотезы H_0 принимают случайную величину $t = \dfrac{t - \beta_0}{S_b}$ которая имеет распределение Стьюдента с $v = n - 2$ степенями свободы.

Вычисляется наблюдаемое значение критерия t. По таблице критических точек распределения Стьюдента по заданному уровню значимости α и числу степеней свободы $v = n - 2$ находят критическую точку $t_{\text{êð}}$.

Сравнивая наблюдаемое значение критерия критическим, можно принять или отвергнуть нулевую гипотезу.

Результаты оценивание регрессии совместимы не только с конкретной гипотезой $H_0 : \beta = \beta_0$, но и с некоторым их множеством.

Любое значение β, совместимое с оценкой b, удовлетворяет условию

$$\left| \frac{b - \beta}{S_b} \right| < t_{\text{êð}}, \text{ или } -t_{кр} < \frac{\beta - b}{S_b} < t_{кр}.$$

Решив это неравенство относительно β, получим

$$b - t_{кр} S_b < \beta < b + t_{кр} S_b,$$

т.е. доверительный интервал для величины β.

Посредине интервала лежит величина b. Границы интервала одинаково относят b, зависят от выбора уровня значимости и являются случайным числами. Доверительный интервал покрывает значение параметра β с заданной вероятностью $(1-\alpha)$, т.е.

$$P\left(b - t_{кр} S_b < \beta < b + t_{кр} S_b\right) = 1-\alpha$$

Проверка гипотезы $H_0 : \beta = 0$. Пусть по выборке получена оценка коэффициента регрессии b. Гипотеза $H_0 : \beta = 0$ используется для установления значимости коэффициента регрессии b.

Соответствующая t- статистика есть оценка коэффициента регрессии, деленная на ее стандартную ошибку, т.е.

$$t = \frac{b-0}{S_b} = \frac{b}{S_b}.$$

Величина t имеет распределение Стьюдента с $v = n-2$ степенями свободы.

Наблюдаемому (расчетному) значению критерия t соответствует определенная *значимости* t, которую можно вычислить в Excel помощью функции *Значимость* t *СТЬЮДРАСП* (t; v; 2). Из сравнения значимости t с заданным стандартным уровнем значимости получаем:

· если значимость t больше стандартного уровня, то b незначим;

если значимость t меньше стандартного уровня, то b значим [9].

Пример 3.2. Зависимость между стоимостью валовой продукции y и баллом бонитета земли x по данным примера 9.1 имеет вид

$$\hat{y} = -340{,}05 + 13{,}46x$$

Оценим значимость коэффициента регрессии $b = 13{,}46$ и построим доверительный интервал для β при 5%- ном уровне значимости.

Наблюдаемое значение критерия $t = \dfrac{b}{S_b} = \dfrac{13{,}46}{4{,}859959478} = 2{,}76$.

Значимость $t = 0{,}0277$, соответствующую расчетному значению критерия $t = 2{,}76$, определяем с помощью функции Значимость t= *СТЬЮДРАСП* (t; v; 2), где $v = 7$. Поскольку *значимость* $t = 0{,}0277 < 0{,}05$, то коэффициент регрессии $b = 13{,}46$ значим.

При $\alpha = 0{,}05$ критическое значение критерия $t_{кр} = 2{,}3646$ определяем с помощью функции $t_{ёð}$ =*СТЬЮДРАСП* $(\alpha;\ \ v)$. Доверительный интервал для β $13{,}46 - 2{,}3646 \cdot 4{,}85 < \beta < 13{,}46 + 2{,}3646 \cdot 4{,}85,$ или $1{,}99169 < \beta < 24{,}92831$.

3.7. Нелинейные модели

Часто приходится использовать нелинейные функции регрессии двух видов:

1. Регрессии, нелинейные относительно включенных в анализ объясняющих переменных:

Полином второй, редко третьей степени $y = a + bx + cx^2 + \varepsilon$

Гипербола $y = a + b/x + \varepsilon$.

Эти модели сводятся к линейным заменой переменных: $z = x^2$ для полинома и $z=1/x$ для гиперболы. После этого можно использовать функцию ЛИНЕЙН и сервис Регрессия, выделяя в качестве влияющих переменных x и z для полинома и z для гиперболы.

2. Регрессии, нелинейные по оцениваемым параметрам:

Степенная $y = ax^b\varepsilon$;

Показательная $y = ab^x\varepsilon$;

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

Здесь $\varepsilon=1+u$. Эти модели могут быть линеаризованы логарифмированием, после чего можно использовать функцию ЛИНЕЙН и сервис *Регрессия*. Например, показательная функция преобразуется в $ln(y) = ln(a) + xln(b) + ln(\varepsilon)$, или, *после переименования* $z=A+cx+v$.

После нахождения коэффициентов A и c можно вычислить z^=$A+cx$ и

y^=exp(z^).

3.8. Пакет анализа EXCEL (программа «Регрессия»)

Построение линейной регрессии, оценивание ее параметров и их значимости можно выполнить значительно быстрее при использовании пакета анализа Excel (Программа «Регрессия»).

Пример 3.3. Имеются данные о стоимости валовой продукции y и балл бонитета земли x в фермерских хозяйствах (усл. ед.):

x	60	53	54	61	55	46	58	52
y	375	360	350	600	420	280	390	410

Рассмотрим интерпретацию полученных результатов в общем случае (k объясняющих) по данным примера 3.6.

I

Регрессионная статистика	
множественный R	0,72
R - квадрат	0,5228
нормированный R - квадрат	0,454
стандартная ошибка	64,88
наблюдения	9

В таблице *Регрессионная статистика* приводятся значения:

1. *Множественный* R - коэффициент множественной корреляции $R = \sqrt{R^2}$.

2. *D- коэффициент* детерминации R^2.

3. *Нормированный* R*-квадрат*- скорректированный R^2 с поправкой на число степеней свободы.

4. *Стандартная ошибка*- стандартная ошибка регрессии S .

5. *Наблюдения*- число наблюдений n .

II

Дисперсионный анализ					
		df *SS*	*MS*	*F*	Значимость *F*
Регрессия		32289,3	32289,3	7,67	0,027710423
Остаток		29466,27	4209,467		
Итого		61755,56			

В таблице *Дисперсионный анализ* приведены:

1. Столбец df - число степеней свободы, равное:

$df = k$ для строки *Регрессия;*

$df = n - k - 1$ для строки *Остаток*;

$df = n - 1$ для строки *Итого*.

2. Столбец SS - сумма квадратов отклонений, равная:

$\sum(\hat{y} - \bar{y})^2$ для строки *Регрессия*;

$\sum(y - \hat{y})^2$ для строки *Остаток*;

$\sum(y - \bar{y})^2$ для строки *Итого*.

3. Столбец MS - дисперсии, определяемые по формуле $MS = SS / df$:

- *факторная* для строки *Регрессия*;

-*остаточная* для строки *Остаток*;

4. Столбец F - расчетное значение F - критерия, вычисляемое по формуле $F = MS$ *(регрессия)*/ MS *(остаток)*

5. Столбец *Значимости* F -значение уровня значимости, соответствующее вычисленной F - статистике: *Значимость* F *=FРАСП (F-статистика;* df *(регрессия);* df *(остаток)).* Если значимость F меньше стандартного уровня значимости, то R^2 статистически значим.

III

	Коэффициент	Стандартная ошибка	$t-$ статистика	$P-$ значение	Нижние 95%	Верхние 95%
y	-340,04	265,48	-1,28	0,241	-967,81	287,71
x	13,46	4,85996	2,7695	0,0277	1,968	24,952

В этой таблице указаны:

1. *Коэффициенты*- значения коэффициентов a, b .

2. *Стандартная ошибка* – стандартные ошибки коэффициентов регрессии $S_{a,}$ S_b .

3. t *-статистика*- расчетные значения t - критерия, вычисляемые по формуле t *-статистика = Коэффициенты /Стандартная ошибка.*

4. P - значение (значимость t)-это значение уровня значимости, соответствующее вычисленной t - статистике:

P *- значение = СТЬЮДРАСП (*t *- статистика;* df *(остаток)).*

Если P*-значимость меньше стандартного уровня значимости, то соответствующий коэффициент статистически значим.*

1. *Нижние 95% и Верхние* 95%-нижние и верхние границы 95%-них доверительных интервалов для коэффициентов теоретического уровня линейной регрессии.

IV

Вывод остатка		
Наблюдение	Предсказанное y	Остатки e
1	467,556	-92,556
2	373,335	-23,335
3	386,795	-26,795
4	481,016	118,983
5	400,255	19,744
6	279,114	0,885
7	440,635	-50,635
8	359,875	50,124
9	346,415	3,584

В таблице *Вывод остатка* указаны:

1. *Наблюдение*- номер наблюдения.
2. *Предсказанное* y - расчетные значения зависимой переменной.
3. *Остатки* e - разница между наблюдаемыми и расчетными значениями зависимой переменной.

Используя результаты работы пакета анализа Excel (Программа «Регрессия»), проанализируем зависимость расходов на питание от величины душевого дохода.

Результаты регрессионного анализа принято записывать в виде

$$\hat{y} = -340{,}05 + 13{,}46x, \quad R^2 = 0{,}52$$

Коэффициенты регрессии $a = -340{,}05$ и $b = 13{,}46$. Направление связи между y и x определят знак коэффициента регрессии $b = 13{,}46$ т.е. связь является прямой и положительной. Коэффициент $b = 13{,}46$ показывает, что при увеличении балла бонитета на 1 балл стоимость валовой продукции увеличивается на 13,46 у.е.

Оценим значимость коэффициентов полученной модели. Значимость коэффициентов (a, b) проверяется по t - тесту:

P-значение $(a) = 0{,}241 < 2{,}76$

P-значение $(b) = 0{,}0277 < 2.76$

Следовательно, коэффициенты (a, b) значимы при 1%-ном уровне, а тем более при 5%-ном уровне значимости. Таким образом, коэффициенты регрессии значимы и модель адекватна исходным данным.

Результаты оценивания регрессии совместимы не только с полученными значениями коэффициентов регрессии, но и с некоторым их множеством (доверительным интервалом). С вероятностью 95% доверительные интервалы для коэффициентов есть (-967,81-287,71) для a и (1,968-24,952) для b.

Качество модели оценивается коэффициентом детерминации R^2. Величина $R^2 = 0{,}52$ означает, что фактором балла бонитета можно объяснить 52% вариации стоимости валовой продукции.

Значимость R^2 проверяется по F - тесту:

Значимость $F = 0{,}0277 < 2{,}76$.

Следовательно, R^2 значим при 1% -ном уровне, а тем более при 5%- ном уровне значимости.

В случае парной линейной регрессии коэффициент корреляции можно определить, как $r = \sqrt{R^2} = 0{,}72$. Полученное значение коэффициента корреляции свидетельствует, что связь между расходами на питание и душевыми доходом тесная.

Взаимозависимости критериев. В парном регрессионном анализе эквивалентны t - критерий для $\boldsymbol{H_0: \beta = 0}$; t - критерий для $\boldsymbol{H_0: \rho = 0}$; F - критерий для R^2.

$$t_b = \frac{b}{S_b}, \quad t_r = r\sqrt{\frac{n-2}{1-r^2}}, \quad F = \frac{R^2(n-2)}{1-R^2}$$

Связь между критериями выражается равенством

$$t_b = t_r = \sqrt{F},$$

причем для критических значений критериев при любом уровне значимости $(t_b)_{кр} = (t_r)_{кр} = \sqrt{F_{кр}}$

и эти критерии дают один и тот же результат.

Вывод. Поверки значимости коэффициента b в парной линейной регрессии, коэффициента корреляции r и коэффициента детерминации R^2 – эквивалентны [9].

3.9. Матричный метод наименьших квадратов (МНК)

МНК (метод наименьших квадратов) в матричном виде основан на представлении множеств X, Y, остатков E и параметров линейной модели в виде векторов, над которыми затем проводятся операции [2].

Рассмотрим более подробно те случаи, что одна переменная зависит от другой.

$$y = \sum_{i=1}^{m} \beta_i x_i + \varepsilon_i$$

где: $x_1, x_2, \ldots, x_m$ — независимые переменные, $y_1, y_2, \ldots, y_m$ — зависимые переменные и непостоянные части модели.

Записанную в векторном виде или в виде системы линейных уравнений, называют *схемой Гаусса-Маркова*.

Пусть $(x_1, x_2, \ldots, x_i ; i = 1,2, \ldots, p)$ будет вектором независимых переменных, состоящим из наблюдаемых величин в m.

Пусть $(y_1, y_2, \ldots, y_p)$ - вектор будет значением переменной Y в эксперименте p. В таком случае стандартный вид регрессионной модели выглядит следующим образом:

$$Y_i = \sum_{i=1}^{m} \beta_k x_k + \varepsilon_i (i = 1, p)$$

в этой модели мы предполагаем, что $x_i = 1, i = \overline{1,p}$, т.е. β_k - свободный член.

Значение параметра метода наименьших квадратов состоит из значений $(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_m)$. Векторы $(\beta_1, \beta_2, \ldots, \beta_m)$ должны быть такими, чтобы сумма квадратов была минимальной:

$$S = \sum_{i=1}^{p} \varepsilon_i^2 = \sum_{i=1}^{p} (Y_i - \sum_{k=1}^{m} \beta_k x_k)^2$$

Модель регрессии в матричной форме выглядит следующим образом:

$$Y = XB + E$$

здесь: $Y = \begin{bmatrix} Y_1 \\ . \\ . \\ Y_p \end{bmatrix}; X = \begin{bmatrix} X_{11} & . & X_{1m} \\ . & . & . \\ X_{p1} & . & X_{pm} \end{bmatrix}; B = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}; \mathrm{E} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{bmatrix};$

Уравнение в виде матрицы выглядит следующим образом:

$$S = \mathrm{E}^t\mathrm{E} = (Y - XB)^t(Y - XB)$$

здесь, E^t - состоит из транспонированной матрицы Е.

Сумма остатков регрессии на основе условия минимизации

$$\begin{cases} \dfrac{\partial S}{\partial \beta_i} = 0; (i = 1,2,\ldots,m); \\ \dfrac{\partial S}{\partial B} = -2(x^t y - x^t x B) = 0 \end{cases}$$

из этого получим $(x^t x)B = x^t y$, здесь $B = (x^t x)^{-1} x^t y$. Здесь $(x^t x)^{-1}$ - обратная матрица к $x^t x$.

3.10. Теорема Гаусса-Маркова

Метод наименьших квадратов, согласно *теореме Гаусса-Маркова*, приведенный к линейному преобразованию матриц или к системе линейных уравнений, обеспечивает наилучшую несмещенную, эффективную и сходящуюся к пределу (“состоятельную”) оценку вектора параметров, т.е. наилучшее качество линейной модели, если соблюдаются условия (по [1]):

1. Линейная модель соответствует действительности.
2. Существует дисперсия регрессора.
3. Математическое ожидание возмущения равно нулю: $E(u_i) = 0$.
4. Возмущение имеет нормальное распределение.
5. Равенство ожидаемых значений дисперсий возмущений в разных диапазонах X: $E(u^2) = Const$. Это свойство называется *гомоскедастичностью*, его несоблюдение – *гетероскедастичностью*. Отклонение от гомоскедастичности проверяется по тесту *Голдфелда-Квандта*

$GQ = \Sigma e_1^2 / \Sigma e_2^2$

где Σe_1^2 и Σe_2^2 – суммы квадратов остатков (отклонений) в первой и последней трети (или в половинах) диапазона X; *большая сумма делится на меньшую!* GQ сравнивают с критерием Фишера для заданных уровня значимости и количества измерений; гипотеза о гомоскедастичности принимается при $GQ < 4,35$.

6. Отсутствие *автокорреляции*, т.е. взаимозависимости возмущений. Ее оценивают, вычисляя *статистику Дарбина-Уотсона остатков е:*

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=2}^{n} e_i^2}$$

для которой вычислены критические значения при различных уровнях значимости и числе измерений. Приблизительно DW=0…1 означает положительную автокорреляцию, 3…4 отрицательную автокорреляцию, DW=1,5…2,5 позволяет принять гипотезу об отсутствии автокорреляции, DW=1…1,5 и DW=2,5…3 не позволяют принять гипотезу о наличии или отсутствии автокорреляции. Наличие автокорреляции означает, что аппроксимирующая функция подобрана неверно, или же требуется применение других методов и моделей.

Статистику Дарбина-Уотсона можно вычислить по формуле $DW = 2(1-R_{авт})$, где $R_{авт}$ - коэффициент автокорреляции, вычисляемый с помощью функции КОРРЕЛ: задать в окне *Массив1* диапазон остатков с номерами 1 : n-1, а в окне *Массив2* диапазон 2 : n.

Понятия “гетероскедастичность” и “автокорреляция” актуальны, если массивы данных упорядочены, что имеет место для временных рядов. “Пространственные” данные можно искусственно упорядочить, например, отсортировав их по возрастанию какой-либо переменной; при этом можно выявить кластеры с аномальной дисперсией остатков, что может означать неоднородность выборки или неадекватность модели.

Считается, что гетероскедастичность может привести к снижению эффективности оценок коэффициентов, и надо ее искусственно подавлять: делить остатки в таблице 3.3 на их стандартные отклонения в диапазонах, а затем минимизировать сумму их квадратов. Эта технология называется Взвешенный метод наименьших квадратов (ВМНК) и обычно используется в матричном варианте МНК. При обнаружении автокорреляции остатков применяется *Обобщенный метод наименьших квадратов ОМНК*, основанный на преобразовании матриц, но с учетом корреляций остатков.

Опорные слова

Эконометрика, математическая статистика, объясняющие переменные, фактор, параметр, линейная зависимость, формула связи, коэффициент корреляции, ковариация, число степеней свободы, метод наименьших квадратов, объясняющая переменная, линии регрессии, число наблюдений, независимая переменная, уравнение регрессии, коэффициент детерминации.

Контрольные вопросы

1. Какие модели называется эконометрическими?
2. Что изучает эконометрика?
3. Какое отличие между эконометрическими и не эконометрическими моделями?
4. Что изучают при помощи корреляционных моделей?
5. Как определяется коэффициент корреляции?
6. Как определяется значимость коэффициента корреляции?
7. Что изучают при помощи регрессионных моделей?
8. Как определяются неизвестные параметры?
9. Объясните сущность метода наименьших квадратов.
10. Как определяется значимость из уравнения регрессии?
11. В чем разница между анализами парной и множественной регрессии?
12. Что такое мультиколлинеарность и автокорреляция?
13. Как можно предотвратить мультиколлинеарность и автокорреляцию?
14. Каким образом прогнозируют с помощью эконометрической модели?

Контрольные задания

1. По 50 предприятиям рассчитана модель парной регрессии, характеризующая зависимость стоимости реализованной продукции (y, у.е.) от стоимости основных фондов (x, у.е.):

$Y = -62711 + 0{,}93x + e$,

$mb_0 = -19562$, $mb_1 = -0{,}086$.

Известно также, что $r_{xy} = 0{,}842$.

1) Определите коэффициент детерминации и сделайте по его величине вывод.

2) Постройте доверительный интервал для коэффициента регрессии и свободного члена уравнения:

а) с доверительной вероятностью 0,95;

б) с доверительной вероятностью 0,99.

3) Оцените значимость параметров и уравнения регрессии в целом.

2. Изучается зависимость потребления продукта C(Y) от среднедушевого дохода (x) по данным 20 семей. При оценке регрессионной модели были получены промежуточные результаты:

$$\sum\left(y - \hat{y}_x\right)^2 = 1{,}2$$

$$\sum\left(y - \bar{y}\right)^2 = 6{,}3.$$

Задание:

1) Определите индексы корреляции, детерминации и сделайте выводы.

2) Постройте таблицу дисперсионного анализа для расчета значения F-критерия Фишера.

3) Сравните фактическое и табличное значение F-критерия и сделайте выводы.

3. Изучается зависимость валовой продукции сельского хозяйства *y* от площади сельскохозяйственных угодий *x*. По 61 наблюдению были получены следующие варианты уравнения регрессии:

1) $Y = -30706 + \underset{(10{,}74)}{34{,}332}x + e$;

2) $\ln y = 2{,}578 + \underset{(11{,}68)}{1{,}086}\ln x + e$, $R^2 = 0{,}698$;

3) $y = 50889 + \underset{(1{,}003)}{13{,}23}x + \underset{(1{,}65)}{0{,}0012}x^2 + e$, $R^2 = 0{,}677$;

4) $y = 428573 - 1246743688/x + e$, $R^2 = 0{,}435$.

В скобках указаны фактически наблюдаемые значения *t*-критерия Стьюдента.

1) Определите коэффициент детерминации для первого уравнения.

2) Найдите коэффициенты корреляции по каждому уравнению.

3) Определите коэффициенты эластичности для каждого из уравнений при площади сельскохозяйственных угодий 5000 и 10000 га.

4) Выберите наилучший вариант уравнения регрессии для данной задачи.

4. Сделайте выводы по полученным результатам. Для продукции трех видов А1, А2 и А3 построены следующие модели зависимости удельных постоянных расходов от объема выпускаемой продукции:

$Y_{A1} = 200,$

$Y_{A2} = 70 + 0,3x,$

$Y_{A3} = 30x^{0,2}.$

1) Определите коэффициенты эластичности по каждому виду продукции.

2) Сравните при x =150 эластичность затрат для продукции видов А2 и А3.

3) Каким должен быть объем выпускаемой продукции x, чтобы коэффициенты эластичности для продукции А2 и А3 были равны.

5. По 60 сельскохозяйственным предприятиям изучалась зависимость выручки от реализации продукции (тыс. у.е.) от численности работающих (чел.). Получены следующие уравнения регрессии:

а) y= 106377 + 435 x; r = 0.764;

(4,29) (9,03)

б) lny =7,81 + 0,801lnx; r = 0,787.

(16,08) (9,73)

В скобках приведены фактически наблюдаемые значения t-критерия Стьюдента.

а) Определите коэффициенты детерминации по каждому уравнению.

б) Определите коэффициенты эластичности, учитывая, что x = 402.

в) Оцените значимость уравнений регрессии при *a* = 0,05.

г) Обоснуйте выбор лучшего уравнения регрессии. Сделайте выводы по полученным результатам.

6. Для 15 наблюдений изучалась зависимость вида: y=ax+b. Получены следующие данные для преобразованных в логарифмах переменных:

ΣXY = 2,1927; EX² = 1,1693; ΣX = 2,0792;

CY = 4,9255; Σ (y-ýx)² =1,2; gy = 2,45.

а) Найдите параметр b.

б) Найдите показатель корреляции.

в) Оцените значимость уравнения регрессии при а = 0,05.

Сделайте выводы по полученным результатам.

7. По группе 25 предприятий, производящих однородную продукцию, получено уравнение регрессии себестоимости единицы продукции у (тыс. у.е.) от уровня технической оснащенности х (млн. у.е.): $y = 30 + \frac{600}{x}$

Доля остаточной дисперсии в общей дисперсии составила 0,16.

Определите:

1) коэффициент эластичности, предполагая, что стоимость активных производственных фондов составляет 25 млн. у.е.;

2) индексы корреляции и детерминации:

3) оцените значимость уравнения регрессии при помощи критерия Фишера.

Сделайте выводы по полученным результатам.

8. Зависимость среднегодовой производительности труда от возраста рабочих характеризуется моделью: y-$a+bx+cx^2$. Ее использование привело к результатам, представленным в следующей таблице.

№	*Производительность труда рабочих, млн. у.е., у*	
	Фактическая	***Расчетная***
1.	2,3	2,0
2	1,8	2,0
3.	2,3	2,3
4.	2,5	2,4
5.	2,6	2,5
6.	2,1	2,2
7.	2,2	2,3
8.	1,9	2,0
9.	2,1	2,0
10.	1,8	1,9

1) Оцените качество построенной модели, определив среднюю ошибку аппроксимации.

2) Определите индекс корреляции, детерминации и средний коэффициент эластичности.

3) С помощью F-критерия Фишера при а = 0,05 оцените значимость уравнения регрессии.

Сделайте выводы по полученным результатам.

9. По 30 наблюдениям зависимость спроса на товар А от цены на него характеризуется следующим уравнением: *lg y = 2,15-0,451gx*.

Доля остаточной дисперсии в общей дисперсии составила 15%.

1) Запишите данное уравнение в виде степенной функции.

2) Оцените эластичность спроса на товар в зависимости от его цены.

3) Определите индекс корреляции.

4) Оцените значимость уравнения регрессии через F-критерий Фишера.

Сделайте выводы по полученным результатам.

10. Известна линейная модель регрессии, характеризующая зависимость между результативной переменной y и независимой переменной x:

y=26,89 +0,365x,

(0,0483)

Известно также, что n =35, *r* = 0,796.

В скобках приведена стандартная ошибка для коэффициента регрессии.

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

а) с вероятностью 0,9;

б) с вероятностью 0,95.

2) Охарактеризуйте статистическую надежность уравнения регрессии с использованием F- критерия Фишера при уровне значимости а = 0,05.

3) Оцените значимость коэффициента корреляции по *t* - критерию Стьюдента при уровне значимости α = -0.01.

Сделайте выводы по полученным результатам.

11. Линейная модель регрессии, характеризующая зависимость между результативной переменной y (удой молока на корову за год, ц) и независимой переменной x (среднегодовое поголовье коров, гол.):

$$y=26{,}609+0{,}014x + \mathcal{E}.$$

Известно также, что n=15, $\bar{x} = 1092, \bar{y} = 61{,}5, r = 0{,}657, m_b = 0{,}0043$

1) Построить доверительный интервал для коэффициента регрессии в линейной модели: а) с вероятностью 99%, б) с вероятностью 95%.

2) Охарактеризуйте статистическую надежность уравнения регрессии с использованием F-критерия Фишера при уровне значимости 0,01 и 0,05.

3) Оцените значимость коэффициента корреляции по t - критерию Стьюдента при уровне значимости 0,01 и 0,05.

4) Определите средний коэффициент эластичности для данной линейной модели. Сделайте выводы по полученным результатам.

12. Зависимость расходов предприятия у (тыс. у.е.) от объема производства х (шт.) для двух видов продукции А и В характеризуется данными, представленными в таблице.

Уравнение регресссии	Показатель корреляции	Число наблюдений
$y_A=120+0.6x$	0,80	35
$y_B=40x^{0.5}$	0,75	30

1) Сравните эластичность расходов от объема производства для продукции А и В при выпуске продукции в 400 единиц.

2) Определите, каким должен быть выпуск продукции А, чтобы эластичность расходов совпадала с эластичностью расходов на продукцию В.

3) Поясните смысл величин 0,6 и 0,5 в уравнениях регрессии.

4) Оцените значимость каждого уравнения регрессии с помощью F – критерия Фишера. Сделайте выводы по полученным результатам.

13. Зависимость объема производства у (млн. у.е..) от численности занятых на производстве рабочих х (чел.) по 60 предприятиям характеризуется следующим уравнением регрессии:

$$Y=-1,3334+1,0083x-0,000482x^2,$$
$$(0,336) \quad (0,0002)$$

доля остаточной дисперсии в общей дисперсии составляет 22,0%.

В скобках указаны средние ошибки параметров уравнения регрессии. 1) Определите коэффициенты эластичности, предполагая, что численность занятых на производстве рабочих составляет 800 и 1200 человек.

2) Оцените значимость параметров уравнения регрессии при помощи t - критерия Стьюдента при уровне значимости 0,01 и 0,05.

3) Определите индекс корреляции.

4) Оцените значимость уравнения регрессии при помощи F-критерия Фишера при уровне значимости 0,01 и 0,05.

14. По 20 семьям получена информация, представленная в следующей таблице:

Показатель	Среднее значение	Коэффициент вариации
Годовой расход на личное потребление, *у.е.*	3378	32
Годовой располагаемый доход, *у.е.*	3508	28

Фактическое значение F-критерия Фишера составило 72.

1) Постройте уравнение линейной регрессии.

2) Найдите средний коэффициент эластичности.

3) Определите линейный коэффициент детерминации. 4) С вероятностью 0,95 постройте доверительный интервал ожидаемого значения годового расхода в предположении роста годового дохода на 10% от своего среднего уровня.

Сделайте выводы по полученным результатам.

15. По 22 торговым предприятиям района зависимость объема продаж y (млн. у.е.) от расходов на рекламу товара x (млн. у.е.) характеризуется уравнением регрессии y=58,5 +4,5x и следующими данными:

$\sigma_x = 2{,}9, \sigma_у = 8{,}1, \sigma_{ост} = 3{,}5.$

1) Определите коэффициенты корреляции и детерминации.

2) Найдите стандартную ошибку оценки коэффициента регрессии.

3) Оцените значимость коэффициента регрессии через *t*-критерий Стьюдента при уровне значимости 0,01.

4) Оцените значимость уравнения регрессии в целом.

5) Определите доверительный интервал для коэффициента регрессии.

Сделайте выводы по полученным результатам.

16. Имеются данные о материальных затратах (тыс. у.е.) и валовой продукции на 1 га сельхозугодий (тыс. у.е.) по 15 сельскохозяйственным предприятиям.

№	Валовая продукция на 1 га сельхозугодий (тыс. у.е.)	Материальные затраты на 1 га сельхозугодий (тыс. у.е.)
1.	70,36	36,44
2	97,73	52,25
3.	52,36	23,97
4.	63,45	32,11
5.	76,59	35,13
6.	62,79	27,38
7.	104,56	59,44
8.	72,84	39,61
9.	59,37	19,97
10.	80,70	46,62
11.	88,86	56,31
12.	83,10	44,46
13.	99,77	49,29
14.	59,10	22,83
15.	67,81	37,24

1) Постройте график зависимости между переменными.

2) Рассчитайте параметры линейного уравнения регрессии методом наименьших квадратов.

3) Оцените качество уравнения регрессии с помощью средней ошибки аппроксимации.

4) Оцените тесноту связи между переменными с помощью показателей корреляции и детерминации. Найдите средний коэффициент эластичности.

5) Охарактеризуйте статистическую надежность результатов регрессионного анализа с использованием F - критерия Фишера при уровне значимости 0,01 и 0,05.

6) Оцените значимость коэффициентов регрессии и корреляции по *t*-критерию Стьюдента при уровне значимости 0,01 и 0,05.

7) Определите прогнозное значение результативного признака, если возможное значение факторного признака составит 1,2 от его среднего уровня.

8) Постройте доверительный интервал прогноза, определите диапазон прогноза. Сделайте выводы по полученным результатам.

17. Изучается зависимость выручки от реализации на 1 га сельскохозяйственных угодий, тыс. у.е. (у) от стоимости основных фондов на 1 га сельскохозяйственных угодий, тыс. у.е. (х). По 20 наблюдениям были получены следующие результаты.

σ_x= 16,38; σ_y= 7,509; $\sum x = 878{,}1$; $\sum y = 858{,}4$; $\sum xy = 39647{,}0$

1) Постройте парное уравнение регрессии.

2) Рассчитайте коэффициенты корреляции и детерминации для линейного уравнения.

3) Определите средний коэффициент эластичности.

4) Оцените значимость коэффициента корреляции по *t* - критерию Стьюдента при уровне значимости о 0,05.

5) Охарактеризуйте статистическую надежность данного уравнения регрессии с использованием F- критерия Фишера при уровне значимости 0,05.

18. Моделирование зависимости розничного товарооборота (млн. у.е.) магазинов среднесписочного числа работников по уравнению y = alnx +b привело к следующим результатам, представленным ниже.

Среднесписочная численность работников, заанятых в магазинах торговой сети, тыс.чел.	Розничный товарооборот, млн. у.е., y	
	Фактическая	*Расчетаная*
91	193,153	212,204
109	227,331	237,036
131	274,84	262,330
150	309,217	280,963
175	315,146	302,171
231	332,033	340,368
251	344,675	351,792
270	352,174	361,831

1) Оцените качество модели, определив среднюю ошибку аппроксимации.

2) Найдите показатель тесноты связи y с исследуемым в данной модели фактором.

3) Рассчитайте F-критерий Фишера.

4) Определите прогнозное значение результативного признака, если возможное значение факторного признака составляет 1,3 от его среднего уровня.

5) Сделайте выводы по полученным результатам.

Тесты для главы №3

1. Суть метода наименьших квадратов состоит в минимизации:

а) суммы отклонений фактических значений результативного признака от теоретических значений, найденных по уравнению регрессии;

б) дисперсии результативного признака;

в) суммы абсолютных отклонений фактических значений результативного признака от теоретических значений, найденных по уравнению регрессии;

г) суммы квадратов отклонений фактических значений результативного признака от теоретических значений, найденных по уравнению регрессии.

2. Спецификация эконометрической модели заключается в (выберите по крайней мере один ответ):

а) установлении в математической форме типа уравнения связи между переменными;

б) обосновании состава зависимых и независимых переменных;

в) определении параметров уравнения связи;

г) оценке тесноты связи между переменными.

3. Классический подход к оцениванию параметров уравнения регрессии основан на:

а) методе наименьших квадратов;

б) методе максимального правдоподобия;

в) методе абсолютных отклонений;

г) пошаговом регрессионном анализе.

4. Наиболее наглядным способом выбора типа уравнения парной регрессии является:

а) аналитический;

г) экспериментальный.

б) графический;

в) описательный;

5. Если каждому значению одной переменной величины соответствует определенное значение другой переменной величины, то связь между ними является:

а) статистической;

б) корреляционной;

в) функциональной;

г) случайной.

6. Если определенному значению одной переменной величины соответствует распределение значений другой переменной величины, то связь между ними является:

а) статистической;

б) корреляционной;

в) функциональной;

г) случайной.

8. Линейная регрессионная зависимость между уравнением:

а) y = a + bx;

б) y = a. bx;

в) y = a + b lnx;

г) y = lna + lnbx.

9. Коэффициент регрессии линейного парного уравнения:

а) показывает, на сколько единиц в среднем изменится результативный признак, с увеличением факторного признака на одну единицу;

б) оценивает статистическую значимость уравнения регрессии;

в) показывает, на сколько процентов изменится в среднем результативный признак если факторный признак изменится на 1%;

г) показывает, на сколько единиц изменится в среднем результативный признак, если факторный признак изменится на 1%.

10. Коэффициент эластичности линейного парного уравнения:

а) показывает, на сколько единиц в среднем изменится результативный признак с изменением факторного признака на одну единицу;

б) оценивает статистическую значимость уравнения регрессии;

в) показывает, на сколько процентов изменится в среднем результативный признак, если факторный признак увеличится на 1%;

г) показывает, на сколько единиц изменится в среднем результативный признак, если факторный признак изменится на 1%.

11. Значение коэффициента по формуле r =

а) коэффициента детерминации;

б) парного коэффициента корреляции;

в) частного коэффициента корреляции;

г) множественного коэффициента корреляции.

12. Коэффициент детерминации, рассчитанный для линейного уравнения парной регрессии, составил 0,64. Следовательно, коэффициент корреляции равен (выберите, по крайней мере, один ответ):

а)-0,8, если $b <0$;

б)-0,36, если $b < 0$;

в) 0,8, если $b > 0$;

г) 0,36, если $b > 0$.

13. Коэффициент детерминации рассчитывается как отношение дисперсии результативного признака, объясненной регрессией, к дисперсии результативного признака.

а) средней;

б) остаточной;

в) общей;

г) факторной.

14. Коэффициент корреляции является статистически значимым (существенным), если:

а) он отличен от нуля;

б) вероятность того, что он равен нулю мала;

в) вероятность того, что он не равен нулю мала;

г) он равен коэффициенту регрессии.

15. Если абсолютное значение коэффициента корреляции, рассчитанное по линейному уравнению регрессии, равно единице, то связь между переменными:

а) нелинейная и функциональная;

б) линейная и функциональная;

в) отсутствует;

г) прямая.

16. Если значение коэффициента корреляции, рассчитанное по линейному уравнению регрессии, равно нулю, то:

а) связь между переменными нелинейная и функциональная;

б) связь между переменными линейная и функциональная:

в) связь между переменными отсутствует;

г) отсутствует линейная связь между переменными.

17. Значение линейного коэффициента корреляции может изменяться в промежутке:

а) [-0; +∞];

б) [-1; +1];

В) [0; +∞];

г) [-0; 0].

18.Знак коэффициента корреляции совпадает со знаком коэффициента регрессии и показывает:

а) силу связи между переменными;

б) направление связи между переменными;

в) отсутствие связи между переменными;

г) возможность связи между переменными.

19. Коэффициент детерминации характеризует:

а) качество линейной модели, исходя из относительных отклонений по каждому наблюдению;

б) долю дисперсии результативного признака, объясняемую регрессией, в общей дисперсии результативного признака;

в) тесноту связи между результативным и факторным признаками;

г) долю дисперсии и результативного признака, вызванную влиянием не учтенных линейной модели факторов.

20. Качество эконометрической модели можно оценить с помощью:

а) коэффициента детерминации

б) коэффициента регрессии;

в) средней ошибки аппроксимации А

г) коэффициента эластичности.

21. Значимость парного линейного уравнения регрессии оценивает

а) F-критерий Фишера;

б) коэффициент корреляции r ух;

в) t -критерий Стьюдента;

г) коэффициент детерминации r ух.

22. Если фактически наблюдаемое значение F-критерия Фишера превышает критическое значение, то следует вывод, что:

а) построенная эконометрическая модель является статистически значимой;

б) зависимость между изучаемыми переменными тесная;

в) построенная эконометрическая модель является статистически незначимой;

г) невозможно использовать построенную модель для описания зависимости между переменными.

23. Остаточная сумма квадратов отклонений равна нулю:

а) когда правильно подобрана регрессионная модель;

б) когда признаки независимые;

в) когда между признаками функциональная связь;

г) никогда.

24. Сумма квадратов отклонений, объясненная регрессией, равна нулю:

а) когда правильно подобрана регрессионная модель;

б) когда признаки независимы;

в) когда между признаками функциональная связь;

г) никогда.

25. Сумма квадратов отклонений, объясненная регрессией, в линейной модели парной регрессии имеет число степеней свободы, равное:

а) n;

6) n-1;

в) 1

г) n-2.

26. Остаточная сумма квадратов отклонений в линейной модели имеет число степеней свободы, равное:

а) n;

б) n-1;

в) 1;

г) n-2.

27. Общая сумма квадратов отклонений в линейной модели парной регрессии имеет степеней свободы, равное:

а) n;

б) n-1;

в) 1;

г) n-2.

28. Оценка значимости коэффициента регрессии производится с помощью:

а) F-критерия Фишера;

6) коэффициента корреляции ryx

в) t-критерия Стьюдента;

г) коэффициента детерминации $r^{2}yx$

29. Наблюдаемое значение -критерия для проверки значимости коэффициента регрессии *ь* имеет вид:

а) $\frac{b}{m_b}$;

б) $\frac{m_b}{b}$;

с) $\frac{b}{m_a}$;

г) $\frac{b}{a}$;

30. Интервальная оценка коэффициента регрессии *в* при уровне значимости *а* определяется неравенством:

а) $b - \frac{m_b}{t_a} \leq \beta \leq b + \frac{m_b}{t_a}$;

б) b□$m_b \leq \beta \leq b$□m_b;

с) b-$m_b \leq \beta \leq b$+m_b;

d) $b - t_a \cdot m_b \leq \beta \leq b + t_a \cdot m_b$;

ГЛАВА 4. МНОЖЕСТВЕННЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Множественный регрессионный анализ является развитием парного регрессионного анализа применительно к случаям, когда зависимая переменная гипотетически связана с более чем одной независимой переменной. Большая часть анализа будет непосредственным расширением парной регрессионной модели, но здесь мы сталкиваемся с двумя новыми проблемами. Во-первых, при оценке влияния данной независимой переменной на зависимую переменную нам придется решать проблему разграничения ее воздействия и воздействий других независимых переменных.

Во-вторых, мы должны будем решить проблему спецификации модели. Часто предполагается, что несколько переменных могут оказывать влияние на зависимую переменную, с другой стороны, некоторые переменные могут не подходить для модели. Мы должны решить, какие из них следует включить в уравнение регрессии, а какие – исключить из него. В данной главе мы полагаем, что спецификация проблемы правильна. В большинстве ситуаций мы ограничимся основным случаем, где используются только две независимые переменные.

Начнем с примера, который рассматривался в парной зависимости (Пример 9.1). Теперь определим факторы балла бонитета и количества вносимых удобрений.

Расширим первоначальную модель, и допустим, что истинную зависимость можно выразить следующим образом:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

где y – стоимость валовой продукции (тыс. сум/га), X_1 – балл бонитета, а X_2 – количество вносимых минеральных удобрений. Это, разумеется, является значительным упрощением как с точки зрения состава независимых переменных, включенных в зависимость, так и с точки зрения математической формулы связи. Кроме того, мы неявно предполагаем наличие лишь прямой связи за счет допущения о том, что стоимость валовой продукции не влияет на балл бонитета и вносимые удобрения.

Используя данные по стоимости валовой продукции, балл бонитета и вносимые удобрения, мы получим уравнение регрессии:

$$\hat{y} = -239{,}778 + 1{,}53x_1 + 163{,}08x_2; \quad R^2 = 0{,}85$$

где y- измерен в тыс.сум, x_1 измерен в баллах, а x_2 измерен в ц.д.в.

Полученное уравнение следует интерпретировать следующим образом. При каждом увеличении балла бонитета земли на 1 балл стоимость валовой продукции увеличится на 1,53 тыс. сум. На каждую единицу увеличения вносимых удобрений стоимость валовой продукции увеличится 163,08 тыс. сум. Чистый эффект в любой момент времени будет зависеть не только от этих коэффициентов, но также от размеров изменений x_1 и x_2.

Даже если бы спецификация модели оказалась правильной (разумеется, это является большим упрощением), то между прогнозируемым изменением и полученным результатом будет наблюдаться расхождение. Прежде всего оценки β_1 и β_2 подвержены влиянию ошибки выборки.

4.1. Вывод коэффициентов множественной регрессии

Как и в случае парной регрессии, мы так выбираем значения коэффициентов регрессии, чтобы обеспечить наилучшее соответствие наблюдениям в надежде получить оптимальные оценки для неизвестных истинных значений параметров. Как и прежде, оценка оптимальности соответствия определяется минимизацией S, т.е. суммы квадратов отклонений:

$$S = e_1^2 + \ldots + e_n^2,$$

где e_i является остатком в наблюдении i, разницей между фактическим значением y в этом наблюдении и значением $\hat{y}$, прогнозируемым по уравнению регрессии:

$$\hat{y}_i = a + b_1 x_{1i} + b_2 x_{2i};$$

$$e_i = y_i - \hat{y}_i = y_i - a - b_1 x_{1i} - b_2 x_{2i}.$$

Используя уравнение, мы можем записать:

$$S = \sum e^2 = \sum (y_i - a - b_1 x_{1i} - b_2 x_{2i})^2.$$

Необходимые условия первого порядка для минимума, то есть $\partial S / \partial a = 0$, $\partial S / \partial b_1 = 0$ и $\partial S / \partial b_2 = 0$, дают следующие уравнения:

$$\partial S / \partial a = -2 \sum (y_i - a - b_1 x_{1i} - b_2 x_{2i}) = 0;$$

$$\partial S / \partial b_1 = -2 \sum x_{1i} (y_i - a - b_1 x_{1i} - b_2 x_{2i}) = 0;$$

$$\partial S / \partial b_2 = -2 \sum x_{2i} (y_i - a - b_1 x_{1i} - b_2 x_{2i}) = 0.$$

Следовательно, мы имеем три уравнения с тремя неизвестными: $a, b_1 u b_2$. Первое уравнение можно легко перегруппировать для выражения величины a через b_1, b_2 и данные наблюдений для x и y:

$$a = y - b_1 x_1 - b_2 x_2.$$

Используя это выражение и два других уравнения, путем некоторых преобразований можно получить следующее выражение для b_1:

$$b_1 = \frac{Cov(x_1, y)Var(x_2) - Cov(x_2 y)Cov(x_1, x_2)}{Var(x_1)Var(x_2) - \{Cov(x_1, x_2)\}^2}.$$

Аналогичное выражение для b_2 можно получить путем перестановки x_1 и x_2 в этом уравнении.

Цель данного обсуждения состоит в выделении двух основных моментов. Во-первых, принципы, лежащие в основе вычисления коэффициентов регрессии, в случае множественной и парной регрессии не различаются.

Во-вторых, используемые при этом формулы будут разными, поэтому не следует пытаться использовать выражения, выведенные для парной регрессии, в случае множественной регрессии. Отметим также, что вычисление формул регрессии при двух независимых переменных является более трудоемкой задачей, чем при одной переменной, и нам придется использовать компьютер.

4.2. Свойства коэффициентов множественной регрессии

Пусть множественная регрессия имеет вид: $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + u$. Как и в случае парного регрессионного анализа, коэффициенты регрессии должны рассматриваться как случайные переменные специального вида, случайные компоненты которых обусловлены наличием в модели случайного члена. Каждый коэффициент регрессии вычисляется как функция значений y и независимых переменных в выборке, а y в свою очередь определяется независимыми переменными и случайным членом. Отсюда следует, что коэффициенты регрессии определяются значениями переменных и случайным членом, а их свойства существенно зависят от свойств последнего.

Считаем, что для множественного регрессионного анализа выполняются следующие условия Гаусса – Маркова: 1) математическое ожидание u в любом наблюдении равно нулю; 2) теоретическая дисперсия его распределения одинакова для всех наблюдений; 3) теоретическая ковариация его значений в любых двух наблюдениях равняется нулю; 4) распределение u независимо от распределения любой объясняющей переменной.

Первые три условия идентичны условиям для парного регрессионного анализа, а четвертое условие является обобщением своего аналога. На данный момент мы примем усиленный вариант четвертого условия, допустив, что независимые переменные являются не стохастическими.

Существуют еще два практических требования. Во-первых, нужно иметь достаточное количество данных для проведения линии регрессии, что означает наличие стольких (независимых) наблюдений, сколько параметров необходимо оценить. Во-вторых, между независимыми переменными не должно существовать строгой линейной зависимости. В общем случае можно сказать, что коэффициенты регрессии, скорее всего, являются более точными:

1) чем больше число наблюдений в выборке.
2) чем больше дисперсия выборки объясняющих переменных.
3) чем меньше теоретическая дисперсия случайного члена.
4) чем меньше связаны между собой объясняющие переменные (x_1, x_2).

Если между x_1 *и* x_2 существует нестрогая линейная зависимость, то коэффициент корреляции $r_{x1,x2}$ будет близким к единице, если зависимость положительна, и к минус единице, если зависимость отрицательна, и в обоих случаях $r^2_{x1,x2}$ будет близким к единице.

t – тесты для коэффициентов множественной регрессии выполняются так же, как это делается в парном регрессионном анализе

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

Общая модель. В предыдущем примере мы имели только две независимые переменные. В тех случаях, когда этих переменных более двух, уже невозможно дать геометрическое представление о том, что происходит, но развитие алгебраических выкладок в принципе вполне очевидно. Допустим, что переменная y связана с k независимыми переменными $x_1,\ldots,x_k$ неизвестной истинной зависимостью:

$$y = \alpha + \beta_1 x_1 + ... + \beta_k x_k + u.$$

Оценим уравнение для данного множества n наблюдений для y, $x_1,...,x_k$ по методу наименьших квадратов:

$$\hat{y} = a + b_1 x_1 + ... + b_k x_k.$$

Это вновь означает минимизацию суммы квадратов разностей, а отклонение в наблюдении i выражается как

$$e_i = y_i - \hat{y}_i = y_i - a - b_1 x_{1i} - \ldots - b_k x_{ki}.$$

Это уравнение является обобщением уравнения. Теперь мы выбираем $a, b_1,..., b_k$ так, чтобы свести к минимуму S – сумму квадратов отклонений $\sum e_i^2$. Мы получаем (k+1) условий первого порядка $\partial S / \partial a = 0,$ $\partial S / \partial b_1 = 0$, …, $\partial S / \partial b_k = 0,$ что дает (k + 1) уравнение для нахождения (k + 1) неизвестных. Можно легко показать, что первое из этих уравнений позволяет получить аналог для уравнения (5.11), относящееся к случаю с двумя переменными:

$$a = y - b_1 x_1 - b_2 x_2 - ... - b_k x_k.$$

Выражения для $b_1, b_2, ...b_k$ становятся очень сложными, и математика не будет здесь представлена в явном виде. Вычисления целесообразнее сделать с помощью матричной алгебры, но для этого в книге не приводится теоретических или практических приложений. Для практических примеров вычисления вручную неприемлемы, и для нахождения решений следует использовать компьютер [2].

4.3. Мультиколлинеарность

Мультиколлинеарность – это взаимная зависимость влияющих переменных. Проблема состоит в том, что при ее наличии становится сложно или невозможно разделить влияние регрессоров на зависимую переменную, и коэффициенты теряют экономический смысл предельной функции или эластичности. Дисперсии коэффициентов растут, сами коэффициенты, оцененные по различным выборкам или методом Монте-Карло, коррелируют между собой. Это приводит к тому, что в области настройки модели графики Y и $\hat{Y}$ прекрасно совпадают, R^2 и F высокие, а в области прогноза графики могут совпасть, или расходятся.

Как обнаружить мультиколлинеарность? Проще всего – по корреляционной матрице. Если коэффициенты корреляции регрессоров больше 0,7, значит они взаимосвязаны. Числовой характеристикой мультиколлинеарности может служить определитель корреляционной матрицы. Если он близок к 1, то регрессоры независимы; если к 0, значит они связаны сильно.

Как бороться с мультиколлинеарностью?

1. Смириться, принять во внимание и ничего не делать.

2. Увеличить объем выборки: дисперсии коэффициентов обратно пропорциональны количеству замеров.

3. Удалять из модели регрессоры, слабо коррелирующие с зависимой переменной, или коэффициенты которых имеют малую t-статистику. (А смысл такой: если регрессоры коррелируют, и вы можете ими управлять, например, расходы на станки и рабочих, то придется изменять их пропорционально). F-статистика, то есть качество модели, при этом растет.

4. Использовать в уравнении регрессии агрегаты из коррелирующих переменных: линейные комбинации с коэффициентами, обратно пропорциональными стандартным отклонениям переменных и выравнивающими их масштабы. Такие агрегаты обычно не имеют экономического смысла, но могут повысить адекватность модели [2].

4.4. Использование метода главных компонент для подавления мультиколлинеарности

Применение метода главных компонент рассмотрим к задаче о торговле цыплятами. Предположим, что на экономический процесс влияют два независимых фактора M1 и M2, а компоненты векторов **V** являются линейными комбинациями их компонент:

$$V1^{\wedge} = a_1 + b_{11}\,M1 + b_{12}\,M2$$
$$V2^{\wedge} = a_2 + b_{21}\,M1 + b_{22}\,M2$$
$$V3^{\wedge} = a_3 + b_{31}\,M1 + b_{32}\,M2$$
$$V4^{\wedge} = a_4 + b_{41}\,M1 + b_{42}\,M2$$

Коэффициентов a_i, b_{ik} и компонент векторов M1 и M2 оцениваются с использованием сервиса *Поиск решения*. Для этого были заданы произвольные значения коэффициентов и компонент M1 и M2, вычислено скалярное произведение (M1, M2), векторы **V^** и квадраты остатков $(V - V\wedge)^2$, которые в таблицах не показаны. Целевая ячейка – сумма квадратов остатков, которая минимизируется, изменяемые ячейки – коэффициенты и компоненты векторов M1 и M2, ограничение: скалярное произведение (M1, M2) = 0, что означает ортогональность этих векторов. *Поиск решения* был запущен несколько раз с разными начальными значениями M1 и M2, чтобы добиться минимума $(V - V\hat{})^2$. Затем был запущен сервис *Регрессия* для оценки коэффициентов уравнения

$$Z\hat{} = a + b_1 \text{ M1} + b_2 \text{ M2}$$

Полученные результаты: Z^=3,18 + 0,93 M1 +1,26 M2

t55 9,7 8,25

Таблица 4.1

V1	*V2*	*V3*	*V4*	*Z*	*Z^*
6,20	3,62	4,00	4,35	3,44	3,52
6,27	3,64	4,15	4,38	3,51	3,53
6,33	3,67	4,25	4,39	3,57	3,53
6,44	3,63	4,19	4,43	3,59	3,58
6,50	3,65	4,17	4,45	3,60	3,60
6,58	3,69	4,25	4,54	3,65	3,62
6,64	3,65	4,29	4,66	3,70	3,67
6,74	3,68	4,22	4,65	3,70	3,69
6,82	3,68	4,37	4,74	3,73	3,71
6,84	3,95	4,56	4,82	3,70	3,68
6,93	3,89	4,55	4,85	3,71	3,72
7,06	4,07	4,82	4,96	3,69	3,73
7,21	4,06	4,87	4,97	3,75	3,76
7,28	4,03	4,77	4,94	3,79	3,79
7,36	4,15	4,87	5,11	3,84	3,82
7,47	4,12	4,87	5,31	3,92	3,90
7,60	4,08	4,85	5,39	3,91	3,95
7,72	4,20	4,95	5,40	3,95	3,96

Коэффициент детерминации R^2= 0,9; *F* = 49,4. *t*-статистики коэффициентов существенно выше, чем при использовании обычной множественной регрессии: 8,36; 2,86; 2,92; 1,05; 0,68 , или при исключении *V3* и *V4* : 12,3; 8,39; 3,6 (см. Таблицу 4.1). Вычисленные методом Монте-Карло относительные стандартные отклонения (в процентах) четырех прогнозных значений *Z^* для линеаризованной модели множественной регрессии: 0,64; 1,02; 1,16; 1,07; то же для метода главных компонент: 0,50; 0,66; 0,86; 0,82, то есть на 27% меньше, но все же хуже, чем дает регрессия по доходу и цене цыплят: 0,52; 0,56; 0,71; 0,72 (35%).

Таблица 4.2. Продолжение

		M1	*M2*	*M1*M2*	*V1^*	*V2^*	*V3^*	*V4^*
		-0,13	0,364	-0,05	6,23	3,56	4,05	4,31
		-0,24	0,443	-0,11	6,29	3,63	4,16	4,35
		-0,32	0,501	-0,16	6,33	3,69	4,23	4,38
		-0,08	0,37	-0,03	6,42	3,65	4,18	4,45
		0,006	0,326	0,002	6,47	3,64	4,18	4,49
		0,014	0,34	0,005	6,56	3,69	4,25	4,56
a1	4,769	0,165	0,268	0,044	6,66	3,69	4,26	4,64
b11	3,128	0,304	0,193	0,059	6,71	3,67	4,23	4,68
b12	5,141	0,265	0,235	0,062	6,81	3,74	4,33	4,75
a2	2,675	-0,12	0,485	-0,06	6,87	3,92	4,58	4,78
b21	1,409	0,06	0,386	0,023	6,94	3,89	4,54	4,83
b22	2,931	-0,21	0,581	-0,12	7,09	4,08	4,80	4,93
a3	2,764	-0,13	0,547	-0,07	7,18	4,10	4,84	5,00
b31	2,105	0,055	0,441	0,024	7,21	4,04	4,77	5,03
b32	4,286	0,038	0,478	0,018	7,35	4,13	4,89	5,13
a4	3,267	0,358	0,316	0,113	7,51	4,11	4,87	5,26
b41	2,302	0,599	0,192	0,115	7,63	4,08	4,85	5,36
b42	3,705	0,492	0,273	0,134	7,71	4,17	4,97	5,41
			Сумма	0				

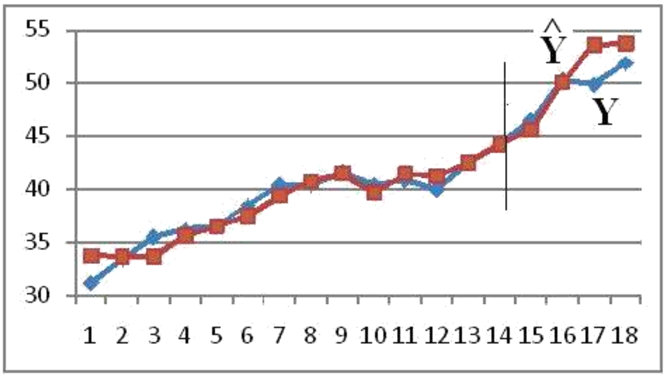


Рис.4.2. Проверка адекватности модели, использующей Метод главных компонент

Проверка модели на адекватность представлена на рисунке 4.2, корреляционная матрица коэффициентов *a, b1* и *b2,* полученная методом Монте-Карло – в таблице 4.3. Таблица 4.3 объясняет, почему статистические ошибки прогнозов существенно меньше, чем у коэффициентов уравнений регрессии: отклонения одних коэффициентов компенсируется противоположными отклонениями других [2].

Таблица 4.3.

	b2	*b1*
b2	1	
b1	0,849	1
A	-0,987	-0,824

Таким образом, метод главных компонент позволяет улучшить качество эконометрических моделей: уменьшить статистические ошибки коэффициентов прогнозов. Его несложно реализовать в среде Excel и использовать для студенческих лабораторных работ, но здесь он представлен для ознакомления, а в реальной работе вы воспользуетесь специализированными программами.

Опорные слова

Эконометрика, множественная регрессия, объясняющие переменные, фактор, параметр, линейная зависимость, формула связи, число степеней свободы, метод наименьших квадратов, объясняющая переменная, линии регрессии, число наблюдений, независимая переменная, уравнение регрессии, коэффициент детерминации, временные ряды, мультиколлинеарность, коэффициент регрессии.

Контрольные вопросы

1. Мультиколлинеарность: чем плоха, как обнаружить и как бороться.
2. Выявление мультиколлинеарности по матрице корреляции экзогенных переменных
3. Что такое и почему возникает ложная корреляция и коинтеграция?
4. Расчетная формула частного коэффициента детерминации.
5. Применение статистики Стьюдента и Фишера в процедуре подбора переменных в модели множественной регрессии .
6. Ошибки от включения в модель незначимых переменных или исключения значимых.

Контрольные задания

1. По данным приложения 1 определите:

1) обобщающие статистические характеристики по всем переменным;

2) парные коэффициенты корреляции между всеми переменными;

3) наличие или отсутствие мультиколлинеарности между признаками.

2. Используя данные приложения Б по одному варианту определите:

1) параметры линейного множественного уравнения регрессии в натуральной и стандартизованной форме;

2) средние коэффициенты эластичности и в-коэффициенты для каждого фактора;

3) коэффициенты частной и множественной корреляции;

4) значимость множественного уравнения регрессии в целом с помощью общего критерия F-Фишера;

5) значимость коэффициентов множественной регрессии с использованием критерия *t* - Стьюдента;

6) доверительные интервалы множественных коэффициентов регрессии при уровне доверительной вероятности 0,95;

Дайте оценку полученным результатам, которые оформите в виде кратких выводов.

По вариантам зависимой переменной (у) является:

а) выручка от реализации продукции на 1 га сельскохозяйственных угодий, тыс. у.е.;

б) валовая продукция на 1 га сельскохозяйственных угодий, тыс. у.е.

Варианты заданий по данным приложения 1

Вариант	1	2	3	4
Переменные	x_1; x_2; x_5; x_6	x_1; x_2; x_3; x_4	x_1; x_2; x_4; x_5	x_1; x_2; x_3; x_4; x_5
Вариант	5	6	7	8
Переменные	x_2; x_3; x_4; x_5	x_2; x_3; x_4; x_6	x_1; x_2; x_3; x_5	x_1; x_2; x_4; x_5; x_6

3. По данным, представленным в приложении 2, изучите влияние совокупности факторов производства молока (vi и уг) в сельскохозяйственных предприятиях.

1) С помощью инструмента пакета анализа (Excel) Описательная статистика рассчитайте обобщающие характеристики вариационных рядов, написав выводы по каждой переменной.

2) С помощью инструмента пакета анализа Корреляция постройте матрицу парных коэффициентов корреляции. Оцените наличие или отсутствие мультиколлинеарности между факторами.

3) С помощью инструмента пакета анализа Регрессия рассчитайте параметры множественного уравнения регрессии с включением всех факторов.

4) Рассчитайте частные коэффициенты корреляции, эластичности и стандартизованные коэффициенты регрессии.

5) Оцените значимость множественного уравнения регрессии и коэффициентов множественной регрессии с помощью критериев Фишера и Стьюдента.

6) Проведите последовательный отсев статистически незначимых факторов и получите модель себестоимости производства молока со всеми статистически значимо влияющими факторами. Рассчитайте частные коэффициенты корреляции, значимость множественного уравнения регрессии и множественных коэффициентов регрессии с помощью критериев Фишера и Стьюдента. Определите среднюю ошибку аппроксимации, эластичность и стандартизованные коэффициенты регрессии. Оцените значимость. Напишите выводы по результатам множественного регрессионного анализа себестоимости производства молока.

4. По данным, приведенным в приложении В, изучите влияние факторов на годовой надой молока на среднегодовую корову (у):

- прямые затраты труда на 1 голову, чел.-ч;

-среднегодовое поголовье коров на предприятии, гол.;

- затраты по оплате труда на 1 чел.-ч., у.е.

- затраты на корма на среднегодовую корову, тыс. у.е.

1) Определите параметры множественного уравнения регрессии с включением в линейную модель всех факторов.

2) Оцените значимость множественных коэффициентов регрессии с помощью t - критерия Стьюдента.

3) Проведите последовательный отсев статистически не значимых факторов.

4) По уравнению со всеми значимо влияющими факторами рассчитайте частные и множественный коэффициенты корреляции и детерминации.

5) Оцените значимость множественного уравнения регрессии с помощью F-критерия Фишера, для чего составить таблицу дисперсионного анализа.

Напишите выводы по результатам расчетов.

5. По данным 16 сельскохозяйственных предприятий изучается влияние факторов на продуктивность коров. Зависимая переменная (у) - годовой надой молока на среднегодовую корову, ц. Объясняющие переменные:

X-1 – производственные затраты на среднегодовую корову, тыс. у.е.;

X-2 – затраты на корма на среднегодовую корову, тыс. у.е.;

X-3 – прямые затраты труда на среднегодовую корову, чел.-ч;

X-4 – среднегодовое поголовье коров на предприятии, гол.;

X-5 – затраты по оплате труда на 1 чел.-ч, у.е.;

X-6 – доля молока в выручке от реализации продукции животноводства, %. В таблицах 4.3 и 4.4 приведены статистические характеристики по изучаемой совокупности предприятий и парные коэффициенты корреляции между переменными.

Таблица 4.3. Статистические характеристики выборочной совокупности сельскохозяйственных предприятий

Показатель	**Y**	x_1	x_2	x_3	x_4	x_5	x_6
Среднее значение	61,761	100,56	44,15	111,12	11133,8 3	2111,58	78,85
Стандартная ошибка	1,413	2,75	1,44	4,49	79,36	10,44	1,91
Медиана	62,682	102,32	42,35	107,35	975,00	181,68	82,95
Среднее квадратическое отклонение	11,483	22,38	11,72	36,48	644,73	84,79	15,55
Дисперсия	131,864	500,68	137,34	1330,91	415672	7189,37	241,66
Эксцесс	0,177	0,47	-0,38	-0,62	0,44	-0,05	3,60
Асимметричность	-0,262	0,37	0,64	0,08	1,09	0,89	-1,89
Интервал	56,024	105,02	45,55	160,13	2587,00	341,52	76,53
Минимум	34,506	52,14	25,50	32,04	403,00	107,71	21,71
Максимум	90,530	157,17	71,06	192,16	2990,00	449,23	98,24
Сумма	4076,24	6636,71	2914,04	7333,73	74833,0	13964,3	5203,86

Таблица 4.4

	y	x_1	x_2	x_3	x_4	x_5	x_6
Y	1						
x_1	0,7294	1					
x_2	0,5640	0,7560	1				
x_3	0,2368	0,3327	0,1095	1			
x_4	0,4347	0,2872	0,2838	0,0284	1		
x_5	0,4611	0,3109	0,1933	-0,2662	0,2980	1	
x_6	0,1665	0,0235	0,0790	0,0774	0,0874	-0,0649	

Варианты заданий

Вариант	1	2	3	4	5	6
Переменные	y; x_1; x_4	y; x_1; x_5	y; x_1; x_6	y; x_2; x_3	y; x_2; x_4	y; x_2; x_5
Вариант	7	8	9	10	11	12
Переменные	y; x_2; x_6	y; x_4; x_5	y; x_4; x_6	y; x_5; x_6	y; x_3; x_4	y; x_3; x_5

1) По одному варианту составьте матрицу парных коэффициентов корреляции между тремя переменными.

2) Определите параметры множественного уравнения регрессии в стандартизированной и естественной форме.

3) Рассчитайте частные коэффициенты эластичности.

4) Рассчитайте частные и множественный коэффициенты корреляции и детерминации.

5) Оцените значимость множественного уравнения регрессии с помощью F критерия Фишера, для чего составить таблицу дисперсионного анализа.

6) С помощью частных F-критериев Фишера оцените целесообразность включения фактора x_1 после x_2 и фактора x_2 после x_1.

7) Оцените значимость множественных коэффициентов регрессии с помощью t-критерия Стьюдента.

6. Напишите выводы по результатам расчетов. По данным, приведенным в приложении 1, проведите регрессионный анализ влияния совокупности факторов на изменение одного из результативных признаков:

а) валовая продукция на одно предприятие;

б) реализованная продукция на одно предприятие;

в) валовая продукция на среднегодового работника;

г) валовая продукция на 1 га сельскохозяйственных угодий;

д) реализованная продукция на среднегодового работника;

е) реализованная продукция на 1 га сельскохозяйственных угодий.

1. Общий вид уравнений парной и множественной регрессии.

2. Нелинейные уравнения регрессии.

3. Формулы для вычисления коэффициентов парной линейной регрессии и их погрешностей.

4. Метод наименьших квадратов (МНК) и система нормальных уравнений парной линейной регрессии.

5. Схема Гаусса-Маркова и Матричный метод МНК.

6. Теорема Гаусса-Маркова: формулировка и условия.

7. Показатели качества эконометрической модели: коэффициент детерминации R^2, статистика Фишера F, t-статистика Стьюдента для коэффициентов уравнений.

8. Показатели качества эконометрической модели: тест Дарбина-Уотсона на автокорреляцию DW [7].

Тесты для главы №4

1. К ошибкам спецификации относятся (выберите, по крайней мере, один ответ):

а) выбор неправильной формы модели;

б) ошибки измерения переменных;

в) неполный набор объясняющих переменных;

г) ошибки выборки.

2. Для получения качественных оценок уравнения регрессии необходимо выполнение следующих предпосылок (выберите, по крайней мере, один ответ):

а) отклонения е должны быть нормально-распределенными случайными величинами с М (е) = 0, D (Е)= - а;

б) отклонения е не коррелированы;

в) отклонения е должны быть показательно-распределенными;

г) отклонения у должны быть равномерно-распределенными.

3. Качество подбора линейного уравнения регрессии можно охарактеризовать на основе

а) коэффициента детерминации;

б) средней ошибки аппроксимации;

в) коэффициента эластичности;

г) коэффициента регрессии.

4. Если наблюдаемое значение F-критерия Фишера меньше критического, то можно сделать вывод о:

а) статистической незначимости построенной модели;

б) статистической значимости построенной модели;

в) незначимости(несущественности) моделируемой зависимости;

г) отсутствии связи между изучаемыми переменными.

5. Если наблюдаемое значение F-критерия Фишера больше критического, то можно делать вывод:

а) статистической незначимости построенной модели;

б) статистической значимости построенной модели;

в) незначимости(несущественности) моделируемой зависимости;

г) отсутствии связи между изучаемыми переменными.

6. Число степеней свободы для общей суммы квадратов отклонений линейного множественного уравнения регрессии с m факторами при n числе наблюдений и составляет:

а) m-1;

б) n-m;

в) n-1;

г) n-m-1.

7. Число степеней свободы для факторной суммы квадратов отклонений в линейной модели при m факторах и n наблюдениях, составляет:

а) m;

б) n-m+1;

в) n-m;

г) n-m-1.

8. Число степеней свободы для остаточной суммы квадратов отклонений в линейной модели при m факторах и n наблюдениях, составляет:

а) m;

б) n-m;

в) n-1;

г) n-m-1.

9. При добавлении новой независимой переменной в уравнение регрессии коэффициент детерминации:

а) уменьшается;

в) не увеличивается:

б) не уменьшается;

г) увеличивается.

10. Множественный коэффициент детерминации определяет:

а) долю дисперсии факторов, объясненную регрессией;

б) долю дисперсии результативного признака, объясненную регрессией;

в) долю дисперсии факторов, не объясненную регрессией;

г) долю дисперсии результативного признака, не объясненную регрессией.

11. Частные коэффициенты корреляции характеризуют:

а) тесноту связи между результативным признаком и факторами, включенными уравнение регрессии;

б) тесноту связи между двумя переменными;

в) количественное влияние факторов на изменение результативного признака;

г) тесноту связи между двумя переменными, исключив влияние других переменных из рассматриваемого набора переменных.

12 Множественный коэффициент корреляции равен 0,8. Значит, множественный коэффициент детерминации равен:

а) 0,8;

б) 0,2;

в) 0,64;

г) 0,36.

13. Множественный коэффициент регрессии показывает:

а) на сколько единиц в среднем изменится результативный признак, при увеличении j-ого фактора на единицу, если остальные факторы закреплены на постоянном уровне;

б) на сколько процентов в среднем изменится результативный признак, при увеличении j-ого фактора на единицу, если остальные факторы закреплены на постоянном уровне;

в) на сколько процентов в среднем изменится результативный признак, при увеличении j-ого фактора на один процент, если остальные факторы закреплены на постоянном уровне;

г) силу связи между результативным признаком и j-ым фактором.

14. Множественный коэффициент эластичности E_j показывает:

а) на сколько единиц в среднем изменится результативный признак, при увеличении j-ого фактора на единицу, если остальные факторы закреплены на постоянном уровне;

б) на сколько процентов в среднем изменится результативный признак, при увеличении j-ого фактора на единицу, если остальные факторы закреплены на постоянном;

в) на сколько процентов в среднем изменится результаты признак, при увеличения j-ого фактора на один процент, если остальные факторы закреплены на постоянном уровне;

г) силу связи между результативным признаком и j-ым фактором.

15. Значимость частных, парных и множественного коэффициентов корреляции проверяется с помощью

а) нормального закона распределения;

б) x^2 - критерия Пирсона;

в) t - критерия Стьюдента;

г) F-критерия Фишера.

16. Множественный коэффициент корреляции изменяется в границах:

а) от 0 до 1;

б) от -1 до +1;

в) от $-\infty$ до $+\infty$;

г) от 0 до $+\infty$.

17. Проверка значимости множественных коэффициентов регрессии проводится для

а) нормального закона распределения,

б) x^2 - критерия Пирсона;

в) t -критерия Стьюдента;

г) F- критерия Фишера.

18. Оценка параметров уравнения регрессии является несмещенной, если:

а) математическое ожидание остатков равно нулю;

б) точность оценки возрастает с увеличением числа наблюдений;

в) математическое ожидание остатков не равно нулю;

г) остатки распределены по нормальному закону.

19. Оценка параметров уравнения регрессии является эффективной, если:

а) математическое ожидание остатков равно нулю;

б) точность оценки возрастает с увеличением числа наблюдений;

в) она имеет наименьшую дисперсию по сравнению с другими оценками;

г) дисперсии остатков равны нулю.

20. Оценка параметров уравнения регрессии является состоятельной если:

а) математическое ожидание остатков равно нулю;

б) точность оценки возрастает с увеличением числа наблюдений;

в) она имеет наименьшую дисперсию по сравнению с другими оценками;

г) остатки распределены по нормальному закону.

21. Гомоскедастичность остатков означает:

а) нормальность распределения остатков;

б) одинаковость дисперсий остатков по каждой единице наблюдения;

в) непостоянство дисперсий остатков по каждой единице наблюдения;

г) остатки распределены по равномерному закону.

22. При наличии гетероскедастичности оценка параметров уравнения регрессии производится:

а) классическим методом наименьших квадратов;

б) методом максимального правдоподобия;

в) обобщенным методом наименьших квадратов;

г) методом минимизации абсолютных отклонений.

23. В степенной функции коэффициенты b_j, являются:

а) коэффициентами регрессии, как и в линейной функции;

б) коэффициентами эластичности;

в) коэффициентами корреляции;

г) весовыми коэффициентами.

24. Получено уравнение множественной регрессии F-критерия по 30 наблюдениям -256+1,4*1+2,6*2. Наблюдаемое значение F-критерия Фишера равно 4,12. Сравнение наблюдаемого значения с критическим значением показывает, что:

а) уравнение не значимо при а = 0,05;

б) уравнение значимо при а = -0,05;

в) уравнение значимо при а = 0,01;

г) переменные независимые.

25. Если абсолютное значение парных коэффициентов корреляции между факторами больше 0,7-0,8, то это может свидетельствовать:

а) о значимости влияния факторов;

б) об отсутствии мультиколлинеарности;

в) о наличии мультиколлинеарности;

г) о наличии автокорреляции.

26. Линейная модель множественной регрессии имеет вид:

а) $yi = \sum_{j=1}^{m} \cdot B_j x_{ij}$;

б) $yi = \beta_0 + \sum_{j=1}^{m} \cdot B_j x_{ij} + \varepsilon_i$

с) $yi = \beta_0 + \Pi^{m}_{j=1} B_j x_{ij} + \varepsilon_i$

г) $yi = \beta_0 \square \Pi^{m}_{j=1} B_j x_{ij} \square \varepsilon_i$

27. Если в уравнение регрессии дополнительно включается новый фактор, то множественный коэффициент корреляции:

а) остается неизменным;

б) уменьшается;

в) возрастает;

г) может как возрастать, так и уменьшаться.

28. Степенная модель множественной регрессии имеет вид:

а) $yi = \sum_{j=1}^{m} \cdot B_j x_{ij}$;

б) $yi = \beta_0 + \sum_{j=1}^{m} \cdot B_j x_{ij} + \varepsilon_i$;

с) $yi = \beta_0 + \Pi^{m}_{j=1} B_j x_{ij} + \varepsilon_i$;

г) $yi = \beta_0 \square \Pi^{m}_{j=1} B_j x_{ij} \square \varepsilon_i$;

29. К ошибкам спецификации относятся (выберите, по крайней мере, один ответ)

а) выбор неправильной формы модели;

б) ошибки измерения переменных;

в) неполный набор объясняющих переменных;

г) ошибки выборки.

30. Для получения качественных оценок уравнения регрессии необходимо выполнение следующих предпосылок (выберите, по крайней мере, один ответ)

а) отклонения должны быть нормально-распределенными случайными величинами с $M(\varepsilon i)=0$, $D(\varepsilon i)=0$;

б) отклонения εi и εj не коррелированы;

в) отклонения εi должны быть показательно-распределенными;

г) отклонения εi должны быть равномерно-распределенными.

31. Качество подбора линейного уравнения регрессии можно охарактеризовать на основе

а) коэффициента детерминации;

б) средней ошибки аппроксимации;

в) коэффициента эластичности;

г) коэффициента регрессии.

32. Если наблюдаемое значение F-критерия Фишера меньше критического, то можно сделать вывод о:

а) статистической незначимости построенной модели;

б) статистической значимости построенной модели;

в) незначимости (несущественности) моделируемой зависимости;

г) отсутствии связи между изучаемыми переменными.

33. Если наблюдаемое значение F-критерия Фишера больше критического, то можно сделать вывод о:

а) статистической незначимости построенной модели;

б) статистической значимости построенной модели;

в) незначимости(несущественности) моделируемой зависимости;

г) отсутствии связи между изучаемыми переменными.

34. Число степеней свободы для суммы квадратов отклонений линейного множественного уравнения регрессии с m факторами при числе наблюдений n составляет:

а) $m-l$;

б) $n-m$;

в) $n-l$;

г) $n-m-l$.

35. Число степеней свободы для факторной суммы квадратов отклонений в линейной модели при m факторах и n наблюдениях, составляет:

а) m;

б) $n-m+l$;

в) $n-m$;

г) $n-m-l$.

36. Число степеней свободы для остаточной суммы квадратов отклонений в линейной модели при *m* факторах и *n* наблюдениях, составляет:

а) *m;*

б) *n-m;*

в) *n-l;*

г) *n-m-l.*

37. При добавлении новой независимой в уравнение регрессии коэффициент детерминации:

а) уменьшается;

б) не уменьшается;

в) не увеличивается;

г) увеличивается.

38. Множественный коэффициент детерминации определяет:

а) долю дисперсии факторов, объясненную регрессией;

б) долю дисперсии результативного признака, объясненную регрессией;

в) долю дисперсии факторов, не объясненную регрессией;

г) долю дисперсии результативного признака, не объясненную регрессией.

39. Частные коэффициенты корреляции характеризует:

а) тесноту связи между результативным признаком и факторами, включенными в уравнение регрессии;

б) тесноту связи между двумя переменными;

в) количественное влияние факторов на изменение результативного признака;

г) тесноту связи между двумя переменными, исключив влияние других переменных из рассматриваемого набора переменных.

40. Множественный коэффициент корреляции равен 0,8. Значит, множественный коэффициент детерминации равен:

а) 0,8;

б) 0,2;

в) 0,64;

г) 0,36.

41. Множественный коэффициент регрессии b_j показывает:

а) на сколько единиц в среднем изменится результативный признак, при увеличении *j*-ого фактора на единицу, если остальные факторы закреплены на постоянном уровне;

б) на сколько процентов в среднем изменится результативный признак, при увеличении j-ого фактора на единицу, если остальные факторы закреплены на постоянном уровне;

в) на сколько процентов в среднем изменится результативный признак, при увеличении j-ого фактора на один процент, если остальные факторы закреплены на постоянном уровне;

г) силу связи между результативным признаком и j-ым фактором.

42. Множественный коэффициент эластичности E_j показывает:

а) на сколько единиц в среднем изменится результативный признак, при увеличении j-ого фактора на единицу, если остальные факторы закреплены на постоянном уровне;

б) на сколько процентов в среднем изменится результативный признак, при увеличении j-ого фактора на единицу, если остальные факторы закреплены на постоянном уровне;

в) на сколько процентов в среднем изменится результативный признак, при увеличении j-ого фактора на единицу, если остальные факторы закреплены на постоянном уровне;

г) силу связи между результативным признаком и j-ым фактором.

43. С помощью частного F-критерия можно проверить значимость j-ого коэффициента чистой регрессии в предложении, что j-й фактор в уравнение множественной регрессии:

а) не был включен;

б) был включен последним;

в) был включен первым;

г) был включен условно.

44. Значимость частных, парных и множественного коэффициента корреляции проверяется с помощью:

а) нормального закона распределения;

б) X^2 – критерия Пирсона;

в) t – критерия Стьюдента;

г) F – критерия Фишера.

45. Множественный коэффициент корреляции изменяется в границах:

а) от 0 до 1;

б) от – 1 до + 1;

в) от - ∞ до + ∞;

г) от 0 до + ∞.

46. Проверка значимости множественных коэффициентов регрессии проводится с использованием:

а) нормального закона распределения;

б) X^2 – критерия Пирсона;

в) t – критерия Стьюдента;

г) F – критерия Фишера.

47. Оценка параметров уравнения регрессии является несмещенной, если:

а) математическое ожидание остатков равно нулю;

б) точность оценки возрастает с увеличением числа наблюдений;

в) математическое ожидание остатков не равно нулю;

г) остатки распределены по нормальному закону.

48. Оценка параметров уравнения регрессии является эффективной, если:

а) математическое ожидание остатков равно нулю;

б) точность оценки возрастает с увеличением числа наблюдений;

в) она имеет наименьшую дисперсию по сравнению с другими оценками;

г) дисперсия остатков равно нулю.

49. Оценка параметров уравнения регрессии является состоятельной, если:

а) математическое ожидание остатков равно нулю;

б) точность оценки возрастает с увеличением числа наблюдений;

в) она имеет наименьшую дисперсию по сравнению с другими оценками;

г) остатки распределены по нормальному закону.

50. Гомоскедастичность остатков означает:

а) нормальность распределения остатков;

б) одинаковость дисперсии остатков по каждой единице наблюдения;

в) непостоянство дисперсии остатков по каждой единице наблюдения;

г) остатки распределены по равномерному закону.

51. При наличии гетероскедастичности оценка параметров уравнения регрессии производится:

а) классическим методом наименьших квадратов;

б) методом максимального правдоподобия;

в) обобщенным методом наименьших квадратов;

г) методом минимизации абсолютных отклонений;

52. В степенной функции $Y = b_0 \cdot x_1^{b_2} \cdot x_2^{b_2} \cdots x_m^{b_m}$ коэффициенты b_j являются:

а) коэффициентами регрессии, как и линейной функции;

б) коэффициентами эластичности;

в) коэффициентами корреляции;

г) весовыми коэффициентами.

53. Получено множественное уравнение регрессии по 30 наблюдениям $\hat{y}=256+1{,}4x_1+2{,}6x_2$. Наблюдаемое значение F-критерия Фишера равно 4,12. Сравнение наблюдаемого значения F-критерия с критическим значением показывает, что:

а) уравнение не значимо при $a = 0{,}05$;

б) уравнение значимо при $a = 0{,}05$;

в) уравнение значимо при $a = 0{,}01$;

г) переменные независимые.

54. Если абсолютное значение парных коэффициентов корреляции между факторами больше 0,7-0,8, то это может свидетельствовать:

а) о значимости влияния факторов;

б) об отсутствии мультиколлинеарности;

в) о наличии мультиколлинеарности;

г) о наличии автокорреляции.

55. Линейная модель множественной регрессии имеет вид:

а) $yi = \sum_{j=1}^{m} B_j x_{ij}$;

б) $yi = \beta_0 + \sum_{j=1}^{m} B_j x_{ij} + \varepsilon_i$

с) $yi = \beta_0 + \Pi^{m}_{j=1} B_j x_{ij} + \varepsilon_i$

г) $yi = \beta_0 \square \Pi^{m}_{j=1} B_j x_{ij} \square \varepsilon_i$

56. Если в уравнение регрессии дополнительно включается новый фактор, то множественный коэффициент корреляции:

а) остается неизменным;

б) уменьшается;

в) возрастает;

г) может как возрастать, так и уменьшаться.

57. Степенная модель множественной регрессии имеет вид [7]:

а) $yi = \sum_{j=1}^{m} B_j x_{ij}$;

б) $yi = \beta_0 + \sum_{j=1}^{m} B_j x_{ij} + \varepsilon_i$

с) $yi = \beta_0 + \Pi^{m}_{j=1} B_j x_{ij} + \varepsilon_i$

г) $yi = \beta_0 \square \Pi^{m}_{j=1} B_j x_{ij} \square \varepsilon_i$

ГЛАВА 5. ПРОИЗВОДСТВЕННЫЕ ФУНКЦИИ

5.1. Понятие производственной функции одной переменной

Производственная функция – это функция, независимая переменная которой принимает значения объемов затрачиваемого или используемого ресурса (фактора производства), а зависимая переменная – значения объемов выпускаемой продукции

$$y = f(x) \qquad (1)$$

В формуле (1) $x(x \geq 0)$ и $y(y \geq 0)$- числовые величины, т.е. $y = f(x)$ есть функция одной переменной x. В связи с этим производственная функция (П.Ф.) f называется одно ресурсной или однофакторной ПФ, ее область определения – множество неотрицательных действительных чисел (т.е. $x \geq 0$). Запись $y = f(x)$ означает, что если ресурс затрачивается или используется в количестве x единиц, то продукция выпускается в количестве $y = f(x)$ единиц. Символ f (знак функции) – является характеристикой производной системы, преобразующей ресурс в выпуск. Символ f связывает между собой независимую переменную x с зависимой переменной y. В микроэкономической теории принято считать, что y – это максимально возможный объем выпускаемой продукции, если ресурс затрачивается или используется в количестве x единиц. В макроэкономике такое понимание не совсем корректно: возможно, при другом распределении ресурсов между структурными единицами экономики выпуск мог бы быть и большим. В этом случае ПФ – это статистически устойчивая связь между затратами ресурса и выпуском. Более правильной является символика $y = f(x, a)$, где a –вектор параметров ПФ.

Пример 1. Возьмем ПФ в виде $f(x) = ax^6$, где x – величина затрачиваемого ресурса (например, рабочего времени), $f(x)$ - объем выпускаемой продукции (например, число готовых к отправке холодильников). Величины a и b – параметры ПФ f. Здесь a и b – положительные числа и число $b \leq 1$, вектор параметров есть двумерный вектор (a, b).

График производственной функции $f(x) = ax^6$ изображен на рис. 5.1. На графике видно, что с ростом величины затрачиваемого ресурса x объем выпуска y растет, однако при этом каждая дополнительная единица ресурса дает все меньший прирост объема y выпускаемой продукции. Отмеченное обстоятельство (рост объема y и уменьшение прироста объема y с ростом величины x) отражает фундаментальное положение экономической теории (хорошо подтверждаемое практикой), называемое законом убывающей эффективности. ПФ $f(x)=ax^6$ является типичным представителем широкого класса однофакторных ПФ.

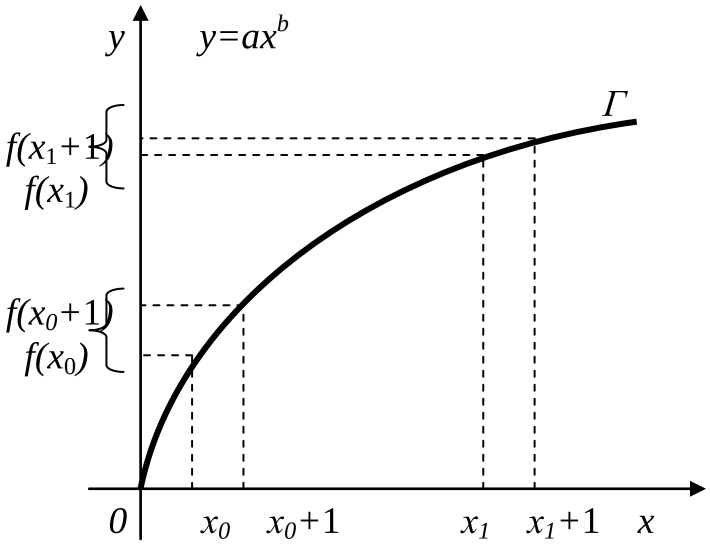


Рис. 5.1.

ПФ могут иметь разные области использования. Сначала остановимся на микроэкономическом уровне. ПФ $f(x)=ax^6$, рассмотренная выше, может быть использована для описания взаимосвязи между величиной затрачиваемого или используемого ресурса x в течении года на отдельном предприятии (фирме) и годовым выпуском продукции y этого предприятия (фирмы). В роли производственной системы здесь выступает отдельное предприятие (фирма) – имеем микроэкономическую ПФ (МИПФ). На макроэкономическом уровне в роли производственной системы может выступать также отрасль, межотраслевой производственный комплекс. МИПФ строятся и используются в основном для решения задач анализа и планирования, а также задач прогнозирования.

ПФ может быть использована для описания взаимосвязи между годовыми затратами труда в масштабе региона или страны в целом и годовым конечным выпуском продукции (или доходом) этого региона или страны в целом. Здесь в роли производственной системы выступает регион или страна в целом (точнее хозяйственная система региона или страны) – имеем макроэкономический уровень и макроэкономическую ПФ (МАПФ). МАПФ строятся и активно используются для решения всех трех типов задач (анализа, планирования и прогнозирования).

На микроэкономическом уровне затраты и выпуск могут быть измерены в человеко-часах (объем человеко-часов – натуральный показатель) или в условных единицах выплаченной заработной платы (ее величина – стоимостной показатель); выпуск продукции может быть представлен в штуках или в других натуральных единицах (тоннах, метрах и т.п.) или в виде своей стоимости.

На макроэкономическом уровне затраты и выпуск измеряются, как правило, в стоимостных показателях и представляют собой стоимостные (ценностные) агрегаты, т.е. суммарные величины произведений объемов затрачиваемых (или используемых) ресурсов и выпускаемых продуктов на их цены.

Производственная функция нескольких переменных – это функция, независимые переменные $x_1,\ldots,x_n$ которых принимают значения объемов затрачиваемых или используемых ресурсов (число переменных n равно числу ресурсов), а значение функции имеет смысл величин объемов выпуска:

$$y = f(x) = f(x_1,\ldots,x_n). \qquad (2)$$

В формуле (2) y ($y \geq 0$) – скалярная, x – векторная величина, $x_1,\ldots,x_n$ координаты вектора x, т.е. $f(x_1,\ldots,x_n)$ есть числовая функция нескольких (многих) переменных $x_1,\ldots,x_n$. В связи с этим ПФ $f(x_1,\ldots,x_n)$ называют много ресурсной или многофакторной ПФ. Более правильной является такая символика $f(x_1,\ldots,x_n,a)$, где a – вектор параметров ПФ.

ПФ $y = f(x_1, x_2)$ называется статической, если ее параметры и ее характеристика f не зависят от времени t, хотя объемы ресурсов и объем выпуска могут зависеть от времени t, т.е. могут иметь представление в виде временных рядов:

$$x_1(0), x_1(1),\ldots, x_1(T); \quad x_2(0), x_2(1),\ldots, x_2(T); \quad y(0), y(1),\ldots, y(T); y(t) = f(x_1(t), x_2(t)).$$

Здесь t – номер года, $t = 0, 1, \ldots, T,$ $t = 0$ - базовый год временного промежутка, охватывающего годы 1, 2, …T.

Пример 2. Для моделирования отдельного региона или страны в целом (т.е. для решения задач на микроэкономическом, а также и на макроэкономическом уровне) часто используется ПФ вида $y = a_0 x_1^{a_1} x_2^{a_2}$, где a_0, a_1, a_2 параметры ПФ. Это положительные постоянные (часто a_1 и a_2 таковы, что $a_1 + a_2 = 1$). ПФ только что приведенного вида называется ПФ Кобба-Дугласа (ПФКД) по имени двух американских экономистов, предложивших ее использовать в 1929 г. ПФКД активно применяется для решения разнообразных теоретических и прикладных задач благодаря своей структурной простоте. ПФКД принадлежит к классу так называемых *мультипликативных* ПФ (МПФ). В приложениях ПФКД $x_1=K$ равно объему используемого основного капитала (объем используемых основных фондов – в отечественной терминологии), $x_2=L$ – затратам живого труда, тогда ПФКД приобретает вид, часто используемый в литературе: $Y = a_0 K^{a_1} L^{a_2}$.

Если число работников и их квалификация остаются неизменными, а число обслуживаемых ими станков (которое уже достаточно велико) увеличивается, например, в два раза, то это естественно не приведет к двойному росту объема выпуска. Отметим, что если $a_1 + a_2 < 1$, то графиком ПФКД является поверхность, которая напоминает выпуклую вверх «горку», крутизна которой падает, если точка (x_1, x_2) перемещается на «северо-восток» по плоскости Ox_1x_2.

Пример 3. Линейная ПФ (ЛПФ) имеет вид: $y = a_0 + a_1 x_1 + a_2 x_2$ (двухфакторная) и $y = a_0 + a_1 x_1 + ... + a_n x_n$ (многофакторная). ЛПФ принадлежит к классу так называемых аддитивных ПФ (АПФ). Переход от мультипликативной ЛПФ к аддитивной осуществляется с помощью операции логарифмирования. Для двухфакторной мультипликативной ПФ

$y = a_0 x_1^{a_1} x_2^{a_2}$ этот переход имеет вид: $\ln y = \ln a_0 + a_1 \ln x_1 + a_2 \ln x_2$.

Полагая $\ln y = w, \quad \ln x_1 = v_1 \quad и \quad \ln x_2 = v_2$ получаем аддитивную ПФ $w = \ln a_0 + a_1 v_1 + a_2 v_2$.

Выполняя обратный переход, из аддитивной ПФ получим мультипликативную ПФ.

Если сумма показателей степеней в ПФ Кобба-Дугласа $y = a_0 K^{a_1} L^{a_2}$ равна единице, то ее можно записать в несколько другой форме:

$$\frac{Y}{L} = \frac{a_0 K^{a_1} L^{a_1}}{L} = \frac{a_0 K^{a_1}}{L^{1-a_1}} = \frac{a_0 K^{a_1}}{L^{a_1}} = a_0 \left[\frac{K}{L}\right]^{a_1}, \quad т.е. \quad \frac{Y}{L} = a_0 \left[\frac{K}{L}\right]^{a_1}$$

Дроби $\frac{Y}{L} = z$ и $\frac{K}{L} = k$ называются соответственно производительностью труда и капиталовооруженностью труда. Используя новые символы, получим $z = a_0 k^{a_1}$, т.е. из двухфакторной ПФКД получим формально однофакторную ПФКД. В связи с тем, что $0 < a < 1$, из последней формулы следует что производительность труда z растет медленнее его капиталовооруженности. Однако этот вывод справедлив для случая статической ПФКД в рамках существующих технологий и ресурсов.

5.2. Формальные свойства производственных функций

Производственная функция $f(x_1, x_2)$ как формальная конструкция определена в неотрицательном октанте двумерной плоскости, т.е. определена при $x_1 \geq 0$, $x_2 \geq 0$. ПФ должна удовлетворять ряду (для каждой конкретной ПФ - своему) свойств:

1. $f(0,0) = 0$;

1”. $f(0, x_2) = f(x_1, 0) = 0$;

2.

$x(1) \geq x(0)$ $\left(x(1) \neq x(0)\right) \Rightarrow f\left(x(1)\right) > f\left(x(k) = (x_1(k), x_2(k), k = 0,1)\right)$

2”. $x > 0 \Rightarrow \frac{\partial f(x)}{\partial x_1} > 0 \quad (i = 1,2),\ x = (x_1, x_2)$;

3. $x > 0 \Rightarrow \frac{\partial^2 f(x)}{\partial {x_1}^2} \leq 0 \quad (i = 1,2),\ x = (x_1, x_2)$;

3”. $x > 0 \Rightarrow \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} \geq 0 \quad x = (x_1, x_2)$;

4. $f(tx_1, tx_2) = t^p f(x_1, x_2)$.

Свойство 1 означает, что без ресурсов нет выпуска. Свойство 1" означает, что при отсутствии хотя бы одного из ресурсов нет выпуска.

Свойство 2 означает, что с ростом затрат хотя бы одного ресурса объем выпуска растет. Свойство 2" (первая частная производная ПФ $\left[\frac{\partial f(x)}{\partial x_i}\right]$ положительна) означает, что с ростом затрат одного ресурса при неизменном количестве другого ресурса объем выпуска растет. Упорядоченная пара (x_1, x_2) чисел x_1 и x_2 для краткости здесь и далее обозначается символом *x*, т.е. $x = (x_1, x_2)$.

Свойство 3 (вторая частная производная ПФ $\left[\frac{\partial f(x)}{\partial x_i}\right]$ неположительная) означает, что с ростом затрат одного (*i*-го) ресурса при неизменном количестве другого ресурса величина прироста выпуска на каждую дополнительную единицу *i*-го ресурса не растет (закон убывающей эффективности). Свойство 3" означает, что при росте одного ресурса предельная эффективность другого ресурса возрастает. Если выполнены условия 3-3", то график ПФ есть поверхность, расположенная в неотрицательном ортанте $x_1 \geq 0, x_2 \geq 0, y \geq 0$ *x1* трехмерного пространства Ox_1x_2y выпуклая вверх. Вообще геометрический образ ПФ должен прежде всего ассоциироваться с выпуклой горкой, крутизна которой убывает, если точка (x_1, x_2) уходит в плоскости Ox_1x_2 на «северо-восток».

Свойство 4 означает, что ПФ является однородной функцией (ОФ) степени $p > 0$. При $p > 1$ с ростом масштаба производства в *t* раз (число $t > 1$), т.е. с переходом от вектора *x* к вектору *tx*, объем выпуска возрастает в t^p $(>t)$ раз, т.е. имеем рост эффективности производства от роста масштаба производства. При $p < 1$ имеем падение эффективности производства от роста масштаба производства. При $p=1$ имеем постоянную эффективность производства при росте его масштаба (или имеем независимость удельного выпуска от масштаба производства – в английской терминологии constant returns to scale).

Для ПФКД $y = a_0 x_1^{a_1} x_2^{a_2} (a_1 + a_2 = 1)$ свойства 1-4 выполняются.

Для ЛПФ $y = a_0 x_1^{a_1} x_2^{a_2} (a_0 > 0, a_1 > 0,\ a_2 > 0)$ свойства 1 и 1" (при $a_0=0$) и свойство 4 не выполняются.

Множество (линия) *lq* уровня $q = f(x_1, x_2)$ ($0<q$ – действительное число) ПФ $y = f(x_1, x_2)$ называется *изоквантой* ПФ. Иными словами, линия уровня *q* - это множество точек, в котором ПФ постоянна и равна *q*.

Различные наборы (v_1, v_2) и (w_1, w_2) затрачиваемых (используемых) ресурсов, принадлежащие одной и той же изокванте (т.е.

$q = f(v_1, v_2) = f(w_1, w_2)$), дают один и тот же объем выпуска q. Изокванта есть линия, расположенная в неотрицательном ортанте $\{(x_1, x_2)|\ x_1 \geq 0,\ x_2 \geq 0\}$ двумерной плоскости Ox_1x_2.

5.3. Предельные (маржинальные) и средние значения производственной функции

Пусть $y = f(x) = f(x_1, x_2)$ - ПФ. Дробь $\frac{f(x)}{x_i}$ $(i = 1,2)$ называется средней производительностью i-го ресурса (фактора производства) (СПФ) или средним выпуском по i-му ресурсу (фактору производства): $A_i = \frac{f(x)}{x_i}$.

Напомним, что в случае двухфакторной ПФКД $Y = a_0 K^{a_1} L^{a_2}$ - для средних производительностей $\frac{K}{Y}$ и $\frac{L}{Y}$ основного капитала и труда были использованы соответственно термины капиталоотдача и производительность труда. Эти термины используют и применительно к любым двухфакторным ПФ, у которых $x_1 = K$ *и* $x_2 = L$.

Пусть $y = f(x) = f(x_1, x_2)$ - ПФ.

Ее первая частная производная $\frac{\partial f(x)}{\partial x_i}$ $(i = 1,2)$ называется предельной (маржинальной) производительностью i-го ресурса (фактора производства) (ППФ) или предельным выпуском по i-му ресурсу (фактору производства): $M_i = \frac{\partial f(x)}{\partial x_i}$. Обозначим символами Δx_1 и $\Delta_i(f(x))(\Delta_1 f(x_1, x_2) = f(x_1 + \Delta x_1, x_2) - f(x_1, x_2))$;

$\Delta f(x_1, x_2) = f(x_1, x_2 + \Delta x_2) - f(x_1, x_2)$ соответственно, приращение переменной x, и соответствующее ей частное приращение ПФ $f(x)$. При малых Δx_i , имеем приближенное равенство $\frac{\partial f(x)}{\partial x_i}(i = 1,2)$.

Следовательно, ППФ (приближенно) показывает, на сколько единиц увеличится объем выпуска y, если объем затрат x i-го ресурса вырастает на одну (достаточно малую) единицу при неизменных объемах другого затрачиваемого ресурса. Здесь предельную величину $\frac{\partial f(x)}{\partial x_1}$, (т.е. ППФ) целесообразно интерпретировать, используя близкое к ней отношение малых конечных величин, т.е. $\Delta f(x)$ и Δx_i.

Отмеченное обстоятельство является ключевым для понимания экономического смысла ППФ $\frac{\partial f(x)}{\partial x_i}$. С другими предельными величинами следует поступать аналогичным образом.

Задача 1. Для ПФКД $y = a_0 x_1^{a_1} x_2^{a_2}$ найти в явном виде A_1, A_2, M_1 и M_2. Решение задачи. Имеем:

$$A_1 = \frac{y}{x_1} = \frac{f(x)}{x_1} = a_0 x_1^{a_1 - 1} x_2^{a_2}; \quad A_2 = \frac{y}{x_2} = \frac{f(x)}{x_2} = a_0 x_1^{a_1} x_2^{a_2 - 1}$$

$$M_1 = \frac{\partial f(x)}{\partial x_1} = a^1 A_1; \quad M_2 = \frac{\partial f(x)}{\partial x_2} = a_2 A_2;$$

$$\frac{M_1}{A_1} = a_1 \leq 1 \Rightarrow M_1 \leq A_1; \quad \frac{M_2}{A_2} = a_2 \leq 1 \Rightarrow M_2 \leq A_2.$$

Для ПФ $y = f(x)$ (не только для ПФКД) неравенства $M_1 \leq A_1$ (i = 1,2) (т.е. предельная производительность *i*-го ресурса не больше средней производительности этого ресурса) обычно выполняются.

Задача 2. Для ЛПФ $y = a_0 + a_1 x_1 + a_2 x_2$ $(a_0 > 0, a_1 > 0,\ a_2 > 0)$ найти в явном виде A_1, A_2, M_1 иM_2.

Решение задачи. Имеем:

$$A_1 = \frac{y}{x_1} = \frac{f(x)}{x_1} = \frac{a_0}{x_1} + a_1 + a_2 \frac{x_2}{x_1};$$

$$A_2 = \frac{y}{x_2} = \frac{f(x)}{x_2} = \frac{a_0}{x_2} + a_1 \frac{x_1}{x_2} + a_2$$

$$M_1 = \frac{\partial f(x)}{\partial x_1} = a_1; \quad M_2 = \frac{\partial f(x)}{\partial x_2} = a_2;$$

$$\frac{M_1}{A_1} \leq 1 \Rightarrow M_1 \leq A_1; \quad \frac{M_2}{A_2} \leq 1 \Rightarrow M_2 \leq A_2.$$

Пусть $y = f(x)$ - ПФ, $x = (x_1, x_2)$. Отношение предельной производительности M_i, *i*-го ресурса к его средней производительности А называется (частной) эластичностью выпуска по *i*-му ресурсу (по фактору производства) (ЭВФ):

$$E_i = \frac{M_1}{A_1} = \frac{x_1}{f(x)} \frac{\partial f(x)}{\partial x_i} \quad (i = 1,2).$$

Сумма $E_1 + E_2 = E_x$ называется эластичностью производства. Поскольку при малом приращении Δx_i имеем приближенное равенство

$$E_i = \left[\frac{\partial f(x)}{\partial x_i}\right] / \left[\frac{f(x)}{x_1}\right] \approx \left[\frac{\Delta_i f(x)}{f(x)}\right] / \left[\frac{\Delta x_i}{x_i}\right]$$

(крайнее правое выражение есть отношение двух относительных величин $\frac{\Delta_i f(x)}{f(x)}$ и $\frac{\Delta x_i}{x_i}$ постольку E_i (приближенно) показывает, на сколько процентов увеличится выпуск *y*, если затраты i-го ресурса увеличатся на один процент при неизменных объемах другого ресурса. Пояснение выражения E_i, содержащего предельную величину $\frac{\partial f(x)}{\partial x_i}$, с помощью выражения, содержащего конечное приближение $\frac{\Delta f(x)}{f(x)}$ предельной величины, является ключевым в понимании экономической сути частной эластичности выпуска по *i*-му ресурсу.

Задача 3. Выписать в явном виде для ПФКД выражения для E_1, E_2 *и* E_x

Решение задачи. Имеем:

$$E_1 = a_1,\ E_2 = a_2, E_x = E_1 + E_2 = a_1 + a_2.$$

Задача 4. Для ЛПФ $y = a_1x_1 + a_2x_2\ (a_0 = 0)$ найти выражения для E_1, E_2 *и* E_x.

Решение задачи. Имеем:

$$E_1 = \frac{x_1}{f(x)}\frac{\partial f(x)}{\partial x_1} = \frac{a_1x_1}{a_1x_1 + a_2x_2}; \quad E_2 = \frac{x_2}{f(x)}\frac{\partial f(x)}{\partial x_2} = \frac{a_2x_2}{a_1x_1 + a_2x_2};$$

$$E_x = E_1 + E_2 = 1$$

Пусть $y = f(x)$ - ПФ, $x = (x_1, x_2)$. Предельной нормой замены (замещения) *i*-го ресурса (фактора производства) *j*-м (аббревиатура: ПНЗФ и символика: R_{ij}) называется выражение:

$$R_{ij} = -\frac{dx_j}{dx_i} \quad (i, j = 1,2) \tag{3}$$

при постоянной *y*.

Обратим внимание на то, что *i* – номер заменяемого ресурса, *j* – номер замещающего ресурса. Используется также термин: предельная технологическая норма замены (замещения) *i*-го ресурса (фактора производства) *j* -м ресурсом (фактором производства). Приведем более краткий (но менее точный) термин: (предельная) норма замены (замещения) ресурсов.

Пусть выпуск *y* является постоянным (т.е. все наборы затрачиваемых ресурсов расположены на одной изокванте), тогда первый полный дифференциал *dy* ПФ *y=f(x)* тождественно равен нулю:

$$0 = dy = \frac{\partial f(x)}{\partial x_1} dx_1 + \frac{\partial f(x)}{\partial x_2} dx_2$$

(здесь dx_1, dx_2 - дифференциалы переменных x_1, x_2), откуда, выражая первый дифференциал dx_j, получим ($i \neq j$)

$$dx_j = -\frac{\frac{\partial f(x)}{\partial x_i}}{\frac{\partial f(x)}{\partial x_j}} dx_i \quad (i, j = 1,2) \tag{4}$$

откуда, поделив на dx_i, получим

$$\frac{dx_j}{dx_i} = -\frac{\frac{\partial f(x)}{\partial x_i}}{\frac{\partial f(x)}{\partial x_j}} \quad (i, j = 1,2) \tag{5}$$

На основании (3), (4) и (5) имеем:

$$R_{ij} = -\frac{dx_j}{dx_i} = \frac{\frac{\partial f(x)}{\partial x_i}}{\frac{\partial f(x)}{\partial x_j}} > 0 \quad (i \neq j, i = 1,2) \tag{6}$$

Непосредственно проверяется, что для двухфакторной ПФ справедливо равенство:

$$R_{12} = \frac{E_1}{E_2} \frac{x_2}{x_1}$$

т.е. (предельная) норма замены первого ресурса вторым равна отношению эластичностей выпуска по первому и второму ресурсам, умноженному на отношение объема второго ресурса к объему первого ресурса. Если $x_1 = K, x_2 = L$, то отношение – $\frac{x_1}{x_2} = \frac{K}{L}$ называется капитало-вооруженностью труда. В этом случае (предельная) норма замены основного капитала трудом равна отношению эластичностей выпуска по основному капиталу и труду, поделенному на капиталовооруженность труда.

Пусть ПФ – двухфакторная. При постоянном выпуске y и малых приращениях Δx_1 и Δx_2, имеем приближенное равенство

$$R_{12} = -\frac{dx_2}{dx_1} \approx -\frac{\Delta x_2}{\Delta x_1} \tag{7}$$

Задача 5. Для ПФКД $y = a_0 x_1^{a_1} x_2^{a_2}$ выписать в явном виде выражения R_{12} и R_{21}. Решение задачи. Имеем:

$$R_{12} = \left[\frac{\partial y}{\partial x_1}\right] / \left[\frac{\partial y}{\partial x_2}\right] = \frac{a_1}{a_2}; \quad R_{21} = \left[\frac{\partial y}{\partial x_2}\right] / \left[\frac{\partial y}{\partial x_1}\right] = \frac{a_2}{a_1}\frac{x_1}{x_2}.$$

Задача 6. Для ПФ $y = a_0 + a_1 x_1 + a_2 x_2$, выписать в явном виде выражения R_{12} и R_{21}.

Решение задачи. Имеем:

$$R_{12} = \left[\frac{\partial y}{\partial x_1}\right] / \left[\frac{\partial y}{\partial x_2}\right] = \frac{a_1}{a_2}; \quad R_{21} = \left[\frac{\partial y}{\partial x_2}\right] / \left[\frac{\partial y}{\partial x_1}\right] = \frac{a_2}{a_1}.$$

5.4. 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.

$$D = f(p). \tag{5.4.1}$$

Функция убывает с ростом *p* (цены). Аналогично была определена и функция предложения

$$S = \varphi(p). \tag{5.4.2}$$

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. 5.1).

Мы анализируем его, чтобы найти ответ на вопрос, является ли распределение ресурсов эффективным в условиях рыночного равновесия и достигает ли общая выгода своего максимального значения. Когда на рынке существует равновесие, равновесная цена определяет продавцов и покупателей, которые участвуют в рынке. На рынке товар покупается такими покупателями, если они оценивают товар выше его рыночной цены (отрезок, представленный сечением SA на кривой спроса); люди, которые оценивают продукт ниже его цены (доля, представленная сечением AE), отказываются покупать его. Точно так же производители, чьи затраты ниже, чем стоимость продукта (представленная сечением DA), производят и продают продукт; фирмы, издержки которых выше рыночной цены (представленной сечением АГ), отказываются его производить.

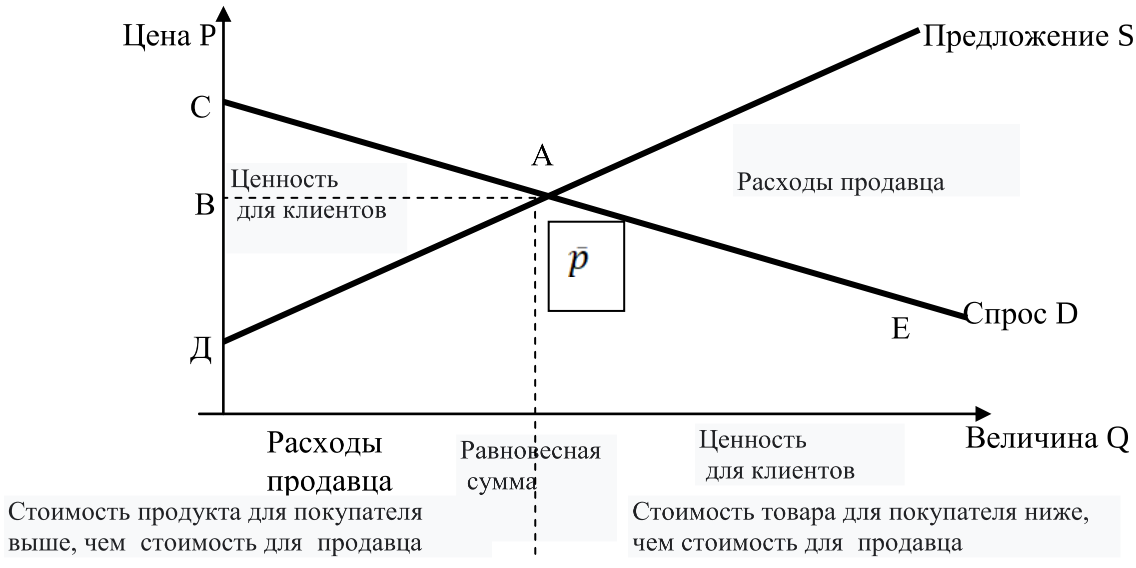


Рис. 5.1.

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);

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:

$$\frac{\Delta p}{p} = \frac{2}{10} \cdot 100 = 20\%.$$

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:

$$\frac{\Delta d}{d} = \frac{-10}{100} \cdot 100 = -10\%.$$

Коэффициент эластичности, вычисляемый как отношение относительных величин приростов спроса и цены, составит:

$$E_{d.p} = \frac{\Delta d}{d} : \frac{\Delta p}{p} = \frac{-10}{100} : \frac{2}{10} = -\frac{1}{2}. \quad (5.4.3)$$

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:

$$E_{S.p} = \frac{\Delta S}{S} : \frac{\Delta p}{p} = \frac{\Delta S p}{\Delta p S}. \quad (5.4.5)$$

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:

$$D_t = f(p_t); \quad (5.4.6)$$

$$S_t = \phi(p_{t-1}). \quad (5.4.7)$$

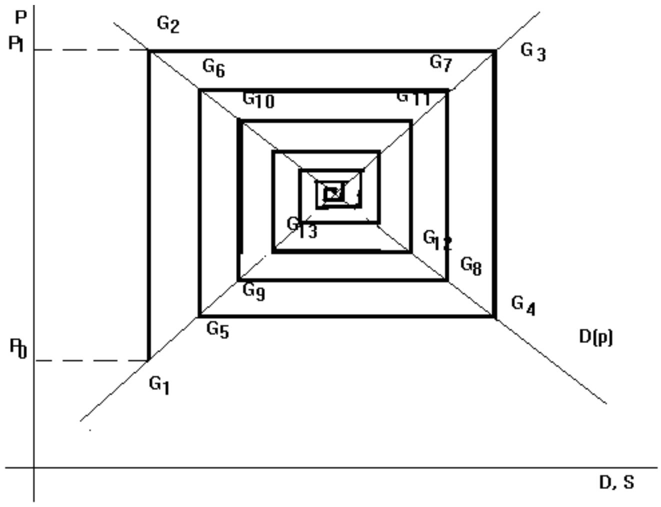


рис. 5.2

Данный вид модели называется *паутинообразной* (рис. 5.2).

В экономической теории важным является понятие равновесия, то есть такого состояния объекта, которое он сохраняет при отсутствии внешних воздействий. Задачи экономической динамики включают как описание процессов выхода к состоянию равновесия, так и процессов трансформации самого этого состояния под воздействием внешних сил.

Эта модель позволяет исследовать устойчивость цен и объемов товаров на рынке, описываемом традиционными кривыми спроса и предложения при наличии запаздывания во времени (лага).

Пусть производители (например, фермерская хозяйства) определяют предложение товара в текущем периоде на основе цен, установившихся в предшествующем периоде, то есть $Q^s\ (t)= S_t\ (p_{t-1})$.

Таким образом, в функцию предложения вклинивается временной лаг продолжительностью в одну единицу времени. Действительно, решение об объеме производства принимается с учетом текущих цен, производственный цикл имеет определенную продолжительность, и соответствующее этому решению предложение появится на рынке по окончании данного цикла (как показано на рис.5.2.).

Кривая спроса характеризует зависимость объема спроса на товар от цены товара в данном периоде, то есть $Q^D(t) = D_1(p_1)$. Таким образом, динамику цены можно описать системой уравнений

$$\{Q_t^s = S_t(p_{t-1}), \quad Q_t^D = D_t(p_t), \quad Q_t^d = Q_t^S\}$$

или одним уравнением

$$D_t(p_t) = S_t(p_{t-1}) \qquad (5.4.8)$$

Из этого уравнения можно найти значение цены p_t в текущий момент времени по известному значению p_{t-1} в предшествующий момент времени. Схема решения очень проста:

$$Q_0 \to p_0 = D^{-1}(Q_0) \to Q_1 = S(p_0) \to p_1 = D^{-1}(Q_1) \to Q_2 = S(p_1) \to \ldots$$

(где D^{-1} – обратная функция спроса).

В качестве частного случая рассмотрим паутинообразную модель, в которой функции спроса и предложения линейны:

$$S(p) = A + \mathrm{B}p_{t-1}, \;\; D(p) = C - Ep_t, S(p) = D(p). \qquad (5.4.9)$$

Здесь Â > 0, так как функция предложения возрастающая; Å > 0, так как функция спроса убывающая; $\mathrm{C} > \mathrm{A} > 0$, то есть $\mathrm{D}(0) > \mathrm{C}(0) > 0$ (считаем, что при нулевой цене спрос превышает предложение). Уравнение, описывающее динамику такой системы, имеет вид

$$D(p_t) = S(p_{t-1}), или C - Ep_t = A + Bp_{t-1}.$$

Найдем сначала равновесную цену $p*$ и равновесный объем производства $Q*$. Они должны удовлетворять уравнениям

$$Q^* = C - Ep^* = A + Bp^*,$$

Откуда

$$p^* = \frac{C - A}{B + E} и Q^* = \frac{BC + AE}{B + E}.$$

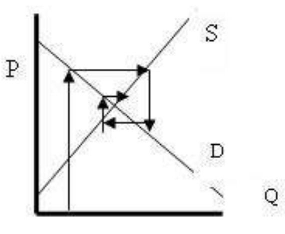


Рис.5.3. *а*

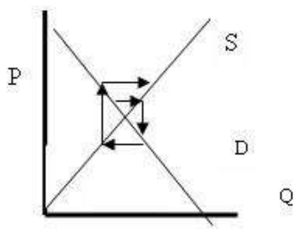


Рис.5.3.б

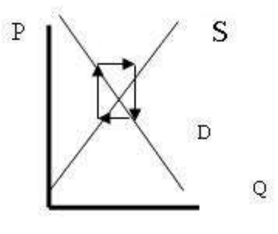


Рис.5.3.в

Далее необходимо исследовать поведения цен и объемов производства в том случае, если начальная точка не совпадает с равновесной точкой. Вначале эту задачу можно решить графически, получив рисунок типа «паутины», подтверждающий ее название. Задав первое первоначальное количество товара и цену, не совпадающие с точкой равновесия, будем последовательно

наносить точки в соответствии с процедурой расчета по модели, соединяя их горизонтальными или вертикальными прямыми линиями. Из графического анализа можно получить следующие результаты. Если кривая предложения наклонена круче, чем кривая спроса, то равновесие на таком рынке будет устойчивым (см. рис.5.3.*а*). Если кривая спроса наклонена круче, чем кривая предложения, то равновесие на рынке будет неустойчивым (см. рис.5.3. б). Наконец, при равном наклоне кривых спроса и предложения цены на рынке будут испытывать регулярные колебания с постоянной амплитудой (см. рис.5.3.в). Теперь перейдем к формальному анализу модели. Выражая p_t через p_{t-1}, имеем следующее рекуррентное соотношение $p_t = \frac{C-A}{E} - \frac{B}{E}$.
Последовательно применяя это соотношение, находим

$$p_1 = \frac{C-A}{E} - \frac{B}{E} \cdot p_0; p_2 = \frac{C-A}{E} - \frac{B}{E} \cdot \left[\frac{C-A}{E} - \frac{B}{E}\right] \cdot p_0$$

Или в общем виде

$$p_1 = \frac{C-A}{E} \cdot \left[1 - \frac{B}{E} + \left[\frac{B}{E}\right]^2 + \ldots + (-1)^{t-1}\left[\frac{B}{E}\right]^{t-1}\right] + (-1)^t \left[\frac{B}{E}\right]^t \cdot p_0$$

Выражение в скобках есть сумма геометрической прогрессии:

$$S_n = a_1 \cdot (1 + q + q^2 + \ldots + q^{n-1}) = a_1 \frac{1-q^n}{1-q}.$$

Если $|q| < 1$, то $\lim\limits_{n\to\infty} S_n = \frac{a_1}{1-q}$.

Для паутинообразной модели $q = -\frac{B}{E}$, $a_1 = \frac{C-A}{E}$. Отсюда получаем выражение для цены p_t в произвольный момент времени *t*.

$$p_t = \frac{C-A}{E} \cdot \frac{1-(-1)^t \left[\frac{B}{E}\right]^t}{1+\frac{B}{E}} + (-1)^t \left[\frac{B}{E}\right]^t \cdot p_{0.} \qquad (10)$$

(5.4.10)

Очевидно при $\frac{B}{E} < 1$ $\left[\frac{B}{E}\right]^t \to 0$ и $p_t \to \frac{C-A}{E} = p^*$, то есть при более крутом наклоне кривой предложения, чем кривой спроса, равновесие является устойчивым. Если $\frac{B}{E} > 1$, то есть более крутой является кривая спроса, то $\left[\frac{B}{E}\right]^t \to \infty$ и процесс расходится (равновесие неустойчиво). При

$\frac{B}{E} = 1$, то есть при B=E, значение p_t чередуется вокруг равновесного значения.

Итак, определяющим моментом для устойчивости системы является менее сильная, сглаживающая реакция на изменения цены той функции, которая имеет временной лаг (здесь – функция предложения).

В реальности при $\frac{B}{E} > 1$ бесконечно возрастающих колебаний, конечно, не будет, так как при больших отклонениях от равновесия линейное приближение становится нереалистичным. В более реалистической нелинейной модели устанавливаются нелинейные колебания большой, но конечной амплитуды, которые являются прообразом экономических циклов подъемов и спада производства.

Самостоятельно предлагается рассмотреть следующую задачу: предположим, что временной лаг, равный 1, присутствует не в функции предложения, а в функции спроса:

$$S_t = A + B \cdot p; D_t = C - E \cdot p_{t-1}; S_t = D_t.$$

Каким станет условие сходимости к равновесной точке? Изобразите этот процесс графически.

Вычисление эластичности

Важная характеристика экономических процессов – ***эластичность***, которая показывает, на сколько процентов изменится зависимая переменная ***Y*** при увеличении влияющей переменной ***X*** на 1 % :

Э =(ΔУ/Y)/(ΔX/X)

Применение компьютера позволяет вычислить эластичность по всему диапазону ***X***, а не только средние значения, как при ручном счете.

В качестве ***X*** и ***Y*** берутся их средние значения на соответствующих интервалах ΔX и ΔУ , расчет ведется по аппроксимирующей функции *Ŷ:*

Э = (Ŷ1 – Ŷ0)/(Ŷ1+ Ŷ0)/(X1 – X0)*(X1+X0)

где индексы 0 и 1 относятся к первым двум значениям переменных ***X*** и ***Ŷ***. Затем формула копируется на весь диапазон, кроме последней ячейки; в

Модели1 расчет начинается с температуры $10^{\mathbf{o}}$. Графики показывают, что расчет эластичности по разным моделям приводит к различным результатам. Обычно экономисты используют среднюю эластичность

$$Э = \frac{\Delta У}{\Delta X} \times \frac{X}{У}$$

Эластичность позволяет изучать влияние добавок ***X*** на изменение ***Y*** при различных значениях влияющей переменной.

Пример вычисления эластичности в компьютере

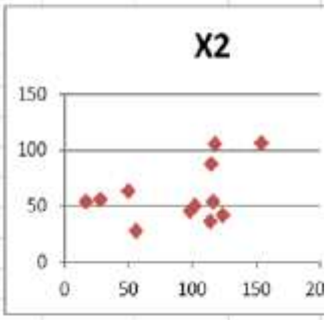
	A	B	C	D
1	X1	X2	y	Y^
2	115	106	202	197,7755
3	27	57	64	81,9073
4	15	52	43	68,9275
5	53	61	114	98,5707
6	52	26	120	39,4928
7	100	49	89	96,443
8	110	51	112	103,626
9	120	44	54	95,716
10	113	34	83	76,2657
11	153	103	238	207,2947
12	116	86	162	164,6184
13	96	47	77	91,5574
14	89,17	59,67	113,17	

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множесті	0,85466
R-квадрат	0,730444
Нормиро	0,670542
Стандарті	34,41801
Наблюде	12

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*ачимость F*
Регрессия	2	28890,27	14445,13	12,19411	0,002741
Остаток	9	10661,4	1184,6		
Итого	11	39551,67			



	Коэффициен	*артная ош*	*татистик*	*Р-Значение*	*ижние 95%*	*ерхние 95%*	*жние 95*
Y-пересеч	-20,1074	29,10478	-0,69086	0,507078	-85,947	45,73223	-85,94
Перемен	0,419109	0,272396	1,538604	0,158283	-0,19709	1,035311	-0,1970
Перемен	1,607321	0,450672	3,566499	0,006057	0,58783	2,626812	0,5878

17	Э	0,67
18	Э(X1; Y)	0,33
19	Э(X2; Y)	0,85

	Столбец 1	*Столбец 2*	*Столбец 3*
Столбец 1	1		
Столбец 2	0,4322631	1	
Столбец 3	0,5911629	0,81212157	1

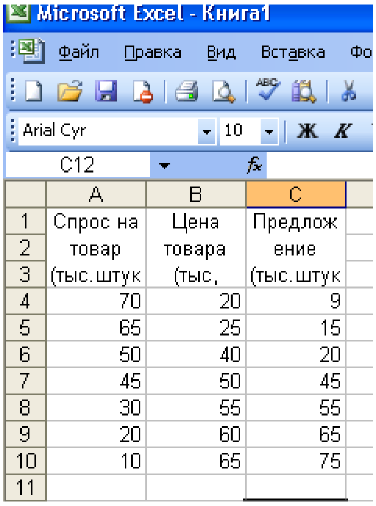
Здесь E=(D3-D2)/(D3+D2)/(A3-A2)*(A3+A2);

E(X1, Y) =G19*A14/C14;

E(X2, Y)= =G20*B14/C14.

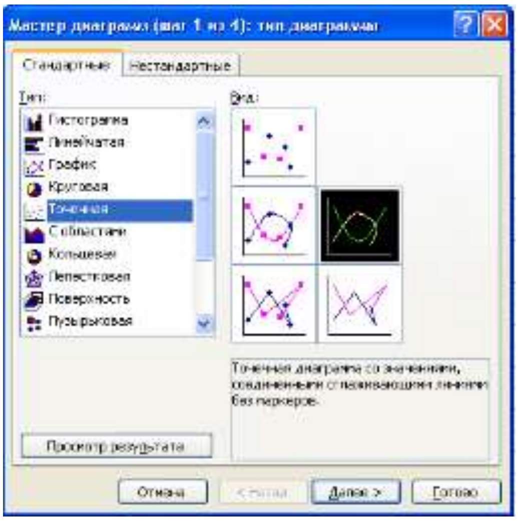
Методические указания для построения графика спроса и предложения в MS Excel

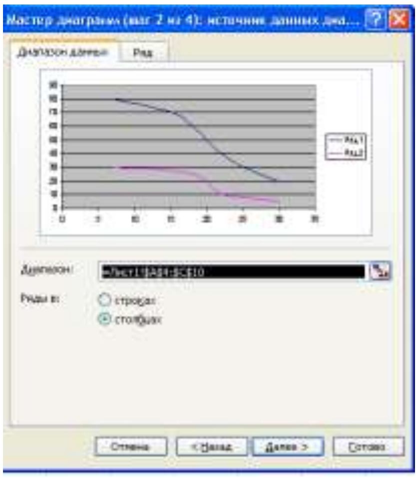
Последовательность построения графика спроса и предложения покажем в программе MS Excel:



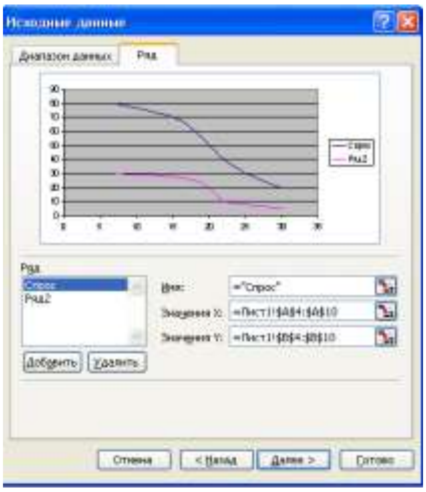
	A	B	C
1	Спрос на	Цена	Предлож
2	товар	товара	ение
3	(тыс.штук	(тыс,	(тыс.штук
4	70	20	9
5	65	25	15
6	50	40	20
7	45	50	45
8	30	55	55
9	20	60	65
10	10	65	75
11			

Выделяем ячейки A4:C10, из панели инструментов нажимаем кнопку  мастера диаграмм, тогда получим диалоговое окно. Построение диаграммы состоит из 4-х шагов. Первый шаг заключается в выборе Типа диаграммы. Из этого окно выбираем Точечная и нажимаем кнопку Далее.

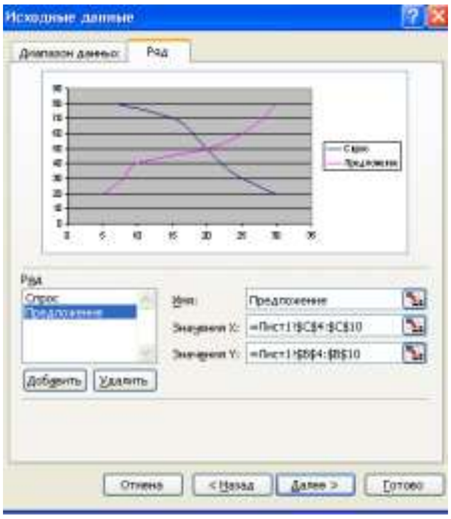




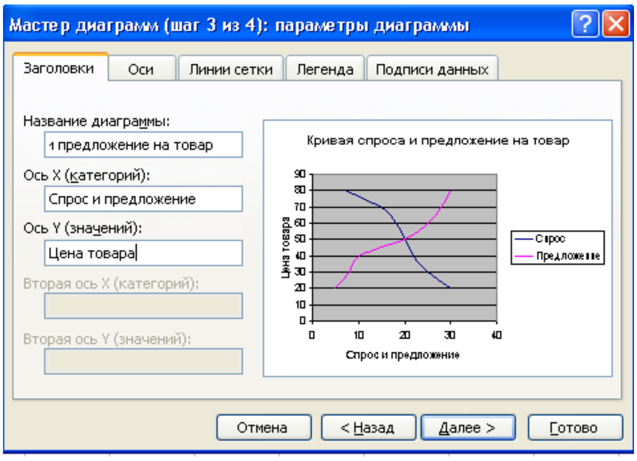
На втором шаге выбираем Ряд, чтобы заменить размещение кривых в оси координат. В разделе Ряд1 размещаем спрос: в разделе Значение X вводим A4: A10, в раздел Значение У вводим B4:B10.



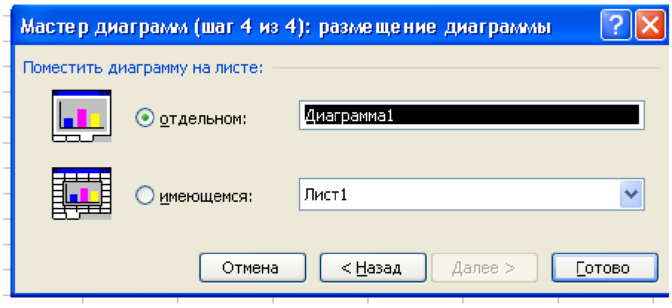
В Ряд2 таким же образом разместим предложение.



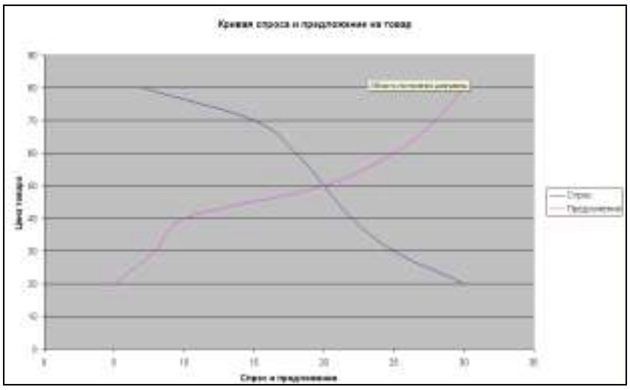
И нажимаем кнопку Далее. В новом окно в разделе Заголовок заполняем заголовки оси X и У.



После этого нажимаем кнопку Далее.



В этом окне мы отмечаем пункт размещения диаграммы на отдельном листе и нажимаем кнопку *Готово*, тогда получим следующий результат:



Опорные слова

Производственная функция, факторы производства, одно ресурсный, однофакторный, независимая переменная, зависимая переменная, многофакторный, мультипликативный, однородная функция, изокванта, маржинальный, средние значения, аддитивный, предельная норма замены, взаимодополняемость, взаимозаменяемость, производительность труда, функция Кобба-Дугласа, статика, динамика, экономическая динамика, непрерывное время, дискретное время, темп роста, прирост, равновесие, паутинообразная модель, функция предложения, функция спроса, кривая спроса, кривая предложения, временной лаг, равновесная точка, сходимость, равновесная цена.

Контрольные вопросы

1. Как определяется (средняя) производительность труда и капитало вооруженность (фондовооруженность) труда? Какие возможны варианты взаимосвязи между ними в случае производственной функции Кобба-Дугласа?
2. Назовите основные свойства, которыми должна обладать производственная функция. Приведите примеры производственных функций, которые обладают всеми основными свойствами.
3. Анализ производства и издержек.
4. Производственные функции и их типы.
5. Свойства производственных функций и их виды.
6. Производственная функция Кобба-Дугласа.
7. Функции издержек.
8. Эконометрический анализ спроса и предложения.
9. Что такое изокванта? В чем ее экономический смысл? Как определяется (средняя) производительность капитала (капиталоотдача)?
10. Дайте содержательную интерпретацию (частной) эластичности выпуска по i-му ресурсу.

ГЛАВА 6. СИСТЕМЫ ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ

6.1. Модель спроса и предложения

Основные экономические модели требуют для своего описания систем взаимосвязанных уравнений. Для построение этих моделей обычно используют временные ряды уровней различных переменных, часть которых принимают за эндогенные, а часть за экзогенные. Выбор переменных определяется исследователем, но обычно экзогенные переменные или не зависят от нас (температура воздуха, курс доллара, цена нефти), или мы можем ими управлять (инвестиции, выпуск продукции).

Для примера берем модель спроса и предложения на конкурентном рынке, а также рыночной цены (***p***) в зависимости от величины дохода (***x***) на душу населения. Ее называют “паутинной моделью”, так как движение спроса и предложения к равновесию в соответствующей системе координат напоминает паутину. Пример взят из учебника [9].

Изменение во времени спроса, предложения и цены на конкурентном рынке закреплено в следующих утверждениях экономической теории:

1) Текущий уровень спроса объясняется текущей ценой товара и текущим располагаемым доходом на душу населения, причем спрос падает с ростом цены и растет с ростом дохода.

2) Текущее предложение объясняется ценой товара в предшествующем периоде и возрастает с ростом этой цены.

3) Текущее значение рыночной цены устанавливается при балансе текущего спроса и текущего предложения товара.

Кратко это можно записать, с учетом случайных возмущений:

$$\text{Спрос} = a_0 + a_1 \cdot \text{цена} + a_2 \cdot \text{доход} + \text{Возмущение1}$$

$$\text{Предложение} = b_0 + b_1 \cdot \text{цена вчера} + \text{Возмущение2}$$

$$\text{Спрос} = \text{предложение(тождество)}.$$

Соответствующая система уравнений и тождеств:

$$\begin{aligned} &d = a_0 + a_1 \cdot p + a_2 \cdot x + \varepsilon_1 \\ &s = b_0 + b_1 \cdot p_{t-1} + \varepsilon_2 \\ &d = s \\ &a_1 < 0, \quad a_2 > 0, \quad b_1 > 0, \end{aligned} \qquad (9.1)$$

В данном случае d, s, p эндогенные, x,$p_{\,t-1}$ предопределенные (экзогенная и лаговая).

Второе уравнение является обычным уравнением регрессии, и его можно настраивать, используя обычный метод наименьших квадратов. А с первым уравнением так поступить нельзя, так как в него входят две эндогенных переменных. Такая модель называется *структурной,* она возникает непосредственно из экономических предпосылок. Требуется преобразовать модель к *приведенному виду*, где в левой части будут стоять эндогенные переменные, а в правой – предопределенные. Можно решать эту задачу путем последовательной замены эндогенных переменных.

Мы рассмотрим метод, основанный на преобразовании матриц. Объединим эндогенные переменные в вектор Y, а предопределенные – в вектор X: Y= (d, s, p); $\mathrm{X} = (1, p_{\,t-1}, x)$

Единица в векторе X появилась, чтобы работать с коэффициентами $a_{\,0}$ и $b_{\,0}$. В матричном виде система уравнений и тождеств выглядит

$$\begin{array}{llr}
\mathrm{AY} & +\quad \mathrm{BX} & =0 \\
1\cdot d+0\cdot s-a_{\,1}\cdot p & +(-a_{\,0})\cdot 1+0\cdot p_{\,t-1}+(-a_{\,2})\cdot x & =0 \\
0\cdot d+1\cdot s+0\cdot p & +(-b_{\,0})\cdot 1+(-b_{\,1})\cdot p_{\,t-1}+0\cdot x & =0 \\
1\cdot d+(-1)\cdot s-0\cdot p & +0\cdot 1+0\cdot p_{\,t-1}\qquad +0\cdot x & =0
\end{array}$$

Здесь матрицы А и В

$$A \quad \begin{matrix} 1 & 0 & -a_{\,1} \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{matrix} \qquad\qquad B \quad \begin{matrix} -a_{\,0} & 0 & -a_{\,2} \\ -b_{\,0} & -b_{\,1} & 0 \\ 0 & 0 & 0 \end{matrix}$$

Приведенная форма модели Y = M X**.** Компоненты матрицы M

$$\mathrm{M} \quad \begin{matrix} \alpha_{\,0} & \alpha_1 & a_{\,2} \\ a_{\,0} & \alpha_1 & \alpha_2 \\ \alpha_{\,0} & \alpha_1 & a_{\,2} \end{matrix}$$

получим преобразованием M = A^{-1}B, где A^{-1} означает обратную к А матрицу.

Приведенная форма модели:

$$d = \alpha_0 + \alpha_1 \cdot p_{t-1} + \alpha_2 \cdot x$$
$$s = \alpha_0 + \alpha_1 \cdot p_{t-1} + \alpha_2 \cdot x$$
$$p = \beta_0 + \beta_1 \cdot p_{t-1} + \beta_2 \cdot x$$

где

$$\alpha_0 = b_0; \qquad \alpha_1 = b_1; \qquad \alpha_2 = 0$$
$$\beta_0 = (b_0 - a_0)/a_1; \qquad \beta_1 = b_1/a_1; \qquad \beta_2 = -a_2/a_1.$$

6.2. Идентифицируемость системы

Интерес представляют коэффициенты не приведенной модели, а структурной, которая имеет экономический смысл. Поэтому после настройки по статистическим данным приведенной модели и оценки ее коэффициентов требуется вычислить по ним структурные коэффициенты. Но это не всегда получается: возникает проблема идентификации – единственности соответствия между приведенной и структурной формами модели. Структурные модели можно подразделить на три вида [2]:

- идентифицируемые;
- неидентифицируемые;
- сверхидентифицируемые.

Модель *идентифицируема*, если структурные коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, то есть число параметров структурной модели равно числу параметров приведенной формы модели.

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

Модель *сверхидентифицируема*, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить несколько значений каждого структурного коэффициента, число структурных коэффициентов меньше числа коэффициентов приведенной формы. Модель может быть практически решена при применении специальных методов.

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых необходимо проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверх идентифицируемая модель содержит хотя бы одно сверх идентифицируемое уравнение.

Чтобы уравнение было идентифицируемо, нужно, чтобы число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного. Условие идентифицируемости модели может быть записано в виде следующего правила:

Предопределенных + 1 = Эндогенных идентифицируемо.

Предопределенных + 1 <Эндогенных неидентифицируемо. *Предопределенных + 1> Эндогенных* сверх идентифицируемо.

Если обозначить число эндогенных переменных в ***j***-м уравнении системы через *Н*, а число предопределенных переменных, которые содержатся в системе, но не входят в данное уравнение, через *D*, то

D +1 = H идентифицируемо.

D+1<H неидентифицируемо.

D +1> H сверх идентифицируемо.

В исследуемой модели *d, p, s* эндогенные; *x, p(t-1)* предопределенные, во всей системе их 2. Исследуем модель:

		Предопр.	D	D+1	H
$d = \alpha_0 + \alpha_1 \cdot p + \alpha_2 \cdot x$	1	1	2	2	идент
$s = \alpha_0 + \alpha_1 \cdot p_{t-1} + \alpha_2 \cdot x$	1	1	2	1	сверх
$p = \beta_0 + \beta_1 \cdot p_{t-1} + \beta_2 \cdot x$	0	2	2	2	идент

Значит, модель сверх идентифицируема.

6.3. Методы решения систем эконометрических уравнений

Идентифицируемую систему эконометрических уравнений можно решить *Косвенным методом наименьших квадратов (КМНК)*. Суть метода в следующем [2]:

- Преобразование структурной формы модели в приведенную;
- Оценка коэффициентов уравнений приведенной формы обычным методом наименьших квадратов;
- Преобразовать их в коэффициенты структурной модели. Если при вычислении коэффициентов структурной формы (6.1) учитывать возмущения,

то они войдут в оба уравнения модели, и принцип независимости эндогенных переменных и остатков будет нарушен. Это приведет к смещению (отсутствию состоятельности) параметров модели. Для подавления этого эффекта, а также для настройки сверх идентифицируемых моделей используется *двухшаговый метод наименьших квадратов ДМНК*.

Основная идея ДМНК – на основе приведенной формы модели получить для сверх идентифицируемого уравнения оцененные значения эндогенных переменных, содержащихся в правой части уравнения. Далее, подставив их в правые части уравнений вместо фактических значений, можно применить обычный МНК к структурной форме сверх идентифицируемого уравнения.

Метод получил название “двухшаговый метод наименьших квадратов”, ибо МНК используется дважды: на первом шаге при определении коэффициентов приведенной формы модели и нахождении на ее основе оценок оцененных значений эндогенных переменных $\hat{Y}$ и на втором шаге применительно к структурному сверхидентифицируемому уравнению при определении структурных коэффициентов модели с использованием оцененных значений эндогенных переменных. Оцененные значения играют роль так называемых *инструментальных переменных (instrumental variables, IV, instruments)* – переменных, которые применяются, если обычные переменные коррелируют с возмущениями. Инструментальные переменные коррелируют с обычными переменными, но не коррелируют с возмущениями, что приводит к состоятельности (consistency) модели. Расчеты с использованием инструментальных переменных включены в статистические пакеты, так что не удивляйтесь, увидев *IV* на распечатке.

Алгоритмы и краткие замечания по КМНК и ДМНК:

Косвенный МНК:

1) Структурная => Приведенная.
2) Коэффициенты по МНК.
3) Преобразовать их в коэффициенты структурной модели.

Нарушение предпосылки независимости факторов приводит к несостоятельности оценок структурных коэффициентов, они могут оказаться бессмысленными.

Двухшаговый МНК

Применяется для сверх идентифицируемых систем уравнений.

1) Структурная => Приведенная.
2) Коэффициенты по МНК.
3) Получить оцененные значения эндогенных переменных.
4) Подставить их в правые части структурной формы.
5) Применить МНК к структурной форме сверх идентифицируемых уравнений. Сверхидентифицируемую модель можно превратить в идентифицируемую путем добавления некоторых переменных или отбрасывания некоторых ограничений на параметры.

В условиях быстро изменяющейся техногенной, политической, экологической, экономической и т.п. обстановки на первое место выходит системный подход к изучению окружающего мира.

Любая реальная система требует решения целого комплекса задач, часто взаимоисключающих друг друга, когда изменение одного факторного признака не может происходить при абсолютной неизменности других. Следовательно, отдельно взятое уравнение множественной регрессии не в состоянии охарактеризовать истинные влияния отдельных признаков на вариацию результирующей переменной. Поэтому важное место занимает проблема описания структуры связей между переменными системы так называемых одновременных уравнений или называемых структурными уравнениями, которые чаще всего рассматриваются на уровне макроэкономических исследований.

Система уравнений в эконометрических исследованиях может быть построена по-разному [7].

1.Система независимых уравнений. В данной системе каждая зависимая переменная (*y)* рассматривается как функция одного и того же набора факторов (*x)*:

$$\begin{cases} y_1 = a_{01} + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m + \varepsilon_1 \\ y_2 = a_{02} + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m + \varepsilon_2 \\ \quad \cdots \\ y_n = a_{0n} + a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m + \varepsilon_m \end{cases} \qquad (3.1)$$

Набор факторов x_j в каждом уравнении может варьироваться. Отсутствие того или иного фактора в уравнении системы может быть следствием как экономической целесообразности его включения в модель, так и несущественности его влияния на результативный признак. Примером такой модели может служить модель экономической эффективности сельскохозяйственного производства, где в качестве зависимых переменных выступает показатели валовой и реализованной продукции, прибыль на 1 га сельскохозяйственных угодий, среднегодового работника, 1 млн.сум основных фондов и т.д., а в качестве факторов – основные фонды на 1 га сельхозугодий, среднегодового работника, обеспеченность трудовыми ресурсами, специализация и концентрация производства и др.

2.*Системы рекурсивных уравнений.* В такой системе зависимая переменная (*y)* одного уравнения выступает в виде фактора (*x)* в другом уравнении:

$$\begin{cases} y_1 = a_{01} + a_{11}\mathrm{x}_1 + a_{12}\mathrm{x}_2 + \cdots + a_{1m}\mathrm{x}_\mathrm{m} + \varepsilon_1 \\ y_2 = a_{02} + a_{21}\mathrm{y}_1 + a_{21}\mathrm{x}_1 + a_{22}\mathrm{x}_2 \ldots + a_{2m}\mathrm{x}_\mathrm{m} + \varepsilon_2 \\ y_3 = a_{03} + b_{31}\mathrm{y}_1 + b_{32}\mathrm{y}_2 + a_{31}\mathrm{x}_1 + a_{32}\mathrm{x}_2 + \cdots + a_{3m}\mathrm{x}_\mathrm{m} + \varepsilon_3 \\ \cdots \\ y_n = a_{0n} + b_{\mathrm{n}1}\mathrm{y}_1 + b_{\mathrm{n}2}\mathrm{y}_2 + \cdots b_{\mathrm{n(n-1)}}\mathrm{y}_{\mathrm{n}-1} + a_{\mathrm{n}m}\mathrm{x}_\mathrm{m} + \varepsilon_n \end{cases} \quad (3.2)$$

Примером такой системы может служить модель производительности труда и фондоотдачи вида:

$$\begin{cases} y_1 = a_{01} + a_{11}\mathrm{x}_1 + a_{12}\mathrm{x}_2 + a_{13}\mathrm{x}_3 + \varepsilon_1 \\ y_2 = a_{02} + a_{21}\mathrm{x}_1 + a_{22}\mathrm{x}_2 + a_{23}\mathrm{x}_3 + \varepsilon_2 \end{cases} \quad (3.3)$$

где y_1 производительность труда;

y_2 – фондоотдача;

x_1 – фондовооруженность труда;

x_2 – энерговооруженность труда;

x_3 – квалификация рабочих.

В этих двух видах систем уравнений каждое уравнение может рассматриваться самостоятельно и его параметры определяются методом наименьших квадратов.

1. *Система взаимозависимых уравнений,* получившая наибольшее распространения в эконометрических исследованиях. В ней одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других уравнениях – в правую часть системы:

$$\begin{cases} y_1 = a_{01} + b_{12}\mathrm{y}_1 + b_{13}\mathrm{y}_3 + \cdots \mathrm{b}_{1n}y_n + a_{11}\mathrm{x}_1 + a_{12}\mathrm{x}_2 + \cdots + a_{1m}\mathrm{x}_\mathrm{m} + \varepsilon_1 \\ y_2 = a_{02} + b_{21}\mathrm{y}_1 + b_{23}\mathrm{y}_3 + \cdots \mathrm{b}_{2n}y_n + a_{21}\mathrm{x}_1 + a_{22}\mathrm{x}_2 + \cdots + a_{2m}\mathrm{x}_\mathrm{m} + \varepsilon_2 \\ \cdots \\ y_n = a_{0n} + b_{\mathrm{n}1}\mathrm{y}_1 + b_{\mathrm{n}2}\mathrm{y}_2 + \cdots \mathrm{b}_{n(n-1)}y_{n-1} + a_{\mathrm{n}1}\mathrm{x}_1 + a_{\mathrm{n}2}\mathrm{x}_2 + \cdots + a_{\mathrm{n}m}\mathrm{x}_\mathrm{m} + \varepsilon_n \end{cases} \quad (3.4)$$

Систему взаимозависимых уравнений называют системой совместных, одновременных уравнений. В эконометрике эта система называется структурной формой модели. В системе одновременных уравнений одни и те же переменные (y) одновременно рассматриваются как зависимые в одних уравнениях, и независимые в других. Уранение не может рассматриваться самостоятельно и МНК для нахождения параметров неприменим. Для этого используется специальные приемы оценивания.

Примером системы одновременных уравнений может служить модель динамиками цены и заработной платы:

$$\begin{cases} y_1 = a_{01} + b_{12}y_2 + a_{11}x_1 + \varepsilon_1 \\ y_2 = a_{02} + b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2 \end{cases} \quad (3.5)$$

где y_1 – темп изменения месячной заработной платы;

y_2 – темп изменения цен;

x_1 – процент безработных;

x_2 – темп изменения постоянного капитала;

x_3 – темп изменения цен на импортное сырье.

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

Эндогенные переменные *(y)* – это зависимые переменные, число которых равно числу уравнений в системе.

Экзогенные переменные *(x)* – это предопределенные переменные, влияющие на эндогенные переменные, но не зависящие от них.

Классификация переменных на эндогенные и экзогенные зависит от теоретической концепции модели. Экономические переменные могут выступать в одних моделях как эндогенные, а в других как экзогенные.

Внеэкономические переменные (например, климатические условия) входят в систему как экзогенные переменные.

Структурная форма модели позволяет увидеть влияние изменения любой экзогенной переменной на значение эндогенной переменной. Целесообразно в качестве экзогенных переменных выбирать такие переменные, которые могут быть объектом регулирования, управляя которыми можно получить целевые значения эндогенных переменных.

Коэффициенты эндогенной переменной и a_{ij} – при экзогенной переменной называются структурными коэффициентами модели. Для определения структурных коэффициентов модели структурная форма модели преобразуется в приведенную форму модели.

Приведенная форма модели – система линейных функций эндогенных переменных от экзогенных:

$$\begin{cases} y_1 = A_{01} + b_{11}\mathrm{x}_1 + b_{12}\mathrm{x}_2 + \cdots \mathrm{b}_{1m}x_m + \varepsilon_{\ 1} \\ y_2 = A_{02} + b_{21}\mathrm{y}_1 + b_{22}x_2 + \cdots \mathrm{b}_{2m}x_m + \varepsilon_{\ 2} \\ \cdots \\ y_n = A_{0n} + b_{\mathrm{n}1}\mathrm{x}_1 + b_{\mathrm{n}2}\mathrm{x}_2 + \cdots \mathrm{b}_{nm}x_m + \varepsilon_{\ n} \end{cases} \qquad (3.6)$$

где b_{ij} коэффициенты приведенной формы модели.

Коэффициенты структурной и приведенной модели связаны нелинейными соотношениями. Покажем на примере простейшей структурной модели:

$$\begin{cases} y_1 = a_{01} + b_{12}\mathrm{y}_2 + a_{11}\mathrm{x}_1 + \varepsilon_{\ 1} \\ y_2 = a_{02} + b_{21}\mathrm{y}_1 + a_{22}\mathrm{x}_2 + a_{23}\mathrm{x}_3 + \varepsilon_{\ 2} \end{cases}$$

для которой приведенная форма модели имеет вид:

$$\begin{cases} y_1 = A_{01} + b_{11}\mathrm{x}_1 + b_{12}\mathrm{x}_2 + \varepsilon_{\ 1} \\ y_2 = A_{02} + b_{21}\mathrm{x}_1 + b_{22}\mathrm{x}_2 + \varepsilon_{\ 2} \end{cases}$$

где

$$\mathrm{A}_{01} = \frac{a_{01} + a_{02}b_{12}}{1 - b_{21}b_{12}}, \quad \mathrm{b}_{11} = \frac{a_{11}}{1 - b_{21}b_{12}}, \quad \mathrm{b}_{12} = \frac{a_{22}b_{12}}{1 - b_{21}b_{12}}, \quad \mathcal{E}_1 = \frac{\varepsilon_1 + \varepsilon_2 b_{12}}{1 - b_{21}b_{12}},$$

$$\mathrm{A}_{01} = \frac{a_{02} + a_{01}b_{21}}{1 - b_{21}b_{22}} \quad \mathrm{b}_{21} = \frac{a_{11}b_{21}}{1 - b_{21}b_{12}} \quad \mathrm{b}_{22} = \frac{a_{22}}{1 - b_{21}b_{12}}, \quad \mathcal{E}_2 = \frac{2 + \varepsilon_1 b_{21}}{1 - b_{21}b_{12}}$$

По своему виду приведенная форма модели не отличается от системы независимых уравнений, параметры которой оцениваются традиционным МНК. Применяя МНК, можно оценить, а затем оценить значения эндогенных переменных через экзогенные. Приведенная форма модели аналитически уступает структурной форме модели, т.к. в ней отсутствуют оценки взаимодействия эндогенных переменных.

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

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

С точки зрения идентифицируемости структурные модели подразделяют на три вида:

- идентифицируемые
- неидентифицируемые

- сверхидентифицируемые.

Модель идентифицируема, если все структурные коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, т.е. число параметров структурной модели равно числу параметров приведенной формы. Тогда структурные коэффициенты модели оцениваются через параметры приведенной формы.

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

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

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых требуется проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Чтобы уравнение было идентифицируемо, необходимо, чтобы число предопределенных (экзогенных) переменных (х), отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных (у) без одного в данном уравнении.

Обозначим число эндогенных переменных в j –ом уравнении системы через H, а число экзогенных (предопределенных) переменных, содержащихся в системе, но не входящих в данное уравнение, - через D, то условие идентифицируемости модели можно записать в виде счетного правила:

$D + 1 = H$ – уравнение идентифицируемо.

$D + 1 < H$ – уравнение неидентифицируемо.

$D + 1 > H$ - уравнение сверх идентифицируемо.

Данное счетное правило есть необходимое, но недостаточное условие идентификации. Более точно условия идентификации определяются, если накладывать ограничения на коэффициенты матриц параметров структурной модели. Уравнение идентифицируемо, если из коэффициентов отсутствующих в нем переменным (эндогенным и экзогенным), можно в других уравнениях системы получить матрицу, определитель которой не равен нулю, а ранг матрицы не меньше числа эндогенных переменных в системе без одного. Целесообразность этой проверки в том, что возможна ситуация, когда для каждого уравнения системы счетное правило выполняется, а определитель матрицы равен нулю. В этом случае соблюдается лишь необходимое, но недостаточное условие идентификации.

Пример 3.1. Оценить следующую структурную модель на идентификацию:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2 \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 \\ y_3 = b_{31}y_1 + a_{32}x_2 + a_{31}x_1 + a_{34}x_4 \end{cases}$$

Решение. Модель имеет три эндогенные переменные (y_1, y_2, y_3) и четыре экзогенные переменные (x_1, x_2, x_3, x_4).

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

Для первого уравнения Н=3(y_1, y_2, y_3) и D=x_3 и x_4 отсутствуют), т.е. необходимое равенство выполнено: *D+1=2+1=3=Н.* Для проверки достаточного условия составим таблицу:

Уравнения	Переменные	
	x_3	x_4
2	a_{23}	a_{24}
3	0	a_{34}

Det A = $a_{23}*a_{34} - a_{24}*0 \neq 0$

Определитель матрицу не равен нулю, ранг матрицы равен 2, следовательно, достаточное условие идентификации выполнено, и первое уравнение точно идентифицируемо.

Для второго уравнения Н=2(y_1, y_2) и D=1 (x_1 отсутствуют), необходимое равенство выполнено: *D+1=1+2=Н.* Для проверки достаточного условия составим таблицу:

Уравнения	Переменные	
	y_3	x_1
1	b_{13}	a_{11}
3	-1	a_{31}

Det A = $b_{13}*a_{31} - a_{11}*(-1) \neq 0$

Определитель матрицы не равен нулю, а ранг равен 2, следовательно, достаточное условие идентификации выполнено, и второе уравнение точно идентифицируемо.

Для третьего уравнения Н=3(y_1, y_2, y_3) и D=2 (x_1 и x_2 отсутствуют), необходимое равенство выполнено: *D+1=2+1=3=H.* Для проверки достаточного условия составим таблицу:

Уравнения	Переменные	
	x_2	x_3
1	a_{12}	0
3	a_{22}	a_{23}

$$Det\ A = a_{12}*a_{23} - 0*\ a_{22} \neq 0$$

Определитель матрицы не равен нулю, а ранг равен 2, следовательно, достаточное условие идентификации выполнено, третье уравнение точно идентифицируемо.

Следовательно, данная система точно идентифицируема и может быть решена косвенным методом.

Коэффициенты структурной модели, в зависимости от вида системы, могут оцениваться разными методами. Наибольшее распространение получили:

- косвенный метод наименьших квадратов *(КМНК);*

- двухшаговый метод наименьших квадратов *(ДМНК);*

- трехшаговый метод наименьших квадратов *(ТМНК);*

- метод максимального правдоподобия с полной информацией *(ММП$_f$);*

- метод максимального правдоподобия с ограниченной информацией *(ММП$_s$).*

При использовании *ММП$_f$* для нахождения статистических оценок неизвестных параметров распределения выбирают те, при которых данные результаты наблюдений «наиболее вероятны». При нормальном распределении признаков результаты *ММП$_f$* совпадают с *МНК.*

Если число уравнений системы достаточно велико, что приводит к сложным вычислительным процедурам, то применяется *ММП$_s$*. В данном методе с параметров, связанных с функционированием системы в целом, снимаются ограничения, что приводит к упрощению решения задачи, но трудоемкость процесса остается высокой.

Более просты в использовании *КМНК* и *ДМНК*.

КМНК применяется для идентифицируемой системы одновременных уравнений. Данный метод предполагает выполнение следующих этапов:

- структурная модель преобразовывается в приведенную форму;
- для каждого уравнения приведенной формы модели обычным *МНК* оценивают приведенные коэффициенты;
- полученные коэффициенты преобразуется в параметры структурной модели.

Если изучаемая система сверхидентифицируема, то применяется *ДМНК*, получивший свое название из-за применения *МНК* дважды. На первом шаге он применяется при определении приведенной формы модели и нахождении на ее основе оценок теоретических значений эндогенной переменной. На втором шаге *МНК* принимается к структурному сверхидентифицируемому уравнению при определении структурных коэффициентов модели по полученным на первом шаге значениям эндогенных переменных.

Если все уравнения системы сверхидентифицируемые, то *ДМНК* используется для оценки структурных коэффициентов каждого уравнения.

Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

ТМНК является дальнейшим развитием *ДМНК*.

Пример 3.2[4]. Рассмотрим определение коэффициентов структурной формы модифицированной модели Кейнса двух шаговым методом наименьших квадратов. Модифицированная модель Кейнса имеет вид:

$C_t=a_t+b_{11}*Y_t+\varepsilon_{1t};$

$I_t=a_2+b_{21}*Y_t+b_{22}*Y_{t-1}+\varepsilon_{2t};$

$Yt=C_t+I_t+G_t$

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

Таблица 3.1 – Данные наблюдений для макроэкономической модели *Кейнса*

Yar	C_t	I_t	*Yt*	Y_{t-1}	G_t
1	1016,6	267,0	1412,7	-	486,1
2	1435,9	376,0	1978,9	1412,7	652,7
3	1776,1	408,8	2292,0	1978,9	839,0
4	2003,8	407,1	2514,4	2292,0	842,1
5	3265,7	670,4	4632,0	2514,4	1258,0
6	4476,9	1165,2	7116,6	4632,0	1960,1
7	5886,9	1504,7	8819,9	7116,6	2419,9
8	7443,2	1762,4	10627,5	8819,9	3422,3
9	9024,8	2186,4	12886,1	10627,5	3964,9
10	11401,4	2865,0	16679,9	12886,1	4669,7
11	14363,5	3611,1	21079,5	16679,9	6820,6
12	17742,6	4585,5	26009,7	21079,5	8375,2

Вывод итогов	
Регрессионная статистика	
Множественный R	0,997456534
R-квадрат	0,994321153
Нормированный R-квадрат	0,992901441
Стандартная ошибка	678,6225682
Наблюдения	11
Дисперсионный анализ	

	df	*SS*	*MS*	*F*	*Значимость F*	
Регрессия	2	645079266,5	3,23E+08	700,3683	1,04E-09	
Остаток	8	3684228,72	460528,6			
Итого	10	648763495,2				
	Коэффициенты	***Стандартная ошибка***	***t-статистика***	***P-значние***	***Нижние 95%***	***Верхние***
Y-пересcечение	412,5083336	336,7301283	1,225041	0,255402	-363,993	1189,009
Y_{t-1}	0,817044334	0,366259434	2,23078	0,056228	-0,02755	1,66164
$G_{t.}$	1,037152085	0,935359119	1,108828	0,299723	-1,11979	3,194094

В исходной модели имеются три эндогенные переменные ***Yt, Ct, It*** и две предопределенные переменные ***Yt-1***. и G_t. Первое уравнение сверх идентифицируемо, так как *H=2, D=2 и D+1>H* второе уравнение идентифицируемо, так как *H=2, D=1 и D+1>H[]*.

Стандартная схема двух шагового МНК состоит из следующих этапов:

1. Исходная система уравнений представляется в приведенной форме, где эндогенные переменные выражаются через предопределенные, параметры которых находятся с помощью МНК.

2. По полученным уравнениям приведенной формы находят расчетные значения инструментальных переменных.

3. С помощью обычного МНК, определяются параметры исходных структурных уравнений (отдельно), где в качестве факторов выступают расчетные значения инструментальных переменных и фактические значения предопределенных переменных.

Опорные слова

Система взаимосвязанных уравнений, эндогенные переменные, экзогенные переменные, модель спроса и предложения, паутина, приведенный вид, матричный, идентифицируемые, неидентифицируемые, сверх идентифицируемые, двух шаговый метод наименьших квадратов, система независимых уравнений.

Контрольные вопросы

1. Какие экономические модели требуют систем взаимосвязанных уравнений?
1. Системы эконометрических уравнений: Модель спроса и предложения.
2. Компактная (матричная) запись структурной и приведенной формы динамической модели из одновременных линейных уравнений
3. Что такое идентифицируемость системы одновременных уравнений.
4. Условие идентифицируемости системы одновременных уравнений.
5. Методы решения систем эконометрических уравнений.
6. Объясните модель спроса и предложения.
7. Приведенный вид модели спроса и предложения?
8. Что такое идентифицируемость системы?
9. Какие виды структурные модели имеются?
10. Какие модели называются сверхидентифицируемое?
11. Какими методами решаются системы эконометрических уравнений?
12. Объясните сущность метода наименьших квадратов.
13. Какими методами решаются сверх идентифицируемых уравнения?

ГЛАВА 7. ИССЛЕДОВАНИЕ ВРЕМЕННЫХ РЯДОВ

Эконометрическую модель можно построить, используя три типа исходных данных [2]:

- данные, характеризующие совокупность различных объектов в определенный момент (период) времени: *cross sectional data*, “пространственные”;
- данные, характеризующие один объект за ряд последовательных моментов (периодов) времени: *временные ряды*, *time series;*
- данные, характеризующие совокупность различных объектов за ряд последовательных моментов времени**:** *panel data*, “панельные”.

Временной ряд – это совокупность значений какого-либо показателя за несколько последовательных моментов (периодов) времени. Он формируется под воздействием большого числа факторов, которые можно условно подразделить на три группы:

- факторы, формирующие тенденцию (*тренд*) ряда;

- факторы, формирующие *циклические* колебания ряда, например сезонный, недельный; для рядов цен на фондовом рынке характерны *непериодические колебания;*
- *случайные* факторы.

В большинстве случаев значения временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент.

В Таблице 7.1 и на рисунке 7.1 приведены уровни розничной торговли в млрд. у.е., в 2002-03г.г. Наиболее простой и достаточно точный способ прогноза – использование *автозаполнения* ячеек Excel. Для этого надо выделить оба столбца данных, поставить курсор на черный квадратик в правом нижнем углу выделенной зоны, чтобы курсор превратился в черный тонкий крест, нажать левую клавишу мыши и протащить курсор на требуемое количество ячеек вправо.

Таблица 7.1

			Прогноз протяжкой		
Месяц	2002	2003	2004	2005	2006
Январь	270,1	324	378,3	432,4	486,5
Февраль	267,1	322	376,9	431,8	486,7
Март	288,2	351	413,8	476,6	539,4
Апрель	292,7	353,8	414,9	476	537,1
Май	291	352,6	414,2	475,8	537,4
Июнь	297,8	356,5	415,2	473,9	532,6
Июль	310,1	368	425,9	483,8	541,7
Август	324,3	379	433,7	488,4	543,1
Сентябрь	326,1	386	445,9	505,8	565,7
Октябрь	339,4	403,4	467,4	531,4	595,4
Ноябрь	346,1	409,7	473,3	536,9	600,5
Декабрь	400,7	477,3	553,9	630,5	707,1

2002 - 2003 год

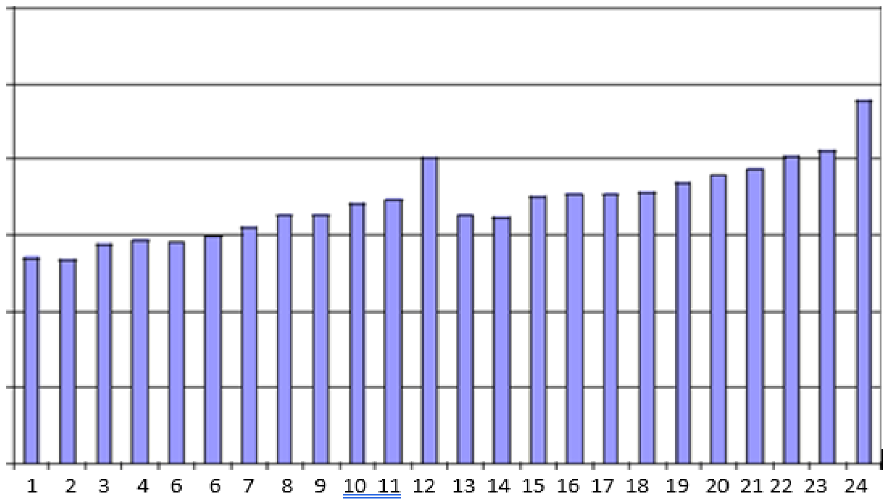


Рис.7.1

Обычно применяется более сложная технология: построение аналитической функции для моделирования тенденции (тренда) временного ряда, или аналитическое выравнивание временного ряда. Для этого чаще всего применяют следующие функции:

- Линейная $Y(t) = a + b *t$;
- Гипербола $Y(t) = a + b / t$;
- Экспонента $Y(t) = exp (a + b*t)$;
- Степенная функция $Y(t) = a * t^{b}$;
- Парабола $Y(t) = a + b_1*t + b_2*t^{2}$

Гиперболу можно линеаризовать заменой $z=1/t$, экспоненту и степенную функцию – логарифмированием, в параболе t и t^2 рассматривать как отдельные переменные множественной регрессии. Тогда параметры трендов можно оценивать обычными средствами МНК: функция ЛИНЕЙН, сервис *Регрессия*.

В качестве независимой переменной выступает время t = 1, 2, ..., n, а в качестве зависимой переменной – уровни (значения) временного ряда Y(t).

Критерием отбора наилучшей формы тренда является наибольшее значение коэффициента детерминации R^2 = 1 – ДИСП ост. / ДИСП Y и соответствующей статистики Фишера [2].

7.1. Временной ряд с сезонными колебаниями

Далее рассмотрена технология расчетов с использованием метода отклонений от тренда, предполагающего вычисление трендовых значений и расчет отклонений от трендов. Для дальнейшего анализа используют не исходные данные, а отклонения от тренда [2].

В таблице 7.2 и на графике 7.2 представлены месячные значения реального располагаемого денежного душевого дохода в некоторой гипотетической стране в 2001 – 2003г.г. в процентах к декабрю 2000 г. Требуется дать прогноз по месяцам 2004 – 2005 г.г. Для этого надо расположить значения по годам в одну строку (или столбец), построить по ним график с линией тренда (щелкнуть правой клавишей мыши по точке графика, *Добавить линию тренда*, *Параметры*,

Показывать уравнение на диаграмме). Ввести строку со сплошной нумерацией месяцев: t = 1, 2, 3, ..., 60 и по коэффициентам a и b уравнения на диаграмме построить линейный тренд $\hat{Y} = a + b*t$. Для 2001 – 2003 г.г. вычислить относительные отклонения $\eta = (Y - \hat{Y}) / \hat{Y}$. Чтобы получить средние η по месяцам 2001 – 2003г.г., надо скопировать эти значения за 2002 год и вставить их в строку под значениями η 2001 года, используя *Специальная вставка – Значения*, затем так же скопировать значения η 2003 года и вставить их в строку под значениями η 2002 года, чтобы получились три строки η за три года.

Вычислите средние значения η по месяцам, используя функцию СРЗНАЧ. Скопируйте полученную строку под тренд 2004 и под тренд 2005 г.г., используя *Специальная вставка – Значения*. Прогнозные значения $\hat{Y}$ по месяцам 2004 – 2005г.г. получим по формуле $\hat{Y}$*прогноз* = $\hat{Y}$ *(1 + η). Технология решения задачи несложная, но требует внимания.

Возможны другие технологии прогнозирования с использованием трендов. В приведенном примере линия тренда проводится по трем суммарным значениям *Y* за 2001, 2002 и 2003 годы. Затем вычислены вклады каждого месяца в годовую сумму (в процентах) и соответствующие средние значения за 3 года.

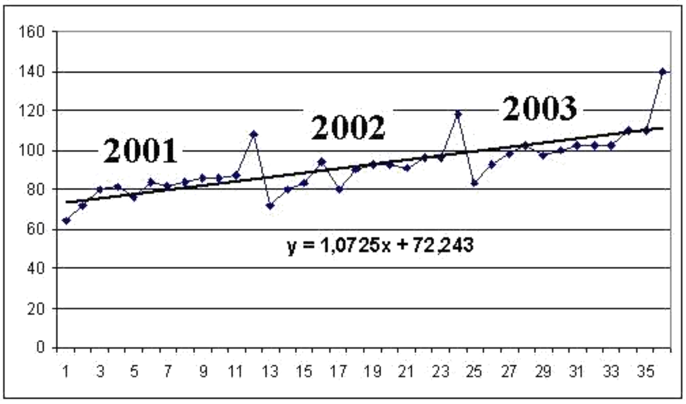


Рис.7.2

Эти же значения можно получить по-другому: просуммировать значения по месяцам за 2001 – 2003 годы и разделить их на сумму за 3 года. Результаты получаются примерно одинаковые. Прогнозные значения по месяцам 2004 – 2005г.г. получаем, умножая полученные средние η по месяцам на полученные по трендам суммарные значения за 2004 и 2005 годы. При решении задачи автозаполнением результаты получаются примерно те же.

Во временных рядах часто последующие значения зависят от предыдущих, т.е. имеет место *автокорреляция в остатках*. В разделе 3.4 была рассмотрена процедура расчета коэффициента автокорреляции как корреляции рядов значений с номерами 1, 2, ..., n - 1 и 2, 3, ... , n. Коэффициенты автокорреляции более высоких уровней m вычисляют как коэффициенты корреляции рядов значений с номерами 1, 2, ..., n - m и m, m + 1, m + 2, ... , n. Возможны другие технологии прогнозирования с использованием трендов. В приведенном примере линия тренда проводится по трем суммарным значениям *Y* за 2001, 2002 и 2003 годы. Затем вычислены вклады каждого месяца в годовую сумму (в процентах) и соответствующие средние значения за 3 года. Последовательность коэффициентов автокорреляции уровней первого, второго и т. д. порядков называют *автокорреляционной функцией временного ряда*. График зависимости ее значений от величины *лага* (порядка коэффициента автокорреляции) называется *коррелограммой*.

Анализ автокорреляционной функции и коррелограммы позволяет определить *лаг* (временной интервал), при котором автокорреляция наиболее высокая, следовательно, лаг, при котором связь между текущим и предыдущими уровнями ряда наиболее тесная, т. е. при помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда. Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию.

Если наиболее высоким оказался коэффициент автокорреляции порядка *m,* ряд содержит циклические колебания периодичностью в *m* моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать предположение относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний, либо ряд содержит сильно нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

Технология построения коррелограммы такова. Если мы имеем *n* уровней ряда (данных) и хотим построить коррелограмму до уровня *m*, то надо вызвать функцию КОРРЕЛ, в верхнем окне указать диапазон ячеек 1: *n-m-1* и “задолларить” его, нажав горячую клавишу F4. В нижнем окне указать диапазон 2: *n-m*. Скопировать функцию на ***m*** ячеек и построить диаграмму. Данные Таблицы 7.2 надо предварительно скопировать в одну строку. Функция выглядит =КОРРЕЛ (C7: Z7; D7:AA7), данные расположены в ячейках C7: AL7. Автокорреляция 12-го порядка близка к 1, то есть через год все повторяется.

Таблица 7.2.

Год	январь	февр	март	апрель	май	июнь	июль	август	сентяб.	октяб.	нояб	дека		
2000												100	Сумма	Тренд
2001	64	72	80	81	76	84	82	84	86	86	87	108	991	980,5
2002	72	80	83	94	80	90	93	93	91	96	96	118	1088	1105
2003	83	93	98	102	97	100	102	102	102	110	110	140	1242	1230
2004	89,18	99,77	106,5	112,9	103	112	113	114	114	119	119	149		1354
2005	97,38	108,9	116,3	123,3	112,5	122	123	124	124	130	130	162,7		1479
				Суммы по месяцам 2001-2003 г.г.										
	219	245	261	277	253	274	277	279	279	292	293	366	3321	
					% от суммы за год									
	6,458	7,265	8,073	8,174	7,669	8,48	8,27	8,48	8,68	8,68	8,78	10,9		
	6,618	7,353	7,629	8,64	7,353	8,27	8,55	8,55	8,36	8,82	8,82	10,85		
	6,683	7,488	7,89	8,213	7,81	8,05	8,21	8,21	8,21	8,86	8,86	11,2		
					Средние по месяцам									
	6, 586	7,369	7,369	8,342	7,611	8,27	8,34	8,41	8,42	8,79	8,82	11,0		

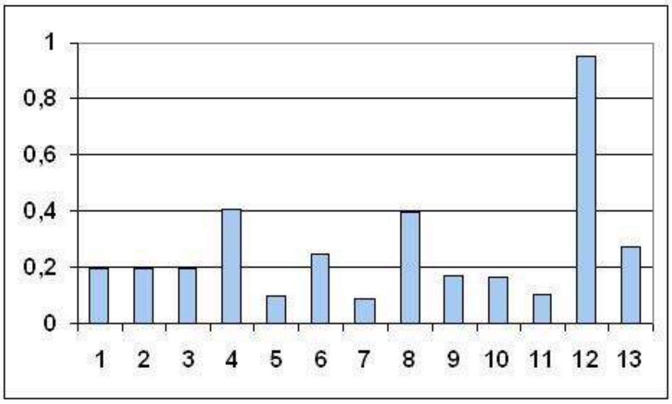


Рис.7.3

Автокорреляционный анализ используется для анализа временных рядов цен биржевых инструментов (акций и т.п.) и валют.

7.2. Тренд и его основные типы. Виды линии тренда

Тренд (от англ. trend – тенденция) – это долговременная тенденция изменения исследуемого временного ряда. Тренды могут быть описаны различными уравнениями – линейными, логарифмическими, степенными и так далее. Фактический тип тренда устанавливают на основе подбора его функциональной модели статистическими методами либо сглаживанием исходного временного ряда.

Тренд в экономике – это направление преимущественного движения показателей. Обычно рассматривается в рамках технического анализа, где подразумевают направленность движения цен или значений индексов. Чарльз Доу отмечал, что при восходящем тренде последующий пик на графике должен быть выше предыдущих, при нисходящем тренде последующие спады на графике должны быть ниже предыдущих.

Различают следующие их виды:

- Повышательный (восходящий, бычий) – рынок растет;
- Понижательный (нисходящий, медвежий) – рынок падает;
- Горизонтальный, боковой – тренд отсутствует – движение наблюдается в горизонтальном диапазоне.

Выделяют тренды восходящий, нисходящий и боковой. На графике часто рисуют линию тренда, которая на восходящем тренде соединяет две или более впадины цены (линия находится под графиком, визуально его поддерживая и подталкивая вверх), а на нисходящем тренде соединяет два или более пика цены (линия находится над графиком, визуально его ограничивая и придавливая вниз). Трендовые линии являются линиями поддержки (для восходящего тренда) и сопротивления (для нисходящего тренда).

Типы тренда

Основной (первичный) – длится 1-3 года.

Вторичный (промежуточный, среднесрочный) – от 3-х недель до 3-6 месяцев.

Незначительный (краткосрочный) – меньше трех недель.

Методы оценки тренда

1. Параметрические

Рассматривают временной ряд как гладкую функцию от t: $X_t = f(t)$, $t = 1…n$;. При этом сначала выявляют один либо несколько допустимых типов функций f(t); затем различными методами (например, МНК) оценивают параметры этих функций, после чего на основе проверки критериев адекватности выбирают окончательную модель тренда. Важное значение для практических приложений имеют линеаризуемые тренды, то есть тренды, приводимые к линейному виду относительно параметров использованием тех или иных алгебраических преобразований.

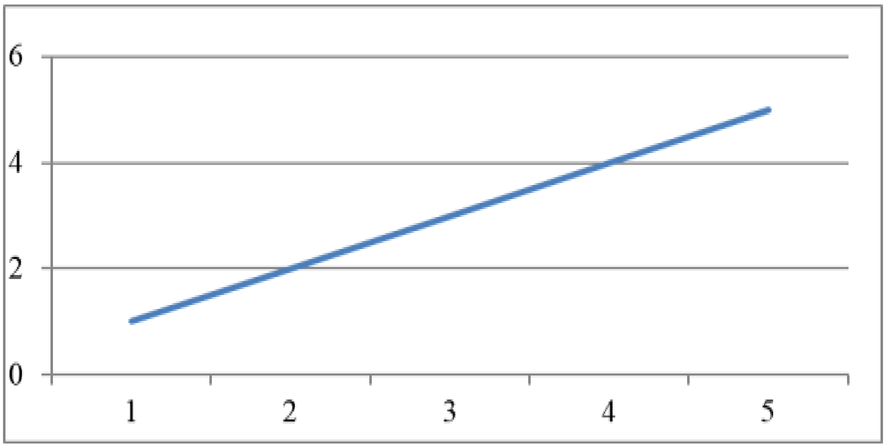
2. Непараметрические

Это разные методы сглаживания исходного временного ряда – скользящие средние (простая, взвешенная), экспоненциальное сглаживание. Эти методы применяются как для оценки тренда, так и для прогнозирования. Они полезны в случае, когда для оценки тренда не удается подобрать подходящую функцию.

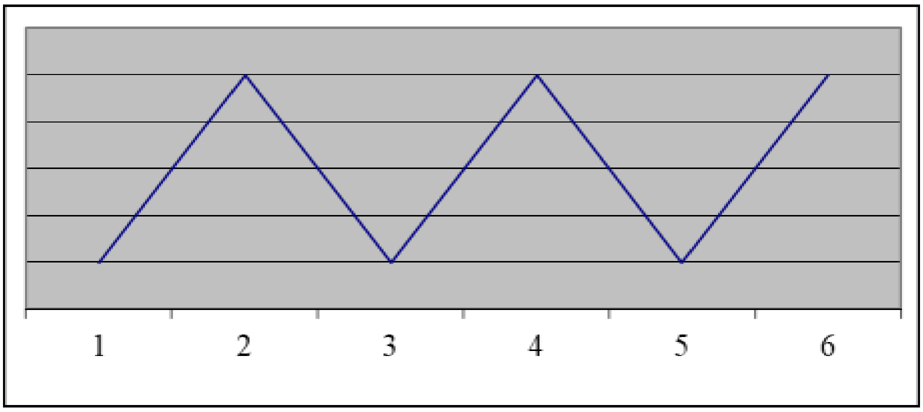
Линии тренда

Трендовые линии широко используются в техническом анализе. На данный момент существует множество методов их построения и интерпретации. Линия тренда – это прямая линия, соединяющая как минимум два пика цен на графике движения курса валюты (актива). Также нужно отметить, что в пределах развития основного тренда, идущего по одной линии, может формироваться множество второстепенных трендов, формирующихся по дополнительным трендовым линиям. Трендовые линии могут пробиваться ценной также как уровни поддержки и сопротивления, показывая этим окончания текущего тренда.

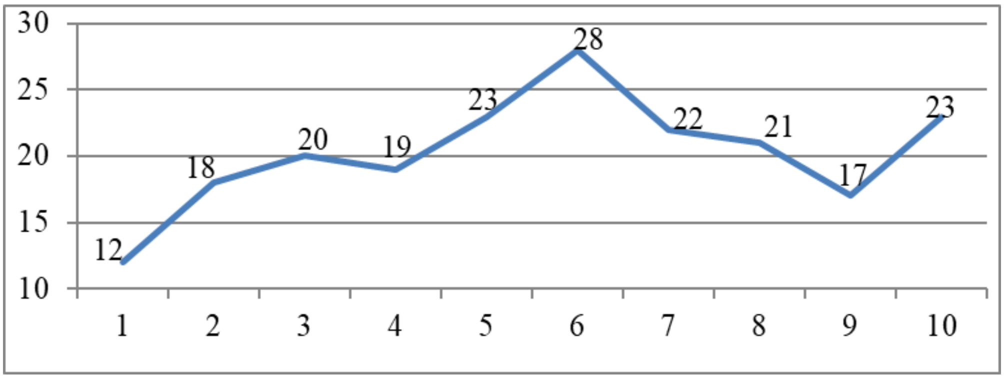
Существует три вида линий тренда:



Тренд (основной тренд динамического ряда)



Циклические (периодические) вариации



Случайные факторы

Линии тренда классифицируются по степени важности при помощи четырех показателей:

Временной масштаб. Чем на более высоком временном масштабе строится линия тренда, тем более важной она является. То есть, линия тренда, построенная на дневном графике показывает более продолжительный и устойчивый тренд, чем линия тренда построенная на часовом графике.

Длительность. Чем длиннее трендовая линия, тем она более надежна. Потому что показывает настроение трейдеров на более длительном промежутке времени.

Число касаний. Чем больше раз цена коснулась линии тренда, тем устойчивей считается этот тренд. Трендовая линия, от которой цена отскочила три и более раз, считается более устойчивой к пробитию, чем линия, которая имеет два отскока.

Угол наклона. Чем больше угол наклона между линией тренда и горизонталью от которой она строится, тем сильнее тренд определяющийся этой линией. Если линия идет под большим углом – это говорит нам о сильном импульсном движении. Если линии формируются полого, значит тренд слабый формирующий как правило коррекционную волну.

Трендовая линия является актуальной до тех пор, пока цена не пробивает ее в противоположную текущему тренду сторону. Показывая тем самым окончание текущего тренда.

7.3. Прогнозирование на основе регрессионных моделей

Под прогнозированием в эконометрике понимается построение оценки зависимой переменной для некоторого набора независимых переменных, которых нет в исходных наблюдениях [9].

Различают *точечное* и *интервальное* прогнозирование. В первом случае оценка – некоторое число, во втором – интервал, в котором находится истинное значение зависимой переменной с заданным уровнем значимости. Рассмотрим регрессионную модель

$$y = \alpha + \beta x + \varepsilon$$

Действительное значение зависимой переменной при $x = x_p$

$$y_p = \alpha + \beta x_p + \varepsilon_p$$

где $M(\varepsilon_p) = 0$, $D(\varepsilon_p) = \sigma^2$ Значения α, β, ε_p неизвестны.

Предсказанным значением является оценка y_p (точечный прогноз):

$$\hat{y}_p = a + bx_p$$

Ошибка предсказания равна разности между предсказанными и действительным значениями:

$$\Delta_p = \hat{y}_p - y_p$$

Ошибка предсказания имеет нулевое математическое ожидание:

$$M(\Delta_p) = 0$$

Действительно,

$$M(\Delta_p) = M(\hat{y}_p) - M(\Delta y_p) = M(a + bx_p) - M(\alpha + \beta x_p + \varepsilon_p) = 0$$

Вычислим дисперсию прогноза. Учитывая, что в случае парной регрессии

$$\hat{y}_p = \bar{y} + b(x_p - \bar{x}), \quad D(\bar{y}) = \frac{\sigma^2}{n}, \quad D(y_p) = D(\varepsilon_p) = \sigma^2,$$

для дисперсии прогноза получим

$$D(\Delta_p) = \left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\,var(x)}\right]\sigma^2.$$

Из формулы следует, что чем больше x_p отклоняется от выборочного среднего $\bar{x}$, тем больше дисперсия ошибки предсказания, и чем больше объем выборки n, тем меньше дисперсия.

Заменяя в дисперсии прогноза σ^2 на ее оценку S^2 и извлекая квадратный корень, получим стандартную ошибку предсказания

$$S_p = S\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\,var(x)}}.$$

Доверительный интервал для действительного значения y_p определяется выражением

$$\hat{y}_p - t_{кр} S_p < y_p < \hat{y}_p + t_{кр} S_p$$

где $t_{кр}$ - критическое значение t –статистики при заданном уровне значимости и числе степеней свободы.

На рис. 7.1 в общем виде показано соотношение между доверительным интервалом предсказания и значением объясняющей переменной. Отрезок, отмеченный на рисунке стрелками, определяет доверительный интервал предсказания в точке x_p.

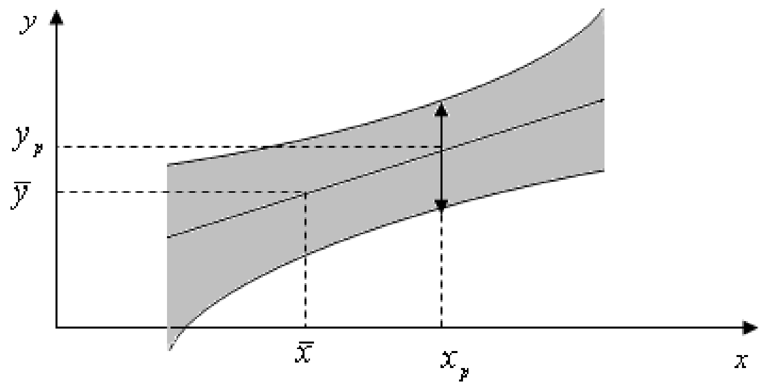


Рис.7.1.

Пример 7.4. По данным о зависимости объема продаж *y* фирмы от затрат на рекламу *x* оценить объем продаж при затратах на рекламу, равных 5,5 усл.ед. Найти стандартную ошибку предсказания и 99%-ный доверительный интервал для полученной оценки.

Исходные данные (усл. ед.):

x	5	8	6	5	3	9	12	4	3	10
y	72	76	78	70	68	80	82	65	62	90

Объем продаж фирмы $y_p = 71{,}87$ при затратах на рекламу $x_p = 5{,}5$ можно определить с помощью статистической функции Excel

y_p = ПРЕДСКАЗ (x_p; массив x; массив y).

Значения $\bar{x} = 6{,}5$, $var(x) = 9{,}61$, $S = 4{,}24$, $t_{кр} = 3{,}35$ можно получить с помощью функций:

$\bar{x}$ = СРЗНАЧ (массив x);

$var(x)$ = ДИСПР (массив x);

S = СТОШУХ (массив y; массив x);

$t_{кр}$ =СТЬЮДРАСПОБР (1- $\alpha; v$);

где $n = 10$, $v = n - 2 = 8$, $\alpha = 1 - 0{,}99 = 0{,}01$.

Вычисляем стандартную ошибку предсказания и доверительный интервал для полученной оценки:

$$S_p = 4{,}24\sqrt{1 + \frac{1}{10} + \frac{(5{,}5 - 6{,}5)^2}{10 \cdot 9{,}61}} = 4{,}47,$$

$$71{,}78 - 3{,}35 \cdot 4{,}47 < y_p < 71{,}87 + 3{,}35 \cdot 4{,}47,$$

или $56{,}9 < y_p < 86{,}84$.

Опорные слова

Момент (период) времени, точечная коррелограмма, автокорреляция в остатках, прогнозирование, непериодические колебания, случайные факторы, циклические колебания ряда, тренд, временные ряды, time series, cross sectional data.

Контрольные вопросы

1. Свойства временных рядов экономических переменных.
2. Прогноз по временному ряду с сезонными колебаниями.
3. Автокорреляция случайного возмущения. Причины. Последствия.
4. Стационарные и нестационарные стохастические процессы.
5. Прогноз, виды прогноза.
6. Как можно определить ошибку предсказания?
7. Как можно определить дисперсию прогноза?

Контрольные задания

1. Для линейной модели, описывающей зависимость переменной *y* от переменных, получена последовательность остатков:

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
e_t	2	0	-2	-3	-4	-1	0	1	-3	-1	1	2	2	0	-1

При уровне значимости *α=0.005*, проверить с помощью критерия Дарбина-Уотсона, гипотезу об автокорреляции между случайными отклонениями, сдвинутыми по времени на одну единицу.

2. Изучается динамика инвестиции в основной капитал и прибыль предыдущего года по предприятию за 10 лет.

1) Постройте уравнения линейного тренда по каждой переменной и дайте интерпретацию их параметров.

2) Определите коэффициент корреляции между временными рядами, используя:

3) Постройте линейное уравнение регрессии по временным рядам $\boldsymbol{y_t}$ и $\boldsymbol{x_t}$. С помощью критерия Дарбина – Уотсона сделайте вывод относительно автокорреляции в остатках в простроенном уравнении.

4) Постройте линейное уравнение регрессии, используя первые разности уровней исходных динамических рядов. Поясните экономический смысл коэффициента регрессии. С помощью критерия Дарбина – Уотсона сделайте вывод относительно автокорреляции в остатках построенном уравнении.

Годы	1	2	3	4	5	6	7	8	9	10
Инвестиции в основной капитал, млн.у.е.,y_t	0,5	0,8	1	1,2	1,6	1,9	2,2	2,4	2,8	3,1
Прибыль предыдушего года, млн. у.е., x_t	2,8	3,2	3,6	3,9	4,4	4,8	5,1	5,4	5,9	6,3

3. Изучается зависимость потребления фруктов и ягод на душу населения за год от среднемесячного душевого дохода населения по следующим данным:

Год	Потребление фруктов и ягод на душу населения, в год, кг	Среднемесячный душевой доход населения, тыс. у.е
1998	30	1,0102
1999	27	1,6589
2000	32	2,2811
2001	35	3,0620
2002	39	3,9472
2003	39	5,1674
2004	43	6,3990
2005	46	8,0883
2006	48	10,1548
2007	51	12,5402
2008	53	14,8636
2009	55	16,8950
2010	58	18,9584
2011	60	20,7800
2012	61	23,2211
2013	64	25,9282
2014	64	27,7666
2015	61	30,4736

1) Постройте уравнения линейного по каждой переменной и дайте интерпретацию их параметров.

2) Определите коэффициент корреляции между временными рядами, используя: а) непосредственно исходные уровни б) отклонения от тренда

3) Постройте линейное уравнение регрессии по временным рядам   и. С помощью критерия Дарбина – Уотсона сделайте вывод относительно автокорреляции в остатках построенном уравнении.

4. Изучается динамика урожайности озимых зерновых культур и цен реализации зерна в сельскохозяйственных организациях за 2000-2015 годы.

Год	Цена реализации зерна (без кукурузы) за 1 ц. у.е.,y_t	Урожайность озимых зерновых культур с 1 га, ц, x_t.
2000	179	38,2
2001	180	42,7
2002	155	45,0
2003	253	39,7
2004	254	43,1
2005	239	46,7
2006	324	42,7
2007	528	45,1
2008	496	55,3
2009	437	45,7
2010	396	49,7
2011	526	55,1
2012	741	39,8
2013	703	50,1
2014	759	54,7
2015	925	57,5

1) Постройте уравнения линейного тренда по каждой переменной и дайте интерпретацию их параметров.

2) Определите коэффициенты корреляции и детерминации по линейным трендам.

3) Постройте уравнения регрессии и оцените тесноту связи между уровнями временных рядов, между первыми разностями. С помощью критерия Дарбина – Уотсона сделайте вывод относительно автокорреляции в остатках в построенных уравнениях.

4) Постройте уравнение множественной регрессии влияния урожайности на уровень цены зерна с включением фактора времени и урожайности предыдущего года. Сравните полученные модели и выберите лучшее из них.

Имеются следующие данные о величине среднедушевого дохода населения и расхода на товар А

Показатель	2010 г.	2011 г.	2012 г.	2013 г.	2014 г.	2015 г.
Расходы на товар А, у.е	33	37	40	43	48	53
Среднедушевой доход населения % к 2010 г.	100,0	109,6	122,5	136,8	146,5	160,7

Требуется:

1) Определить ежегодные абсолютные приросты доходов и расходов и сделать выводы о тенденции развития каждого ряда.

2) Перечислить основные пути устранения тенденции для построения модели спроса на товар А в зависимости от дохода.

3) Построить линейную модель спроса, используя первые разности уровней исходных динамических рядов.

4) Пояснить экономический смысл коэффициента регрессии.

5) Построить линейную модель спроса на товар А, включив в нее фактор времени. Интерпретировать полученные результаты.

Тесты для главы №7

1. Временной ряд является нестационарным, если:

а) среднее значение его членов постоянно

б) его случайная составляющая зависит от времени

в) его члены не зависят от времени

г) его неслучайная составляющая зависит от времени

2. Если регрессионные остатки в эконометрической модели статистически взаимозависимы, то ее называют моделью с:

а) параллельными остатками

б) автокоррелированными остатками

б) гомоскедастичными остатками

в) картезианскими остатками

3. Модель называется аддитивной, если:

а) $Y=T*S*Ɛ$;

б) $Y=T+S+\varepsilon$;

в) $Y=(T*S)+\varepsilon$;

г) $Y=(T+S)*\varepsilon$;

4. Модель называется мультипликативной, если:

а) $Y=T*S*\varepsilon$;

б) $Y=T+S+\varepsilon$;

в) $Y=(T*S)+\varepsilon$;

г) $Y=(T+S)*\varepsilon$;

5. Автокорреляционная функция временного ряда – это:

а) последовательность коэффициентов автокорреляции уровней временного ряда

б) коррелограмма

в) последовательность уровней временного ряда

6. Если независимые переменные имеют ярко выраженный тренд, то они оказываются:

а) некоррелированными

б) малозначимыми

в) тесно коррелированными

г) слабо коррелированными

7. Автокорреляция первого порядка – ситуация, когда коррелируют уровни временного ряда:

а) нечетных

б) последовательных

в) первых и последних

г) четных

8. Какие точки исключаются из временного ряда процедурой сглаживания:

а) стоящие в начале временного ряда

б) стоящие в начале и в конце временного ряда

в) стоящие в конце временного ряда

г) стоящие в середине временного ряда

9. При автокорреляции оценка коэффициентов регрессии становится:

а) смещенной

б) невозможной

в) неэффективной

г) равной 0

10. Если регрессионные остатки в эконометрической модели статистически взаимозависимы, то ее называют моделью с:

а) параллельными остатками

б) автокоррелированными остатками

в) гомоскедастичными остатками

г) картезианскими остатками

11. При отрицательной автокорреляции критерий Дарбина – Уотсона:

а) равен 0

б) меньше 2

с) больше 2

г) больше 1

12. Какой из перечисленных методов не может быть применен для обнаружения автокорреляции?

а) метод рядов

б) критерий Дарбина – Уотсона

в) тест ранговой корреляции Спирмена

г) тест Уайта

13. Критерий Дарбина – Уотсона применяется для:

а) обнаружения автокорреляции в остатках

б) обнаружения циклической составляющей

в) для проверки подчинения случайного компонента нормальному закону распределения

14. При положительной автокорреляции критерий Дарбина – Уотсона:

а) равен 0

б) меньше 2

в) больше 2

г) равен 1

15. Значения статистики Дарбина – Уотсона находятся в промежутке:

а) [-1 и1];

б) [0 и 6];

в) [-2 и 2];

г) [0 и 4];

16. К зоне неопределенности в тесте Дарбина–Уотсона *(DW)* относится случай, при котором _______(d_1 , d_2- нижняя и верхняя границы) а) $DW>d_2$;

б) $DW>d_1$;

в) $d_1<DW<d_2$;

г) $DW=0$;

17. Чем больше число наблюдений, тем __________ зона неопределенности для критерия Дарбина – Уотсона:

а) левее расположена

б) уже

в) шире

г) правее расплоложена

18. Методы исключения тенденции из временного ряда:

а) наименьших квадратов

б) отклонений от трендов

в) главных компонент

г) последовательных разностей

д) включение фактора в модель регрессии

19. Для устранения автокорреляции в остатках можно применить:

а) метод наименьших квадратов

б) обобщенный метод наименьших квадратов

в) метод Алмон

г) метод Койка

20. Если во временном ряду имеется четко выраженная линейная тенденция, то ее можно устранить, переходя от исходных уровней ряда к

а) первым разностям

б) вторым разностям

в) третьим разностям

г) скользящим средним

21. 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

ОСНОВНАЯ ЛИТЕРАТУРА

1. C. Dougherty. Introduction to Econometrics. Oxford, University Press, 2007.

2. Н.В.Катаргин. Прикладная эконометрика. Учебное пособие. 2013

3. В.А. Бывшев. Эконометрика. М.: Финансы и статистика, 2008.

4. Гмурман В.Е. Теория вероятностей и математическая статистика. М., Высшая школа, 1991.

5. Л.О. Бабешко. Основы эконометрического моделирования. – М.: Комкнига, 2001.

6. Эконометрика. Под редакцией И.И.Елисеевой. – М.: Финансы и статистика,2005.

7. Практикум по эконометрике. Под редакцией профессора И.А. Кацко. Кнорус, Москва, 2019.

8. Замков О.О. и др. Математические методы в экономике. М., изд-во ДИС, 1998.

9. Г. Шадманова, С.С. Мирзаев. Экономико-математические методы и модели, ТИИМ,2011.

10. А.И.Новиков: Эконометрика. М. «Дашко и К».2021.

11. А.В.Гладилин и др. Кнорус, Москва, 2009.

12. Виктор Абатуров, ЦЭИР. Экономическое обозрение №12 (250) 2020.

Дополнительная литература

1. Полунин И.Ф. Курс математического программирования. Минск, 1975.

2. Вентцель Е.С. Исследование операций. М., 1972.

3. Кади Дж. Количественные методы в экономике. М,1977.

4. Практикум по эконометрике. Под редакцией И.И.Елисеевой. – М.: Финансы и статистика, 2005.

5. https://www.financialguide.ru/encyclopedia/trend

ОТВЕТЫ

Глава №1

Вопрос	1	2	3	4	5	6	7	8	9	10
Ответ	г	а,б	а	б	в	а	б	а	а	в
Вопрос	11	12	13	14	15	16	17	18	19	20
Ответ	б	а,б	в	б	б	г	б	б	б	а,в
Вопрос	21	22	23	24	25	26	27	28	29	30
Ответ	а,в	а,б	в	б	в	г	б	в	а	г

Глава №2

Вопрос	1	2	3	4	5	6	7	8	9	10
Ответ	а,в	а,б	а,б	а	б	в	а	г	б	б
Вопрос	11	12	13	14	15	16	17	18	19	20
Ответ	г	в	а	в	а	г	а	в	а	в
Вопрос	21	22	23	24	25	26	27	28	29	
Ответ	б	в	в	б	б	в	б	в	г	

Тема 3

Вопрос	1	2	3	4	5	6	7	8	9	10
Ответ	в	а	а	б	б	в	а	а	а	б
Вопрос	11	12	13	14	15	16	17	18	19	20
Ответ	а	г	б	а,б,в	б	г	г	а	а	б
Вопрос	21	22	23	24	25	26	27	28	29	30
Ответ	а	а	а	а	б	а	а	г	г	б

Тема 4

Вопрос	1	2	3	4	5	6
Ответ	б	б	г	в	а	а

Тема 5

Вопрос	1	2	3	4	5	6	7	8	9	10
Ответ	а	а	д	а	б	а	б	в	в	б

Глава №7

Вопрос	1	2	3	4	5	6	7	8	9	10
Ответ	г	б	б	а	а	а	в	б	б	в
Вопрос	11	12	13	14	15	16	17	18	19	20
Ответ	б	г	б	а	б	г	в	б	б,г,д	б
Вопрос	21	22								
Ответ	а	а								

Тема 7

Вопрос	1	2	3	4	5	6	7	8	9
Ответ	а	а	а	а,г	б	а,г	а,б	а	б

ГЛОССАРИЙ

А

Автокорреляция — явление взаимосвязи между рядами: первоначальным и этим же рядом, сдвинутым относительно первоначального положения на h моментов времени.

Авторегрессионная модель — разновидность динамической эконометрической модели, которая содержит в качестве факторных переменных лаговые значения эндогенных переменных.

Авторегрессия — регрессия, учитывающая влияние предыдущих уровней ряда на последующие.

Адаптивных ожиданий модель — разновидность динамической эконометрической модели, в которой учитывается ожидаемое значение факторного признака x_{t+1}.

Аддитивная модель временного ряда — модель, в которой все компоненты ряда динамики представлены как сумма этих составляющих $y_t = u_t + v_t + \varepsilon_t$.Ее применяют в случае, когда амплитуда сезонных колебаний со временем не меняется.

Б

Бета-коэффициент показывает, на какую часть своего среднего квадратического отклонения изменится в среднем значение результативного признака при изменении факторного признака на величину своего среднеквадратического отклонения.

В

Верификация модели — проверка истинности модели, определение соответствия построенной модели реальному экономическому явлению.

Временный лаг — сдвиг, временное смещение уровней временного ряда относительно первоначального положения на h моментов времени.

Временной ряд — ряд последовательно расположенных во времени числовых показателей, которые характеризуют уровень, состояния и изменения явления или процесса.

Временные данные — набор сведений, характеризующий один и тот же объект за разные периоды времени.

Г

Графический метод — способ распознавания типа тренда, при котором временные интервалы откладывают на оси абсцисс, величины уровней — по оси ординат. При этом по каждой оси следует установить такой масштаб, чтобы ширина графика была примерно в 1,5 раза больше его высоты.

Д

Двухшаговый метод наименьших квадратов — один из способов решения систем одновременных уравнений, который применяется как для идентифицируемых, так и для сверхидентифицируемых моделей.

Динамическая эконометрическая модель учитывает в данный момент времени значения входящих в нее переменных, относящихся к текущему и к предыдущему моменту времени.

Долгосрочный мультипликатор — показатель модели авторегрессии, который определяет общее абсолютное изменение результата в долгосрочном периоде.

И

Идентификация модели — проведение статистического анализа модели и оценивания качества ее параметров; установление соответствия между приведенной и структурной формами модели.

Идентифицируемая модель — разновидность структурной модели системы одновременных уравнений, в которой все структурные коэффициенты однозначно определяются через приведенные коэффициенты.

Интервальный ряд динамики — ряд последовательно расположенных показателей за определенный период.

К

Ковариация характеризует сопряженность вариации двух признаков и представляет собой статистическую меру взаимодействия двух случайных переменных.

Коинтеграция — причинно-следственная связь в уровнях двух или более временных рядов, которая выражается в совпадении или противоположной направленности их тенденций и случайной колеблемости.

Корреляционная зависимость это связь, при которой каждому значению независимой переменной *x* соответствует *определенное математическое ожидание (среднее значение)* зависимой переменной *y.*

Корреляционный анализ заключается в количественном определении тесноты связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи).

Корреляция — это статистическая зависимость между случайными величинами, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

Косвенный метод наименьших квадратов — один из способов решения систем одновременных уравнений, основанный на получении состоятельных и несмещенных оценок параметров структурной формы модели по оценкам параметров приведенной формы.

Койка метод — оценивание эконометрических моделей с бесконечным числом лагов.

Коэффициент эластичности показывает, на сколько процентов изменяется результативный признак у при изменении факторного признака х на один процент.

Криволинейная зависимость — это связь, при которой с возрастанием величины факторного признака возрастание (или убывание) результативного признака происходит неравномерно (выражаются уравнениями кривых линий).

Л

Лаги Алмон — один из видов модели с распределенным лагом, который характеризуется полиномиальной структурой и конечной величиной лага.

Лаговые (экзогенные или эндогенные) — это такие переменные модели, которые датируются предыдущими моментами времени и находятся в уравнении с текущими переменными.

М

Метод разности средних двух частей одного и того же ряда **—**один из критериев проверки на наличие тренда, где проверяется гипотеза о существовании разности средних $H_0: \overline{Y_1} = \overline{Y_2}$. Для этого временной ряд разбивают на две равные или почти равные части. В качестве критерия проверки гипотезы принимают критерий Стьюдента. Если $t_{факт} \geq t_{теор}$, то гипотеза об отсутствии тренда отвергается; если $t_{факт} < t_{табл}$ то гипотеза H^0 принимается.

Метод Фостера—Стюарта — критерий проверки на наличие тренда, где определяется наличие тенденции явления и тренд дисперсии уровней временного ряда. Часто этот метод используют в случае детального анализа временного ряда и построения по нему прогнозов. Вычисление критерия проводится поэтапно: проводится сравнительная оценка каждого уровня временного ряда со всеми предыдущими уровнями; вычисляют значения величин ***q*** и ***d;*** определяют критерий Стьюдента и сравнивают его с табличным значением. Величина ***d*** характеризует тенденцию изменения средней и имеет два предела: нижний и верхний. Величина ***q*** характеризует тенденцию изменения дисперсии временного ряда и принимает значения в пределах: $0 \leq q \leq n-1$

Множественная корреляция — это зависимость между результативным признаком и двумя и более факторными признаками, включенными в исследование.

Многофакторная (множественная) зависимость — это связь между несколькими факторными признаками и результативным признаком (факторы действуют комплексно, т.е. одновременно и во взаимосвязи).

Множественная регрессия характеризует связь между результативным признаком и двумя и более факторными признаками.

Моментный ряд динамики — это ряд последовательно расположенных показателей на определенную дату.

Модель временного ряда — разновидность эконометрической модели, в которой результативный признак является функцией переменной времени или переменных, относящихся к другим моментам времени.

Модель регрессионная с одним уравнением — имеет вид Y = M*x*(Y)+ε, где результативный признак является функцией от факторных признаков Y =

$f(X_1, X_2, ..., X_k) + \varepsilon$, а объясненная составляющая $f(X_1, X_2, ..., X_k)$ представляет собой ожидаемое значение результата Y при заданных значениях факторов $X_1, X_2, ..., X_k$.

Мультиколлинеарность — это тесная зависимость между факторными признаками, включенными в модель.

Мультипликативная модель временного ряда — модель, в которой факторы влияния представлены в виде произведения составляющих $t = u_t \cdot v_t \cdot \varepsilon_t$. Такую модель применяют в случае, если происходят существенные сезонные изменения.

Н

Неидентифицируемая модель — разновидность структурной модели системы одновременных уравнений, в которой структурные коэффициенты невозможно найти по приведенным коэффициентам.

Неполной (частичной) корректировки модель — разновидность динамическом эконометрической модели, в которой учитывается ожидаемое значение результативного признака Y_{t+1}

О

Однофакторная (парная) *зависимость* — это связь между одним признаком-фактором и результативным признаком (при абстрагировании влияния других).

П

Параметризация — определение вида экономической модели, выражение в математической форме взаимосвязи между ее переменными, формулирование исходных предпосылок и ограничений модели.

Парная корреляция — это связь между двумя признаками (результативным и факторным или двумя факторными).

Парный коэффициент регрессии показывает, на какую величину в среднем изменится результативный признак *у,* если переменную х увеличить на единицу измерения.

Парная регрессия характеризует связь между двумя признаками: результативным и факторным.

Парный коэффициент детерминации показывает, какая доля вариации переменной *у* учтена в модели и обусловлена влиянием на нее переменной х.

Поведенческие уравнения описывают взаимодействие между экзогенными и эндогенными переменными в структурной форме системы одновременных уравнений.

Предопределенные переменные — лаговые и текущие экзогенные, а также — лаговые эндогенные переменные модели.

Приведенная форма модели — один из способов записи системы одновременных уравнений, в котором каждая эндогенная переменная определена в виде линейной функции от всех предопределенных переменных.

Проверка статистических гипотез о типе тренда — метод распознавания типа тренда, при котором проводится: сглаживание ряда уровней (скользящая средняя); вычисляют цепные абсолютные изменения $\Delta_i = Y_{i+1} - Y_i$ (для параболы — ускорения, для экспоненты — темпы роста); расчет по равным или примерно равным подпериодам средней величины того параметра, постоянство которого подтверждает выдвинутую гипотезу о типе тренда (средний абсолютный прирост — для прямой, среднее ускорение — для параболы, средний темп — для экспоненты); проверяется методом дисперсионного анализа или по t-критерию существенность различия средних значений параметра в разных подпериодах исходного ряда. Если различия средних признаются существенными, гипотеза о данном типе тренда отвергается и выдвигается следующая гипотеза в порядке усложнения: после отклонения прямой линии — об экспоненте; после отклонения экспоненты — о параболе; при отклонении параболы — о других типах линий.

Промежуточный мультипликатор — показатель модели авторегресси, который определяет общее абсолютное изменение результата в момент времени $(t+1)$.

Пространственные данные — набор сведений по разным объектам, взятым за один и тот же период времени.

Прямолинейная зависимость — это связь, при которой с возрастанием величины факторного признака происходит равномерное возрастание (или убывание) величин результативного признака.

Р

Регрессионный анализ заключается в определении аналитической формы связи, в которой изменение результативного признака обусловлено влиянием одного или нескольких факторных признаков, а множество всех прочих факторов, также оказывающих влияние на результативный признак, принимается за постоянные и средние значения.

Регулярная компонента u_t — составляющая временного ряда, которая характеризует общую тенденцию ряда.

Ряд динамики — это ряд последовательно (в хронологическом порядке) расположенных статистических показателей, изменение которых имеет определенную тенденцию развития изучаемого явления. Он содержит лаговую составляющую.

Ряд Фурье — в гармониках Фурье исходным рядом является не первичный ряд за несколько лет, а усредненные значения месячных уровней, в которых исключены тренд и случайная компонента.

С

Сверхидентифицируемая модель — разновидность структурной модели системы одновременных уравнений, в которой структурные коэффициенты, выраженные через приведенные коэффициенты, имеют два и более числовых значений.

Связные временные ряды — временные ряды, показывающие зависимость результативного признака от одного или нескольких факторных.

Сезонная волна — это графическое изображение полученных индексов сезонности.

Сезонная компонента v_t — компонента временного ряда, которая характеризует внутригодичные колебания показателя. В общем виде является циклической составляющей.

Система независимых уравнений — одна из разновидностей систем эконометрических уравнений, в которой каждый результативный признак является функцией одной и той же совокупности факторов; набор факторов в каждом уравнении системы может варьировать в зависимости от изучаемого явления.

Система одновременных уравнений — одна из разновидностей эконометрических моделей, состоящая из тождеств и регрессионных уравнений, в которых наряду с факторными признаками включены результативные признаки из других уравнений системы.

Система рекурсивных уравнений — одна из разновидностей систем эконометрических уравнений, в которой результативный признак одного уравнения системы в каждом последующем уравнении является фактором наряду с одной и той же совокупностью факторов.

Статистическая зависимость это связь, при которой каждому значению независимой переменной x соответствует множество значений зависимой переменной y, причем неизвестно заранее, какое именно значение примет y.

«Смыкание рядов» — это объединение в один более длинный динамический ряд двух (или нескольких) рядов динамики, уровни которых исчислены по различной методологии или по различным границам территорий. Для смыкания необходимым условием является наличие за один период данных, рассчитанных по разной методологии (или в разных драницах).

С распределенным лагом модель содержит наряду с текущими значениями факторных переменных их лаговые значения.

Структурная форма модели — один из способов записи системы одновременных уравнений, который отражает реальный экономический объект или явление и показывает, как изменение любой экзогенной переменной определяет значения эндогенной переменной модели.

Т

Тенденция автокорреляции — вид тенденции временного ряда, который характеризует связь между отдельными уровнями ряда динамики.

Тенденция дисперсии — вид тенденции временного ряда, который характеризует направление изменения отклонений между эмпирическими уровнями и детерминированной компонентой ряда.

Тенденция среднего уровня — вид тенденции временного ряда, который выражается обычно с помощью математического уравнения линии, вокруг которой варьируют фактические уровни исследуемого явления. Уравнение тенденции имеет вид: $Y_t = f_t + \varepsilon_t$. Смысл этой функции заключается в том, что значения тренда в отдельные моменты времени выступают математическими ожиданиями ряда динамики.

Тождество — одна из разновидностей структурных уравнений модели, которая устанавливает соотношение между эндогенными переменными; не содержит случайных составляющих и структурных коэффициентов.

Тренд — это основная достаточно устойчивая тенденция во временном ряду, более или менее свободная от случайных колебаний.

Ф

Функциональная зависимость — это связь, при которой каждому значению независимой переменной *x* соответствует точно *определенное значение* зависимой переменной *y*.

Ч

Частная корреляция — **это** зависимость между результативным и одним факторным признаками или двумя факторными признаками при фиксированном значении других факторных признаков.

Частные показатели временного ряда характеризуют явления изолированно, односторонне.

Э

Экзогенные (независимые) **—** это переменные, значения которых задаются извне модели.

Эконометрика **—** это наука, предметом изучения которой является количественное выражение взаимосвязей экономических явлений и процессов.

Эндогенные (зависимые) — это переменные, значения которых определяются внутри модели.

Приложение 1. Статистические показатели по крупным и средним сельскохозяйственным организациям

№	y_1	y_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	56,033	64,630	43,118	3,67	29,905	262,87	0,694	21,510	46,353
2	70,087	103,323	66,328	5,18	60,231	282,97	0,908	47,888	75,174
3	60,066	74,543	43,336	3,57	45,364	268,46	1,047	39,698	62,966
4	50,731	66,284	53,156	4,33	53,436	266,60	1,005	28,693	49,382
5	50,937	70,356	62,080	0,58	33,750	266,04	0,544	22,472	48,163
6	80,942	97,731	87,999	5,96	89,582	243,28	1,018	52,248	73,420
7	43,558	52,358	50,017	2,38	22,427	250,97	0,448	23,972	42,561
8	53,721	62,452	34,791	4,69	24,288	190,71	0,698	32,115	52,034
9	67,633	76,591	52,539	5,62	84,502	195,21	1,608	35,130	57,703
10	50,926	62,790	45,466	3,91	25,414	244,55	0,559	27,386	45,626
11	59,066	104,561	82,215	8,19	56,305	238,33	0,685	59,436	91,435
12	56,632	72,839	39,308	3,22	38,420	254,36	0,977	39,609	57,118
13	49,937	59,365	38,605	3,24	26,726	252,89	0,692	19,977	42,870
14	66,474	80,969	83,559	7,90	97,844	225,08	1,171	46,624	48,050
15	49,528	88,864	49,097	5,83	38,702	220,27	0,788	56,311	82,291
16	57,854	83,101	72,252	4,68	57,553	327,38	0,797	44,462	69,844
17	67,589	99,770	55,525	6,39	45,270	257,67	0,815	49,289	90,147
18	53,876	59,105	28,178	3,28	44,882	221,92	1,593	22,835	40,464
19	51,234	67,814	23,398	6,13	23,963	187,08	1,024	37,240	64,545
20	40,520	62,802	22,193	4,62	33,969	200,75	1,531	31,885	59,290
21	58,888	99,118	43,444	4,07	59,139	250,76	1,361	40,927	85,268
22	45,951	59,314	23,542	3,80	27,399	212,86	1,164	32,270	47,453
23	48,342	86,096	45,530	5,04	32,590	202,75	0,716	59,685	77,127
24	67,525	69,820	58,578	3,78	51,519	218,23	0,879	33,151	58,127
25	78,398	116,104	86,891	5,63	101,645	318,18	1,170	54,311	87,060
26	54,972	68,291	66,332	3,42	46,651	235,90	0,703	31,825	51,837
27	55,136	98,352	44,453	6,40	41,903	237,89	0,943	40,705	76,760
28	28,458	30,315	29,482	2,62	22,407	152,62	0,760	14,730	26,687
29	34,776	51,060	34,823	2,15	75,212	217,22	2,160	26,791	39,856
30	41,580	37,590	16,676	2,76	35,146	220,74	2,108	18,193	30,323
31	42,218	49,785	22,068	3,57	25,791	230,98	1,169	25,261	41,299
32	44,563	47,49	32,621	4,31	31,249	172,62	0,958	19,543	35,100
33	90,292	110,717	76,894	7,37	76,432	352,67	0,994	36,124	86,433
34	45,595	60,248	50,106	3,73	37,998	338,16	0,758	23,970	44,251
35	35,521	53,672	28,424	4.00	57,414	242,84	1,844	27,622	44,800
36	47,427	87,569	25,539	2,99	56,845	246,63	2,226	26,929	65,200
37	45,161	75,786	25,646	5,96	23,202	167,44	0,905	31,653	55,023
38	29,611	37,416	35,115	2,65	36,959	244,53	1,053	22,709	34,717
39	25,181	32,827	20,724	2,93	36,367	132,80	1,755	19,724	32,772
40	55,921	71,668	60,349	3,93	42,078	223,40	0,697	27,309	46,880
41	65,386	67,146	54,770	2,06	41,921	262,85	0,765	22,626	53,460
42	63,359	82,285	58,643	6,54	59,182	203,74	1,009	40,455	63,309
43	50,285	52,988	52,779	3,41	48,551	259,78	0,920	22,421	37,939

№	y_1	y_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7
44	80,097	84,401	64,855	4,57	70,186	249,15	1,082	34,924	65,826
45	43,410	54,695	29,563	3,25	26,219	267,22	0,887	23,423	37,564
46	73,076	87,552	67,157	4,71	69,679	239,04	1,038	47,448	71,221
47	66,618	83,616	45,478	4,70	95,029	235,71	2,090	26,945	60,394
48	45,646	59,144	52,611	2,11	42,632	301,31	0,810	24,853	37,755
49	53,933	64,442	19,886	1,11	35,262	159,35	1,773	20,658	43,126
50	51,193	57,971	50,218	3,83	36,192	211,45	0,721	27,374	50,238
51	21,955	30,482	23,333	2,65	17,126	128,97	0,731	12,081	25,782
52	59,217	72,521	61,811	4,21	34,476	240,30	0,558	29,870	47,952
53	69,962	83,882	76,108	4,21	63,901	261,25	0,840	34,743	56,494
54	30,843	45,088	42,180	3.00	30,902	214,55	0,733	23,273	37,087
55	58,214	74,259	55,890	3,33	54,918	205,36	0,983	29,348	50,496
56	42,019	65,724	26,644	4,54	27,664	197,71	1,038	41,914	54,328
57	79,249	77,814	55,351	5,64	55,771	244,37	1,008	51,990	66,990
58	61,542	90,323	60,749	3,93	68,638	246,72	1,130	34,487	62,867
59	39,837	48,729	49,494	3,39	30,859	280,39	0,623	23,980	40,585
60	71,845	100,084	73,858	6,40	62,703	262,76	0,849	65,097	71,412
61	59,554	74,426	52,663	3,52	35,508	259,16	0,674	28,675	49,751
62	47,710	68,226	52,048	3,86	51,774	275,05	0,995	25,273	46,567
63	53,267	66,969	48,428	3,90	40,555	248,33	0,837	24,369	47,033
64	38,027	41,698	55,163	2,67	30,626	235,34	0,555	19,150	35,284
65	43,846	50,151	45,771	4,29	29,670	282,85	0,648	24,846	45,099
66	37,037	39,126	35,006	1,63	22,525	219,59	0,643	20,405	28,053

Обозначения

y1 - Выручка от продажи на 1 га сельскохозяйственных угодий, тыс. у.е.;

y2 - Валовая продукция на 1 га сельскохозяйственных угодий, тыс. у.е.;

x1 - среднегодовая стоимость основных средств на 1 га сельскохозяйственных угодий, тыс. у.е.;

x2 - среднегодовая численность работников на 100 га сельскохозяйственных угодий, чел.;

x3 - среднегодовая стоимость оборотных средств;

x4 - годовая оплата труда 1 работника, тыс. у.е.;

x5 - приходится оборотных средств на 1 усл.ден.ед. основных средств;

x6 - материальные затраты на 1 га сельскохозяйственных угодий, тыс. у.е.;

x7 - затраты на 1 га сельскохозяйственных угодий, тыс. у.е.;

Приложение 2 Данные по производству молока на сельскохозяйственных предприятиях

№	y_1	y_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	1383	1383	61,8	0,73	1362	320	569	89,2	35,1
2	1436	1436	73,7	1,08	1686	239	567	81,1	41,8
3	1299	1299	68,3	2,10	800	272	575	79,9	39,3
4	1377	1424	64,7	1,26	2200	124	595	92,0	38,5
5	1850	1850	59,2	1,73	760	186	743	29,7	44,0
6	1486	1489	64,5	1,06	700	356	679	87,1	43,8
7	1516	1516	43,6	2,35	1650	127	591	73,9	25,8
8	1447	1575	61,9	3,10	536	157	584	78,7	36,2
9	1676	1677	61.0	1,50	2600	209	798	85,2	48,7
10	1391	1389	61,9	1,24	520	169	487	91,9	30,1
11	1465	1465	72,4	2,42	1091	225	531	52,4	38,4
12	1688	1690	50,8	1,70	1100	341	863	77,8	43,9
13	1641	1639	62,8	2,24	780	165	772	85,5	48,5
14	1711	1812	68,8	2,02	1400	298	545	78,2	37,5
15	1746	1746	59,3	2,14	1100	210	673	80,3	40,0
16	1537	1653	63,5	2,24	1538	165	609	62,0	38,7
17	1783	1906	56,4	2,19	600	162	721	78,7	40,6
18	1983	2153	53,4	2,84	1650	247	584	70,2	31,1
19	2138	2165	35,0	5,21	800	109	800	40,4	28,0
20	1701	1706	64,6	1,65	1025	111	1003	90,6	64,8
21	1608	1608	53,8	1,43	700	180	691	91,4	37,1
22	2032	2036	54,3	2,76	500	166	699	50,7	38,0
23	1439	1592	80,1	1,38	1450	334	502	83,6	40,2
24	2008	2007	60,7	3,11	800	192	600	64,0	36,4
25	1853	1900	42,2	3,16	420	108	715	71,8	30,2
26	1818	1869	45,4	4,01	450	149	716	79,6	32,5
27	1902	1911	39,3	2,61	713	146	995	73,0	39,1
28	1647	1647	66,5	1,85	650	156	856	89,8	57,0
29	1950	1952	38,6	3,38	644	152	796	84,2	30,8
30	1504	1750	80,0	1,29	2990	449	496	79,9	39,7
31	1597	1683	73,0	2,17	600	243	545	78,8	39,8
32	1582	1582	70,8	1,41	500	210	734	88,7	51,9
33	1522	1524	67,9	1,85	1440	158	762	86,5	51,7
34	1647	1650	63,4	2,63	667	152	803	98,2	50,9
35	1915	1920	53,6	2,15	1148	191	1105	80,8	59,2
36	2049	2082	53,5	2,85	499	126	1123	73,1	60,1
37	1491	1489	71,6	1,90	1900	109	599	83,4	42,9
38	1422	1420	90,5	1,29	521	215	770	62,8	69,7
39	1341	1436	59,5	2,56	1800	132	540	80,4	32,1
40	1347	1347	63,1	1,21	1051	217	696	93,2	44,0
41	1861	1861	65,2	1,51	2350	156	1026	86,1	66,9
42	1521	1525	68,0	1,40	1750	184	644	80,8	43,8
43	1337	1337	62,5	1,01	950	201	704	92,1	44,0

№	y_1	y_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7
44	1622	1657	60,2	1,66	800	269	746	87,7	44,9
45	1472	1531	52,6	1,46	483	109	653	84,8	34,3
46	1695	1829	60,4	1,06	874	215	998	79,0	60,3
47	1255	1301	68,5	1,11	1414	303	550	88,7	37,7
48	1536	1633	67,8	0,86	517	570	802	38,6	54,4
49	2005	2153	50,1	2,99	413	127	866	43,3	43,4
50	1536	1634	66,0	2,22	730	140	667	87,6	44,0
51	1714	1783	79,0	1,05	460	374	852	52,6	67,3
52	1420	1487	81,1	0,70	2522	283	876	88,7	71,1
53	1639	1639	65,9	2,12	1132	176	824	90,7	54,2
54	1852	1855	52,7	4,61	600	128	629	93,3	33,2
55	1608	1610	78,4	2,14	757	154	552	86,4	43,3
56	1374	1374	74,0	1,36	654	156	647	82,9	47,9
57	1404	1405	59,3	1,95	2600	286	525	83,3	31,1
58	1342	1511	71,6	1,88	1000	189	524	89,3	37,5

Обозначение

y1 - производственная, себестоимость 1ц молока, у.е.;

y2 - рольная себестоимость1 ц молока, у.е.;

x1 - годовой надой молока на среднегодовую корову, кг;

x2 - прямые затраты труда на 1 ц, чел.-ч;

x3 - среднегодовое поголовье коров на предприятии, гол.;

x4 - затраты по оплате труда на 1 чел. -ч.; у.е.;

x5 - затраты на корма на 1 ц молока, у.е.;

x6 - доля молока в выручке от реализации продукции животноводства, %;

x7 - затраты на корма на среднегодовую корову, тыс. у.е.

Приложение 3. t-распределение: критические значения t

Число	Тесты	Уровень значимости					
степеней	Двусторонний	10%	5%	2%	1%	0,2%	0,1%
свободы	Односторонний	5%	2,5%	1%	0,5%	0,1%	0,05%
1		6,314	12,706	31,821	63,657	318,31	636,62
2		2,92	4,303	6,965	9,925	22,327	31,598
3		2,353	3,182	4,541	5,841	10,214	12,924
4		2,132	2,776	3,747	4,604	7,173	8,61
5		2,015	2,571	3,365	4,032	5,893	6,869
6		1,943	2,447	3,143	3,707	5,208	5,959
7		1,895	2,365	2,998	3,499	4,785	5,408
8		1,86	2,306	2,896	3,355	4,501	5,041
9		1,833	2,262	2,821	3,25	4,297	4,781
10		1,812	2,228	2,764	3,169	4,144	4,587
11		1,796	2,201	2,718	3,106	4,025	4,437
12		1,782	2,179	2,681	3,055	3,93	4,318
13		1,771	2,16	2,65	3,012	3,852	4,221
14		1,761	2,145	2,624	2,977	3,787	4,14
15		1,753	2,131	2,602	2,947	3,733	4,073
16		1,746	2,12	2,583	2,921	3,686	4,015
17		1,74	2,11	2,567	2,898	3,646	3,965
18		1,734	2,101	2,552	2,878	3,61	3,922
19		1,729	2,093	2,539	2,861	3,579	3,883
20		1,725	2,086	2,528	2,845	3,552	3,85
21		1,721	2,08	2,518	2,831	3,527	3,819
22		1,717	2,074	2,508	2,819	3,505	3,792
23		1,714	2,069	2,5	2,807	3,485	3,767
24		1,711	2,064	2,492	2,797	3,467	3,745
25		1,708	2,06	2,485	2,787	3,45	3,725
26		1,706	2,056	2,479	2,779	3,435	3,707
27		1,703	2,052	2,473	2,771	3,421	3,69
28		1,701	2,048	2,467	2,763	3,408	3,674
29		1,699	2,045	2,462	2,756	3,396	3,659
30		1,697	2,042	2,457	2,75	3,385	3,646
40		1,684	2,021	2,423	2,704	3,307	3,551
60		1,671	2	2,39	2,66	3,232	3,46
120		1,658	1,98	2,358	2,617	3,16	3,373
		1,645	1,96	2,326	2,576	3,09	3,291

Приложение 4. F-распределение: критические значения F с v1 и v2 степенями свободы, уровень значимости в 1%

v/v	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	@
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98,5	99	99,2	99,25	99,3	99,33	99,36	99,37	99,39	99,4	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,5
3	34,1	30,8	29,5	28,71	28,24	27,91	27,67	27,49	27,35	27,2	27,05	26,87	26,69	26,6	26,5	26,41	26,32	26,22	26,13
4	21,2	18	16,7	15,98	15,52	15,21	14,98	14,8	14,66	14,6	14,37	14,2	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	16,3	13,3	12,1	11,39	10,97	10,67	10,46	10,29	10,16	10,1	9,89	9,72	9,55	9,47	9,38	9,29	9,2	9,11	9,02
6	13,8	10,9	9,78	9,15	8,75	8,47	8,26	8,1	7,98	7,87	7,72	7,56	7,4	7,31	7,23	7,14	7,06	6,97	6,88
7	12,3	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,2	5,12	5,03	4,95	4,86
9	10,6	8,02	6,99	6,42	6,06	5,8	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,4	4,31
10	10	7,56	6,55	5,99	5,64	5,39	5,2	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,4	4,25	4,1	4,02	3,94	3,86	3,78	3,69	3,6
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,5	4,39	4,3	4,16	4,01	3,86	3,78	3,7	3,62	3,54	3,45	3,36
13	9,07	6,7	5,74	5,21	4,86	4,62	4,44	4,3	4,19	4,1	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,8	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4	3,89	3,8	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	8,53	6,23	5,29	4,77	4,44	4,2	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,1	3,02	2,93	2,84	2,75
17	8,4	6,11	5,18	4,67	4,34	4,1	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3	2,92	2,83	2,75	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,6	3,51	3,37	3,23	3,08	3	2,92	2,84	2,75	2,66	2,57
19	8,18	5,93	5,01	4,5	4,17	3,94	3,77	3,63	3,52	3,43	3,3	3,15	3	2,92	2,84	2,76	2,67	2,58	2,49
20	8,1	5,85	4,94	4,43	4,1	3,87	3,7	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,4	3,31	3,17	3,03	2,88	2,8	2,72	2,64	2,55	2,46	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,5	2,4	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,3	3,21	3,07	2,93	2,78	2,7	2,62	2,54	2,45	2,35	2,26
24	7,82	5,61	4,72	4,22	3,9	3,67	3,5	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,4	2,31	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,7	2,62	2,54	2,45	2,36	2,27	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,81	2,66	2,58	2,5	2,42	2,33	2,23	2,13
27	7,68	5,49	4,6	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,2	2,1
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,9	2,75	2,6	2,52	2,44	2,35	2,26	2,17	2,06
29	7,6	5,42	4,54	4,04	3,73	3,5	3,33	3,2	3,09	3	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	7,56	5,39	4,51	4,02	3,7	3,47	3,3	3,17	3,07	2,98	2,84	2,7	2,55	2,47	2,39	2,3	2,21	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,8	2,66	2,52	2,37	2,29	2,2	2,11	2,02	1,92	1,8
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,5	2,35	2,2	2,12	2,03	1,94	1,84	1,73	1,6
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38

Приложение 5. F-распределение: критические значения F с v1 и v2 степенями свободы, уровень значимости в 5%

v/v	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	@
1	161	200	216	225	230	234	237	239	240,5	242	244	246	248	249	250	251	252	253,3	254,3
2	18,5	19	19,2	19,3	19,3	19,3	19,4	19,4	19,38	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,49	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,7	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6	5,96	5,91	5,86	5,8	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,5	4,46	4,43	4,4	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,1	4,06	4	3,94	3,87	3,84	3,81	3,77	3,74	3,7	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,3	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,5	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,9	2,86	2,83	2,79	2,75	2,71
10	4,96	4,1	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,7	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,2	3,09	3,01	2,95	2,9	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,4
12	4,75	3,89	3,49	3,26	3,11	3	2,91	2,85	2,8	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,3
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,6	2,53	2,46	2,42	2,38	2,34	2,3	2,25	2,21
14	4,6	3,74	3,34	3,11	2,96	2,85	2,76	2,7	2,65	2,6	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,9	2,79	2,71	2,64	2,59	2,54	2,48	2,4	2,33	2,29	2,25	2,2	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,2	2,96	2,81	2,7	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,1	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,9	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,1	2,87	2,71	2,6	2,51	2,45	2,39	2,35	2,28	2,2	2,12	2,08	2,04	1,99	1,95	1,9	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,1	2,05	2,01	1,96	1,92	1,87	1,81
22	4,3	3,44	3,05	2,82	2,66	2,55	2,46	2,4	2,34	2,3	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,8	2,64	2,53	2,44	2,37	2,32	2,27	2,2	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,4	3,01	2,78	2,62	2,51	2,42	2,36	2,3	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,6	2,49	2,4	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,9	1,85	1,8	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,2	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,2	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,7	2,55	2,43	2,35	2,28	2,22	2,18	2,1	2,03	1,94	1,9	1,85	1,81	1,75	1,7	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4	3,15	2,76	2,53	2,37	2,25	2,17	2,1	2,04	1,99	1,92	1,84	1,75	1,7	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,5	1,43	1,35	1,25

Приложение 6. d-статистика Дарбина-Уотсона: dl и du, уровень значимости в 1%

N	k=1		k=2		k=3		k=4		k=5	
	dl	du	dl	du	dl	du	dl	du	dl	du
15	0,81	1,07	0,7	1,25	0,59	1,46	0,49	1,7	0,39	1,96
16	0,84	1,09	0,74	1,25	0,63	1,44	0,53	1,66	0,44	1,9
17	0,87	1,1	0,77	1,25	0,67	1,43	0,57	1,63	0,48	1,85
18	0,9	1,12	0,8	1,26	0,71	1,42	0,61	1,6	0,52	1,8
19	0,93	1,13	0,83	1,26	0,74	1,41	0,65	1,58	0,56	1,77
20	0,95	1,15	0,86	1,27	0,77	1,41	0,68	1,57	0,6	1,74
21	0,97	1,16	0,89	1,27	0,8	1,41	0,72	1,55	0,63	1,71
22	1	1,17	0,91	1,28	0,83	1,4	0,75	1,54	0,66	1,69
23	1,02	1,19	0,94	1,29	0,86	1,4	0,77	1,53	0,7	1,67
24	1,04	1,2	0,96	1,3	0,88	1,41	0,8	1,53	0,72	1,66
25	1,05	1,21	0,98	1,3	0,9	1,41	0,83	1,52	0,75	1,65
26	1,07	1,22	1	1,31	0,93	1,41	0,85	1,52	0,78	1,64
27	1,09	1,23	1,02	1,32	0,95	1,41	0,88	1,51	0,81	1,63
28	1,1	1,24	1,04	1,32	0,97	1,41	0,9	1,51	0,83	1,62
29	1,12	1,25	1,05	1,33	0,99	1,42	0,92	1,51	0,85	1,61
30	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
31	1,15	1,27	1,08	1,34	1,02	1,42	0,96	1,51	0,9	1,6
32	1,16	1,28	1,1	1,35	1,04	1,43	0,98	1,51	0,92	1,6
33	1,17	1,29	1,11	1,36	1,05	1,43	1	1,51	0,94	1,59
34	1,18	1,3	1,13	1,36	1,07	1,43	1,01	1,51	0,95	1,59
35	1,19	1,31	1,14	1,37	1,08	1,44	1,03	1,51	0,97	1,59
36	1,21	1,32	1,15	1,38	1,1	1,44	1,04	1,51	0,99	1,59
37	1,22	1,32	1,16	1,38	1,11	1,45	1,06	1,51	1	1,59
38	1,23	1,33	1,18	1,39	1,12	1,45	1,07	1,52	1,02	1,58
39	1,24	1,34	1,19	1,39	1,14	1,45	1,09	1,52	1,03	1,58
40	1,25	1,34	1,2	1,4	1,15	1,46	1,1	1,52	1,05	1,58
45	1,29	1,38	1,24	1,42	1,2	1,48	1,16	1,53	1,11	1,58
50	1,32	1,4	1,28	1,45	1,24	1,49	1,2	1,54	1,16	1,59
55	1,36	1,43	1,32	1,47	1,28	1,51	1,25	1,55	1,21	1,59
60	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,6
65	1,41	1,47	1,38	1,5	1,35	1,53	1,31	1,57	1,28	1,61
70	1,43	1,49	1,4	1,52	1,37	1,55	1,34	1,58	1,31	1,61
75	1,45	1,5	1,42	1,53	1,39	1,56	1,37	1,59	1,34	1,62
80	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,6	1,36	1,62
85	1,48	1,53	1,46	1,55	1,43	1,58	1,41	1,6	1,39	1,63
90	1,5	1,54	1,47	1,56	1,45	1,59	1,43	1,61	1,41	1,64
95	1,51	1,55	1,49	1,57	1,47	1,6	1,45	1,62	1,42	1,64
100	1,52	1,56	1,5	1,58	1,48	1,6	1,46	1,63	1,44	1,65

Содержание

Шадманова Гулчера

Мирзаев Сайибджан Сабитович

ПРИКЛАДНАЯ ЭКОНОМЕТРИКА

Учебное пособие

Редактор З.С.Абдуллаев

Подписано в печать: _15.11.2022_г. Формат 60x84 - 1/16.
Объём: _11,0_ п.л. Тираж: _50_экз. Заказ №____.
Отпечатано в типографии НИУ “ТИИИМСХ”.
Ташкент 100000, ул. Кари-Ниязова, 39.