

Министерство сельского хозяйства Российской Федерации

ФГБОУ ВО Пензенский ГАУ

О.В. Ментюкова

ЭКОНОМЕТРИКА

УЧЕБНОЕ ПОСОБИЕ

Пенза 2020

Министерство сельского хозяйства Российской Федерации

ФГБОУ ВО Пензенский ГАУ

О.В. Ментюкова

ЭКОНОМЕТРИКА

Учебное пособие
для студентов, обучающихся
по направлению подготовки 38.03.01 Экономика

Пенза 2020

УДК 330.43
ББК 65в6я73
М 50

Рецензент – кандидат экономических наук, доцент кафедры «Бухгалтерский учет, анализ и аудит» О. В. Лаврина.

Рекомендовано к изданию методической комиссией экономического факультета _____.2020 г, протокол № ____.

Ментюкова, Оксана Викторовна
М 50 Эконометрика: учебное пособие / О.В. Ментюкова. – Пенза: РИО ПГАУ, 2020. – 140 с.

Учебное пособие содержит краткий курс лекций по дисциплине «Эконометрика», включая описание задач эконометрического моделирования и его основных этапов, методов построения и статистического анализа эконометрических моделей. Предназначено для студентов, обучающихся по направлению подготовки 38.03.01 Экономика.

© ФГБОУ ВО
Пензенский ГАУ, 2020
© О.В. Ментюкова, 2020

СОДЕРЖАНИЕ

Предисловие	5
1 Введение в эконометрику	6
1.1 Предмет, цели и задачи эконометрики	6
1.2 Типы зависимостей.....	8
1.3 Типы данных и классы моделей.....	9
1.4 Основные этапы эконометрического моделирования.....	11
<i>Контрольные вопросы</i>	14
2 Парный регрессионный анализ	15
2.1 Суть регрессионного анализа	15
2.2 Парная линейная регрессия	16
2.3 Коэффициент корреляции	21
2.4 Основные предпосылки регрессионного анализа. Теорема Гаусса-Маркова	24
2.5 Оценка значимости параметров линейного уравнения регрессии	26
2.6 Оценка значимости уравнения регрессии.....	30
2.7 Проверка общего качества уравнения регрессии. Коэффициент детерминации	33
2.8 Интервалы прогноза по линейному уравнению регрессии...	35
2.9 Парная нелинейная регрессия	38
2.10 Коэффициент ранговой корреляции Спирмена	42
<i>Задачи для самостоятельного решения</i>	44
<i>Контрольные вопросы</i>	48
3 Множественный регрессионный анализ	49
3.1 Множественная линейная регрессия	49
3.2 Множественная корреляция	55
3.3 Оценка статистической значимости коэффициентов уравнения множественной линейной регрессии	57
3.4 Оценка значимости множественной регрессии	58
3.5 Оценка уравнения линейной регрессии в MS Excel.....	59
3.6 Мультиколлинеарность	65
3.7 Частная корреляция	68
3.8 Фиктивные переменные в регрессионных моделях	70
3.9 Нелинейные модели регрессии. Производственная функция Кобба-Дугласа	71

<i>Задачи для самостоятельного решения</i>	72
<i>Контрольные вопросы</i>	78
4 Временные ряды и прогнозирование	80
4.1 Общие сведения о временных рядах	80
4.2 Автокорреляция уровней временного ряда	82
4.3 Аналитическое выравнивание временного ряда	84
4.4 Прогнозирование на основе моделей временных рядов	87
<i>Задачи для самостоятельного решения</i>	89
<i>Контрольные вопросы</i>	90
5 Гетероскедастичность и автокорреляция остатков	91
5.1 Суть гетероскедастичности	91
5.2 Тесты на гетероскедастичность	92
5.3 Метод взвешенных наименьших квадратов	97
5.4 Суть и причины автокорреляции	101
5.5 Обнаружение автокорреляции	103
5.6 Методы устранения автокорреляции	107
<i>Задачи для самостоятельного решения</i>	112
<i>Контрольные вопросы</i>	115
6 Системы эконометрических уравнений	116
6.1 Системы уравнений, используемые в эконометрике	116
6.2 Структурная и приведенная формы модели	118
6.3 Проблема идентифицируемости	119
6.4 Оценивание параметров структурной модели	120
<i>Контрольные вопросы</i>	122
Тестовые задания	123
Список литературы	136
Приложения	137

ПРЕДИСЛОВИЕ

Современные экономические теории и исследования, опирающиеся в значительной степени на использование математических моделей и методов анализа, требуют от экономистов достаточно свободного владения математическим аппаратом изучения статистических данных. Поэтому в последнее время заметно возрос интерес к эконометрике – науке, изучающей количественные взаимосвязи экономических явлений и процессов. Эконометрические методы широко используются в маркетинговых исследованиях, финансовом анализе, банковской деятельности и других областях.

Настоящее учебное пособие представляет собой лаконичный конспект лекций и одновременно практикум по эконометрике, что позволяет использовать его для методического сопровождения как лекционного курса, так и практических занятий.

Пособие состоит из введения, основного учебного материала и приложений. Рассмотрены основные начальные разделы эконометрики: парные и множественные линейные регрессионные модели, нелинейные модели; системы линейных одновременных уравнений; модели временных рядов. Учебный материал сопровождается иллюстрирующими его примерами и задачами. В заключение каждой темы для проверки полученных знаний приводятся контрольные вопросы и задачи для самостоятельного решения. Список литературы в конце пособия включает в себя книги, содержащие достаточно полное теоретическое изложение основ эконометрики. Для удобства читателя в приложении представлены необходимые математико-статистические таблицы.

Учебное пособие предназначено студентам, обучающимся по направлению подготовки 38.03.01 Экономика.

1 ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ

В 1930 г. в экономической науке выделилось новое направление – эконометрика. Термин «эконометрика» подчеркивает специфику, содержание эконометрики как науки, центральной проблемой которой является построение эконометрической модели и определение возможностей ее использования для описания, анализа и прогнозирования реальных экономических процессов.

Известный эконометрист Цви Гриллихес писал: «Эконометрика является одновременно нашим телескопом и нашим микроскопом для изучения окружающего экономического мира». Это высказывание подчеркивает значение эконометрического подхода как на микроуровне (поведение индивидов, фирм и т. п.), так и на макроуровне. Свидетельством всемирного признания эконометрики является присуждение за наиболее выдающиеся разработки в этой области Нобелевских премий по экономике: Р. Фришу и Я. Тинбергу за разработку математических методов анализа экономических процессов (1969 г.); Л. Клейну за создание эконометрических моделей и их применение к анализу экономических колебаний и экономической политике (1980 г.); Т. Хаавельмо за прояснение вероятностных основ эконометрики и анализ одновременных экономических структур (1989 г.); Дж. Хекману и Д. Макфаддену за развитие теории и методов анализа селективных выборок и моделей дискретного выбора (2000 г.).

1.1 Предмет, цели и задачи эконометрики

Закономерности в экономике выражаются в виде зависимостей экономических показателей и математических моделей их поведения. Такие зависимости и модели могут быть получены только путем обработки реальных статистических данных с учетом внутренних связей и случайных факторов.

Эконометрика – это наука, изучающая конкретные количественные закономерности и взаимосвязи экономических объектов и процессов с помощью математических методов и моделей.

Цель эконометрики, согласно утверждению Э. Маленво, – *эмпирический вывод экономических законов*. В эконометрике на базе реальных статистических данных строятся, анализируются и совершенствуются математические модели реальных экономических явлений.

Эконометрика позволяет найти количественное подтверждение либо опровержение того или иного экономического закона либо гипотезы.

К основным задачам эконометрики относятся:

- построение эконометрических моделей, т. е. представление экономических моделей в математической форме, удобной для проведения эмпирического анализа;
- оценка параметров построенной модели, делающих выбранную модель наиболее адекватной реальным данным;
- проверка качества найденных параметров модели и самой модели в целом;
- использование построенных моделей для объяснения поведения исследуемых экономических показателей, прогнозирования и предсказания, а также для осмысленного проведения экономической политики.

Эконометрика зародилась и получила развитие на основе слияния статистики, экономической теории и математики. Действительно, *предмет ее исследования – экономические явления*. Но в отличие от экономической теории эконометрика делает упор на количественные, а не на качественные аспекты этих явлений. Например, экономическая теория утверждает, что спрос на товар с ростом его цены убывает. Но при этом практически неисследованным остается вопрос, как быстро и по какому закону происходит это убывание. Эконометрика отвечает на этот вопрос для каждого конкретного случая.

Изучение экономических процессов (взаимосвязей) в эконометрике осуществляется через математические (эконометрические) модели. Но если математическая экономика строит и анализирует эти модели без использования реальных числовых значений, то эконометрика концентрируется на изучении моделей на базе эмпирических данных. Основной базой данных для эконометрических исследований служат данные официальной статистики.

Одной из основных задач экономической статистики является сбор, обработка и представление экономических данных в наглядной форме: в виде таблиц, графиков, диаграмм. Эконометрика также активно пользуется этим инструментарием, но идет дальше, применяя его для анализа экономических взаимосвязей и прогнозирования.

Мощным инструментом эконометрических исследований является аппарат математической статистики. Действительно, большинство экономических показателей носит характер случайных величин,

предсказать точные значения которых практически невозможно. Связи между экономическими показателями обычно не носят строгий функциональный характер, а допускают наличие каких-либо случайных отклонений. Вследствие этого использование методов математической статистики в эконометрике естественно и обоснованно.

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

Широкому внедрению эконометрических методов способствовало появление во второй половине XX в. электронных вычислительных машин и, в частности, персональных компьютеров. Компьютерные эконометрические пакеты сделали эти методы более доступными и наглядными, так как наиболее трудоемкую работу по расчету различных статистик, параметров, характеристик, построению таблиц и графиков в основном стал выполнять компьютер, а эконометристу осталось главным образом творческая работа: постановка задачи, выбор соответствующей модели и метода ее решения, интерпретация результатов.

1.2 Типы зависимостей

Изучение природы и общества требует рассмотрения явлений в их взаимосвязи. Связь двух явлений или признаков математически отображается в форме уравнения связи двух переменных. Между различными явлениями и их признаками различают два типа зависимостей: *функциональную* и *статистическую*.

В случае *функциональной зависимости* значению одной переменной обязательно соответствует *точно заданное* значение другой переменной (например, скорость свободного падения тела в вакууме в зависимости от времени и т. п.). В экономике функциональная зависимость между переменными проявляется редко.

Статистическая зависимость существует, когда каждому значению одной переменной соответствует *множество возможных* значений другой переменной. Иначе говоря, каждому значению одной переменной соответствует определенное (условное) распределение другой переменной.

Возникновение статистической связи обуславливается тем, что зависимая переменная, кроме выделенной переменной, подвержена влиянию ряда неконтролируемых или неучтенных факторов, а также тем, что измерение значений переменных неизбежно сопровождается некоторыми случайными ошибками. Примером статистической связи является зависимость урожайности от количества внесенных удобрений, производительности труда на предприятии от его энерговооруженности и т.п.

Статистическая зависимость называется *корреляционной*, если каждому значению одной переменной соответствует определенное условное математическое ожидание (среднее значение) другой.

Корреляционная зависимость может быть представлена в виде

$$M(Y / X = x) = f(x), \quad (1.1)$$

где $M(Y / X = x)$ – математическое ожидание (среднее значение) случайной переменной Y , вычисленное в предположении, что переменная X приняла значение x .

Статистические связи между переменными можно изучать методами корреляционного и регрессионного анализа.

Основной задачей *регрессионного анализа* является установление формы и изучение зависимости между переменными. Основная задача *корреляционного анализа* – выявление связи между переменными и оценка ее тесноты.

Понятия *корреляции* и *регрессии* появились в середине XIX в. благодаря работам английских статистов Ф. Гальтона и К. Пирсона. Первый термин произошел от латинского «*correlatio*» – соотношение, взаимосвязь. Второй термин (от лат. «*regression*» – движение назад) введен Ф. Гальтоном, который изучая зависимость между ростом родителей и их детей, обнаружил явление «регрессии к среднему» – у детей, родившихся у очень высоких родителей, рост имел тенденцию быть ближе к средней величине.

1.3 Типы данных и классы моделей

Эмпирическим материалом для проведения эконометрического исследования служат данные наблюдений за изучаемым экономическим объектом. Как правило, используют два типа выборочных данных: *пространственные* и *временные*.

Пространственные данные – это данные по какому-либо экономическому показателю, полученные от разных однотипных объектов (фирм, регионов, областей, стран и т.п.), но относящиеся к одному и тому же моменту времени. Например, данные об объеме производства, количестве работников, размере основных производственных фондов разных фирм в один и тот же момент времени.

Временные ряды – это данные, характеризующие один и тот же объект в различные моменты времени. Например, ежегодные данные об уровне инфляции, ежемесячные данные о средней заработной плате, ежедневные данные о курсе доллара.

Любые экономические данные формируются под воздействием множества факторов, не все из которых доступны внешнему контролю. Неучтенные факторы обуславливают случайность данных, которые они определяют.

Главным инструментом эконометрических исследований является *эконометрическая модель* – уравнение (или система уравнений), которое в математической форме описывает основные количественные зависимости между анализируемыми экономическими явлениями или процессами.

Для анализа и (или) прогноза применяются три основных класса моделей: регрессионные модели с одним уравнением, модели временных рядов и системы одновременных уравнений.

1. В *регрессионных моделях с одним уравнением* зависимая (объясняемая) переменная Y представляется в виде функции

$$Y = f(X_1, \dots, X_k, \beta_1, \dots, \beta_\varphi), \quad (1.2)$$

где $X_1, \dots, X_k$ – независимые (объясняющие переменные);
 $\beta_1, \dots, \beta_\varphi$ – параметры.

В зависимости от количества независимых (объясняющих) переменных, включенных в уравнение, принято различать *простую (парную)* и *множественную* регрессии.

Простая (парная) регрессия представляет собой зависимость между двумя переменными.

Множественная регрессия – это регрессия между зависимой переменной и несколькими независимыми переменными.

По виду функции регрессионные модели делятся на *линейные* и *нелинейные*.

2. *Модели временных рядов* – это модели, в которых зависимая (объясняемая) переменная является функцией переменной времени или переменных, относящихся к другим моментам времени.

К моделям временных рядов относятся *модели тренда* и *сезонности*. *Тренд* – устойчивое изменение уровня показателя в течение длительного периода времени. *Сезонность* – регулярное устойчивое колебание уровня показателя (в течение недели, месяца или года). К этому классу относятся и более сложные модели: модель адаптивного прогноза, модель авторегрессии, скользящего среднего и другие. Общей чертой моделей этого класса является то, что они объясняют поведение временного ряда, исходя только из предыдущих значений.

3. *Системы одновременных уравнений* описываются системами уравнений. Системы могут состоять из тождеств и регрессионных уравнений, каждое из которых может кроме объясняющих переменных, включать в себя также объясняемые переменные из других уравнений системы.

1.4 Основные этапы эконометрического моделирования

При эконометрическом исследовании можно выделить шесть основных этапов: постановочный, априорный, этап параметризации, информационный, этапы идентификации и верификации модели.

1-й этап (постановочный). На этом этапе формируется *цель* исследования; определяется *набор* участвующих в модели *экономических переменных*. В качестве цели эконометрического моделирования обычно рассматривают *анализ* исследуемого экономического объекта (процесса); *прогноз* его экономических показателей, *имитацию* развития объекта при различных значениях переменных (отражая их случайный характер, изменение во времени), *выработку управленческих решений*.

При выборе в модель экономических переменных необходимо теоретическое обоснование каждой из них. При этом рекомендуется, чтобы число переменных было не очень большим и, как минимум, в 6 – 8 раз меньше числа наблюдений. Объясняющие переменные не должны быть связаны функциональной или тесной корреляционной зависимостью, так как это может привести к невозможности оценки параметров модели или к получению неустойчивых, не имеющих реального смысла оценок. Но в любом случае определяющим при

включении в модель тех или иных переменных является экономический (качественный) анализ исследуемого объекта.

2-й этап (априорный). На этом этапе проводится предварительный *анализ сущности* изучаемого объекта. Происходит формирование априорной (известной до начала моделирования) информации о данном явлении и ее дальнейшая формализация в виде ряда гипотез и исходных допущений.

3-й этап (параметризация). На этом этапе осуществляется непосредственно *моделирование*, т.е. выбор *общего вида* модели, выявление входящих в нее связей.

Основная задача, решаемая на этом этапе, – выбор вида функции $f(X)$ в эконометрической модели. Весьма важной проблемой на этом (и предыдущих) этапе эконометрического моделирования является проблема *спецификации* модели, в частности: выражение в математической форме обнаруженных связей и соотношений; уточнение состава переменных, формулировка исходных предпосылок и ограничений модели. От того, насколько удачно решена проблема спецификации модели, в значительной степени зависит успех всего эконометрического моделирования.

4-й этап (информационный). Этап посвящен *сбору* необходимой статистической информации о тех показателях, которые были отобраны на постановочном этапе. При этом возможно использование различных статистических сборников, а при необходимости и проведение дополнительных исследований. При сборе статистической информации необходимо обращать внимание на ее достоверность и полноту.

5-й этап (идентификация модели). На этом этапе осуществляется *оценка параметров (коэффициентов) модели* и ее *статистический анализ*.

6-й этап (верификация модели). Этот этап состоит в *проверке истинности, адекватности модели*. Выясняется, насколько удачно решены проблемы спецификации, идентификации модели, какова точность расчетов по данной модели, в конечном счете, насколько соответствует построенная модель моделируемому реальному экономическому объекту или процессу.

Следует заметить, что если имеются статистические данные, характеризующие моделируемый экономический объект в данный и предшествующие моменты времени, то для верификации модели, построенной для прогноза, достаточно сравнить реальные значения пе-

ременных в последующие моменты времени с соответствующими их значениями, полученными на основе рассматриваемой модели по данным предшествующих моментов.

Если модель удовлетворяет всем необходимым требованиям, то она может быть использована для прогнозирования или объяснения исследуемых экономических процессов. Такая модель позволяет предсказать среднее значение исследуемого экономического показателя на основе прогнозируемых значений факторов; предвидеть допустимые отклонения конкретных значений изучаемой величины от предсказываемого по модели значения. Хорошая модель поможет определить, на какие факторы и в каком направлении и объеме следует воздействовать, чтобы значение исследуемого показателя лежало в определенных числовых границах.

Суть и последовательность эконометрических исследований можно наглядно представить в виде схемы (рис. 1.1).

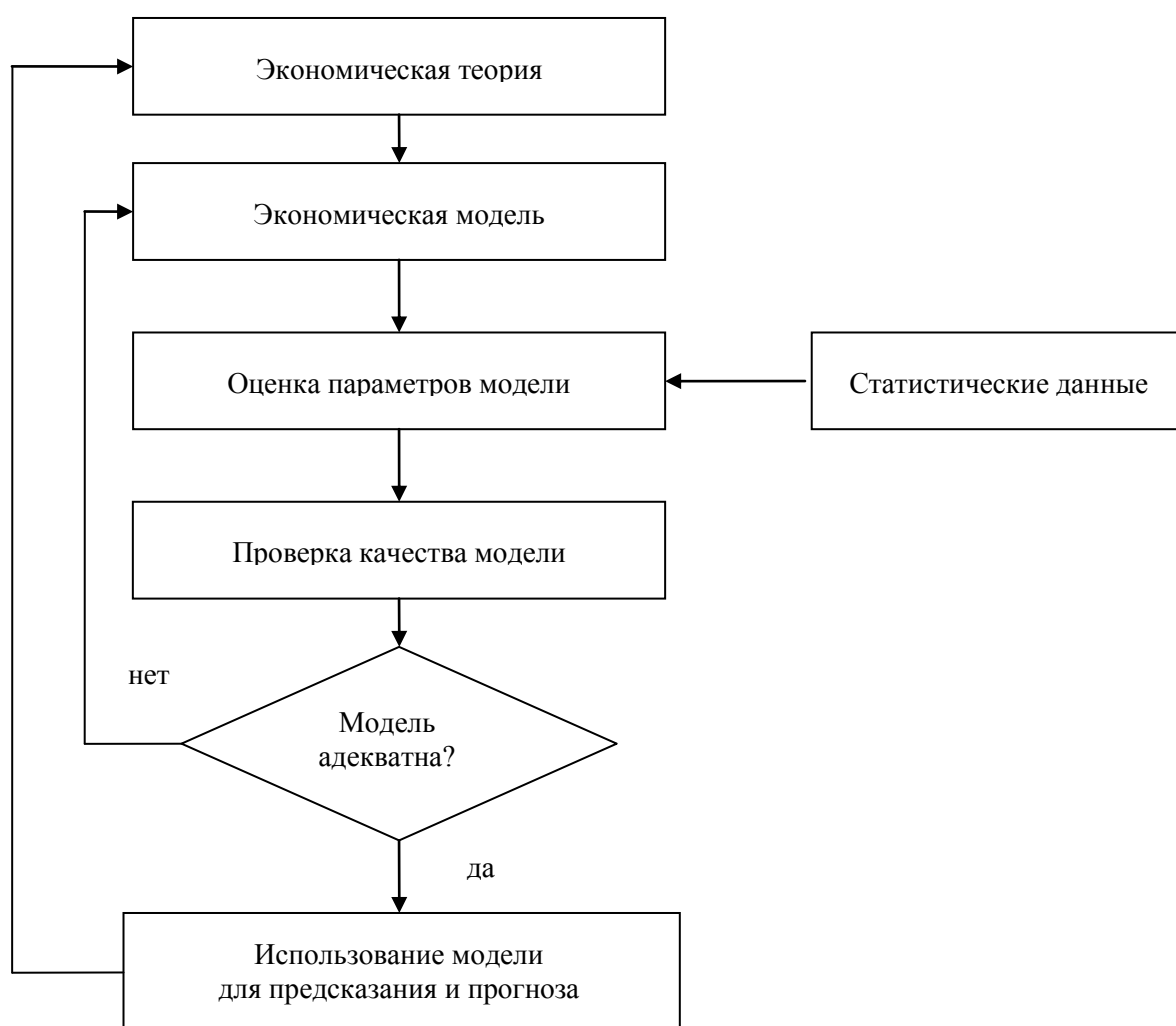


Рисунок 1.1 – Схема эконометрического исследования

Данная схема отражает циклический характер современных экономических исследований: от экономической теории к моделированию; от моделирования к совершенствованию теории и более глубокому пониманию сути происходящих процессов.

Контрольные вопросы

1. Дайте определение эконометрики.
2. Какие цели и задачи решает эконометрика?
3. С какими науками тесно связана эконометрика?
4. В чем состоит особая роль статистики в формировании эконометрического метода?
5. Назовите основные этапы эконометрического моделирования.
6. Какие задачи решаются на этапе спецификации модели?
7. Какие типы данных используются в эконометрическом исследовании?
8. Какие данные называются пространственными? Приведите примеры.
9. Что такое временной ряд? Приведите примеры.
10. Назовите основные классы моделей, используемые в эконометрике.
11. Дайте определение простой и множественной регрессии.
12. Какая зависимость называется функциональной? Приведите примеры.
13. Дайте определение статистической зависимости. Приведите примеры.
14. Дайте определение корреляционной зависимости.
15. Назовите основные задачи, решаемые регрессионным и корреляционным анализом.

2 ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Большую роль в экономике играет исследование зависимостей и взаимосвязей между объективно существующими явлениями и процессами. Оно дает возможность глубже понять сложный механизм причинно-следственных отношений между явлениями.

Поведение и значение любого экономического показателя зависят практически от бесконечного количества факторов, и все учесть нереально. Но в этом и нет необходимости. Обычно лишь ограниченное количество факторов действительно существенно воздействуют на исследуемый экономический показатель. Доля влияния остальных факторов столь незначительна, что их игнорирование не может привести к существенным отклонениям в поведении исследуемого объекта. Часто из всего круга факторов можно выделить лишь один доминирующий фактор, что позволяет использовать при анализе одну из наиболее простых моделей – парную (простую) регрессию.

2.1 Суть регрессионного анализа

В регрессионном анализе рассматривается *односторонняя зависимость случайной переменной Y от одной (или нескольких) неслучайной независимой переменной X* . Такая зависимость может возникнуть, например, в случае, когда при каждом фиксированном значении X соответствующие значения Y подвержены случайному разбросу за счет действия ряда неконтролируемых факторов. Такая зависимость Y от X может быть представлена в виде модельного уравнения регрессии Y по X

$$M(Y / X = x) = f(x), \quad (2.1)$$

где $M(Y / X = x)$ – математическое ожидание случайной переменной Y , вычисленное в предположении, что переменная X приняла значение x .

Зависимую переменную Y называют также *объясняемой, результативным признаком, эндогенной*, а независимую переменную X – *объясняющей, фактором, экзогенной*.

Уравнение (2.1) называется *уравнением регрессии*, а функция $f(x)$ – *функцией регрессии*.

При рассмотрении зависимости двух переменных говорят о *парной регрессии*.

В силу воздействия неучтенных случайных факторов и причин реальные (фактические) значения переменной Y будут в большей или меньшей мере отклоняться от функции регрессии $f(X)$. В этом случае уравнение взаимосвязи двух переменных (парная регрессионная модель) может быть представлена в виде:

$$Y = f(X) + \varepsilon, \quad (2.2)$$

где ε – случайная переменная, характеризующая отклонение реального значения зависимой переменной от теоретического значения, найденного по уравнению регрессии.

Переменную ε называют *возмущением*. Причинами присутствия в регрессионной модели случайной величины ε могут быть:

- невключение в модель всех объясняющих переменных;
- неправильный выбор функциональной формы модели;
- ошибки измерений;
- ограниченность статистических данных;
- непредсказуемость человеческого фактора и другие.

2.2 Парная линейная регрессия

Наиболее распространенным и простым видом зависимости между экономическими переменными является модель парной линейной регрессии.

Теоретическая линейная регрессионная модель представляется в виде

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (2.3)$$

где β_0, β_1 – параметры уравнения регрессии.

Для определения значений параметров регрессии (β_0, β_1) необходимо знать и использовать все значения переменных X и Y генеральной совокупности, что практически невозможно. Как правило, исследователь располагает лишь выборкой пар значений $(x_i; y_i)$ ограниченного объема n . Таким образом, задача линейного регрессионного анализа состоит в том, чтобы по имеющимся статистическим данным $(x_i; y_i)$, $i = 1, 2, \dots, n$ для переменных X и Y получить наилучшие оценки неизвестных параметров β_0 и β_1 .

По выборке ограниченного объема n , можно построить так называемое *выборочное уравнение регрессии*

$$\hat{y} = b_0 + b_1 x, \quad (2.4)$$

где b_0, b_1 – оценки параметров β_0 и β_1 ;

$\hat{y}$ – условная (групповая) средняя переменной Y при конкретном значении переменной X .

Для нахождения параметров b_0 и b_1 используется **метод наименьших квадратов (МНК)**, согласно которому *неизвестные параметры b_0 и b_1 выбираются таким образом, чтобы сумма квадратов отклонений фактических значений y_i от значений $\hat{y}_i$ найденных по уравнению регрессии (2.4), была минимальной:*

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 \rightarrow \min. \quad (2.5)$$

Необходимым условием существования минимума функции двух переменных $S = S(b_0, b_1)$ (2.5) является равенство нулю ее частных производных по неизвестным параметрам b_0 и b_1 , т. е.

$$\begin{cases} \frac{\partial S}{\partial b_0} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) = 0; \\ \frac{\partial S}{\partial b_1} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) x_i = 0. \end{cases}$$

После преобразований получим *систему нормальных уравнений*:

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i; \\ b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \end{cases} \quad (2.6)$$

Разделив обе части уравнений (2.6) на n , получим систему нормальных уравнений в виде

$$\begin{cases} b_0 + b_1 \bar{x} = \bar{y}; \\ b_0 \bar{x} + b_1 \bar{x}^2 = \overline{xy}, \end{cases} \quad (2.7)$$

где соответствующие средние значения определяются по формулам:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}; \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}; \quad \bar{x}^2 = \frac{\sum_{i=1}^n x_i^2}{n}; \quad \overline{xy} = \frac{\sum_{i=1}^n x_i y_i}{n}$$

Решая систему (2.7), найдем:

$$\begin{cases} b_0 = \bar{y} - b_1 \bar{x}; \\ b_1 = \frac{\overline{xy} - \bar{x} \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\text{cov}(x, y)}{\sigma_x^2}, \end{cases} \quad (2.8)$$

где $\text{cov}(x, y)$ – выборочная ковариация переменных X и Y :

$$\text{cov}(x, y) = \overline{xy} - \bar{x} \bar{y} \quad (2.9)$$

σ_x^2 – выборочная дисперсия переменной X :

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 \quad (2.10)$$

Отметим, что линия регрессии проходит через точку $(\bar{x}; \bar{y})$.

Коэффициент b_1 называется *коэффициентом регрессии* Y по X . Его величина показывает, на сколько единиц в среднем изменяется переменная Y при увеличении переменной X на одну единицу. Формально b_0 – значение зависимой переменной Y при $x = 0$. Если переменная X не имеет и не может иметь нулевого значения, то такая трактовка коэффициента b_0 не имеет смысла.

► **Пример 2.1.** Для анализа зависимости между урожайностью зерновых культур Y (ц/га) и количеством внесенных органических удобрений X (т/га) получены данные по 20 сельскохозяйственным предприятиям (таблица 2.1).

Таблица 2.1 – Исходные данные: урожайность зерновых культур Y (ц/га) и количество внесенных органических удобрений X (т/га)

i	x_i	y_i	i	x_i	y_i
1	5,0	13,8	11	8,3	16,8
2	5,5	14,0	12	8,5	17,0
3	5,6	14,3	13	6,0	17,2
4	6,5	14,8	14	7,5	17,6
5	7,0	15,1	15	8,5	17,9
6	8,0	15,6	16	6,0	18,1
7	7,0	15,8	17	8,0	18,3
8	7,3	16,1	18	9,5	19,5
9	6,5	16,4	19	10,0	20,3
10	7,5	16,6	20	10,9	20,8

Необходимо: 1) определить вид зависимости между X и Y ;
2) оценить по МНК параметры уравнения линейной регрессии Y по X .

Решение.

1) Самым простым способом определения вида зависимости между двумя переменными является построение *корреляционного поля* (*поля корреляции*). С этой целью отобразим каждую пару наблюдений $(x_i; y_i)$ в виде точки на координатной плоскости (рис. 2.1).

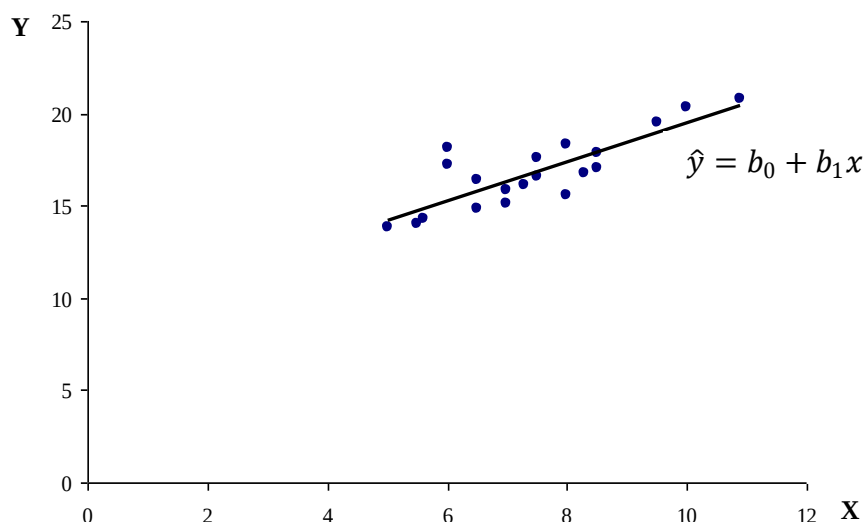


Рисунок 2.1 – Графическое изображение зависимости

По характеру расположения точек на корреляционном поле предполагаем наличие *линейной* корреляционной (регрессионной) зависимости между переменными X и Y , т. е. зависимость между урожайностью зерновых культур и количеством внесенных удобрений на 1 га посева можно описать уравнением регрессии вида $\hat{y} = b_0 + b_1 x$.

2) Найдем значения параметров b_0 и b_1 уравнения линейной регрессии. Для наглядности вычислений представим исходные данные и расчетные показатели в виде таблицы 2.2.

По формулам (2.8) – (2.10) находим выборочные характеристики и параметры уравнения регрессии:

$$\text{cov}(x, y) = 127,6635 - 7,455 \cdot 16,8 = 2,4195;$$

$$\sigma_x^2 = 57,8675 - 7,455^2 = 2,2905;$$

$$b_1 = \frac{2,4195}{2,2905} = 1,056; \quad b_0 = 16,8 - 1,056 \cdot 7,455 = 8,93.$$

Таблица 2.2 – Корреляционная зависимость между урожайностью зерновых культур Y (ц/га) и количеством внесенных органических удобрений X (т/га)

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	5,0	13,8	25,00	190,44	69,00
2	5,5	14,0	30,25	196,00	77,00
3	5,6	14,3	31,36	204,49	80,08
4	6,5	14,8	42,25	219,04	96,20
5	7,0	15,1	49,00	228,01	105,70
6	8,0	15,6	64,00	243,36	124,80
7	7,0	15,8	49,00	249,64	110,60
8	7,3	16,1	53,29	259,21	117,53
9	6,5	16,4	42,25	268,96	106,60
10	7,5	16,6	56,25	275,56	124,50
11	8,3	16,8	68,89	282,24	139,44
12	8,5	17,0	72,25	289,00	144,50
13	6,0	17,2	36,00	295,84	103,20
14	7,5	17,6	56,25	309,76	132,00
15	8,5	17,9	72,25	320,41	152,15
16	6,0	18,1	36,00	327,61	108,60
17	8,0	18,3	64,00	334,89	146,40
18	9,5	19,5	90,25	380,25	185,25
19	10,0	20,3	100,00	412,09	203,00
20	10,9	20,8	118,81	432,64	226,72
Сумма	149,1	336,0	1157,35	5719,44	2553,27
Среднее	7,455	16,8	57,8675	285,972	127,6635
	$\bar{x}$	$\bar{y}$	$\overline{x^2}$	$\overline{y^2}$	$\overline{xy}$

Итак, уравнение парной линейной регрессии имеет вид $\hat{y} = 8,93 + 1,056x$.

Коэффициент $b_1 = 1,056$ показывает, что при увеличении на 1 т количества удобрения, внесенного на 1 га посева, урожайность зерновых культур увеличится в среднем на 1,056 ц/га. Свободный член в данном уравнении является положительным ($b_0 = 8,93$) и означает среднюю урожайность зерновых культур при отсутствии органических удобрений, т. е. при $x = 0$. ►

2.3 Коэффициент корреляции

В случае линейной регрессии для оценки тесноты и направления корреляционной зависимости используют *выборочный коэффициент корреляции*

$$r_{xy} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y} = b_1 \frac{\sigma_x}{\sigma_y}, \quad (2.11)$$

где σ_x – среднее квадратическое отклонение переменной X :

$$\sigma_x = \sqrt{\overline{x^2} - \bar{x}^2}, \quad (2.12)$$

σ_y – среднее квадратическое отклонение переменной Y :

$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2}. \quad (2.13)$$

Коэффициент корреляции является безразмерной величиной и характеризует степень линейной зависимости двух переменных.

Коэффициент корреляции r_{xy} совпадает по знаку с коэффициентом регрессии b_1 . Если $r_{xy} > 0$ ($b_1 > 0$), то корреляционная связь между переменными является *прямой*, если $r_{xy} < 0$ ($b_1 < 0$) – *обратной*. При прямой (обратной) связи увеличение одной из переменных ведет к увеличению (уменьшению) условной средней другой.

Коэффициента корреляции обладает следующими свойствами:

Свойство 1. Симметричность, т. е. $r_{xy} = r_{yx}$.

Свойство 2. Коэффициент корреляции принимает значения на отрезке $[-1; 1]$.

В зависимости от того, насколько $|r_{xy}|$ приближается к 1, различают связь слабую, умеренную, заметную, тесную и весьма тесную, т. е. чем ближе $|r_{xy}|$ к 1, тем теснее связь (таблица 2.3).

Таблица 2.3 – Оценка тесноты связи

Теснота связи	Значение коэффициента корреляции при наличии	
	прямой связи	обратной связи
Слабая	0,1÷0,3	(-0,1)÷(-0,3)
Умеренная	0,3÷0,5	(-0,3)÷(-0,5)
Заметная	0,5÷0,7	(-0,5)÷(-0,7)
Тесная	0,7÷0,9	(-0,7)÷(-0,9)
Весьма тесная	0,9÷0,99	(-0,9)÷(-0,99)

На рис. 2.2 приведены две корреляционные зависимости переменной Y от X . В случае $a)$ точки корреляционного поля концентрируются, а в случае $b)$ разбросаны около линии регрессии. Очевидно, в случае $a)$ зависимость между переменными более тесная и коэффициент корреляции должен быть больше, чем в случае $b)$, т. е. $r_a) > r_b)$.

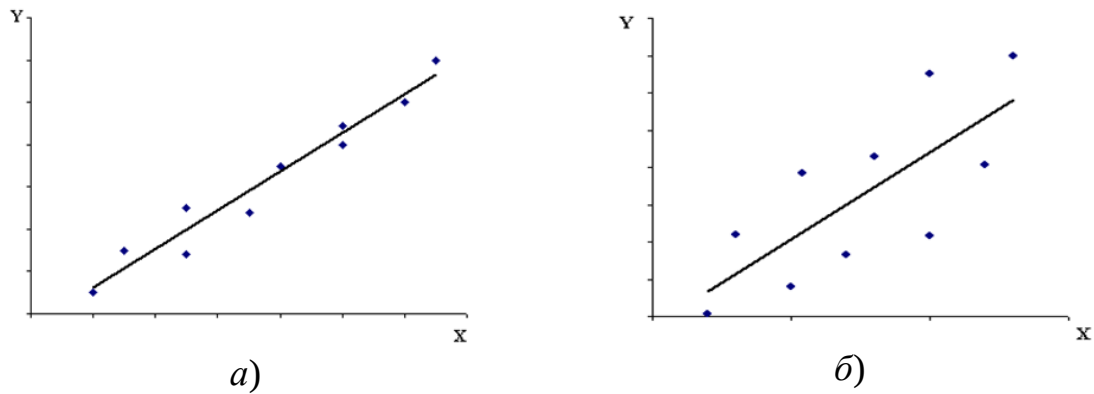


Рисунок 2.2 – Корреляционная зависимость переменной Y от X :
 $a)$ – тесная корреляция; $b)$ – слабая корреляция

Свойство 3. При $r_{xy} = \pm 1$ корреляционная связь представляет линейную функциональную зависимость. При этом все наблюдаемые значения располагаются на прямой линии.

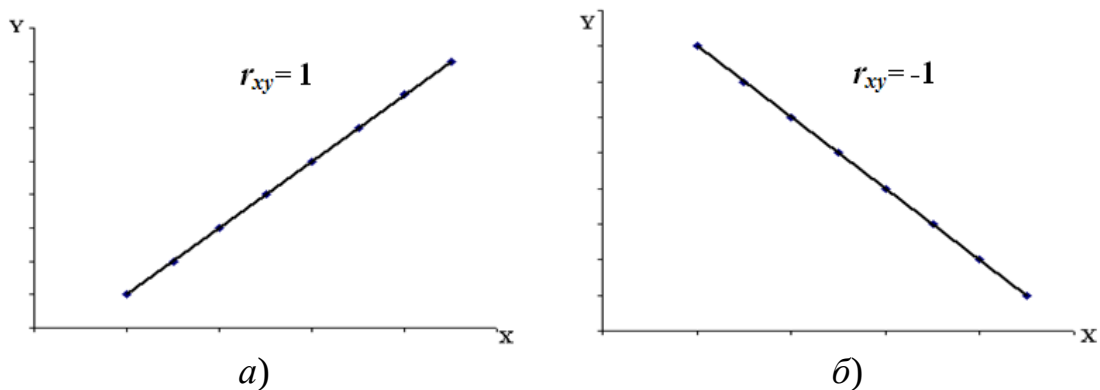


Рисунок 2.3 – Линейная функциональная зависимость:
 $a)$ – прямая связь; $b)$ – обратная связь

Свойство 4. При $r_{xy} = 0$ линейная корреляционная связь отсутствует.

Близкая к нулю величина коэффициента корреляции говорит лишь об отсутствии *линейной* корреляционной зависимости переменных, но не об отсутствии корреляционной зависимости между ними вообще (рис. 2.4, $b)$).

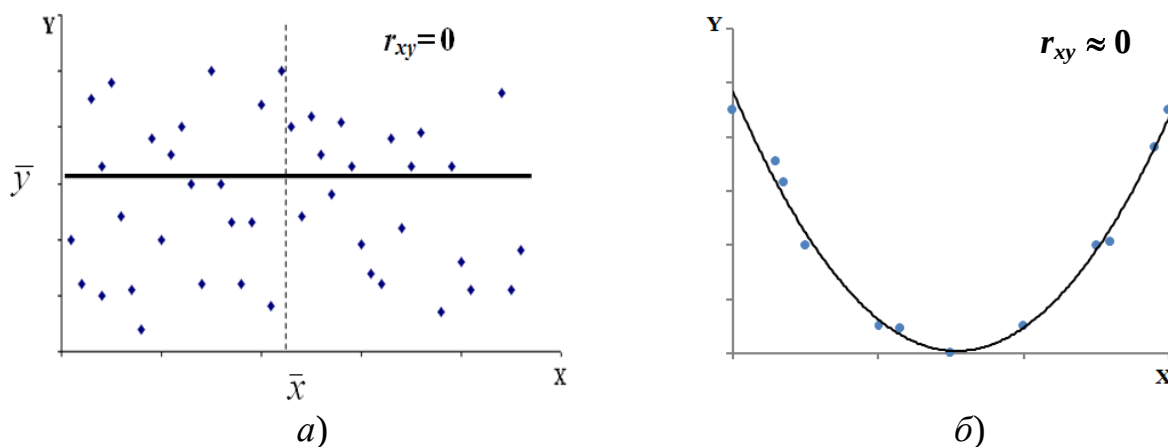


Рисунок 2.4 – Отсутствие линейной корреляционной зависимости

Проверка значимости коэффициента корреляции. Выборочный коэффициент корреляции r_{xy} является оценкой теоретического (генерального) коэффициента корреляции ρ_{xy} между переменными X и Y . Так как r_{xy} вычисляется по значениям переменных, случайно попавшим в выборку из генеральной совокупности, то он является случайной величиной.

Если по данным выборки объема n получено значение $r_{xy} \neq 0$, то возникает вопрос, объясняется ли это действительно существующей линейной связью между переменными X и Y в генеральной совокупности или является следствием случайности отбора переменных в выборку (т. е. при другом отборе возможно, например, $r_{xy} = 0$ или изменение знака r_{xy}). В этом случае проверяется гипотеза H_0 об отсутствии линейной корреляционной связи между переменными в генеральной совокупности, т. е. гипотеза о равенстве нулю истинного значения коэффициента корреляции ($H_0: \rho_{xy} = 0$). При справедливости этой гипотезы статистика

$$t = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \quad (2.14)$$

имеет t – распределение Стьюдента с $k = n - 2$ степенями свободы. Поэтому гипотеза H_0 отвергается, т. е. коэффициент корреляции r_{xy} значимо (существенно) отличается от нуля, и зависимость является достоверной, если $|t| > t_{1-\alpha; n-2}$, где $t_{1-\alpha; n-2}$, – табличное значение t – критерия Стьюдента, определенное на уровне значимости α при числе степеней свободы $(n - 2)$.

Замечание. В практических эконометрических исследованиях наиболее часто используется 5% и 1% уровни значимости. Выбор уровня значимости определяется исследователем.

► **Пример 2.2.** По данным таблицы 2.1 вычислить коэффициент корреляции между переменными X и Y и проверить его значимость на уровне значимости $\alpha = 0,05$.

Решение. В примере 2.1 были рассчитаны

$$\bar{y} = 16,8; \quad \overline{y^2} = 285,972; \quad \text{cov}(x, y) = 2,4195; \quad \sigma_x^2 = 2,2905$$

Вычислим выборочную дисперсию переменной Y :

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2 = 285,972 - 16,8^2 = 3,732$$

По формуле (2.11)
$$r_{xy} = \frac{2,4195}{\sqrt{2,2905 \cdot 3,732}} \approx 0,83.$$

Данное значение коэффициента корреляции позволяет сделать вывод о тесной (прямой) линейной зависимости между урожайностью зерновых культур и количеством внесенного удобрения на 1 га посева. Это также подтверждается расположением точек на корреляционном поле (рис. 2.1).

Для проверки значимости r_{xy} по формуле (2.14) вычислим

$$t = \frac{0,83\sqrt{20-2}}{\sqrt{1-0,83^2}} = 6,3$$

Найдем по таблице II приложений $t_{0,95;18} = 2,10$. Поскольку $|t| > t_{0,95;18}$, коэффициент корреляции между Y и X значимо отличается от нуля. ►

2.4 Основные предпосылки регрессионного анализа.

Теорема Гаусса-Маркова

Пусть для оценки параметров линейной регрессии взята выборка, содержащая n пар значений переменных $(x_i; y_i)$, где $i = 1, 2, \dots, n$. В этом случае линейная парная регрессионная модель имеет вид

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (2.15)$$

Для того чтобы регрессионный анализ, основанный на МНК, давал наилучшие из всех возможных результаты, должны выполняться определенные условия относительно возмущения ε .

Основные предпосылки регрессионного анализа:

1. Математическое ожидание возмущения ε_i в любом наблюдении должно быть равно нулю, т. е.

$$M(\varepsilon_i) = 0, \quad i = 1, 2, \dots, n \quad (2.16)$$

2. Дисперсия возмущения ε_i должна быть постоянной для всех наблюдений:

$$D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2 \quad (2.17)$$

Выполнимость данного условия называется *гомоскедастичностью*. Невыполнимость предпосылки называется *гетероскедастичностью*.

3. Возмущение ε_i и ε_j должны быть некоррелированы между собой:

$$M(\varepsilon_i \varepsilon_j) = 0, \quad (i \neq j) \quad (2.18)$$

4. В модели (2.15) возмущение ε_i есть величина случайная, а объясняющая переменная x_i – величина неслучайная.

В этом случае модель (2.15) называется *классической нормальной линейной регрессионной моделью*.

Наряду с предпосылками 1 – 4 обычно предполагается, что возмущение ε_i имеет нормальное распределение.

Оценкой модели (2.15) по выборке является уравнение регрессии (2.4): $\hat{y} = b_0 + b_1 x$.

Выборочной оценкой возмущения ε_i является отклонение e_i значения y_i переменной Y от значения $\hat{y}_i$, найденного по уравнению регрессии (2.4), т. е. $e_i = y_i - \hat{y}_i$.

Дисперсия возмущений $D(\varepsilon_i)$ заменяется ее несмещенной оценкой – *выборочной остаточной дисперсией*:

$$S^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2} = \frac{\sum_{i=1}^n e_i^2}{n - 2} \quad (2.19)$$

Отметим, что корень квадратный из остаточной дисперсии, т. е.

$$S = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n-2}}, \text{ называется стандартной ошибкой регрессии.}$$

Теорема Гаусса-Маркова. Если условия 1 – 4 регрессионного анализа выполняются, то оценки (b_0, b_1) , сделанные с помощью МНК, являются наилучшими линейными несмещенными оценками, т. е. обладают следующими свойствами:

- 1) *несмещенности*: $M(b_0) = \beta_0, M(b_1) = \beta_1$;
- 2) *эффективности*: имеют наименьшую дисперсию в классе всех линейных несмещенных оценок;
- 3) *состоятельности*: $\lim_{n \rightarrow \infty} D(b_0) = 0, \lim_{n \rightarrow \infty} D(b_1) = 0$, т. е. при достаточно большом n оценки (b_0, b_1) близки к (β_0, β_1) .

2.5 Оценка значимости параметров линейного уравнения регрессии

Параметры b_0 и b_1 уравнения линейной регрессии (2.4) определяются на основе конечного числа статистических данных (выборки) и являются случайными величинами. Поэтому в линейной регрессии обычно оценивается статистическая значимость ее параметров. В этом случае выдвигаются нулевые гипотезы H_0 о случайной природе формирования параметров, т. е. о незначимом их отличии от нуля. Отклонение нулевой гипотезы H_0 означает статистическую значимость соответствующего параметра и, наоборот, невозможность отклонения нулевой гипотезы H_0 означает статистическую незначимость параметра.

Значимость параметров регрессии проверяется путем сопоставления их значений с величинами стандартных ошибок.

Правило проверки значимости параметра b_1 . Стандартная ошибка для параметра b_1 вычисляется по формуле

$$m_{b_1} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (2.20)$$

Находится отношение параметра b_1 к величине его стандартной ошибки:

$$t_{b_1} = \frac{b_1}{m_{b_1}} \quad (2.21)$$

Данное отношение называется t – *статистикой* и имеет распределение Стьюдента с числом степеней свободы $k = n - 2$, где n – количество наблюдений.

Из таблицы распределения Стьюдента с $(n - 2)$ степенями свободы по заданному уровню значимости α выбирается критическое (табличное) значение $t_{1-\alpha;n-2}$.

Сравниваются фактическое (t_{b_1}) и табличное ($t_{1-\alpha;n-2}$) значения t – статистики:

- 1) если $|t_{b_1}| > t_{1-\alpha;n-2}$, то гипотезу H_0 следует отклонить и, следовательно, признать параметр b_1 *статистически значимым* на уровне α .
- 2) если $|t_{b_1}| < t_{1-\alpha;n-2}$, то гипотезу H_0 следует принять и признать параметр b_1 *статистически незначимым*.

Правило проверки значимости параметра b_0 . С целью оценки статистической значимости параметра b_0 определяется его *стандартная ошибка* по формуле

$$m_{b_0} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} \cdot \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (2.22)$$

Рассчитывается статистика:

$$t_{b_0} = \frac{b_0}{m_{b_0}}, \quad (2.23)$$

которая распределена по закону Стьюдента с $(n - 2)$ степенями свободы.

Из таблицы распределения Стьюдента с $(n - 2)$ степенями свободы по заданному уровню значимости α выбирается критическое значение $t_{1-\alpha;n-2}$. Тогда:

1) если $\left| t_{b_0} \right| > t_{1-\alpha;n-2}$, то гипотезу H_0 следует отклонить и, следовательно, признать параметр b_0 статистически значимым на уровне α .

2) если $\left| t_{b_0} \right| < t_{1-\alpha;n-2}$, то гипотезу H_0 следует принять и признать параметр b_0 статистически незначимым.

Замечание. Для парной регрессии более важным является анализ статистической значимости коэффициента b_1 , так как именно в нем скрыто влияние объясняющей переменной X на зависимую переменную Y .

Величины стандартных ошибок m_{b_0} и m_{b_1} совместно с t – статистикой Стьюдента применяются для интервальной оценки параметров линейного уравнения регрессии. Формулы для расчета доверительных интервалов параметров имеют следующий вид:

$$[b_1 - t_{1-\alpha;n-2} m_{b_1}; \quad b_1 + t_{1-\alpha;n-2} m_{b_1}] \quad (2.24)$$

$$[b_0 - t_{1-\alpha;n-2} m_{b_0}; \quad b_0 + t_{1-\alpha;n-2} m_{b_0}] \quad (2.25)$$

Фактически доверительный интервал определяет значение теоретических коэффициентов регрессии β_0 и β_1 , которые будут приемлемы с надежностью $(1 - \alpha)$ при найденных оценках b_0 и b_1 .

► **Пример 2.3.** По данным таблицы 2.1: 1) оценить на уровне $\alpha = 0,05$ значимость параметров уравнения регрессии Y по X ;

2) рассчитать 95%-е доверительные интервалы для параметров регрессии.

Решение. 1) Составим вспомогательную таблицу 2.4 с учетом того, что $\bar{x} = 7,455$, а значения $\hat{y}$ определяются по полученному в примере 2.1 уравнению регрессии: $\hat{y} = 8,93 + 1,056x$.

Выше (в примере 2.1) была найдена $\sum_{i=1}^{20} x_i^2 = 1157,35$. По формулам (2.20) – (2.23) вычислим стандартные ошибки параметров и фактические значения t – статистик:

$$m_{b_1} = \sqrt{\frac{23,524/(20-2)}{45,81}} = 0,1689; \quad t_{b_1} = \frac{1,056}{0,1689} = 6,25;$$

$$m_{b_0} = \sqrt{\frac{23,524}{20-2} \cdot \frac{1157,35}{20 \cdot 45,81}} = 1,2849; \quad t_{b_0} = \frac{8,93}{1,2849} = 6,95.$$

По таблицам t -распределения (таблица II приложений) $t_{0,95;18} = 2,10$. Так как $|t_{b_1}| > t_{0,95;18}$, то коэффициент регрессии b_1 статистически значим. Поскольку $|t_{b_0}| > t_{0,95;18}$, то параметр b_0 также значим на уровне $\alpha = 0,05$.

Таблица 2.4 – Расчетные величины по уравнению регрессии

i	x_i	y_i	$(x_i - \bar{x})^2$	$\hat{y}_i = 8,93 + 1,056x_i$	$(y_i - \hat{y}_i)^2$
1	5,0	13,8	6,027	14,210	0,168
2	5,5	14,0	3,822	14,738	0,545
3	5,6	14,3	3,441	14,844	0,296
4	6,5	14,8	0,912	15,794	0,988
5	7,0	15,1	0,207	16,322	1,493
6	8,0	15,6	0,297	17,378	3,161
7	7,0	15,8	0,207	16,322	0,272
8	7,3	16,1	0,024	16,639	0,290
9	6,5	16,4	0,912	15,794	0,367
10	7,5	16,6	0,002	16,850	0,063
11	8,3	16,8	0,714	17,695	0,801
12	8,5	17,0	1,092	17,906	0,821
13	6,0	17,2	2,117	15,266	3,740
14	7,5	17,6	0,002	16,850	0,563
15	8,5	17,9	1,092	17,906	0,000
16	6,0	18,1	2,117	15,266	8,032
17	8,0	18,3	0,297	17,378	0,850
18	9,5	19,5	4,182	18,962	0,289
19	10,0	20,3	6,477	19,490	0,656
20	10,9	20,8	11,868	20,440	0,129
Сумма	149,1	336,0	45,81	336,0	23,52

2) По формулам (2.24) – (2.25) найдем 95%-е доверительные интервалы:

для параметра b_1 : $[1,056 - 2,1 \cdot 0,1689; 1,056 + 2,1 \cdot 0,1689] = [0,70; 1,41]$;

для параметра b_0 : $[8,93 - 2,1 \cdot 1,2849; 8,93 + 2,1 \cdot 1,2849] = [6,23; 11,63]$. ►

2.6 Оценка значимости уравнения регрессии

Проверить *значимость уравнения регрессии* – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, наблюдаемым процессам и явлениям и достаточно ли включенных в уравнение объясняющих переменных для описания зависимой переменной.

Проверка значимости уравнения регрессии проводится на основе *дисперсионного анализа*, согласно основной идее которого общую сумму квадратов отклонений значений зависимой переменной от среднего по выборке можно разложить на две части – объясненную регрессией и остаточную (необъясненную):

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.26)$$

где $\sum_{i=1}^n (y_i - \bar{y})^2$ – общая сумма квадратов отклонений зависимой переменной от среднего по выборке;

$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ – сумма квадратов, объясненная регрессией;

$\sum_{i=1}^n (y_i - \hat{y}_i)^2$ – остаточная сумма квадратов, характеризующая влияние неучтенных факторов.

Очевидно, что если сумма квадратов отклонений, обусловленная регрессией, будет больше остаточной суммы квадратов, то уравнение регрессии статистически значимо, и объясняющая переменная X оказывает существенное воздействие на зависимую переменную Y .

Для сравнения указанных сумм необходимо рассчитать дисперсию на одну степень свободы, путем деления каждой суммы квадратов на соответствующее ей число степеней свободы:

$$S_R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{m}; \quad S^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m - 1},$$

где n – число наблюдений; m – число параметров при объясняющих переменных в уравнении регрессии (в случае парной линейной регрессии $m = 1$).

Выдвигаются нулевая H_0 и альтернативная H_1 гипотезы:

H_0 : (объясненная дисперсия) = (остаточная дисперсия).

H_1 : (объясненная дисперсия) > (остаточная дисперсия).

Чтобы отвергнуть гипотезу H_0 и подтвердить значимость уравнения регрессии в целом, необходимо, чтобы объясненная дисперсия превышала остаточную в несколько раз.

Из соотношения значений объясненной и остаточной дисперсий, рассчитанных на одну степень свободы, строится F – статистика

$$F = \frac{S_R^2}{S^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / m}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - m - 1)} \quad (2.27)$$

При выполнении предпосылок регрессионного анализа построенная F -статистика имеет распределение Фишера с числом степеней свободы $k_1 = m$, $k_2 = n - m - 1$. Английским статистиком Снедекором разработаны таблицы критических точек распределения Фишера при разных уровнях значимости и различных степенях свободы. Эти таблицы содержат максимальные значения отношения дисперсий, при которых нулевая гипотеза подтверждается.

Для проверки гипотезы H_0 выполняется сравнение фактического и определенного на уровне значимости α при k_1 , k_2 степенных свободы табличного $F_{\alpha; k_1; k_2}$ значений F -статистики Фишера.

Если $F > F_{\alpha; k_1; k_2}$, то H_0 отклоняется с пользу H_1 . Это означает, что объясненная дисперсия существенно больше остаточной дисперсии, и, следовательно, уравнение регрессии *статистически значимо*

и достаточно качественно отражает динамику изменения зависимой переменной.

Если $F < F_{\alpha; k_1; k_2}$, то нет оснований для отклонения H_0 . Значит, объясненная дисперсия соизмерима с дисперсией, вызванной случайными факторами. Это дает основание считать, что влияние объясняющей (объясняющих) переменной модели несущественно, а общее качество модели невысоко, т. е. признается *статистическая незначимость*, ненадежность уравнения регрессии.

Схему дисперсионного анализа можно представить в виде таблицы 2.5.

Таблица 2.5 – Дисперсионный анализ результатов регрессии

Компоненты дисперсии	Сумма квадратов отклонений	Число степеней свободы	Дисперсия на одну степень свободы	F-статистика	
				фактическое значение	табличное значение
Объясненная	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	m	$S_R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{m}$	$F = \frac{S_R^2}{S^2}$	$F_{\alpha; k_1; k_2}$
Остаточная	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$n - m - 1$	$S^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m - 1}$		
Общая	$\sum_{i=1}^n (y_i - \bar{y})^2$	$n - 1$	—		

► **Пример 2.4.** По данным таблицы 2.1 оценить на уровне $\alpha = 0,05$ значимость уравнения Y по X .

Решение. Выше, в примерах 2.1, 2.3 были найдены:

$$\sum_{i=1}^{20} y_i = 336; \quad \sum_{i=1}^{20} y_i^2 = 5719,44; \quad \sum_{i=1}^{20} (y_i - \hat{y}_i)^2 = 23,52$$

Вычислим необходимые суммы квадратов.

При расчете общей суммы квадратов полезно иметь в виду, что

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i \right)^2}{n}$$

Получим:

$$\sum_{i=1}^{20} (y_i - \bar{y})^2 = 5719,44 - \frac{336^2}{20} = 74,64;$$

$$\sum_{i=1}^{20} (\hat{y}_i - \bar{y})^2 = 74,64 - 23,52 = 51,12.$$

По формуле (2.27) $F = \frac{51,12}{23,52/18} = 39,11$

По таблице F -распределения (таблица I приложений) $F_{0,05;1;18}=4,41$. Так как $F > F_{0,05;1;18}$, то уравнение регрессии значимо. ►

2.7 Проверка общего качества уравнения регрессии. Коэффициент детерминации

Одной из эффективных оценок адекватности регрессионной модели, мерой качества уравнения регрессии является *коэффициент детерминации*, определяемый по формуле

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.28)$$

Величина R^2 показывает, какая часть (доля) вариации зависимой переменной обусловлена вариацией объясняющей переменной.

Коэффициент детерминации имеет следующие свойства:

Свойство 1. R^2 принимает значения от 0 до 1 включительно. Чем ближе R^2 к единице, тем лучше регрессия соответствует наблюдаемым данным. Если R^2 не меньше 0,5, то уравнение регрессии может считаться приемлемым. Регрессионные модели с R^2 выше 0,8 признаются хорошими.

Свойство 2. Если $R^2 = 1$, то между переменными X и Y существует функциональная зависимость. В этом случае все эмпирические точки наблюдений $(x_i; y_i)$ лежат на линии регрессии.

Свойство 3. Если $R^2 = 0$, то вариация зависимой переменной полностью обусловлена воздействием неучтенных в модели переменных.

В случае парной линейной регрессии коэффициент детерминации равен квадрату коэффициента корреляции, т. е.

$$R^2 = r_{xy}^2 \quad (2.29)$$

Если известно значение R^2 , то критерий значимости уравнения регрессии или самого коэффициента детерминации может быть записан в виде

$$F = \frac{R^2(n-m-1)}{(1-R^2)m} > F_{\alpha; k_1; k_2} \quad (2.30)$$

где $F_{\alpha; k_1; k_2}$ – табличное значение F –статистики Фишера, определенное на уровне значимости α при $k_1 = m$ и $k_2 = n - m - 1$ степенях свободы.

► **Пример 2.5.** По данным таблицы 2.1 найти коэффициент детерминации и пояснить его смысл.

Решение. *1-й способ.* В примере 2.4 было получено

$$\sum_{i=1}^{20} (\hat{y}_i - \bar{y})^2 = 51,12; \quad \sum_{i=1}^{20} (y_i - \bar{y})^2 = 74,64$$

По формуле (2.28) $R^2 = \frac{51,12}{74,64} = 0,68.$

2-й способ. Учитывая, что в примере 2.2 был вычислен коэффициент корреляции $r_{xy} = 0,83$, то по формуле (2.29) получим $R^2 = 0,83^2 = 0,68.$

Таким образом, вариация зависимой переменной Y – урожайности зерновых культур – на 68% объясняется изменчивостью объясняющей переменной X – количеством удобрения, внесенным на 1 га посева. ►

Оценка качества соответствия уравнения регрессии фактическим данным может проводиться с помощью *средней ошибки аппроксимации*, которая вычисляется по формуле

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\% \quad (2.31)$$

Значение этой ошибки не более 8 – 10% свидетельствует о хорошем подборе уравнения к исходным данным.

2.8 Интервалы прогноза по линейному уравнению регрессии

Одной из задач эконометрического моделирования является предсказание (прогнозирование) значений зависимой переменной при определенных значениях объясняющих переменных. Здесь возможен двоякий подход: либо предсказать условное математическое ожидание зависимой переменной при определенных значениях объясняющих переменных (предсказание среднего значения), либо прогнозировать некоторое конкретное значение зависимой переменной.

Пусть построено уравнение парной регрессии $\hat{y} = b_0 + b_1x$, на основе которого необходимо предсказать условное математическое ожидание $M(Y / X = x_p)$ переменной Y при $X = x_p$. В данном случае значение $\hat{y}_p = b_0 + b_1x_p$ является оценкой $M(Y / X = x_p)$. Для определения того, как сильно может уклониться модельное среднее значение $\hat{y}_p$, рассчитанное по выборочному уравнению регрессии, от соответствующего условного математического ожидания, необходимо получить интервальную оценку, построенную с заданной надежностью $(1 - \alpha)$ при конкретном значении x_p объясняющей переменной.

Для построения доверительного интервала вычисляется *стандартная ошибка* $m_{\hat{y}_p}$ групповой средней $\hat{y}$ по формуле

$$m_{\hat{y}_p} = \sqrt{S^2 \left(\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)} = S \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (2.32)$$

где S^2 – выборочная остаточная дисперсия на одну степень свободы;

$S = \sqrt{S^2}$ – стандартная ошибка регрессии.

Доверительный интервал для условного математического ожидания $M(Y / X = x_p)$ имеет вид

$$[\hat{y}_p - t_{1-\alpha; n-2} m_{\hat{y}_p}; \quad \hat{y}_p + t_{1-\alpha; n-2} m_{\hat{y}_p}] \quad (2.33)$$

где $t_{1-\alpha; n-2}$ табличное значение t -критерия Стьюдента с $(n - 2)$ степенями свободы и уровнем значимости α .

Из формул (2.32) и (2.33) видно, что длина доверительного интервала зависит от значения объясняющей переменной X : при $x_p = \bar{x}$ она минимальна, а по мере удаления x_p от $\bar{x}$ величина доверительного интервала увеличивается, т. е. чем больше разность между x_p и $\bar{x}$, тем больше ошибка, с которой предсказывается среднее значение Y при заданном значении x_p . Таким образом, прогноз значений зависимой переменной Y по уравнению регрессии оправдан, если значение x_p объясняющей переменной не выходит за диапазон наблюдаемых значений (причем тем более точный, чем ближе x_p к $\bar{x}$). Использование кривой регрессии вне пределов обследованного диапазона значений объясняющей переменной может привести к значительным погрешностям.

Построенная доверительная область для $M(Y / X = x_p)$ определяет местоположение модельной линии регрессии, но не отдельных возможных значений зависимой переменной, которые отклоняются от средней. Поэтому при определении доверительного интервала для индивидуальных значений y_p^* зависимой переменной необходимо учитывать еще один источник вариации – рассеяние вокруг линии регрессии. Индивидуальные значения могут отклоняться на величину случайной ошибки ε , дисперсия которой оценивается как остаточная дисперсия на одну степень свободы S^2 . Поэтому ошибка предсказываемого индивидуального значения m_{y_p} должна включать кроме $m_{\hat{y}_p}$ случайную ошибку S :

$$m_{y_p} = S \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (2.34)$$

Доверительный интервал для прогнозов индивидуальных значений y_p^* будет определяться по формуле

$$[\hat{y}_p - t_{1-\alpha; n-2} m_{y_p}; \quad \hat{y}_p + t_{1-\alpha; n-2} m_{y_p}] \quad (2.35)$$

Заметим, что данный интервал шире доверительного интервала для условного математического ожидания.

► **Пример 2.6.** По данным таблицы 2.1: 1) спрогнозировать урожайность зерновых культур при внесении на 1 га посева 9 т органи-

ческих удобрений; 2) построить 95%-ый доверительный интервал для данного прогноза.

Решение. 1) Уравнение регрессии Y по X было получено в примере 2.1: $\hat{y} = 8,93 + 1,056x$. По уравнению регрессии выполним прогноз урожайности, если количество внесенного удобрения на 1 га посева составит $x_p = 9$ (т):

$$\hat{y}_p = 8,93 + 1,056 \cdot 9 = 18,4 \text{ (ц/га)}.$$

2) В примере 2.3 были найдены:

$$\sum_{i=1}^{20} (y_i - \hat{y}_i)^2 = 23,52; \quad \sum_{i=1}^{20} (x_i - \bar{x})^2 = 45,81$$

По формуле (2.19) имеем: $S^2 = \frac{23,52}{20-2} = 1,307$; $S = \sqrt{1,307} = 1,143$.

С учетом того, что $\bar{x} = 7,455$, вычислим по (2.34) ошибку прогноза:

$$m_{y_p} = 1,143 \sqrt{1 + \frac{1}{20} + \frac{(9 - 7,455)^2}{45,81}} = 1,2 \text{ (ц/га)}$$

По таблице II приложений $t_{0,95;18} = 2,10$. Теперь по (2.35) получим искомый доверительный интервал $[18,4 - 2,10 \cdot 1,2; 18,4 + 2,10 \cdot 1,2]$ или $[15,9; 20,9]$.

Таким образом, при внесении на 1 га посева 9 т органических удобрений урожайность зерновых культур с вероятностью 0,95 будет находиться в пределах от 15,9 до 20,9 ц/га. ►

При прогнозировании на основе регрессионного уравнения, и особенно парного, следует соблюдать следующие ограничения:

- должно выполняться условие стабильности или, по крайней мере, малой изменчивости других факторов и условий изучаемого процесса. Если резко изменяется «внешняя среда» протекающего процесса, прежнее уравнение регрессии потеряет свое значение. Например, в засушливый год доза удобрений может не оказать влияния на урожайность сельскохозяйственной культуры, так как последнюю лимитирует недостаточная влагообеспеченность.

- нельзя подставлять значения фактора, значительно отличающиеся от входящих в базисную информацию, по которой получено уравнение регрессии.

2.9 Парная нелинейная регрессия

При изучении социально-экономических явлений исследователь часто сталкивается с соотношениями, которые нельзя выразить линейной функцией, так как при этом могут возникнуть неоправданно большие ошибки. Например, увеличение количества выпавших осадков приводит (при прочих равных условиях) к увеличению урожайности до некоторого предела, после чего урожайность будет снижаться.

Нелинейные соотношения между экономическими явлениями выражаются с помощью соответствующих нелинейных функций.

Нелинейные регрессии делятся на два класса:

- регрессии, *нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам*:

- полиномы различных степеней: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$;

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon \text{ и т. д.};$$

- гиперболическая функция – $Y = \beta_0 + \frac{\beta_1}{X} + \varepsilon$

- регрессии, *нелинейные по оцениваемым параметрам*:

- степенная функция $Y = \beta_0 X^{\beta_1} \varepsilon$;

- показательная функция $Y = \beta_0 \beta_1^X \varepsilon$;

- экспоненциальная функция $Y = e^{\beta_0 + \beta_1 X} \varepsilon$.

Выбор вида уравнения регрессии производится на основании опыта предыдущих исследований, других соображений профессионально-теоретического характера, а также визуального наблюдения расположения точек корреляционного поля.

Для оценки параметров нелинейных моделей используется подход, основанный на *линеаризации* модели, и заключается в том, что с помощью преобразований исходных переменных исследуемую зависимость представляют в виде линейного соотношения между преобразованными переменными. Затем для оценки параметров полученной линейной модели применяют обычный метод наименьших квадратов. Нелинейность по переменным устраняется путем *замены переменных*. Нелинейность по параметру часто устраняется путем *логарифмического преобразования* уравнения.

В случае, когда нелинейную модель невозможно привести к линейному виду, используются итеративные процедуры. Однако модели такого вида достаточно редко применяются в эконометрике.

► **Пример 2.7.** Изучается зависимость между надоем молока на одну корову и затратами кормов на одну голову в год по 16 сельскохозяйственным предприятиям.

Таблица 2.6 – Исходные данные: надой молока на 1 корову Y (л) и затраты кормовых единиц на 1 корову в год X (ц)

i	x_i	y_i	i	x_i	y_i
1	31,2	1711	9	37,3	2556
2	38,4	2722	10	32,4	1869
3	32,4	2025	11	33,1	2160
4	33,0	2158	12	37,6	2595
5	32,0	1831	13	37,4	2490
6	38,3	2603	14	36,1	2126
7	38,8	2848	15	37,9	2599
8	37,4	2395	16	35,5	2162

По данным таблицы 2.6 рассчитать параметры уравнений:

- 1) гиперболической парной регрессии;
- 2) степенной парной регрессии.

Решение. 1) Нелинейное уравнение гиперболической парной регрессии вида $\hat{y} = b_0 + b_1 \frac{1}{x}$ после замены переменной $z = \frac{1}{x}$ становится линейным: $\hat{y} = b_0 + b_1 z$. Параметры полученного линейного уравнения найдем обычным методом наименьших квадратов. Для расчетов используем данные таблицы 2.7.

По формуле (2.8):

$$b_1 = \frac{\overline{zy} - \bar{z} \bar{y}}{\overline{z^2} - \bar{z}^2} = \frac{64,47939461 - 0,02828782 \cdot 2303,1}{0,00080482 - 0,02828782^2} \approx -145114,4;$$

$$b_0 = \bar{y} - b_1 \bar{z} = 2303,1 - (-145114,4) \cdot 0,02828782 = 6408,1.$$

Итак, уравнение гиперболической парной регрессии имеет вид:

$$\hat{y} = 6408,1 - \frac{145114,4}{x}.$$

Таблица 2.7 – Гиперболическая зависимость надоя молока на 1 корову Y и затрат кормовых единиц на 1 корову в год X

i	x_i	y_i	$z_i = \frac{1}{x_i}$	$z_i y_i$	z_i^2
1	31,2	1711	0,03205128	54,83974359	0,00102728
2	38,4	2722	0,02604167	70,88541667	0,00067817
3	32,4	2025	0,03086420	62,50000000	0,00095260
4	33,0	2158	0,03030303	65,39393939	0,00091827
5	32,0	1831	0,03125000	57,21875000	0,00097656
6	38,3	2603	0,02610966	67,96344648	0,00068171
7	38,8	2848	0,02577320	73,40206186	0,00066426
8	37,4	2395	0,02673797	64,03743316	0,00071492
9	37,3	2556	0,02680965	68,52546917	0,00071876
10	32,4	1869	0,03086420	57,68518519	0,00095260
11	33,1	2160	0,03021148	65,25679758	0,00091273
12	37,6	2595	0,02659574	69,01595745	0,00070733
13	37,4	2490	0,02673797	66,57754011	0,00071492
14	36,1	2126	0,02770083	58,89196676	0,00076734
15	37,9	2599	0,02638522	68,57519789	0,00069618
16	35,5	2162	0,02816901	60,90140845	0,00079349
Сумма	568,8	36850	0,45260511	1031,67031373	0,01287713
Среднее	35,55	2303,1	0,02828782	64,47939461	0,00080482

2) При построении степенного уравнения регрессии $\hat{y} = b_0 x^{b_1}$ следует иметь в виду, что оно нелинейно относительно параметра b_1 . Путем логарифмирования уравнение может быть приведено к линейному виду

$$\lg y = \lg b_0 + b_1 \lg x \text{ или } Y = C + b_1 X,$$

где $Y = \lg y$; $X = \lg x$; $C = \lg b_0$.

Параметры b_1 и C линейного уравнения находим по формулам:

$$C = \bar{Y} - b_1 \bar{X}; \quad b_1 = \frac{\overline{XY} - \bar{X} \bar{Y}}{\overline{X^2} - \bar{X}^2}$$

Используя данные таблицы 2.8, рассчитаем:

$$b_1 = \frac{5,20505 - 1,54963 \cdot 3,35761}{2,40242 - 1,54963^2} = 1,87;$$

$$C = 3,35761 - 1,87 \cdot 1,54963 = 0,46.$$

Таблица 2.8 – Степенная зависимость надоя молока на 1 корову Y и затрат кормовых единиц на 1 корову в год X

i	x_i	y_i	$X_i = \lg x_i$	$Y_i = \lg y_i$	$X_i Y_i$	X_i^2
1	31,2	1711	1,49415	3,23325	4,83098	2,23250
2	38,4	2722	1,58433	3,43489	5,44200	2,51011
3	32,4	2025	1,51055	3,30643	4,99450	2,28175
4	33,0	2158	1,51851	3,33405	5,06280	2,30588
5	32,0	1831	1,50515	3,26269	4,91084	2,26548
6	38,3	2603	1,58320	3,41547	5,40737	2,50652
7	38,8	2848	1,58883	3,45454	5,48868	2,52439
8	37,4	2395	1,57287	3,37931	5,31521	2,47393
9	37,3	2556	1,57171	3,40756	5,35569	2,47027
10	32,4	1869	1,51055	3,27161	4,94191	2,28175
11	33,1	2160	1,51983	3,33445	5,06780	2,30988
12	37,6	2595	1,57519	3,41414	5,37791	2,48122
13	37,4	2490	1,57287	3,39620	5,34179	2,47393
14	36,1	2126	1,55751	3,32756	5,18270	2,42583
15	37,9	2599	1,57864	3,41481	5,39075	2,49210
16	35,5	2162	1,55023	3,33486	5,16979	2,40321
Сумма	568,8	36850	24,79411	53,72181	83,28072	38,43871
Среднее	35,55	2303,1	1,54963	3,35761	5,20505	2,40242

Получим линейное уравнение $Y = 0,46 + 1,87 X$.

Выполнив его потенцирование, получим уравнение степенной парной регрессии $\hat{y} = 10^{0,46} \cdot x^{1,87} = 2,9x^{1,87}$. ►

В экономическом анализе часто используется эластичность функции. *Эластичность функции* $y = f(x)$ рассчитывается как относительное изменение Y к относительному изменению X , т. е.

$$\mathcal{E} = \left(\frac{dy}{y} \right) / \left(\frac{dx}{x} \right) = f'(x) \frac{x}{y} \quad (2.36)$$

Эластичность показывает, на сколько процентов изменяется функция $y = f(x)$ при изменении независимой переменной на 1%.

В силу того, что эластичность почти всех функций не является постоянной величиной и зависит от значения x , обычно рассчитывается средний показатель эластичности по формуле

$$\bar{\mathcal{E}} = f'(x) \frac{\bar{x}}{\bar{y}} \quad (2.37)$$

Замечание. Для степенной функции $y = b_0 x^{b_1}$ эластичность представляет собой постоянную величину равную b_1 .

2.10 Коэффициент ранговой корреляции Спирмена

На практике встречаются случаи, когда необходимо установить тесноту связи между *порядковыми* переменными (например, качество жилищных условий, тестовые баллы, экзаменационные оценки и т.п.). В этом случае объекты анализа *ранжируют* (упорядочивают) по степени выраженности измеряемых переменных. При этом каждому объекту присваивается определенный номер, называемый *рангом*. Например, объекту с наименьшим значением признака присваивается ранг 1, следующему за ним – ранг 2 и т. д. Если объекты ранжированы по двум признакам, то имеется возможность оценить тесноту связи между переменными, основываясь на рангах, т. е. тесноту *ранговой корреляции*.

Коэффициент ранговой корреляции Спирмена находится по формуле

$$\rho = 1 - \frac{6 \sum_{i=1}^n (r_i - s_i)^2}{n(n^2 - 1)} \quad (2.38)$$

где r_i и s_i – ранги i -го объекта по переменным X и Y ;
 n – число пар наблюдений.

Если ранги всех объектов равны ($r_i = s_i$, $i = 1, 2, \dots, n$), то $\rho = 1$, т.е. при полной прямой связи $\rho = 1$. При полной обратной связи, когда ранги объектов по двум переменным расположены в обратном порядке, можно показать, что $\rho = -1$. Во всех остальных случаях $|\rho| < 1$.

При ранжировании бывают случаи, когда невозможно найти существенные различия между объектами по величине проявления рассматриваемого признака: объекты, как говорят, оказываются *связанными*. Связанным объектам приписывают одинаковые средние ранги, такие, чтобы сумма всех рангов оставалась такой же, как и при отсутствии связанных рангов. Например, если четыре объекта оказались равнозначными в отношении рассматриваемого признака и невозможно определить, какие из четырех рангов (4, 5, 6, 7) припи-

сать этим объектам, то каждому объекту приписывается средний ранг, равный $(4 + 5 + 6 + 7) / 4 = 5,5$.

При проверке значимости ρ исходят из того, что в случае справедливости нулевой гипотезы об отсутствии корреляционной связи между переменными статистика

$$t = \frac{\rho \sqrt{n-2}}{\sqrt{1-\rho^2}} \quad (2.39)$$

имеет t -распределение Стьюдента с $(n-2)$ степенями свободы. Поэтому коэффициент ρ значим на уровне α , если $|t| > t_{1-\alpha; n-2}$, где $t_{1-\alpha; n-2}$ – табличное значение t – критерия Стьюдента, определенное на уровне значимости α при числе степеней свободы $(n-2)$.

► **Пример 2.5.** При приеме на работу семи кандидатам было предложено два теста. Результаты тестирования приведены в таблице 2.9. Вычислить коэффициент ранговой корреляции Спирмена между результатами тестирования по двум тестам и проверить его значимость на уровне $\alpha=0,05$.

Таблица 2.9 – Результаты тестирования кандидатов (в баллах)

Тест	1-й	2-й	3-й	4-й	5-й	6-й	7-й
1	31	82	25	26	53	30	29
2	21	55	8	27	32	42	26

Решение. Представим исходные данные и расчетные показатели в виде таблицы.

Таблица 2.10 – Расчет коэффициента корреляции Спирмена

№ п/п	тест 1	тест 2	ранг r_i	ранг s_i	$(r_i - s_i)^2$
1	31	21	3	6	9
2	82	55	1	1	0
3	25	8	7	7	0
4	26	27	6	4	4
5	53	32	2	3	1
6	30	42	4	2	4
7	29	26	5	5	0
Сумма	–	–	–	–	18

По формуле (2.38) $\rho = 1 - \frac{6 \cdot 18}{7^3 - 7} = 0,68$.

Для проверки значимости ρ по формуле (2.39) вычислим

$$t = 0,68 \frac{\sqrt{7-2}}{\sqrt{1-0,68^2}} = 2,1$$

Так как $|t| < t_{0,95; 5} = 2,57$, то коэффициент ранговой корреляции ρ незначим на 5%-ном уровне. ►

Замечание. Ранговый коэффициент корреляции ρ может быть использован и для оценки тесноты связи между обычными количественными переменными. Достоинство ρ здесь заключается в том, что нахождение этого коэффициента не требует нормального распределения переменных, линейной связи между ними.

Задачи для самостоятельного решения

2.6. Данные наблюдений за случайными величинами X и Y представлены следующими таблицами:

a)

X	1	2	3	4	5
Y	0	2	3	5	6

б)

X	1	3	5	7	9
Y	10	7	8	5	3

в)

X	0	1	2	3	4	5	6
Y	9	4	1	0	1	4	9

Постройте корреляционное поле; вычислите коэффициент корреляции; сделайте вывод о линейной зависимости между переменными (о силе и направлении).

2.7. По выборке объемом 10 наблюдений получены следующие результаты: $\sigma_x^2 = 34,39$; $\sigma_y^2 = 66,93$; $\sum x_i = 100$; $\sum y_i = 200$; $\sum x_i y_i = 21000$; $\sum x_i^2 = 12000$; $\sum y_i^2 = 45000$. Оцените коэффициент корреляции r_{xy} . Проверьте его значимость.

2.8. Исследуется зависимость между количеством (N) покупателей в ювелирном магазине и количеством (Q) проданных товаров. По наблюдениям получены следующие данные:

N	50	61	72	43	60	65	76	55	62	40
Q	10	12	20	9	15	15	21	14	18	7

Оцените наличие и степень линейной зависимости между N и Q . Будет ли коэффициент корреляции статистически значимым?

2.9. По выборке объемом 10 наблюдений получены следующие результаты: $\sum x_i = 100$; $\sum y_i = 200$; $\sum x_i y_i = 21000$; $\sum x_i^2 = 12000$; $\sum y_i^2 = 45000$. Оцените по методу наименьших квадратов коэффициенты парной линейной регрессии Y на X .

2.10. Вычислите коэффициенты выборочной парной линейной регрессии, если известно, что $\bar{x} = 21$; $\bar{y} = 112,45$; $\sigma_x^2 = 64$; $\sigma_y^2 = 225$; $r_{xy} = 0,8$.

2.11. Вычислите для парной линейной регрессии значения коэффициентов детерминации R^2 и корреляции r_{xy} , если известно, что $n = 100$; $\sum e_i^2 = 1100$; $\sum (y_i - \bar{y})^2 = 22500$. Проверьте значимость r_{xy} .

2.12. Имеются данные об уровне механизации работ X (%) и производительности труда Y (т/ч) для 14 однотипных предприятий:

X	32	30	36	40	41	47	56	54	60	55	61	67	69	76
Y	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Оцените тесноту и направление связи между переменными с помощью коэффициента корреляции. Найдите уравнение линейной регрессии Y по X .

2.13. При исследовании корреляционной зависимости между ценой на нефть X и индексом нефтяных компаний Y получены следующие данные:

$$\bar{x} = 16,2(\text{ден.ед.}), \quad \bar{y} = 4000(\text{усл.ед.}), \quad \sigma_x^2 = 4, \quad \sigma_y^2 = 500, \quad \text{cov}(X, Y) = 40.$$

Составьте уравнение линейной регрессии Y по X .

2.14. В следующей выборке представлены данные по цене P некоторого блага и количеству Q данного блага, приобретаемого домохозяйствами.

P	10	20	15	25	30	35	40	35	25	40	45	40
Q	110	75	100	80	60	55	40	80	60	30	40	30

Постройте корреляционное поле и по его виду определить формулу зависимости между P и Q . Оцените параметры уравнения линейной регрессии. Рассчитайте выборочный коэффициент корреляции. Проинтерпретируйте результаты.

2.15. Даны регрессии, рассчитанные по 25 годовым наблюдениям:

а) $y_t = -30 + 0,18x_t$ (y_t – расходы на оплату жилья, x_t – доход);

б) $y_t = 50 + 4,5t$ (y_t – расходы на оплату жилья, t – время, год).
 Дайте экономическую интерпретацию построенных регрессий.
 Согласуются ли они друг с другом?

2.16. Проводится анализ взаимосвязи количества населения X (млн. чел.) и количества практикующих врачей Y (тыс. чел.).

X	10,0	10,3	10,4	10,55	10,6	10,7	10,75	10,9	10,9	11,0
Y	12,1	12,6	13,0	13,8	14,9	16,0	18,0	20,0	21,0	22,0

Рассчитайте коэффициент корреляции. Оцените коэффициенты линейного уравнения регрессии $Y = a + bX$.

2.17. По выборке объемом 12 наблюдений получены следующие результаты:

$$\bar{x} = 122,167; \quad \bar{y} = 125,25; \quad \sum (x_i - \bar{x})^2 = 2135,679; \quad \sum (y_i - \bar{y})^2 = 2216,168; \\ \sum (x_i - \bar{x})(y_i - \bar{y}) = 2115.$$

Рассчитайте параметры линейного уравнения регрессии. Найдите коэффициент детерминации и поясните его смысл. Проверьте значимость построенного уравнения регрессии на 5%-ом уровне с помощью F – критерия Фишера.

2.18. Вычислите для парной линейной регрессии значения коэффициентов детерминации R^2 и корреляции r_{xy} , если известно, что $n = 30$; $\sum e_i^2 = 103627,8$; $\bar{y} = 406,7$, $\bar{x} = 10,6$; $\sigma_x = 4,34$; $\sigma_y = 120$. $\sum (y_i - \bar{y})^2 = 417700,5$; $\sum (x_i - \bar{x})^2 = 547,2$. Проверьте значимость r_{xy} . Вычислите коэффициенты выборочной парной линейной регрессии. Проверьте статистическую значимость коэффициента b_1 и постройте для него доверительный интервал. Постройте точечный прогноз для значения $x_p = 17$ и доверительный интервал прогноза.

2.19. Зависимость между сменной добычей угля на одного рабочего Y (т) и мощностью пласта X (м) по данным, характеризующим добычу угля в 10 шахтах, дана в таблице:

X	8	11	12	9	8	8	9	9	8	12
Y	5	10	10	7	5	6	6	5	6	8

Найдите линейное уравнение регрессии Y по X . Оцените качество уравнения регрессии с помощью коэффициента детерминации и F – критерия Фишера. Проверьте статистическую значимость коэффи-

циентов регрессии при уровне значимости $\alpha=0,05$. Рассчитайте 95% доверительные интервалы для коэффициентов регрессии.

2.20. По территории некоторой области известны данные по двум показателям:

Район	Потребительские расходы на душу населения, усл. ед, Y	Средняя заработная плата и выплаты социального характера, усл. ед, X
1	464	912
2	524	809
3	298	748
4	351	847
5	624	1087
6	584	1074
7	425	1008

1) Постройте поле корреляции, сформулируйте гипотезу о форме связи между изучаемыми показателями.

2) Рассчитайте параметры уравнений линейной, степенной, экспоненциальной, гиперболической парных регрессий.

3) Для каждого уравнения рассчитайте средний коэффициент эластичности.

4) Оцените качество уравнений с помощью коэффициента детерминации.

5) Оцените статистическую значимость уравнений регрессии.

6) Выберите лучшее уравнение регрессии и дайте обоснование.

2.21. Изучается зависимость материалоемкости продукции от размера предприятия по 10 однородным заводам.

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

Рассчитайте параметры уравнения гиперболической парной регрессии. Охарактеризуйте эластичность изменения материалоемкости продукции. Сделайте вывод о значимости уравнения регрессии.

Контрольные вопросы

1. В чем состоит отличие теоретического и выборочного уравнений регрессии?
2. Назовите основные причины присутствия в регрессионной модели случайного возмущения.
3. В чем суть метода наименьших квадратов?
4. Докажите справедливость формул вычисления МНК-оценок параметров парной линейной регрессии.
5. Дайте интерпретацию уравнению регрессии $\hat{y} = 3 + 2x$, где Y – объём продукции (в млн. руб.), X – объём инвестиций в ИТ-технологии (в сотнях тыс. долларов).
6. В чем состоят основные предпосылки регрессивного анализа?
7. Как вычисляется выборочный коэффициент корреляции?
8. Сформулируйте основные свойства коэффициента корреляции.
9. Как оценивается значимость коэффициента корреляции?
10. Как связан коэффициент линейной регрессии с выборочным коэффициентом корреляции?
11. Объясните суть коэффициента детерминации.
12. Перечислите свойства коэффициента детерминации.
13. Как связаны между собой коэффициент парной корреляции и коэффициент детерминации?
14. Как определяются стандартные ошибки регрессии и коэффициентов регрессии?
15. Сформулируйте правила проверки статистической значимости коэффициентов парной линейной регрессии.
16. В чем состоит идея проверки статистической значимости уравнения регрессии? Сформулируйте правило проверки.
17. Назовите модели, нелинейные относительно включаемых переменных; оцениваемых параметров.
18. Как определяется коэффициент эластичности?
19. Опишите схему построения нелинейной регрессии.

3 МНОЖЕСТВЕННЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

На любой экономический показатель чаще всего оказывает влияние не один, а несколько факторов. В связи с этим возникает задача исследования зависимости одной объясняемой переменной Y от нескольких объясняющих переменных $X_1, X_2, \dots, X_p$. Эта задача решается с помощью множественного регрессионного анализа.

Уравнение множественной регрессии может быть представлено в виде

$$Y = f(\beta, X) + \varepsilon \quad (3.1)$$

где Y – зависимая (объясняемая) переменная;
 $X = (X_1, X_2, \dots, X_p)$ – вектор объясняющих переменных;
 β – вектор параметров, подлежащих определению;
 ε – случайный член (возмущение).

Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

Наиболее употребляемой и простой моделью множественной регрессии является модель множественной линейной регрессии.

3.1 Множественная линейная регрессия

Множественная линейная регрессия с p объясняющими переменными имеет вид

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon, \quad (3.2)$$

где $\beta_0, \beta_1, \dots, \beta_p$ – параметры модели.

Предполагается, что объясняющие переменные некоррелированы друг с другом, и случайный член ε удовлетворяет предпосылкам регрессионного анализа.

На основе n наблюдений оценивается выборочное уравнение регрессии

$$\hat{y} = b_0 + b_1 x_1 + \dots + b_p x_p, \quad (3.3)$$

где $b_0, b_1, \dots, b_p$ – оценки параметров $\beta_0, \beta_1, \dots, \beta_p$.

Для оценки параметров регрессии используется МНК, согласно которому минимизируется сумма квадратов отклонений фактических значений зависимой переменной от ее значений, полученных по уравнению регрессии.

Обозначим i -ое наблюдение зависимой переменной y_i , а объясняющих переменных – $x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip}$ ($i = 1, 2, \dots, n$), тогда в соответствии с МНК:

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_p x_{ip})^2 \rightarrow \min \quad (3.4)$$

Необходимым условием минимума функции нескольких переменных $S(b_0, b_1, \dots, b_p)$, является равенство нулю всех ее частных производных по $b_0, b_1, \dots, b_p$. Приравнявая к нулю частные производные, приходим к системе из $(p + 1)$ линейного уравнения с $(p + 1)$ неизвестным, называемой *системой нормальных уравнений*:

$$\left\{ \begin{array}{l} \frac{\partial S}{\partial b_0} = \sum_{i=1}^n 2(y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_p x_{ip}) \cdot (-1) = 0, \\ \frac{\partial S}{\partial b_1} = \sum_{i=1}^n 2(y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_p x_{ip}) \cdot (-x_{i1}) = 0, \\ \dots \\ \frac{\partial S}{\partial b_p} = \sum_{i=1}^n 2(y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_p x_{ip}) \cdot (-x_{ip}) = 0. \end{array} \right. \quad (3.5)$$

Преобразуя (3.5), получим систему в виде:

[illegible]

Решение системы (3.6) может быть найдено с помощью метода Крамера:

$$b_j = \frac{\Delta b_j}{\Delta}, \quad j = 0, 1, 2, \dots, p, \quad (3.7)$$

где Δ – основной определитель системы линейных уравнений,
 Δb_j – частые определители.

При этом

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \dots & \sum_{i=1}^n x_{ip} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i2} x_{i1} & \dots & \sum_{i=1}^n x_{ip} x_{i1} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^n x_{ip} & \sum_{i=1}^n x_{i1} x_{ip} & \sum_{i=1}^n x_{i2} x_{ip} & \dots & \sum_{i=1}^n x_{ip}^2 \end{vmatrix},$$

а определители Δb_j получаются из основного определителя путем замены j -го столбца на столбец свободных членов:

$$\begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n y_i x_{i1} \\ \dots \\ \sum_{i=1}^n y_i x_{ip} \end{pmatrix}.$$

► **Пример 3.1.** Имеются данные (условные) о сменной добыче угля на одного рабочего Y (т), мощности пласта X_1 (м) и уровне механизации работ X_2 (%), характеризующие процесс добычи угля в 10 шахтах (таблица 3.1).

Найти аналитическое выражение линейной корреляционной зависимости между переменными Y , X_1 и X_2 (уравнение регрессии Y по X_1 и X_2).

Таблица 3.1 – Сменная добыча угля на одного рабочего Y (т), мощность пласта X_1 (м) и уровень механизации работ X_2 (%)

i	x_{i1}	x_{i2}	y_i	i	x_{i1}	x_{i2}	y_i
1	8	5	5	6	8	8	6
2	11	8	10	7	9	6	6
3	12	8	10	8	9	4	5
4	9	5	7	9	8	5	6
5	8	7	5	10	12	7	8

Решение. Запишем систему нормальных уравнений (3.6) для двух объясняющих переменных:

$$\begin{cases} b_0 n + b_1 \sum_{i=1}^n x_{i1} + b_2 \sum_{i=1}^n x_{i2} = \sum_{i=1}^n y_i, \\ b_0 \sum_{i=1}^n x_{i1} + b_1 \sum_{i=1}^n x_{i1}^2 + b_2 \sum_{i=1}^n x_{i2} x_{i1} = \sum_{i=1}^n y_i x_{i1}, \\ b_0 \sum_{i=1}^n x_{i2} + b_1 \sum_{i=1}^n x_{i1} x_{i2} + b_2 \sum_{i=1}^n x_{i2}^2 = \sum_{i=1}^n y_i x_{i2}. \end{cases}$$

Для удобства вычислений составим вспомогательную таблицу.

Таблица 3.2 – Расчетная таблица для линейной корреляционной зависимости между переменными Y , X_1 и X_2

i	x_{i1}	x_{i2}	y_i	x_{i1}^2	x_{i2}^2	$x_{i1}x_{i2}$	$y_i x_{i1}$	$y_i x_{i2}$
1	8	5	5	64	25	40	40	25
2	11	8	10	121	64	88	110	80
3	12	8	10	144	64	96	120	80
4	9	5	7	81	25	45	63	35
5	8	7	5	64	49	56	40	35
6	8	8	6	64	64	64	48	48
7	9	6	6	81	36	54	54	36
8	9	4	5	81	16	36	45	20
9	8	5	6	64	25	40	48	30
10	12	7	8	144	49	84	96	56
Сумма	94	63	68	908	417	603	664	445

Получим систему нормальных уравнений:

$$\begin{cases} 10b_0 + 94b_1 + 63b_2 = 68, \\ 94b_0 + 908b_1 + 603b_2 = 664, \\ 63b_0 + 603b_1 + 417b_2 = 445. \end{cases}$$

Найдем определители:

$$\Delta = \begin{vmatrix} 10 & 94 & 63 \\ 94 & 908 & 603 \\ 63 & 603 & 417 \end{vmatrix} = 3738; \quad \Delta b_0 = \begin{vmatrix} 68 & 94 & 63 \\ 664 & 908 & 603 \\ 445 & 603 & 417 \end{vmatrix} = -13230;$$

$$\Delta b_1 = \begin{vmatrix} 10 & 68 & 63 \\ 94 & 664 & 603 \\ 63 & 445 & 417 \end{vmatrix} = 3192; \quad \Delta b_2 = \begin{vmatrix} 10 & 94 & 68 \\ 94 & 908 & 664 \\ 63 & 603 & 445 \end{vmatrix} = 1372.$$

По формуле (3.7) вычислим коэффициенты уравнения регрессии

$$b_0 = \frac{-13230}{3738} = -3,54; \quad b_1 = \frac{3192}{3738} = 0,85; \quad b_2 = \frac{1372}{3738} = 0,37.$$

Получим уравнение множественной линейной регрессии вида

$$\hat{y} = -3,54 + 0,85x_1 + 0,37x_2.$$

Уравнение показывает, что при увеличении только мощности пласта X_1 (при неизменном значении уровня механизации работ X_2) на 1 м добыча угля на одного рабочего Y увеличивается в среднем на 0,85 т, а при увеличении только уровня механизации работ X_2 (при неизменном значении мощности пласта X_1) – в среднем на 0,37 т. ►

Коэффициенты регрессии $b_1, b_2, \dots, b_p$ характеризуют среднее изменение зависимой переменной Y с изменением соответствующей объясняющей переменной на единицу при неизменном значении других переменных, закрепленных на среднем уровне.

Одна из задач анализа уравнения множественной регрессии заключается в проведении сравнения силы влияния различных факторов на зависимую переменную. Если объясняющие переменные не-соизмеримы (имеют разные единицы измерения), то выводы, осно-

ванные на простом сравнении величин коэффициентов $b_1, b_2, \dots, b_p$, будут ошибочны.

Для корректного сравнения силы влияния объясняющих переменных на зависимую переменную можно использовать *стандартизованных коэффициентов регрессии* b'_j и *средние коэффициенты эластичности* $\bar{\mathcal{E}}_j$ ($j=1, 2, \dots, p$):

$$b'_j = b_j \frac{\sigma_{x_j}}{\sigma_y} \quad (3.8)$$

$$\bar{\mathcal{E}}_j = b_j \frac{\bar{x}_j}{\bar{y}} \quad (3.9)$$

Стандартизованный коэффициент регрессии b'_j показывает, на сколько величин σ_y изменится в среднем зависимая переменная Y при увеличении только j -й объясняющей переменной на одну σ_{x_j} , а коэффициент эластичности $\bar{\mathcal{E}}_j$ – на сколько процентов (от средней) изменится в среднем Y при увеличении только X_j на 1%.

Считается, что более сильное влияние на зависимую переменную оказывает та переменная, которой соответствует бóльший по абсолютной величине стандартизованный коэффициент или коэффициент эластичности.

► **Пример 3.2.** По данным примера 3.1 сравнить отдельное влияние на сменную добычу угля двух факторов – мощности пласта и уровня механизации работ.

Решение. Для сравнения влияния объясняющих переменных по формуле (3.8) вычислим стандартизованные коэффициенты регрессии, а по формуле (3.9) – коэффициенты эластичности:

$$b'_1 = 0,85 \cdot \frac{1,56}{1,83} = 0,72; \quad b'_2 = 0,37 \cdot \frac{1,42}{1,83} = 0,29,$$

$$\bar{\mathcal{E}}_1 = 0,85 \cdot \frac{9,4}{6,8} = 1,18; \quad \bar{\mathcal{E}}_2 = 0,37 \cdot \frac{6,3}{6,8} = 0,34,$$

где $\bar{x}_1 = 9,4$; $\bar{x}_2 = 6,3$; $\bar{y} = 6,8$; $\sigma_{x_1} = 1,56$; $\sigma_{x_2} = 1,42$; $\sigma_y = 1,83$.

Таким образом, увеличение мощности пласта и уровня механизации работ только на одно σ_{x_1} или на одно σ_{x_2} увеличивает в среднем сменную добычу угля на одного рабочего соответственно на $0,72\sigma_y$ или на $0,29\sigma_y$, а увеличение этих переменных на 1% (от своих средних значений) приводит в среднем к росту добычи угля соответственно на 1,18% и 0,34%. Итак, по обоим показателям на сменную добычу угля большее влияние оказывает фактор «мощность пласта» по сравнению с фактором «уровень механизации работ». ►

3.2 Множественная корреляция

Тесноту связи всего набора объясняющих переменных $X_1, X_2, \dots, X_p$ с исследуемой зависимой переменной Y характеризует *индекс множественной корреляции*. Независимо от формы связи индекс множественной корреляции может быть найден по формуле

$$R_{yx_1x_2\dots x_p} = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (3.10)$$

Значение индекса множественной корреляции лежит в пределах от 0 до 1. Чем ближе его значение к 1, тем теснее связь зависимой переменной со всем набором исследуемых объясняющих переменных.

Для линейной регрессии $\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$ индекс множественной корреляции может быть определен с помощью формулы *линейного коэффициента множественной корреляции*:

$$R_{yx_1x_2\dots x_p} = \sqrt{b'_1r_{yx_1} + b'_2r_{yx_2} + \dots + b'_pr_{yx_p}} = \sqrt{\sum_{j=1}^p b'_j r_{yx_j}} \quad (3.11)$$

где b'_j – стандартизированные коэффициенты регрессии;

r_{yx_j} – парные коэффициенты корреляции зависимой переменной Y с каждой объясняющей переменной.

Для проверки общего качества уравнения регрессии используется *коэффициент детерминации* R^2 , который рассчитывается по формуле

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (3.12)$$

R^2 определяет долю общего разброса значений зависимой переменной Y , объясненного уравнением регрессии. В общем случае $0 \leq R^2 \leq 1$. Чем ближе этот коэффициент к единице, тем лучше уравнение объясняет поведение Y .

Недостатком коэффициента R^2 является то, что он обычно увеличивается при добавлении новых объясняющих переменных, хотя это не всегда означает улучшение качества уравнения регрессии. Для компенсации такого увеличения R^2 вводится *скорректированный коэффициент детерминации* $\bar{R}^2$ с поправкой на число степеней свободы:

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - m - 1)}{\sum_{i=1}^n (y_i - \bar{y})^2 / (n - 1)}, \quad (3.13)$$

где m – число параметров при объясняющих переменных.

В линейной зависимости величина m совпадает с количеством объясняющих переменных p , включенных в уравнение регрессии.

Соотношение (3.13) может быть представлено в виде

$$\bar{R}^2 = 1 - \left(1 - R^2\right) \frac{n - 1}{n - m - 1} \quad (3.14)$$

Из формулы (3.14) следует, что чем больше величина m , тем сильнее различия между R^2 и $\bar{R}^2$.

В отличие от R^2 скорректированный коэффициент $\bar{R}^2$ может уменьшаться при введении в модель новых объясняющих переменных, не оказывающих существенного влияния на зависимую переменную Y . Поэтому добавление в модель новых объясняющих переменных осуществляется до тех пор, пока растет скорректированный коэффициент детерминации $\bar{R}^2$. При слабых связях зависимой переменной с объясняющими переменными величина $\bar{R}^2$ может стать отрицательной. В этом случае он должен считаться равным нулю.

3.3 Оценка статистической значимости коэффициентов уравнения множественной линейной регрессии

Статистическая значимость коэффициентов b_j ($j = 1, 2, \dots, p$) множественной линейной регрессии с p объясняющими переменными проверяется на основе t – статистики вида

$$t_{b_j} = \frac{b_j}{m_{b_j}}, \quad (3.15)$$

имеющей распределение Стьюдента с числом степеней свободы $k = n - m - 1$.

Величина m_{b_j} является *стандартной ошибкой коэффициента* регрессии b_j и может быть определена по формуле

$$m_{b_j} = \frac{\sigma_y \sqrt{1 - R_{yx_1 \dots x_p}^2}}{\sigma_{x_j} \sqrt{1 - R_{x_j x_1 \dots x_p}^2}} \cdot \frac{1}{\sqrt{n - m - 1}}, \quad (3.16)$$

где σ_y – среднее квадратическое отклонение для Y ;

σ_{x_j} – среднее квадратическое отклонение для X_j ;

$R_{yx_1 \dots x_p}^2$ – коэффициент детерминации для уравнения регрессии;

$R_{x_j x_1 \dots x_p}^2$ – коэффициент детерминации для зависимости X_j со всеми другими объясняющими переменными уравнения множественной регрессии.

При требуемом уровне значимости α фактическое значение t_{b_j} сравнивается с табличным значением $(t_{1-\alpha; n-m-1})$ t – критерия Стьюдента, определенным при числе степеней свободы $k = n - m - 1$.

1) Если $|t_{b_j}| > t_{1-\alpha; n-m-1}$, то коэффициент b_j считается *статистически значимым*.

2) Если $|t_{b_j}| < t_{1-\alpha; n-m-1}$, то коэффициент b_j считается *статистически незначимым*. Это означает, что переменная X_j линейно не связана с зависимой переменной Y . Не оказывая серьезного влияния на зависимую переменную, она лишь искажает реальную картину взаимосвязи. Поэтому рекомендуется исключить переменную X_j из

уравнения регрессии. Это не приведет к существенной потере качества уравнения регрессии, но сделает его более конкретным.

Для коэффициентов уравнения регрессии можно построить доверительные интервалы с заданным уровнем доверия, в качестве которого обычно выбирают вероятность 0,95, соответствующую уровню значимости 0,05. Доверительный интервал для коэффициента b_j определяется по формуле

$$[b_j - m_{b_j} t_{1-\alpha; n-m-1}; b_j + m_{b_j} t_{1-\alpha; n-m-1}] \quad (3.17)$$

3.4 Оценка значимости множественной регрессии

Значимость уравнения множественной регрессии в целом оценивается с помощью F – критерия Фишера. При этом выдвигается гипотеза об одновременном равенстве нулю всех коэффициентов регрессии при объясняющих переменных:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

Если данная гипотеза не отклоняется, то делается вывод о том, что совокупное влияние всех p объясняющих переменных $X_1, X_2, \dots, X_p$ на зависимую переменную Y можно считать статистически несущественным, а общее качество уравнения регрессии невысоким. Принятие нулевой гипотезы равнозначно также статистической незначимости коэффициента множественной детерминации R^2 .

Проверка гипотезы H_0 осуществляется на основе дисперсионного анализа сравнения объясненной и остаточной дисперсий. Для чего строится F – статистика

$$F = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / m}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - m - 1)}, \quad (3.18)$$

которая имеет распределение Фишера с числами степеней свободы $k_1 = m$ и $k_2 = n - m - 1$.

Если $F > F_{\alpha; k_1; k_2}$ ($F_{\alpha; k_1; k_2}$ – табличное значение F – критерия Фишера), то H_0 отклоняется, и уравнение регрессии считается статистически значимым.

Если $F < F_{\alpha; k_1; k_2}$, то нет оснований для отклонения H_0 , следовательно, построенное уравнение регрессии признается *статистически незначимым*.

Критерий значимости уравнения регрессии (3.18) может быть записан в виде

$$F = \frac{R^2 (n - m - 1)}{(1 - R^2) m} > F_{\alpha; k_1; k_2}. \quad (3.19)$$

При оценивании адекватности уравнения регрессии фактически существующей зависимости возможны следующие ситуации:

1) Модель по F -критерию адекватна, все коэффициенты уравнения регрессии статистически значимы. Такая модель может быть использована для принятия решений и осуществления прогнозов.

2) Модель по F -критерию адекватна, но часть коэффициентов уравнения регрессии статистически незначима. Модель пригодна для принятия некоторых решений, но не для прогнозирования.

3) Модель по F -критерию адекватна, но все коэффициенты уравнения регрессии статистически незначимы. Модель в целом считается неадекватной, на ее основе не принимаются решения, и не осуществляются прогнозы.

3.5 Оценка уравнения линейной регрессии в MS Excel

Для построения базовых моделей регрессионного анализа хорошо подходит пакет MS Excel. Для расчета параметров уравнения линейной регрессии методом наименьших квадратов и проверки его адекватности исследуемому процессу служит инструмент анализа данных **Регрессия**.

► **Пример 3.3.** Изучается влияние стоимости основных и оборотных средств на величину прибыли предприятий. Данные по 15 предприятиям приведены в таблице, сформированной на рабочем листе MS Excel (рис. 3.1). Необходимо определить параметры уравнения линейной регрессии и провести его анализ.

	A	B	C	D
1	Номер предприятия	Прибыль Y , млн. руб.	Стоимость основных фондов X_1 , млн. руб.	Величина оборотных средств X_2 , млн. руб.
2	1	188	510	129
3	2	78	190	64
4	3	93	240	69
5	4	152	470	87
6	5	55	110	47
7	6	161	420	102
8	7	135	405	68
9	8	81	171	66
10	9	66	130	54
11	10	123	312	95
12	11	172	487	130
13	12	145	436	78
14	13	103	225	72
15	14	87	180	65
16	15	149	463	91

Рисунок 3.1 – Данные о прибыли предприятий Y , стоимости основных фондов X_1 и величине оборотных средств X_2

Решение. Для решения используем режим работы **Регрессия**, вызов которого выполняется по команде: **Данные / Анализ данных / Регрессия**.

В диалоговом окне данного режима задаются следующие параметры:

Входной интервал Y – вводится ссылка на ячейки, содержащие данные по зависимой переменной. Диапазон должен состоять из одного столбца.

Входной интервал X – вводится ссылка на ячейки, содержащие факторные признаки.

Метки – устанавливается флажок, если первая строка во входном диапазоне содержит заголовки.

Уровень надежности – устанавливается флажок в активное состояние, если в поле, расположенное напротив флажка, необходимо ввести уровень надежности, отличный от уровня 95%, применяемого по умолчанию.

Константа-ноль – устанавливается флажок, если требуется построить уравнение без свободного члена b_0 .

Выходной интервал – вводится ссылка на левую верхнюю ячейку выходного диапазона.

Новый рабочий лист – устанавливается переключатель, чтобы открыть новый лист, в который, начиная с ячейки A1, вставляются результаты анализа.

Новая рабочая книга – устанавливается переключатель, чтобы открыть новую книгу, на первом листе которой, начиная с ячейки A1, вставляются результаты анализа.

Остатки – устанавливается флажок, чтобы включить остатки в выходной диапазон.

Стандартизированные остатки – устанавливается флажок, чтобы включить стандартизированные остатки в выходной диапазон.

График остатков – устанавливается флажок, чтобы построить диаграмму остатков для каждой независимой переменной.

График подбора – устанавливается флажок, чтобы построить диаграммы наблюдаемых и предсказанных значений для каждой независимой переменной.

График нормальной вероятности – устанавливается флажок, чтобы построить диаграмму нормальной вероятности.

Значения параметров, установленных в диалоговом окне Регрессия, представлены на рис. 3.2.

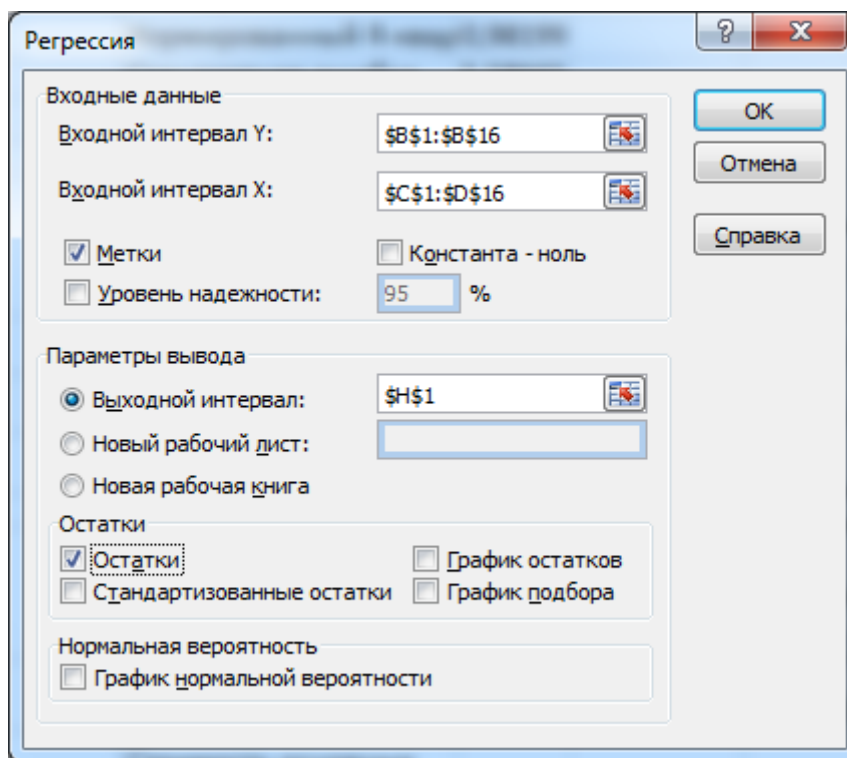


Рисунок 3.2 – Диалоговое окно ввода параметров режима Регрессия

Показатели регрессионного анализа, рассчитанные в данном режиме, представлены на рис. 3.3 – 3.6.

Результаты по регрессионной статистике (рис. 3.3) соответствуют следующим показателям:

- *Множественный R* – коэффициент корреляции;
- *R-квадрат* – коэффициент детерминации R^2 ;
- *Нормированный R-квадрат* – скорректированный коэффициент детерминации $\bar{R}^2$;
- *Стандартная ошибка* – остаточное стандартное отклонение, вычисляемое по формуле

$$S = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m - 1}};$$

- *Наблюдения* – число наблюдений.

	Н	І
1	ВЫВОД ИТОГОВ	
2		
3	Регрессионная статистика	
4	Множественный R	0,992
5	R-квадрат	0,985
6	Нормированный R-квадрат	0,982
7	Стандартная ошибка	5,587
8	Наблюдения	15

Рисунок 3.3 – Результаты по регрессионной статистике

В таблицу *Дисперсионный анализ* (рис. 3.4) включены все основные результаты расчетов, проводимых при проверке значимости регрессионной модели (§ 2.6). Столбцы таблицы имеют следующую интерпретацию:

- df – число степеней свободы;
- SS – сумма квадратов отклонений (для строки *Регрессия* – объясненная, для строки *Остаток* – остаточная, для строки *Итого* – общая).
- MS – дисперсии, рассчитанные на одну степень свободы;
- F – расчетное значение F -критерия Фишера;
- *Значимость F* – значение уровня значимости, соответствующее вычисленному значению F -критерия Фишера. Если *Значимость*

F будет меньше заданного уровня значимости, то уравнение регрессии считается статистически значимым.

	Н	І	Ј	К	Л	М
10	Дисперсионный анализ					
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
12	Регрессия	2	23881,873	11940,936	382,592	1,35501E-11
13	Остаток	12	374,527	31,211		
14	Итого	14	24256,400			

Рисунок 3.4 – Результаты дисперсионного анализа

В таблице на рис. 3.5 представлены значения коэффициентов регрессии и их статистические оценки. Столбцы таблицы имеют следующую интерпретацию:

- *Коэффициенты* – значения коэффициентов b_j ;
- *Стандартная ошибка* – стандартные ошибки коэффициентов m_{b_j} ;
- *t-статистика* – расчетные значения t – критерия.
- *P-значение* – значения уровней значимости, соответствующие вычисленным значениям t – критерия. Если *P-значение* окажется меньше заданного уровня значимости, то соответствующий коэффициент считается значимым.
- *Нижние 95%* и *Верхние 95%* – соответственно нижние и верхние границы доверительных интервалов для коэффициентов регрессии b_j .

	Н	І	Ј	К	Л	М	Н
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
17	У-пересечение	12,136	5,279	2,299	0,04029	0,633	23,639
18	Стоимость основных фондов X1, млн. руб.	0,214	0,018	11,852	5,551E-08	0,175	0,254
19	Величина оборотных средств X2, млн. руб.	0,483	0,106	4,569	0,00064	0,253	0,714

Рисунок 3.5 – Коэффициенты регрессии и их статистические оценки

В таблице на рис. 3.6 представлены теоретические значения $\hat{y}_i$ зависимой переменной Y и значения остатков. Последние вычисляют-

ся как разность между фактическими y_i и теоретическими $\hat{y}_i$ значениями зависимой переменной Y .

23	ВЫВОД ОСТАТКА		
24			
25	<i>Наблюдение</i>	<i>Предсказанная прибыль</i>	<i>Остатки</i>
26	1	183,7825	4,2175
27	2	83,7854	-5,7854
28	3	96,9186	-3,9186
29	4	154,9152	-2,9152
30	5	58,4238	-3,4238
31	6	151,4456	9,5544
32	7	131,8024	3,1976
33	8	80,6792	0,3208
34	9	66,0929	-0,0929
35	10	124,9141	-1,9141
36	11	179,3357	-7,3357
37	12	143,2789	1,7211
38	13	95,1529	7,8471
39	14	82,1252	4,8748
40	15	155,3475	-6,3475

Рисунок 3.6 – Теоретические значения Y и значения остатков

Перейдем к анализу сгенерированных таблиц.

Рассчитанные коэффициенты регрессии b_j позволяют построить уравнение множественной линейной регрессии, выражающее зависимость прибыли предприятий Y от стоимости основных фондов X_1 и величины оборотных средств X_2 :

$$\hat{y} = 12,136 + 0,214 x_1 + 0,483 x_2 .$$

Коэффициент b_1 при стоимости основных фондов положителен, что согласуется с теоретическим положением о росте прибыли с ростом стоимости основных средств. Увеличение только основных фондов на 1 млн. руб. ведет к росту прибыли в среднем на 0,214 млн. руб.

Коэффициент b_2 при величине оборотных средств положителен, что согласуется с теоретическим положением о росте прибыли с ростом оборотных средств. Увеличение оборотных средств на 1 млн. руб. ведет к росту прибыли в среднем на 0,483 млн. руб.

Коэффициент $b_0 = 12,136$ признается *статистически значимым*, так как $|t_{b_0}| = 2,299 > t_{0,95;12} = 2,179$. К тому же P -значение $= 0,0403$ меньше заданного уровня значимости $\alpha=0,05$, а доверительный интервал $[0,633; 23,639]$ не содержит нулевое значение.

Аналогичные рассуждения относительно коэффициентов b_1 и b_2 позволяют сделать вывод об их *статистической значимости*.

Значение коэффициента детерминации $R^2 = 0,985$ показывает, что 98,5% вариации зависимой переменной объясняется вариацией факторов X_1 и X_2 . Значит, включенные в уравнение регрессии факторы существенно влияют на прибыль.

Стандартная ошибка $S = 5,587$ мала по сравнению со средним значением зависимой переменной Y ($\bar{y} = 119,2$), что говорит о малом разбросе выборочных данных относительно регрессии.

Уравнение регрессии в целом статистически значимо, так как $F = 382,592 > F_{0,05; 2; 12} = 3,88$. К тому же *Значимость* $F = 1,355E-11$ меньше заданного уровня значимости $\alpha = 0,05$.

Таким образом, проверка статистической значимости построенного уравнения регрессии подтверждает его адекватность. ►

3.6 Мультиколлинеарность

Одной из проблем при построении множественной линейной регрессии по МНК является *мультиколлинеарность* – высокая коррелированность двух или нескольких объясняющих переменных в уравнении регрессии.

Мультиколлинеарность может привести к целому ряду отрицательных эффектов:

- большие стандартные ошибки оценок, что затрудняет нахождение истинных значений определяемых величин и расширяет интервальные оценки, ухудшая их точность;
- уменьшаются t -статистики коэффициентов, что может привести к неоправданному выводу о существенности влияния соответствующей объясняющей переменной на зависимую переменную;
- возможно получение неверного знака у коэффициента регрессии;

- оценки коэффициентов по МНК и их стандартные ошибки становятся очень чувствительны к малейшим изменениям данных, т.е. они становятся неустойчивыми;

- затрудняется определение вклада каждой из объясняющих переменных в объясняемую уравнением регрессии дисперсию зависимой переменной.

Для определения наличия или отсутствия мультиколлинеарности используют несколько подходов. Один из них заключается в анализе *корреляционной матрицы* между объясняющими переменными $X_1, X_2, \dots, X_p$ и выявлении пар переменных, имеющих высокие коэффициенты корреляции (обычно более 0,7). Если такие переменные существуют, то говорят о мультиколлинеарности между ними.

Другой подход состоит в исследовании множественных коэффициентов детерминации между одной из объясняющих переменных и некоторой группой из них. Наличие высокого множественного коэффициента детерминации (обычно более 0,6) свидетельствует о мультиколлинеарности.

Полезно также находить множественный коэффициент детерминации $R^2_{yx_1 \dots x_p}$. Если коэффициент детерминации достаточно высок, но при этом некоторые из коэффициентов регрессии статистически незначимы (имеют низкие t -статистики), то это говорит, скорее всего, о наличии мультиколлинеарности.

Для устранения или уменьшения мультиколлинеарности применяют ряд методов.

1. *Исключение переменной(ых) из модели*

Этот метод состоит в том, что из двух объясняющих переменных, имеющих высокий коэффициент корреляции (более 0,7), одну переменную исключают из рассмотрения. При этом, какую переменную оставить, а какую удалить из анализа, решают в первую очередь на основании экономических соображений. Если с экономической точки зрения ни одной из переменных нельзя отдать предпочтение, то оставляют ту из двух переменных, которая имеет больший коэффициент корреляции с зависимой переменной.

2. *Получение дополнительных данных или новой выборки*

Поскольку мультиколлинеарность напрямую зависит от выборки, то, возможно, при другой выборке мультиколлинеарности не будет, либо она не будет столь серьезной.

Иногда для уменьшения мультиколлинеарности достаточно увеличить объем выборки. Например, при использовании ежегодных данных можно перейти к поквартальным данным. Увеличение количества данных сокращает дисперсии коэффициентов регрессии и тем самым увеличивает их статистическую значимость. Однако получение новой выборки или расширение старой не всегда возможно или связано с серьезными издержками. Эти проблемы ограничивают возможность использования данного метода.

3. Изменение спецификации модели

Проблема мультиколлинеарности может быть решена путем изменения спецификации модели: либо изменяется форма модели, либо добавляются объясняющие переменные, не учтенные в первоначальной модели, но существенно влияющие на зависимую переменную. Если данный метод имеет основания, то его использование уменьшает сумму квадратов отклонений, тем самым сокращая стандартную ошибку регрессии, что приводит к уменьшению стандартных ошибок коэффициентов.

4. Преобразование переменных

В ряде случаев минимизировать либо вообще устранить проблему мультиколлинеарности можно с помощью преобразования переменных.

Например, пусть выборочное уравнение регрессии имеет вид

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2,$$

причем X_1 и X_2 – коррелированные переменные. В этой ситуации можно попытаться определить регрессионные зависимости относительных величин:

$$\frac{\hat{Y}}{X_1} = b_0 + b_1 \frac{X_2}{X_1}, \quad \frac{\hat{Y}}{X_2} = b_0 + b_1 \frac{X_1}{X_2} \quad (3.20)$$

Вполне вероятно, что в моделях, аналогичных (3.20), проблема мультиколлинеарности будет отсутствовать.

Возможны и другие преобразования. Например, если в уравнении рассматриваются взаимосвязи номинальных экономических показателей, то для снижения мультиколлинеарности можно попытаться перейти к реальным показателям.

3.7 Частная корреляция

Для оценки тесноты связи между двумя переменными часто используют выборочный коэффициент линейной корреляции (§ 2.3). Однако при изучении многомерных связей в ряде случаев коэффициенты линейной корреляции могут давать совершенно неверные представления о характере связи между переменными. Например, между двумя переменными X и Y может быть высокий положительный коэффициент корреляции не потому, что одна из них стимулирует изменение другой, а оттого, что обе эти переменные изменяются в одном направлении под влиянием других переменных, как учтенных в модели, так и, возможно, неучтенных. В связи с этим возникает необходимость исследовать частную корреляцию между переменными при исключении влияния одной или нескольких переменных.

Выборочным частным коэффициентом корреляции между переменными X_i и X_j при фиксированных значениях остальных $(p - 2)$ переменных называется выражение

$$r_{ij.12\dots p} = \frac{-q_{ij}}{\sqrt{q_{ii} q_{jj}}}, \quad (3.21)$$

где q_{ij} и q_{jj} – алгебраические дополнения элементов r_{ij} и r_{jj} матрицы выборочных коэффициентов корреляции

$$q_p = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix} \quad (3.22)$$

Коэффициент корреляции r_{ij} определяется по формуле (2.11), а алгебраическое дополнение q_{ij} элемента r_{ij} равно

$$q_{ij} = (-1)^{i+j} M_{ij}, \quad (3.23)$$

где M_{ij} – минор элемента r_{ij} – определитель матрицы $(p - 1)$ -го порядка, полученной из матрицы q_p вычеркиванием i -й строки и j -го столбца.

Частный коэффициент корреляции $r_{ij.12\dots p}$, как и парный коэффициент r_{ij} , может принимать значения на отрезке от -1 до $+1$.

В случае трех переменных из общей формулы (3.21) следует, что

$$r_{ij.k} = \frac{r_{ij} - r_{ik} r_{jk}}{\sqrt{(1 - r_{ik}^2)(1 - r_{jk}^2)}}. \quad (3.24)$$

► **Пример 3.4.** Для исследования зависимости между производительностью труда (X_1), возрастом (X_2) и производственным стажем (X_3) была произведена выборка из 100 рабочих одной и той же специальности. Вычисленные парные коэффициенты корреляции составили: $r_{12} = 0,20$; $r_{13} = 0,41$; $r_{23} = 0,82$. Вычислить частные коэффициенты корреляции.

Решение. По формуле (3.24) частные коэффициенты корреляции

$$r_{12.3} = \frac{0,20 - 0,41 \cdot 0,82}{\sqrt{(1 - 0,41^2)(1 - 0,82^2)}} = -0,26$$

и аналогично $r_{13.2} = 0,44$; $r_{23.1} = 0,83$.

Сравнивая частные коэффициенты корреляции $r_{ij.k}$ с соответствующими парными коэффициентами, видим, что за счет «очищения связи» наибольшему изменению подвергся коэффициент корреляции между производительностью труда (X_1) и возрастом (X_2) рабочих (изменилось не только его значение, но и знак: $r_{12} = 0,20$; $r_{12.3} = -0,26$). По значению $r_{12} = 0,20$ можно сделать вывод, что между производительностью труда (X_1) и возрастом (X_2) рабочих существует прямая корреляционная связь. Если же устранить влияние переменной «производственный стаж» (X_3), то в чистом виде производительность труда (X_1) находится в обратной по направлению (и слабой по тесноте) связи с возрастом рабочих (X_2) ($r_{12.3} = -0,26$). Это вполне объяснимо, если рассматривать возраст только как показатель работоспособности организма на определенном этапе его жизнедеятельности. Подобным образом могут быть интерпретированы и другие частные коэффициенты корреляции. ►

3.8 Фиктивные переменные в регрессионных моделях

В регрессионных моделях в качестве объясняющих переменных часто используют не только количественные переменные, но и качественные признаки, имеющие два или несколько уровней (градаций). К числу таких признаков можно отнести: пол (мужской, женский), образование (начальное, среднее, высшее), фактор сезонности (зима, весна, лето, осень) и т. п.

При исследовании влияния качественных признаков в модель можно вводить *фиктивные переменные*, принимающие, как правило, два значения: единица, если данный признак присутствует в наблюдении, и ноль при его отсутствии. Если включаемый в рассмотрение качественный признак имеет не два, а несколько уровней, то целесообразно использовать несколько фиктивных переменных. *Число вводимых фиктивных переменных должно быть на единицу меньше числа уровней качественного признака.*

При назначении фиктивных переменных исследуемая совокупность по числу уровней качественного признака разбивается на группы. Одну из групп выбирают как эталонную и определяют фиктивные переменные остальных. Например, при изучении зависимости объема потребления некоторого продукта Y от фактора сезонности в регрессионную модель можно ввести три фиктивные переменные Z_1, Z_2, Z_3 :

$$Z_1 = \begin{cases} 1, & \text{если месяц зимний,} \\ 0, & \text{в остальных случаях,} \end{cases}$$

$$Z_2 = \begin{cases} 1, & \text{если месяц весенний,} \\ 0, & \text{в остальных случаях,} \end{cases}$$

$$Z_3 = \begin{cases} 1, & \text{если месяц летний,} \\ 0, & \text{в остальных случаях.} \end{cases}$$

и оценивать уравнение: $Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \varepsilon$.

Оценки $\beta_1, \beta_2, \beta_3$ показывают средние сезонные отклонения в объеме потребления по отношению к осенним месяцам.

3.9 Нелинейные модели регрессии. Производственная функция Кобба-Дугласа

Во многих практических случаях моделирование экономических зависимостей линейными уравнениями дает вполне удовлетворительный результат и может использоваться для анализа и прогнозирования. Однако в силу многообразия и сложности экономических процессов ограничиться рассмотрением лишь линейных регрессионных моделей невозможно. Многие экономические зависимости не являются линейными по своей сути, и поэтому их моделирование линейными уравнениями регрессии, безусловно, не дает положительного результата. Например, нелинейными оказываются производственные функции (зависимости между объемом произведенной продукции и основными факторами производства – трудом, капиталом и т.п.), функции спроса (зависимость между спросом на товары и услуги и их ценами или доходом) и другие.

Широкую известность имеет *производственная функция Кобба-Дугласа*:

$$Y = AK^{\alpha}L^{\beta}, \quad (3.25)$$

где A, α, β – параметры функции;
 Y – объем производства;
 K – затраты капитала;
 L – затраты труда.

Показатели α и β являются коэффициентами эластичности объема производства Y по затратам капитала и труда соответственно. Это означает, что при увеличении только затрат капитала на 1% объем производства увеличится на α %, а увеличение одних только затрат труда на 1% приведет к росту объема производства на β %.

Сумма $(\alpha + \beta)$ является важным экономическим показателем и характеризует эффект от масштаба производства. Так, при росте затрат каждого из ресурсов K, L в λ раз объем производства возрастает в $\lambda^{(\alpha + \beta)}$ раз. Это означает следующее:

- если $\alpha + \beta = 1$, то говорят о постоянной отдаче от масштаба, т.е. во сколько раз увеличиваются затраты капитала и труда, во столько же раз возрастает объем производства;
- если $\alpha + \beta < 1$, то имеет место убывающая отдача от масштаба, т.е. увеличение объема производства меньше увеличения затрат ресурсов;

■ если $\alpha + \beta > 1$, то имеет место возрастающая отдача от масштаба, т.е. увеличение объема производства больше увеличения затрат ресурсов.

Учитывая влияние случайных возмущений, присущих каждому экономическому явлению, функцию Кобба-Дугласа можно представить в виде

$$Y = AK^\alpha L^\beta \varepsilon \quad (3.26)$$

Функция Кобба-Дугласа легко сводится к линейной путем логарифмирования обеих частей уравнения:

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L + \ln \varepsilon \quad (3.27)$$

При этом обычно предполагается, что ошибки $\ln \varepsilon$ обладают свойствами, необходимыми для оценивания линейной регрессионной модели.

Функция Кобба-Дугласа с учетом технического прогресса имеет вид

$$Y = AK^\alpha L^\beta e^{rt} \varepsilon, \quad (3.28)$$

где t – время;

r – темп прироста объема производства благодаря техническому прогрессу.

Модель (3.28) приводится к линейному виду аналогично модели (3.26).

Задачи для самостоятельного решения

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

$$\hat{y} = 21,1 - 6,2 x_1 + 0,95 x_2 + 3,57 x_3$$

(1,8) (0,54) (0,83)

где Y – цена объекта, тыс. дол.;

X_1 – расстояние до центра города, км;

X_2 – полезная площадь объекта, кв. м;

X_3 – число этажей в доме, ед.

В скобках указаны значения стандартных ошибок для коэффициентов множественной регрессии.

Коэффициент множественной корреляции равен 0,84.

Проверьте статистическую значимость полученного уравнения регрессии в целом и его параметров.

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

$$\hat{y} = 1,4 - 0,52 x_1 + 0,17 x_2 + 11,16 x_3 - 0,38 x_4 - 4,75 x_5, \quad (R^2=0,6),$$

(−5,90) (4,34) (3,91) (−0,79) (−2,71)

где Y – темпы экономического роста (темпы роста среднедушевого ВВП в % к базисному периоду);

X_1 – реальный среднедушевой ВВП, %;

X_2 – бюджетный дефицит, % к ВВП;

X_3 – объем инвестиций, % к ВВП;

X_4 – внешний доход, % к ВВП;

X_5 – уровень инфляции, %.

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

Проверьте статистическую значимость полученного уравнения регрессии в целом и его параметров.

3.7. По 72 наблюдениям построено уравнение множественной регрессии $\hat{y} = -12,23 + 0,91x_1 - 2,1x_2$, где Y – объем потребления (в сотнях руб.), X_1 – располагаемый доход (в тыс. руб.), X_2 – процент банковской ставки по вкладам. Известны значения t -статистик коэффициентов: $t_{b_1} = 123,7$; $t_{b_2} = -3,2$ и значение коэффициента детерминации $R^2 = 0,976$. Дайте интерпретацию коэффициентам регрессии. Проверьте статистическую значимость коэффициентов. Найдите стандартные ошибки коэффициентов. Проверьте статистическую значимость уравнения в целом по критерию Фишера.

3.8. По 20 предприятиям легкой промышленности получено уравнение регрессии, характеризующее зависимость объема выпуска продукции Y (млн. руб.) от количества отработанных за год человеко-часов X_1 (тыс. чел.-ч.) и среднегодовой стоимости производственного оборудования X_2 (млн. руб.): $\hat{y} = 35 + 0,06 x_1 + 2,5 x_2$. Значение множественного коэффициента корреляции 0,92. Известно, что сумма квадратов отклонений расчетных значений результата от фактических равна 3300. Определите коэффициент детерминации. Составьте таблицу дисперсионного анализа. Сделайте выводы.

3.8. Предполагается, что объем предложения Y некоторого товара для функционирующей в условиях конкуренции фирмы зависит линейно от цены X_1 данного товара и заработной платы X_2 сотрудников фирмы, производящих товар. Статистически данные по 16 наблюдениям приведены в таблице 3.3.

Таблица 3.3 – Исходные данные для задачи 3.8

i	y_i	x_{i1}	x_{i2}	i	y_i	x_{i1}	x_{i2}
1	20	10	12	9	105	38	4
2	35	15	10	10	110	55	5
3	30	20	9	11	120	50	3
4	45	25	9	12	130	35	1
5	60	40	8	13	130	40	2
6	69	37	8	14	130	55	3
7	75	43	6	15	135	45	1
8	90	35	4	16	140	65	2

Необходимо: а) оценить по МНК коэффициенты уравнения регрессии; б) проверить гипотезы о том, что при прочих равных условиях рост цены товара увеличивает предложение; рост заработной платы снижает предложение; в) оценить общее качество уравнения регрессии.

3.9. Постройте уравнение множественной линейной регрессии для выборки, представленной в таблице 3.4, оцените полученный результат.

Таблица 3.4 – Исходные данные для задачи 3.9

i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	y_i	i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	y_i
1	1	8	5	9	21	11	9	2	5	6	29
2	2	6	3	4	13	12	4	5	8	5	16
3	3	3	2	7	27	13	7	6	6	1	18
4	4	2	5	6	24	14	6	1	3	2	23
5	5	4	6	5	22	15	5	2	2	3	31
6	8	9	1	1	18	16	1	3	4	6	15
7	6	4	2	2	31	17	2	4	9	3	19
8	3	7	3	3	14	18	3	5	4	5	20
9	2	5	4	6	31	19	6	8	7	8	27
10	4	3	5	3	27	20	3	9	6	9	29

3.10. Изучается влияние стоимости основных и оборотных средств на величину валового дохода торговых предприятий. Для этого по 12 торговым предприятиям были получены данные, приведенные в таблице 3.5.

Таблица 3.5 – Исходные данные для задачи 3.10

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

Необходимо построить линейные уравнения множественной регрессии и пояснить экономический смысл его параметров. Оцените полученное уравнение на основе коэффициента детерминации и F -критерия Фишера.

3.11. Имеются данные о выработке литья на одного работающего X_1 (т), браке литья X_2 (%) и себестоимости 1 т литья Y (руб.) по 25 литейным цехам заводов.

Необходимо: а) найти уравнение множественной регрессии Y по X_1 и X_2 , оценить значимость этого уравнения и его коэффициентов на уровне $\alpha = 0,05$; б) определить коэффициент детерминации и пояснить его смысл; в) сравнить раздельное влияние на зависимую переменную каждой из объясняющих переменных, используя стандартизированные коэффициенты регрессии и коэффициенты эластичности.

Таблица 3.6 – Исходные данные для задачи 3.11

i	x_{i1}	x_{i2}	y_i	i	x_{i1}	x_{i2}	y_i
1	14,6	4,2	239	14	75,8	3,4	172
2	13,5	6,7	254	15	27,6	1,1	201
3	21,5	5,5	262	16	88,4	0,1	130
4	17,4	7,7	251	17	16,6	4,1	251
5	44,8	1,2	158	18	33,4	2,3	195
6	111,9	2,2	101	19	17,0	9,3	282
7	20,1	8,4	259	20	33,1	3,3	196
8	28,1	1,4	186	21	30,1	3,5	186
9	22,3	4,2	204	22	65,2	1,0	176
10	25,3	0,9	198	23	22,6	5,2	238
11	56,0	1,3	170	24	33,4	2,3	204
12	40,2	1,8	173	25	19,7	2,7	205
13	40,6	3,3	197				

3.12. Для уравнения регрессии $\hat{y} = -73,52 + 1,62x_1 - 2,25x_2$, отражающего зависимость среднего душевого дохода (руб.) от средней заработной платы в день (руб.) и среднего возраста (год), построенного по 30 наблюдениям, известны выборочные коэффициенты корреляции $r_{yx_1} = 0,84$; $r_{yx_2} = -0,21$; $r_{x_1x_2} = -0,116$. Дайте интерпретацию коэффициентам регрессии. Вычислите частные коэффициенты корреляции, сравните их с парными.

3.13. Вычислите частные коэффициенты корреляции между Y и факторами, проверьте их значимость, если известна матрица парных коэффициентов корреляции:

	y	x_1	x_2
y	1	0,87	0,84
x_1	0,87	1	0,71
x_2	0,84	0,71	1

3.14. Имеются данные о потреблении некоторого продукта Y (усл. ед.) в зависимости от уровня урбанизации (доли городского населения) X_1 , относительного образовательного уровня X_2 и относительного заработка X_3 для 9 районов.

Таблица 3.7 – Исходные данные для задачи 3.14

i	x_{i1}	x_{i2}	x_{i3}	y_i	i	x_{i1}	x_{i2}	x_{i3}	y_i
1	42,2	11,2	31,9	167,1	6	44,5	10,8	8,5	174,6
2	48,6	10,6	13,2	174,4	7	39,1	10,7	24,3	163,7
3	42,6	10,6	28,7	160,8	8	40,1	10,0	18,6	174,5
4	39,0	10,4	26,1	162,0	9	45,9	12,0	20,4	185,7
5	34,7	9,3	30,1	140,8					

Известна корреляционная матрица:

	X_1	X_2	X_3	Y
X_1	1			
X_2	0,684	1		
X_3	-0,616	-0,172	1	
Y	0,802	0,768	-0,629	1

Используя пошаговую процедуру отбора наиболее информативных объясняющих переменных, определите подходящую регрессионную модель, исключив при этом мультиколлинеарность.

3.15. Имеются следующие данные о весе Y (в фунтах) и возрасте X (в неделях) 13 индеек, выращенных в областях A , B , C .

Таблица 3.8 – Исходные данные для задачи 3.15

i	x_i	y_i	Область происхождения	i	x_i	y_i	Область происхождения
1	28	12,3	A	8	26	11,8	B
2	20	8,9	A	9	21	11,5	C
3	32	15,1	A	10	27	14,2	C
4	22	10,4	A	11	29	15,4	C
5	29	13,1	B	12	23	13,1	C
6	27	12,4	B	13	25	13,8	C
7	28	13,2	B				

Есть основание полагать, что на вес индеек оказывает влияние не только их возраст, но и область происхождения. Необходимо:

а) найти уравнение парной регрессии Y по X и оценить его значимость;

б) введя соответствующие фиктивные переменные, найти общее уравнение множественной регрессии Y по всем объясняющим переменным, включая фиктивные;

в) оценить значимость общего уравнения множественной регрессии по F -критерию и значимость его коэффициентов по t -критерию на уровне $\alpha = 0,05$;

г) проследить за изменением скорректированного коэффициента детерминации при переходе от парной к множественной регрессии;

д) получить из общего уравнения множественной регрессии уравнения регрессий для каждой области.

3.16. В таблице 3.9 приведены данные по объемам выпуска Y , затрат капитала K и труда L в некоторой отрасли за 20 лет.

Таблица 3.9 – Исходные данные для задачи 3.16

i	Y	K	L	i	Y	K	L
1	46000	2	2	11	70500	2	4
2	59000	5,6	2	12	70500	5,6	2
3	37500	2	4	13	108000	5,6	4
4	107000	5,6	4	14	90500	2	6
5	130000	2	6	15	74000	10,4	2
6	128000	10,4	2	16	160000	5,6	6
7	154000	5,6	6	17	225000	10,4	4
8	226500	10,4	4	18	167500	10,4	6
9	146500	10,4	6	19	88500	5,6	4
10	31500	2	2	20	54000	2	2

Оцените производственную функцию Кобба-Дугласа вида $Y = AK^\alpha L^\beta$.

Контрольные вопросы

1. Запишите модель множественной линейной регрессии и назовите ее основные компоненты
2. Что характеризуют коэффициенты множественной линейной регрессии?
3. Дайте интерпретацию уравнению регрессии $\hat{y} = 3 + 2x_1 - 4x_2$, где y – доход (в млн. руб.), X_1 – объем инвестиций в ин-

формационные технологии (в сотнях тысяч долларов), X_2 – объем заработной платы (в сотнях тысяч рублей).

4. Дайте определение коэффициентов эластичности. Каков смысл этих величин?
5. Какие коэффициенты используются для оценки сравнительной силы воздействия объясняющих переменных на зависимую переменную?
6. Приведите формулы расчета стандартных ошибок коэффициентов регрессии.
7. Сформулируйте правило проверки статистической значимости коэффициентов множественной линейной регрессии.
8. Как строятся интервальные оценки коэффициентов множественной линейной регрессии?
9. В чем суть коэффициента детерминации? Опишите свойства коэффициента детерминации.
10. Чем отличается обычный коэффициент детерминации от скорректированного коэффициента детерминации?
11. Каким образом можно оценить качество уравнения регрессии?
12. Как используется F – статистика в регрессионном анализе?
13. Приведите формулу для расчета частного коэффициента корреляции.
14. Чем отличается парный коэффициент корреляции от частного коэффициента корреляции?
15. Объясните значение термина «мультиколлинеарность»; опишите последствия наличия мультиколлинеарности на результат построения модели.
16. Как можно обнаружить мультиколлинеарность?
17. Назовите методы устранения мультиколлинеарности.
18. Каковы основные причины использования фиктивных переменных в регрессионном анализе?
19. В чем состоит основное правило для определения количества вводимых фиктивных переменных в регрессионную модель?
20. Дайте описание производственной функции Кобба-Дугласа.

4 ВРЕМЕННЫЕ РЯДЫ И ПРОГНОЗИРОВАНИЕ

При анализе многих экономических показателей часто используют ежегодные, ежеквартальные, ежемесячные, ежедневные данные. Например, это могут быть годовые данные по ВВП, инфляции, месячные данные по объему продажи продукции, ежедневные объемы выпуска какого-либо предприятия. Для рационального анализа необходимо упорядочить данные по времени их получения и построить так называемые *временные ряды*.

В этой главе мы остановимся на некоторых общих понятиях и вопросах, связанных с временными рядами, использованием регрессионных моделей временных рядов для прогнозирования. При анализе точности этих моделей и определении интервальных ошибок прогноза на их основе, будем полагать, что рассматриваемые в главе регрессионные модели временных рядов удовлетворяют условиям классической модели.

4.1 Общие сведения о временных рядах

Временной ряд (динамический ряд) – это последовательность наблюдений некоторого признака (случайной величины) Y в последовательные моменты времени. Отдельные наблюдения называются *уровнями ряда*, которые будем обозначать y_t ($t = 1, 2, \dots, n$), где n – число уровней.

В таблице 4.1 приведены данные, отражающие валовой сбор сахарной свеклы (фабричной) в хозяйствах некоторого района за семилетний период (тыс. т), т. е. временной ряд валового сбора y_t .

Таблица 4.1 – Динамика валового сбора сахарной свеклы (тыс. т)

Год, t	1	2	3	4	5	6	7
Валовой сбор, y_t	32,4	42,1	57,1	65,9	74,8	86,2	110,3

Каждый уровень y_t временного ряда формируется под совместным влиянием длительных, кратковременных и случайных факторов. Длительные, постоянно действующие факторы оказывают на изучаемое явление определяющее влияние и формируют основную тенденцию ряда – *тренд* T_t . Кратковременные, периодические факторы формируют циклические (сезонные) колебания ряда S_t . Случайные факторы отражаются случайными изменениями уровней ряда ε_t .

Компоненты T_t и S_t необязательно присущи каждому временному ряду. Например, могут быть временные ряды, в которых имеется тренд, но отсутствуют периодические колебания, или наоборот. Наличие случайной составляющей ε_t является обязательной.

Рассматриваемые компоненты позволяют представить уровень временного ряда в виде аддитивной или мультипликативной моделей:

$$\text{Аддитивная модель: } y_t = T_t + S_t + \varepsilon_t. \quad (4.1)$$

$$\text{Мультипликативная модель: } y_t = T_t \cdot S_t \cdot \varepsilon_t. \quad (4.2)$$

Выбор вида модели зависит от характера периодических колебаний. Аддитивная модель используется в случае, когда амплитуда циклической составляющей не зависит от времени t (рис. 4.1а). Если же амплитуда циклической составляющей с ходом времени изменяется в том же направлении, что и тенденция (возрастает или убывает), то применяется мультипликативная модель (рис. 4.1б).

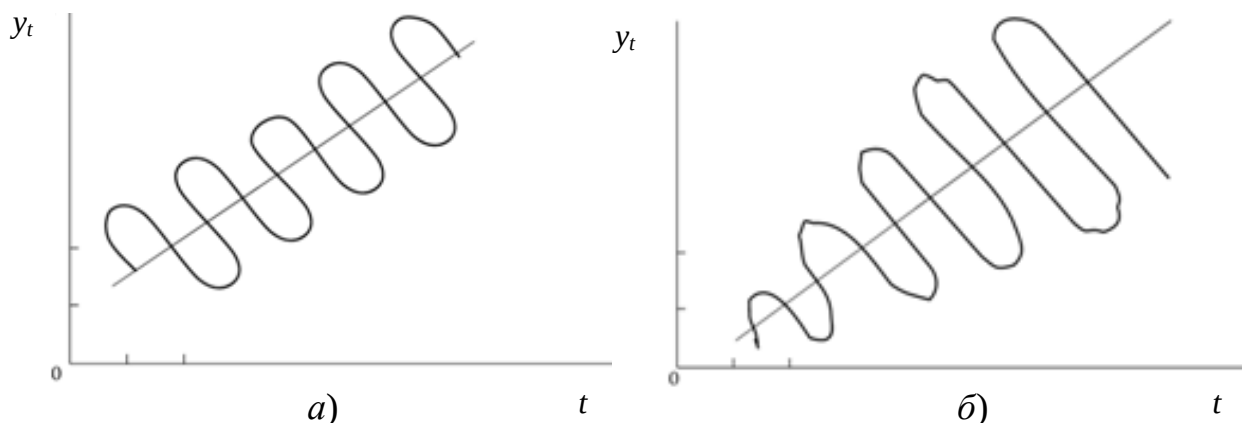


Рисунок 4.1 – Временной ряд с циклическими колебаниями:

а) – амплитуда колебаний во времени постоянна;

б) – амплитуда колебаний изменяется во времени

В качестве основных этапов анализа временных рядов можно выделить следующие:

- графическое представление и описание поведения временного ряда;
- выделение и удаление закономерных (неслучайных) составляющих временного ряда;
- сглаживание и фильтрация (удаление низко- или высокочастотных составляющих временного ряда);
- исследование случайной составляющей временного ряда;
- прогнозирование на основе построенной модели.

Важнейшей задачей при исследовании экономических временных рядов является выявление и статистическая оценка основной тенденции развития изучаемого процесса и отклонений от нее.

4.2 Автокорреляция уровней временного ряда

Автокорреляцией уровней временного ряда называют корреляционную зависимость между последовательными уровнями временного ряда.

Степень тесноты связи между уровнями исходного временного ряда y_t и уровнями этого ряда, сдвинутыми на τ шагов во времени $y_{t-\tau}$, может быть определена с помощью линейного коэффициента корреляции $r(\tau)$:

$$r(\tau) = \frac{\text{cov}(y_t, y_{t-\tau})}{\sigma_{y_t} \sigma_{y_{t-\tau}}} \quad (4.3)$$

Коэффициент $r(\tau)$ называют *коэффициентом автокорреляции порядка τ* , а число периодов τ – *лагом*. Его величина изменяется в пределах от -1 до $+1$.

С увеличением лага τ число пар значений, по которым рассчитывается коэффициент автокорреляции $r(\tau)$, уменьшается. Для обеспечения статистической достоверности коэффициентов автокорреляции необходимо, чтобы $\tau \leq n / 4$.

► **Пример 4.1.** По данным таблицы 4.1 для временного ряда y_t найти коэффициент автокорреляции первого порядка (для лага $\tau = 1$).

Решение. Найдем коэффициент автокорреляции $r(\tau)$ временного ряда (для лага $\tau=1$), т. е. коэффициент корреляции между последовательностями шести пар наблюдений y_t и y_{t-1} ($t = 2, 3, \dots, 7$).

y_t	42,1	57,1	65,9	74,8	86,2	110,3
y_{t-1}	32,4	42,1	57,1	65,9	74,8	86,2

Вычислим необходимые суммы:

$$\sum y_t = 42,1 + 57,1 + 65,9 + 74,8 + 86,2 + 110,3 = 436,4;;$$

$$\sum y_t^2 = 42,1^2 + 57,1^2 + 65,9^2 + 74,8^2 + 86,2^2 + 110,3^2 = 34567,2;$$

$$\sum y_{t-1} = 32,4 + 42,1 + 57,1 + 65,9 + 74,8 + 86,2 = 358,5;$$

$$\sum y_{t-1}^2 = 32,4^2 + 42,1^2 + 57,1^2 + 65,9^2 + 74,8^2 + 86,2^2 = 23450,87;$$

$$\sum y_t y_{t-1} = 42,1 \cdot 32,4 + 57,1 \cdot 42,1 + 65,9 \cdot 57,1 + 74,8 \cdot 65,9 + 86,2 \cdot 74,8 + 110,3 \cdot 86,2 = 28415,78.$$

Найдем средние значения:

$$\bar{y}_t = \frac{\sum y_t}{6} = 72,73; \quad \bar{y}_{t-1} = \frac{\sum y_{t-1}}{6} = 59,75; \quad \overline{y_t^2} = \frac{\sum y_t^2}{6} = 5761,2;$$

$$\overline{y_{t-1}^2} = \frac{\sum y_{t-1}^2}{6} = 3908,48; \quad \overline{y_t y_{t-1}} = \frac{\sum y_t y_{t-1}}{6} = 4735,96.$$

По формуле (4.3) коэффициент автокорреляции

$$r(1) = \frac{4735,96 - 72,73 \cdot 59,75}{\sqrt{(5761,2 - 72,73^2)(3908,48 - 59,75^2)}} = 0,977. \blacktriangleright$$

Последовательность коэффициентов автокорреляции первого, второго и более высоких порядков называется *автокорреляционной функцией* временного ряда. График зависимости ее значений от величины лага называется *коррелограммой*.

Автокорреляционная функция дает представление о внутренней структуре временного ряда.

Достаточно высокое значение коэффициента автокорреляции первого порядка означает наличие тенденции в ряде динамики. Для временного ряда с монотонной тенденцией к возрастанию (или убыванию) уровней автокорреляционная функция имеет значения близкие к +1, которые медленно снижаются с возрастанием величины лага. Если же ряд характеризуется сменой тенденций, то автокорреляционная функция примет значения, стремительно уменьшающиеся с возрастанием величины лага, сопровождаемые иногда сменой знака коэффициента автокорреляции.

С помощью автокорреляционной функции можно определить наличие или отсутствие в ряду динамики периодических колебаний и соответственно величину периода колебаний: если наиболее высоким оказался коэффициент автокорреляции порядка $\tau > 1$, то ряд содержит циклические колебания с периодом τ .

4.3 Аналитическое выравнивание временного ряда

Одной из важнейших задач исследования экономического временного ряда является выявление основной тенденции изучаемого процесса, выраженной неслучайной составляющей $f(t)$ (тренда либо тренда с циклической (сезонной) компонентой).

Для решения этой задачи вначале необходимо выбрать вид функции $f(t)$. Наиболее часто используются следующие функции:

линейная – $f(t) = b_0 + b_1 t$;

полиномиальная – $f(t) = b_0 + b_1 t + b_2 t^2 + \dots + b_n t^n$;

экспоненциальная – $f(t) = e^{b_0 + b_1 t}$;

Гомперца – $\log_c f(t) = a - br^t$, где $0 < r < 1$;

логистическая – $f(t) = \frac{a}{1 + be^{-ct}}$

При выборе функции $f(t)$ можно использовать:

- содержательный анализ, который помогает установить характер динамики процесса;
- визуальные наблюдения на основе графического изображения временного ряда;
- сравнение коэффициентов автокорреляции уровней ряда:
 - если временной ряд имеет линейную тенденцию, то его соседние уровни y_t и y_{t-1} тесно коррелируют; коэффициент автокорреляции $r(1)$ должен быть высоким;
 - если временной ряд содержит нелинейную тенденцию, то коэффициент $r(1)$ по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда.

Из двух функций предпочтение обычно отдается той, при которой меньше сумма квадратов отклонений фактических данных от расчетных на основе этих функций. При прочих равных условиях предпочтение следует отдавать более простым функциям.

Для выявления основной тенденции чаще всего используется *метод наименьших квадратов*.

Значения временного ряда y_t рассматриваются как зависимая переменная, а время t – объясняющая:

$$y_t = f(t) + \varepsilon_t, \quad (4.4)$$

где ε_t – возмущения, удовлетворяющие основным предпосылкам регрессионного анализа.

Согласно МНК параметры прямой $\hat{y}_t = f(t) = b_0 + b_1 t$ находятся из системы нормальных уравнений:

$$\begin{cases} b_0 n + b_1 \sum_{t=1}^n t = \sum_{t=1}^n y_t, \\ b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2 = \sum_{t=1}^n y_t t. \end{cases} \quad (4.5)$$

► **Пример 4.2.** По данным таблицы 4.1 найти уравнение тренда для временного ряда y_t , полагая тренд линейным.

Решение. Вычислим необходимые суммы:

$$\sum_{t=1}^7 t = 1 + 2 + 3 + 4 + 5 + 6 + 7 = 28;$$

$$\sum_{t=1}^7 t^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 = 140;$$

$$\sum_{t=1}^7 y_t = 32,4 + 42,1 + 57,1 + 65,9 + 74,8 + 86,2 + 110,3 = 468,8;$$

$$\begin{aligned} \sum_{t=1}^7 y_t t &= 32,4 \cdot 1 + 42,1 \cdot 2 + 57,1 \cdot 3 + 65,9 \cdot 4 + 74,8 \cdot 5 + 86,2 \cdot 6 + 110,3 \cdot 7 = \\ &= 2214,8. \end{aligned}$$

Система нормальных уравнений (4.5) имеет вид:

$$\begin{cases} 7b_0 + 28b_1 = 468,8, \\ 28b_0 + 140b_1 = 2214,8, \end{cases}$$

откуда $b_0 = 18,457$; $b_1 = 12,129$; уравнение тренда $\hat{y} = 18,457 + 12,129t$.

Параметры линейного тренда можно интерпретировать так:
 b_0 – начальный уровень временного ряда в момент времени $t = 0$;
 b_1 – средний за период абсолютный прирост уровней ряда.

Применительно к данному временному ряду можно сказать, что валовой сбор сахарной свеклы за семь лет увеличивался от уровня 18,457 тыс.т со средним за год абсолютным приростом, равным 12,129 тыс.т.

Рассчитанное в соответствии с формулой (2.27) значение F - критерия подтвердило значимость уравнения тренда на 5%-ом уровне, так как $F = 202,1 > F_{0,05; 1; 5} = 6,61$. ►

При применении МНК для оценки параметров экспоненциальной, логистической функций или функции Гомперца возникают сложности с решением получаемой системы нормальных уравнений, поэтому предварительно, до получения соответствующей системы, прибегают к некоторым преобразованиям этих функций (например, логарифмированию).

Другим методом выравнивания (сглаживания) временного ряда, т. е. выделения неслучайной составляющей, является *метод скользящих средних*. Он основан на переходе от начальных значений членов ряда к их средним значениям на интервале времени, длина которого определена заранее. При этом сам выбранный интервал времени «скользит» вдоль ряда. Получаемый таким образом ряд скользящих средних ведет себя более гладко, чем исходный ряд из-за усреднения отклонений ряда. Для усреднения могут быть использованы средняя арифметическая (простая и с некоторыми весами), медиана и другие показатели.

► **Пример 4.3.** Провести сглаживание временного ряда y_t по данным таблицы 4.1 методом скользящих средних, используя простую среднюю арифметическую с интервалом сглаживания $m = 3$ года.

Решение. Скользящие средние рассчитывают за одинаковый период, последовательно сдвигая их на одну дату. Первая средняя определяется за 1 – 3 год, вторая – за 2 – 4 год, третья – за 3 – 5 год и т. д. При этом средние будут относиться к срединному году периода – 2, 3, 4 и т. д. В итоге число средних скользящих всегда меньше числа исходных уровней на $(m - 1)$, т. е. на два года.

Находим скользящие средние:

$$\tilde{y}_2 = \frac{1}{3}(y_1 + y_2 + y_3) = \frac{1}{3}(32,4 + 42,1 + 57,1) = 43,87;$$

$$\tilde{y}_3 = \frac{1}{3}(y_2 + y_3 + y_4) = \frac{1}{3}(42,1 + 57,1 + 65,9) = 55,03;$$

$$\tilde{y}_4 = \frac{1}{3}(y_3 + y_4 + y_5) = \frac{1}{3}(57,1 + 65,9 + 74,8) = 65,93;$$

$$\tilde{y}_5 = \frac{1}{3}(y_4 + y_5 + y_6) = \frac{1}{3}(65,9 + 74,8 + 86,2) = 75,63;$$

$$\tilde{y}_6 = \frac{1}{3}(y_5 + y_6 + y_7) = \frac{1}{3}(74,8 + 86,2 + 110,3) = 90,43.$$

В результате получим сглаженный ряд:

t	1	2	3	4	5	6	7
$\tilde{y}_t$	–	43,87	55,03	65,93	75,63	90,43	–

Сопоставление средних скользящих позволяет сделать вывод о четкой тенденции роста валового сбора. ►

4.4 Прогнозирование на основе моделей временных рядов

Одна из важнейших задач анализа временного ряда состоит в *прогнозировании* на его основе развития изучаемого процесса. При этом исходят из того, что тенденция развития, установленная в прошлом, может быть распространена (экстраполирована) на будущий период.

Задача ставится так: имеется временной ряд y_t ($t = 1, 2, \dots, n$) и требуется дать прогноз уровня этого ряда на момент $n + \tau$.

Если рассматривать временной ряд как регрессионную модель изучаемого признака по переменной «время», то к нему могут быть применены рассмотренные выше методы анализа. Следует вспомнить, что одна из основных предпосылок регрессионного анализа состоит в том, что возмущения ε_t ($t = 1, 2, \dots, n$) представляют собой независимые случайные величины с математическим ожиданием (средним значением), равным нулю. При работе с временными рядами такое допущение оказывается во многих случаях неверным.

► **Пример 4.4.** По данным таблицы 4.1 дать точечную и с надежностью 0,95 интервальную оценки прогноза значения валового сбора сахарной свеклы на момент $t = 8$ (восьмой год). Полагая, что тренд линейный, а возмущения удовлетворяют требованиям классической модели.

Решение. В примере 4.2 получено уравнение регрессии $\hat{y} = 18,457 + 12,129t$, т. е. ежегодно валовой сбор сахарной свеклы увеличивался в среднем на 12,129 тыс. т. Таким образом, при сохранении тенденции, которая наблюдается в течение последних семи лет, можно ожидать, что валовой сбор сахарной свеклы в последующий год составит

$$\hat{y}_{t=8} = 18,457 + 12,129 \cdot 8 = 115,5 (\text{тыс.т}).$$

Для нахождения интервальной оценки прогноза рассчитаем среднюю ошибку прогнозируемого значения по формуле (2.34)

$$m_{y_p}^2 = S^2 \left(1 + \frac{1}{n} + \frac{(t_p - \bar{t})^2}{\sum_{i=1}^n (t_i - \bar{t})^2} \right)$$

Учитывая величину стандартной ошибки ($m_{y_p} = 5,9$) и табличное значение t -критерия Стьюдента ($t_{0,95;5} = 2,57$), вычислим по формуле (2.35) интервальную оценку прогноза:

$$115,5 - 2,57 \cdot 5,9 \leq y^*(8) \leq 115,5 + 2,57 \cdot 5,9 \quad \text{или}$$

$$100,3 \leq y^*(8) \leq 130,7$$

С надежностью 0,95 значение валового сбора сахарной свеклы на восьмой год будет заключено в пределах от 100,3 до 130,7 тыс. т. ►

Прогноз развития изучаемого процесса на основе экстраполяции временных рядов может оказаться эффективным, как правило, в рамках краткосрочного, в крайнем случае, среднесрочного периода прогнозирования.

Задачи для самостоятельного решения

4.5. Имеются данные об уровне инфляции y_t (%) за 8 месяцев:

Месяц	1	2	3	4	5	6	7	8
y_t	8,8	8,6	8,4	8,1	7,9	7,6	7,4	7,0

Определите коэффициенты автокорреляции уровней этого ряда первого и второго порядка. Обоснуйте выбор уравнения тренда и определите его параметры. Интерпретируйте полученные результаты.

4.6. Используя данные таблицы 4.2, необходимо: а) найти коэффициенты автокорреляции уровней временного ряда первого и второго порядка; б) найти уравнение линейного тренда и проверить его значимость на уровне 0,05; в) провести сглаживание временно ряда методом скользящих средних, используя простую среднюю арифметическую с интервалом сглаживания $m=3$; $m=5$.

Таблица 4.2 – Урожайность озимой пшеницы за 10 лет

t	1	2	3	4	5	6	7	8	9	10
y_t	16,3	20,3	17,1	7,7	15,3	16,3	19,9	14,4	18,7	20,7

4.7. Администрация банка изучает динамику депозитов физических лиц за ряд лет (млн. долл.). Исходные данные представлены в таблице 4.3.

Таблица 4.3 – Динамика депозитов физических лиц (млн. долл.)

Время, год	1	2	3	4	5	6	7
Депозиты физических лиц, y_t	2	6	7	3	10	12	13

Постройте уравнение линейного тренда и дайте интерпретацию его параметров. Определите коэффициент детерминации для линейного тренда. Верно ли предположение администрации банка, что среднегодовой абсолютный прирост депозитов физических лиц составляет не менее 2,5 млн. долл.

4.8. Имеются данные об урожайности зерновых культур в хозяйствах области:

Год, t	1	2	3	4	5	6	7	8
Урожайность y_t , ц/га	10,2	10,7	11,7	13,1	14,9	17,2	20,0	23,2

Обоснуйте выбор типа уравнения тренда. Рассчитайте параметры уравнения тренда. Дайте прогноз урожайности зерновых культур на следующий год.

4.9. Используя данные таблицы 4.4, найдите уравнение линейного тренда и проверьте его значимость на уровне 0,05.

Таблица 4.4 – Динамика доходов на душу населения (ден. ед.) за восьмилетний период

t	1	2	3	4	5	6	7	8
y_t	1133	1222	1354	1389	1342	1377	1491	1684

Дайте точечный и с надежностью 0,95 интервальный прогнозы доходов на девятый год.

Контрольные вопросы

1. Дайте определение временного ряда.
2. Назовите основные этапы анализа временных рядов.
3. Перечислите основные составляющие временного ряда.
4. Запишите общий вид мультипликативной и аддитивной моделей временного ряда.
5. Что такое автокорреляция уровней временного ряда и как ее можно оценить количественно?
6. Что означает понятие «коррелограмма»?
7. Как используется автокорреляционная функция при моделировании временного ряда?
8. Какие функции часто используются для описания тренда.
9. Какова интерпретация параметров линейного тренда?
10. Как оцениваются параметры линейного тренда?
11. В чем суть метода скользящих средних?
12. Что означает прогнозирование на основе экстраполяции временного ряда?
13. Как получить точечную оценку прогноза на основе модели временного ряда?
14. Приведите формулу для нахождения интервальной оценки прогноза.

5 ГЕТЕРОСКЕДАСТИЧНОСТЬ И АВТОКОРРЕЛЯЦИЯ ОСТАТКОВ

При практическом проведении регрессионного анализа с помощью МНК следует обратить внимание на проблемы, связанные с выполнимостью свойств случайных возмущений моделей. В частности, могут не выполняться предпосылки регрессионного анализа о том, что случайные возмущения модели имеют постоянную дисперсию и не коррелированы между собой. Например, при использовании зависимости расходов на потребление от уровня доходов семей можно ожидать, что в более обеспеченных семьях вариация расходов выше, чем в малообеспеченных, т. е. дисперсии возмущений не одинаковы. При рассмотрении временных рядов часто возникает ситуация, когда наблюдаемые в данный момент значения зависимой переменной коррелируют с их значениями в предыдущие моменты времени, т. е. наблюдается корреляция между возмущениями в разные моменты времени.

При нарушении основных предпосылок регрессионного анализа МНК может давать оценки с плохими статистическими свойствами. В этом случае для получения более точных оценок коэффициентов регрессии необходимо использовать другие методы их определения.

5.1 Суть гетероскедастичности

Одной из ключевых предпосылок регрессионного анализа, основанного на МНК, является условие постоянства дисперсий случайных возмущений: *дисперсия возмущения ε_i должна быть постоянна для всех наблюдений, т. е. $D(\varepsilon_i) = \sigma^2$, $i = 1, 2, \dots, n$.*

Выполнимость данной предпосылки называется *гомоскедастичностью*. Если это условие не соблюдается, то имеет место *гетероскедастичность*.

Наличие гетероскедастичности приводит к неэффективности оценок коэффициентов регрессии, т. е. они не будут иметь наименьшую дисперсию по сравнению с другими оценками данного параметра. Увеличение дисперсии оценок снижает вероятность получения максимально точных оценок. Все выводы, получаемые на основе соответствующих t – и F –статистик, а также интервальные оценки будут ненадежными. Следовательно, статистические выводы, получа-

мые при стандартных проверках качества оценок, могут быть ошибочными и приводить к неверным заключениям по построенной модели. Вполне вероятно, что стандартные ошибки коэффициентов будут занижены, а, следовательно, t -статистики будут завышены. Это может привести к признанию статистически значимыми коэффициентов, которые таковыми на самом деле не являются.

Проблема гетероскедастичности характерна для пространственных данных и довольно редко встречается при рассмотрении временных рядов.

5.2 Тесты на гетероскедастичность

Для обнаружения гетероскедастичности разработано большое количество тестов и критериев для них. Все они используют в качестве нулевой гипотезы H_0 *гипотезу об отсутствии гетероскедастичности*.

Тест ранговой корреляции Спирмена. При использовании данного теста предполагается, что дисперсия случайного возмущения будет либо увеличиваться, либо уменьшаться с увеличением значений X . Поэтому для регрессии, построенной по МНК, абсолютные величины остатков e_i и значения x_i будут коррелированы. Значения $|e_i|$ и x_i ранжируются (упорядочиваются по величинам) и определяется коэффициент ранговой корреляции

$$\rho_{x,e} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}, \quad (5.1)$$

где d_i – разность между рангами $|e_i|$ и x_i ; n – число наблюдений.

Коэффициент ранговой корреляции значим на уровне α , если статистика

$$|t| = \frac{|\rho_{x,e}| \sqrt{n-2}}{\sqrt{1 - \rho_{x,e}^2}} > t_{1-\alpha; n-2}, \quad (5.2)$$

где $t_{1-\alpha; n-2}$ – табличное значение t -критерия Стьюдента, определенное на уровне значимости α и числе степеней свободы $(n - 2)$.

В этом случае необходимо отклонить гипотезу об отсутствии гетероскедастичности.

Тест Голдфелда-Квандта. Тест построен на следующих предположениях:

- среднее квадратическое отклонение возмущения σ_i пропорционально значению x_i объясняющей переменной X в этом наблюдении;
- случайное возмущение ε_i имеет нормальное распределение;
- отсутствует автокорреляция остатков.

Тест Голдфелда-Квандта включает следующие шаги:

1. Все n наблюдений упорядочиваются по возрастанию переменной X .
2. Упорядоченная выборка после этого разбивается на три подвыборки размерностей m , $(n - 2m)$, m соответственно.
3. Оцениваются отдельные регрессии для первой подвыборки (m первых наблюдений) и для третьей подвыборки (m последних наблюдений).
4. Строится F -статистика:

$$F = \frac{S_3 / (m - p - 1)}{S_1 / (m - p - 1)} = \frac{S_3}{S_1}, \quad (5.3)$$

где $S_1 = \sum_{i=1}^m e_i^2$ – остаточная сумма квадратов отклонений в регрессии, оцененной по первой подвыборке;

$S_3 = \sum_{i=n-m+1}^n e_i^2$ – остаточная сумма квадратов отклонений в регрессии, оцененной по третьей подвыборке;

p – количество объясняющих переменных в уравнении регрессии;

$m - p - 1$ – число степеней свободы соответствующих сумм.

При сделанных предположениях относительно случайных возмущений построенная F -статистика имеет распределение Фишера с числами степеней свободы $k_1 = k_2 = m - p - 1$.

5. Если $F = \frac{S_3}{S_1} > F_{\alpha; k_1; k_2}$, то гипотеза H_0 об отсутствии гетероскедастичности отклоняется.

Отметим, что в случае множественной линейной регрессии проверка гетероскедастичности производится по каждой из объясняющих переменных.

Тест Глейзера. По данному методу оценивается регрессионная зависимость модулей отклонений $|e_i|$ от x_i . При этом рассматриваемая зависимость моделируется уравнением регрессии

$$|e_i| = \beta_0 + \beta_1 x_i^k + \nu_i \quad (5.4)$$

Изменяя значение k , можно построить различные регрессии. Обычно $k = \dots -1; -0,5; 0,5; 1; \dots$. Статистическая значимость коэффициента β_1 в каждом конкретном случае фактически означает наличие гетероскедастичности.

Проверка выполнения требования гомоскедастичности случайных возмущений может быть произведена визуально на основе анализа графиков зависимости отклонений e_i от значений объясняющих переменных, входящих в регрессию, либо от значений зависимой переменной, полученных по уравнению регрессии.

В случае гомоскедастичности «облако» случайных отклонений находится в области, параллельной оси абсцисс (рис. 5.1а). Все прочие случаи соответствуют гетероскедастичности (рис. 5.1б).

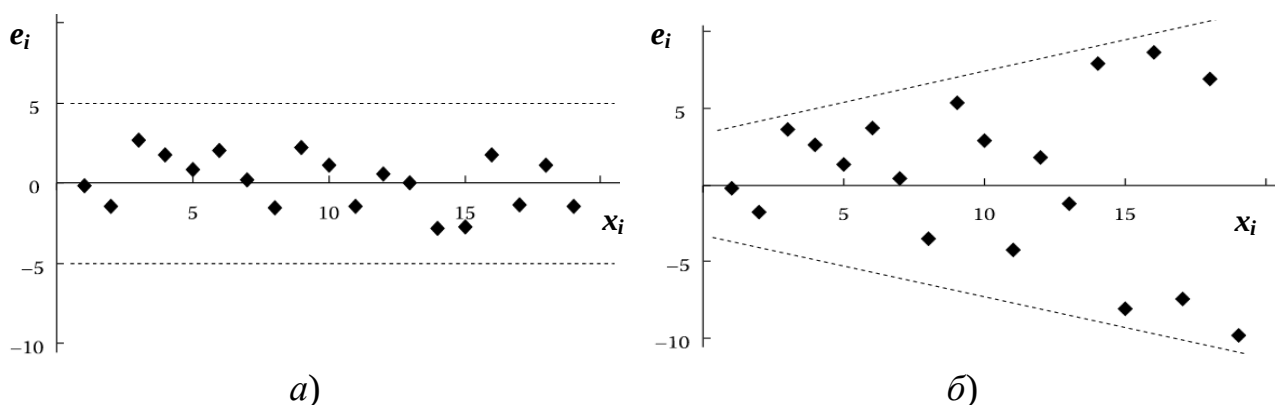


Рисунок 5.1 – Зависимость отклонений e_i от значения x_i :

а) – нет зависимости (гомоскедастичность);

б) – дисперсия отклонений увеличивается с увеличением значения x_i (один из случаев гетероскедастичности)

► **Пример 5.1.** Анализируется связь выпуска инновационной продукции на душу населения Y от величины валового внутреннего продукта X для 17 стран. Исходные данные (усл. ед.) приведены в таблице

(5.1). Построить уравнение линейной регрессии Y по X и проверить наличие гетероскедастичности.

Таблица 5.1 – Зависимость выпуска инновационной продукции на душу населения Y от величины валового внутреннего продукта X

i	x_i	y_i	i	x_i	y_i
1	3	18	10	24	100
2	6	27	11	25	63
3	7	18	12	26	130
4	9	45	13	27	135
5	13	55	14	28	60
6	15	68	15	35	70
7	18	51	16	37	80
8	21	84	17	44	180
9	22	85			

Решение. С помощью обычного МНК получим уравнение регрессии $\hat{y} = 12,844 + 2,918x$, отражающее зависимость выпуска инновационной продукции на душу населения Y от величины валового внутреннего продукта X .

График зависимости отклонений e_i от значения x_i объясняющей переменной имеет вид, представленный на рис. 5.2. Можно отметить определенное увеличение разброса точек с увеличением значения x_i . Такая картина может свидетельствовать о наличии гетероскедастичности.

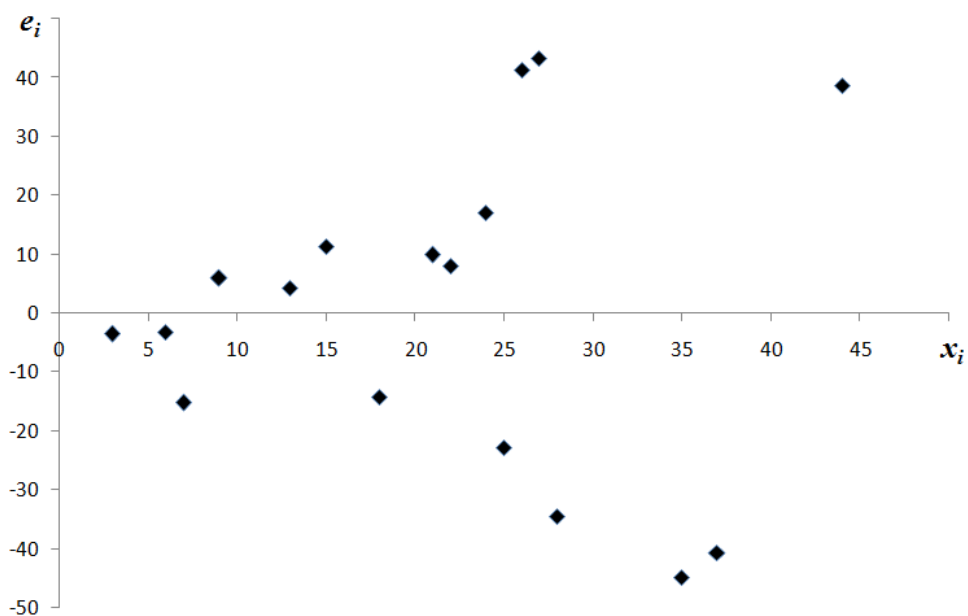


Рисунок 5.2 – Зависимость отклонений e_i от значения x_i

Проверим уравнение на наличие гетероскедастичности, используя тесты Спирмена, Глейзера и Голдфелда-Квандта.

1) Проверка гетероскедастичности по тесту Спирмена.

Рассчитав по уравнению линейной регрессии $\hat{y}_i$, находим остатки e_i , модули остатков $|e_i|$, ранги x_i и $|e_i|$, квадраты разностей между рангами d_i^2 . Результаты представим в таблице 5.2.

Таблица 5.2 – Расчет коэффициента ранговой корреляции

i	x_i	y_i	$\hat{y}_i$	$ e_i $	Ранг x_i	Ранг $ e_i $	d_i	d_i^2
1	3	18	21,60	3,60	17	16	1	1
2	6	27	30,35	3,35	16	17	-1	1
3	7	18	33,27	15,27	15	9	6	36
4	9	45	39,11	5,89	14	14	0	0
5	13	55	50,78	4,22	13	15	-2	4
6	15	68	56,62	11,38	12	11	1	1
7	18	51	65,38	14,38	11	10	1	1
8	21	84	74,13	9,87	10	12	-2	4
9	22	85	77,05	7,95	9	13	-4	16
10	24	100	82,89	17,11	8	8	0	0
11	25	63	85,81	22,81	7	7	0	0
12	26	130	88,72	41,28	6	3	3	9
13	27	135	91,64	43,36	5	2	3	9
14	28	60	94,56	34,56	4	6	-2	4
15	35	70	114,99	44,99	3	1	2	4
16	37	80	120,83	40,83	2	4	-2	4
17	44	180	141,26	38,74	1	5	-4	16
Σ	–	–	–	–	–	–	–	110

По формуле (5.1) находим коэффициент ранговой корреляции Спирмена

$$\rho_{x,e} = 1 - \frac{6 \cdot 110}{17(17^2 - 1)}, = 0,865.$$

Проверим значимость коэффициента $\rho_{x,e}$:

$$|t| = \frac{0,865\sqrt{17-2}}{\sqrt{1-0,865^2}} = 6,67.$$

Так как $t_{0,95;15} = 2,13$, то гипотеза H_0 отклоняется, следовательно, имеет место гетероскедастичность.

2) Проверка гетероскедастичности по тесту Глейзера.

Оценим зависимость модулей $|e_i|$ от x_i уравнением линейной регрессии $|e| = b_0 + b_1x$. Применяя обычный МНК, находим параметры b_0 и b_1 . Получим: $|e| = -3,13 + 1,15x$.

Значимость коэффициента b_1 : $t = 5,89$; $t_{0,95;15} = 2,13$.

Гипотеза H_0 отклоняется, так как b_1 значимо отличается от нуля. В модели имеет место гетероскедастичность.

3) Проверка по тесту Голдфелда-Квандта.

Упорядоченную по возрастанию значений переменной X выборку разбиваем на три подвыборки: 6 – 5 – 6 наблюдений. Для первой и третьей подвыборок по МНК строим уравнения регрессии:

$$\hat{y}_1 = -0,18 + 4,38x;$$

$$\hat{y}_3 = 39,90 + 2,11x.$$

Вычислим остаточные суммы квадратов и их отношение:

$$F = \frac{S_3}{S_1} = \frac{9804,49}{224,39} = 43,69.$$

Табличное значение F -критерия Фишера: $F_{0,05;4;4} = 6,39$. Так как наблюдаемое значение F больше табличного, гипотеза H_0 об отсутствии гетероскедастичности отклоняется.

Таим образом, рассмотренные тесты показали наличие в модели гетероскедастичности. ►

5.3 Метод взвешенных наименьших квадратов

При наличии в модели гетероскедастичности МНК-оценки коэффициентов регрессии будут неэффективными. Для получения более качественных оценок используют *взвешенный (обобщенный) метод наименьших квадратов*.

Данный метод применяется при известных для каждого наблюдения значениях дисперсии σ_i^2 . В этом случае можно устранить гетероскедастичность, разделив каждое наблюдаемое значение на соответствующее ему значение среднего квадратического отклонения.

Для простоты изложения опишем взвешенный МНК на примере парной регрессии

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (5.5)$$

Пусть в исходной модели регрессии (5.5) случайные возмущения гетероскедастичны, т. е.

$$M(\varepsilon_i) = 0; \quad D(\varepsilon_i) = \sigma_i^2 \quad (i = 1, 2, \dots, n)..$$

Разделим каждое наблюдение на известное $\sigma_i = \sqrt{\sigma_i^2}$:

$$\frac{y_i}{\sigma_i} = \frac{\beta_0}{\sigma_i} + \beta_1 \frac{x_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i} \quad (5.6)$$

Для преобразованной модели (5.6) выполняется условие гомоскедастичности:

$$M\left(\frac{\varepsilon_i}{\sigma_i}\right) = 0; \quad D\left(\frac{\varepsilon_i}{\sigma_i}\right) = \frac{D(\varepsilon_i)}{\sigma_i^2} = 1 \quad (i = 1, 2, \dots, n).$$

Оценка параметров преобразованного уравнения обычным МНК приводит к оценке параметров исходного уравнения взвешенным МНК. Для каждого из этих методов необходимо минимизировать сумму квадратов отклонений вида

$$S = \sum \left(\frac{y_i}{\sigma_i} - \frac{b_0}{\sigma_i} - b_1 \frac{x_i}{\sigma_i} \right)^2 = \sum \frac{1}{\sigma_i^2} (y_i - b_0 - b_1 x_i)^2$$

При минимизации этой суммы квадратов отдельные ее слагаемые взвешиваются: наблюдениям с большей дисперсией придается меньший вес $\frac{1}{\sigma_i^2}$. Оценки МНК коэффициентов преобразованной модели дают непосредственно оценки исходной модели.

Для применения взвешенного МНК необходимо знать фактические значения дисперсий σ_i^2 . На практике такие значения известны крайне редко. Следовательно, чтобы применить взвешенный МНК, необходимо сделать реалистические предположения о значениях σ_i^2 . Может оказаться целесообразным предложить, что дисперсии σ_i^2 возмущений ε_i пропорциональны значениям x_i или значениям x_i^2 .

1. Дисперсии σ_i^2 пропорциональны x_i :

$$\sigma_i^2 = \sigma^2 x_i \quad (\sigma^2 - \text{коэффициент пропорциональности}).$$

Тогда уравнение (5.5) преобразуется делением его левой и правой частей на $\sqrt{x_i}$:

$$\frac{y_i}{\sqrt{x_i}} = \frac{\beta_0}{\sqrt{x_i}} + \frac{\beta_1 x_i}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}} \Rightarrow \frac{y_i}{\sqrt{x_i}} = \beta_0 \frac{1}{\sqrt{x_i}} + \beta_1 \sqrt{x_i} + v_i \quad (5.7)$$

Для случайных возмущений $v_i = \frac{\varepsilon_i}{\sqrt{x_i}}$ выполняется условие гомоскедастичности. Следовательно, для регрессии (5.7) применим обычный МНК.

2. Дисперсии σ_i^2 пропорциональны x_i^2

В этом случае преобразованием будет деление уравнения регрессии (5.5) на x_i :

$$\frac{y_i}{x_i} = \beta_0 \frac{1}{x_i} + \beta_1 + \frac{\varepsilon_i}{x_i} \Rightarrow \frac{y_i}{x_i} = \beta_0 \frac{1}{x_i} + \beta_1 + v_i, \quad (5.8)$$

где $v_i = \frac{\varepsilon_i}{x_i}$.

Несложно показать, что для возмущений v_i будет выполняться условие гомоскедастичности. После определения по МНК оценок коэффициентов β_0 и β_1 для уравнения (5.8) возвращаются к исходному уравнению (5.5).

Пример 5.2. По данным примера 5.1 постройте уравнение регрессии, используя взвешенный метод наименьших квадратов.

Решение. Ранее было выявлено, что в модели регрессии $y_i = 12,844 + 2,918x_i + e_i$, построенной по данным таблицы 5.1, случайные отклонения гетероскедастичны. Для устранения гетероскедастичности применим взвешенный метод наименьших квадратов, полагая, что дисперсии возмущений σ_i^2 пропорциональны значениям x_i^2 . Выполним преобразование уравнения в соответствии с формулой (5.8). Ниже приведена таблица, содержащая преобразованные переменные.

Таблица 5.3 – Преобразование переменных при применении взвешенного МНК

i	x_i	y_i	$x_i^* = \frac{1}{x_i}$	$y_i^* = \frac{y_i}{x_i}$
1	3	18	0,3333	6,0000
2	6	27	0,1667	4,5000
3	7	18	0,1429	2,5714
4	9	45	0,1111	5,0000
5	13	55	0,0769	4,2308
6	15	68	0,0667	4,5333
7	18	51	0,0556	2,8333
8	21	84	0,0476	4,0000
9	22	85	0,0455	3,8636
10	24	100	0,0417	4,1667
11	25	63	0,0400	2,5200
12	26	130	0,0385	5,0000
13	27	135	0,0370	5,0000
14	28	60	0,0357	2,1429
15	35	70	0,0286	2,0000
16	37	80	0,0270	2,1622
17	44	180	0,0227	4,0909

По преобразованным данным оценим уравнение регрессии с помощью обычного МНК. Получим следующее уравнение:

$$\hat{y}^* = 7,7823x^* + 3,1978.$$

В данном уравнении параметр при переменной x^* является свободным членом исходной модели, а параметр, равный 3,1978, – коэффициентом регрессии при объясняющей переменной. Таким образом, в исходном виде уравнение регрессии будет иметь вид:

$$\hat{y} = 7,7823 + 3,1978x.$$

Сопоставив полученное уравнение с первоначальным уравнением: $\hat{y} = 12,844 + 2,918x$, можно отметить существенное изменение параметров. ►

5.4 Суть и причины автокорреляции

Автокорреляция (последовательная корреляция) определяется как корреляция между наблюдаемыми показателями, упорядоченными во времени или пространстве.

Автокорреляцией возмущений (остатков) ε_i регрессионной модели называется корреляционная зависимость между текущими и предыдущими значениями остатков.

Независимость значений случайных возмущений ε_i от значений возмущений во всех других наблюдениях является важной предпосылкой построения качественной регрессионной модели по МНК. Отсутствие зависимости гарантирует отсутствие коррелированности между любыми возмущениями и, в частности, между соседними возмущениями.

Различают *положительную* и *отрицательную* автокорреляцию.

Положительная автокорреляция имеет место, когда завышенные (заниженные) значения в предыдущих наблюдениях приводят к завышению (занижению) их в наблюдениях последующих. Графически положительная автокорреляция выражается в чередовании зон, где наблюдаемые значения оказываются выше объясненных регрессией, и зон, где наблюдаемые значения ниже (рис. 5.3а).

Отрицательная автокорреляция встречается в тех случаях, когда наблюдения действуют друг на друга по принципу «маятника» – за положительным отклонением следует отрицательное и наоборот. Графически это выражается в том, что результаты наблюдений «слишком часто» перескакивают через график функции регрессии (рис. 5.3б).

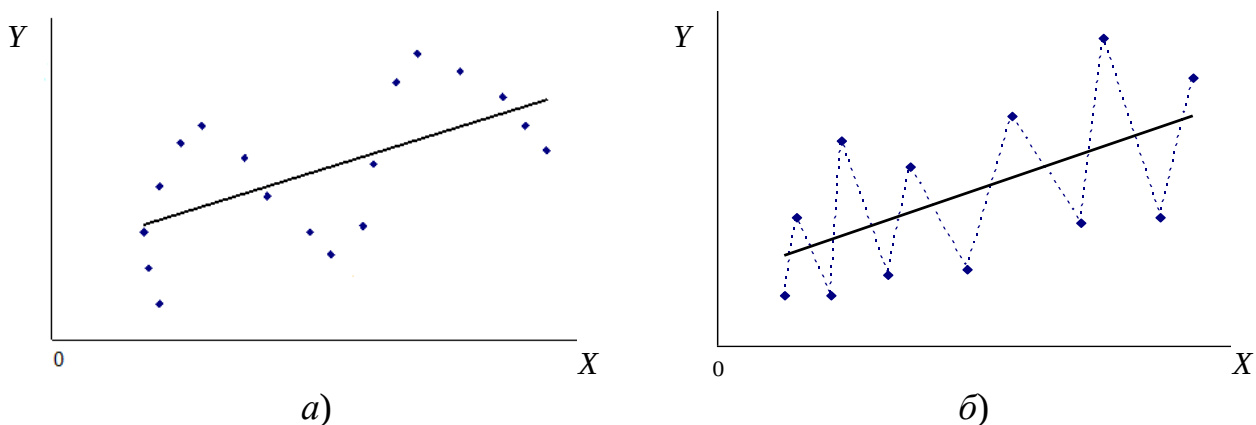


Рисунок 5.3 – Схема рассеяния точек с автокорреляцией возмущений:

а) – положительная автокорреляция;

б) – отрицательная автокорреляция

Среди основных причин, вызывающих появление автокорреляции, можно выделить ошибки спецификации, инерцию в изменении экономических показателей, «эффект паутины», сглаживание данных.

Ошибки спецификации. Неучет в модели какой-либо важной объясняющей переменной либо неправильный выбор формы зависимости обычно приводит к системным отклонениям от линии регрессии.

Инерция в изменении экономических показателей. Многие экономические показатели обладают определенной цикличностью. Например, экономический подъем приводит к росту занятости, сокращению инфляции, увеличению ВВП и т. д. Этот рост продолжается до тех пор, пока изменение конъюнктуры рынка не приведет к замедлению роста, затем остановке и движению вспять рассматриваемых показателей.

Эффект паутины. Во многих производственных сферах экономические показатели реагируют на изменение экономических условий с запаздыванием. Например, предложение сельскохозяйственной продукции реагирует на изменение цены с запаздыванием (равным периоду созревания урожая). Высокая цена сельскохозяйственной продукции в прошлом году вызовет скорее всего ее перепроизводство в текущем году, а следовательно цена на нее снизится.

Сглаживание данных. Часто данные по некоторому продолжительному временному периоду получают усреднением данных по составляющим его подынтервалам, что может привести к определенному сглаживанию колебаний, которые имелись внутри рассматриваемого периода.

Если случайные возмущения подвержены автокорреляции, то оценки параметров, полученные по МНК, оставаясь линейными и несмещенными, перестают быть эффективными, а их стандартные ошибки рассчитываются некорректно (занижаются). Это влечет за собой увеличение t статистик, что может привести к признанию статистически значимыми объясняющие переменные, которые в действительности таковыми не являются. Вследствие этого ухудшаются прогнозные качества модели.

Автокорреляция возмущений обычно встречается в регрессионном анализе при использовании данных временных рядов. В силу этого вместо символа i порядкового номера наблюдения будем использовать символ t , отражающий момент наблюдения.

5.5 Обнаружение автокорреляции

Выводы о наличии или отсутствии автокорреляции осуществляются на основе оценок случайных возмущений e_t ($t = 1, 2, \dots, n$), полученных из выборочного уравнения регрессии.

Рассмотрим возможные методы определения автокорреляции.

Графический метод. Существует несколько вариантов графического определения автокорреляции. Один из них связан с построением последовательно-временных графиков. В этом случае по оси абсцисс обычно откладываются либо время (момент) получения статистических данных, либо порядковый номер наблюдения, а по оси ординат – оценки возмущений e_t . Затем с помощью графика делается вывод о наличии или отсутствии автокорреляции.

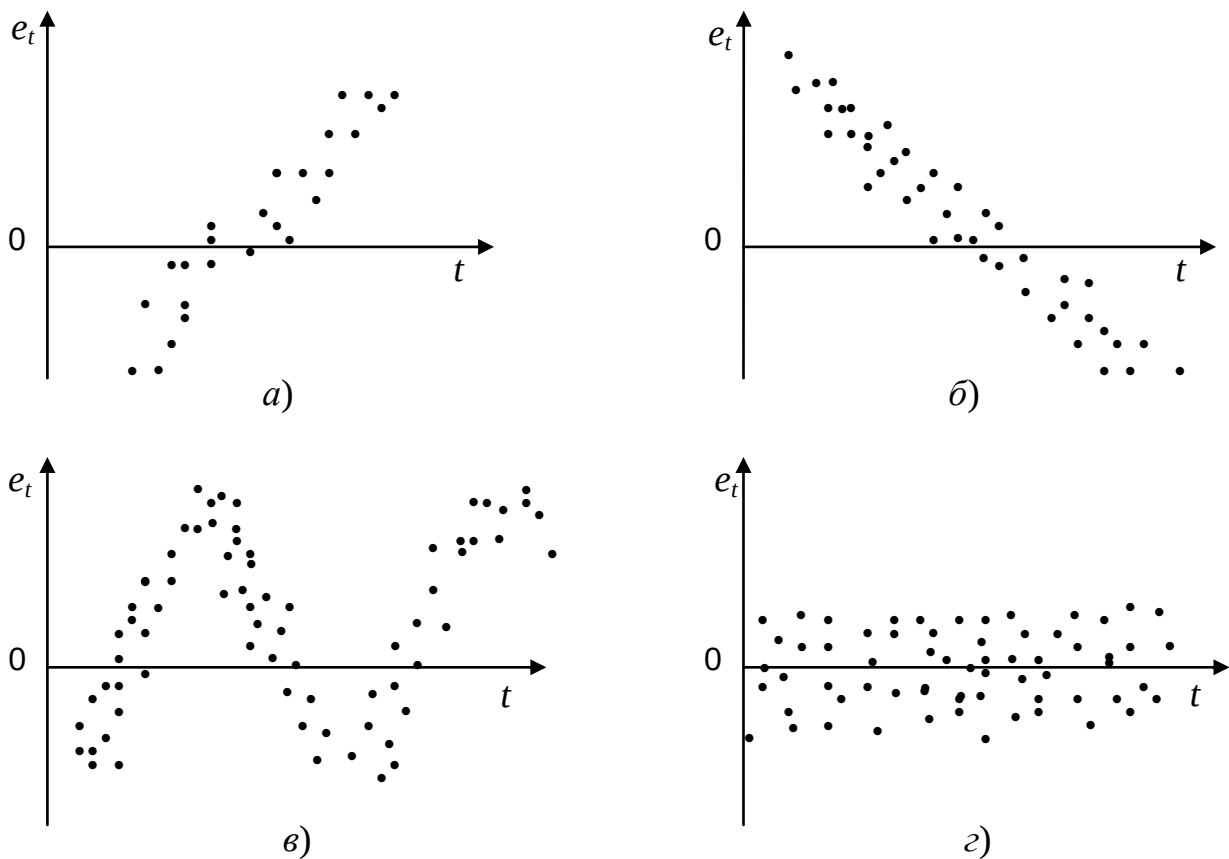


Рисунок 5.4. – Графическое определение автокорреляции:

- а) – нарастающая тенденция в остатках;
- б) – убывающая тенденция в остатках;
- в) – циклические колебания в остатках;
- г) – отсутствие зависимости (отсутствие автокорреляции)

Метод рядов. Метод основан на последовательном определении знаков отклонений e_t ($t = 1, 2, \dots, n$).

Например, $(- - - - -)(+ + + + +)(- - -)(+ + + + +)$, т. е. при 20 наблюдениях 5 «-», 7 «+», 3 «-», 5 «+».

Ряд – непрерывная последовательность одинаковых знаков.

Количество знаков в ряду называется длиной ряда.

Если рядов слишком мало по сравнению с количеством наблюдений n , то вполне вероятна положительная автокорреляция. Так, в рассмотренном примере имеет место положительная автокорреляция.

Если же рядов слишком много, то вероятна отрицательная автокорреляция.

Критерий Дарбина–Уотсона. Это простой критерий для определения наличия автокорреляции между соседними членами.

В тесте Дарбина–Уотсона используется статистика вида

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (5.9)$$

Статистика Дарбина–Уотсона тесно связана с выборочным коэффициентом корреляции $r_{e_t e_{t-1}}$ между соседними наблюдениями:

$$DW \approx 2(1 - r_{e_t e_{t-1}}) \quad (5.10)$$

Если $r_{e_t e_{t-1}} \approx 0$ (автокорреляция отсутствует), то $DW \approx 2$.

Если $r_{e_t e_{t-1}} \approx 1$ (положительная автокорреляция), то $DW \approx 0$.

Если $r_{e_t e_{t-1}} \approx -1$ (отрицательная автокорреляция), то $DW \approx 4$.

Таким образом, $0 \leq DW \leq 4$, и ее значения могут указать на наличие либо отсутствие автокорреляции.

Для более точного определения, какое значение DW свидетельствует об отсутствии автокорреляции, а какое – об ее наличии, была построена таблица критических точек распределения Дарбина–Уотсона. По ней для заданного уровня значимости α , числа наблюдений n и количества объясняющих переменных p определяются два значения: d_H – нижняя граница и d_B – верхняя граница. По этим значениям отрезок $[0; 4]$ разбивается на пять зон. В зависимости от того,

в какую зону попадает расчетное значение критерия, осуществляют выводы по следующему правилу:

- 1) если $0 \leq DW < d_H$ – существует положительная автокорреляция;
- 2) если $d_H \leq DW < d_B$ – вывод о наличии автокорреляции не определен;
- 3) если $d_B \leq DW < 4 - d_B$ – автокорреляция отсутствует;
- 4) если $4 - d_B \leq DW < 4 - d_H$ – вывод о наличии автокорреляции не определен;
- 5) если $4 - d_H \leq DW \leq 4$ – существует отрицательная автокорреляция.

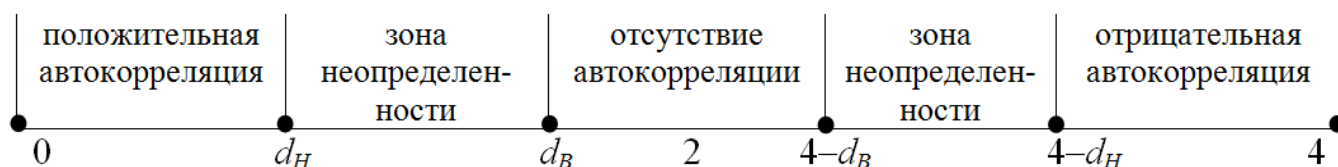


Рисунок 5.5 – Критические области статистики Дарбина–Уотсона

Главным недостатком критерия Дарбина–Уотсона являются зоны неопределенности. Их наличие связано с тем, что распределение DW -статистики зависит не только от числа наблюдений и числа объясняющих переменных, но и от значений объясняющих переменных.

► **Пример 5.3.** Анализируется зависимость между инфляцией Y и безработицей X . Используются статистические данные за 22 года. Необходимо вычислить статистику DW Дарбина–Уотсона и на ее основе определить наличие или отсутствие автокорреляции.

Таблица 5.4 – Исходные данные

t	1	2	3	4	5	6	7	8	9	10	11
x_t	9,10	6,50	4,63	8,50	9,55	3,71	5,80	3,60	6,53	4,32	9,20
y_t	0,70	2,20	2,38	0,90	1,10	5,12	0,93	2,54	1,55	3,45	1,09

t	12	13	14	15	16	17	18	19	20	21	22
x_t	5,75	3,65	7,30	9,65	3,65	9,80	6,28	7,80	8,75	7,22	3,65
y_t	2,15	5,14	1,72	0,74	4,16	0,93	1,79	1,24	1,12	1,28	5,30

Решение. По МНК получим уравнение линейной регрессии:

$$\hat{y} = 5,9139 - 0,5697x.$$

В таблице 5.5 приведен расчет сумм, необходимых для вычисления статистики Дарбина–Уотсона.

Таблица 5.5 – Расчет статистики Дарбина–Уотсона

t	x_t	y_t	$\hat{y}_t$	$e_t = y_t - \hat{y}_t$	e_{t-1}	$(e_t - e_{t-1})^2$	e_t^2
1	9,10	0,70	0,7296	-0,0296	—	—	0,0009
2	6,50	2,20	2,2109	-0,0109	-0,0296	0,0004	0,0001
3	4,63	2,38	3,2762	-0,8962	-0,0109	0,7838	0,8032
4	8,50	0,90	1,0715	-0,1715	-0,8962	0,5252	0,0294
5	9,55	1,10	0,4733	0,6267	-0,1715	0,6371	0,3928
6	3,71	5,12	3,8003	1,3197	0,6267	0,4802	1,7416
7	5,80	0,93	2,6096	-1,6796	1,3197	8,9960	2,8212
8	3,60	2,54	3,8630	-1,3230	-1,6796	0,1272	1,7503
9	6,53	1,55	2,1938	-0,6438	-1,3230	0,4613	0,4144
10	4,32	3,45	3,4528	-0,0028	-0,6438	0,4108	0,00001
11	9,20	1,09	0,6727	0,4173	-0,0028	0,1765	0,1742
12	5,75	2,15	2,6381	-0,4881	0,4173	0,8199	0,2383
13	3,65	5,14	3,8345	1,3055	-0,4881	3,2171	1,7043
14	7,30	1,72	1,7551	-0,0351	1,3055	1,7972	0,0012
15	9,65	0,74	0,4163	0,3237	-0,0351	0,1287	0,1048
16	3,65	4,16	3,8345	0,3255	0,3237	0,0000	0,1060
17	9,80	0,93	0,3308	0,5992	0,3255	0,0749	0,3590
18	6,28	1,79	2,3362	-0,5462	0,5992	1,3118	0,2983
19	7,80	1,24	1,4702	-0,2302	-0,5462	0,0998	0,0530
20	8,75	1,12	0,9290	0,1910	-0,2302	0,1774	0,0365
21	7,22	1,28	1,8007	-0,5207	0,1910	0,5064	0,2711
22	3,65	5,30	3,8345	1,4655	-0,5207	3,9449	2,1477
Сумма						24,6767	13,4482

По формуле (5.9) статистика Дарбина–Уотсона

$$DW = \frac{24,6767}{13,4482} = 1,83.$$

По таблице III приложений при $n = 22$, $p = 1$ критические значения $d_H = 1,24$; $d_B = 1,43$. Фактически найденное значение $DW = 1,83$ находится в пределах от d_B до $4 - d_B$ ($1,43 < 1,83 < 2,57$). Это свидетельствует об отсутствии автокорреляции возмущений. ►

5.6 Методы устранения автокорреляции

При установлении автокорреляции необходимо в первую очередь проанализировать правильность спецификации модели. Если после ряда возможных усовершенствований регрессии (уточнения состава объясняющих переменных либо изменения формы зависимости) автокорреляция по-прежнему имеет место, то, возможно, это связано с внутренними свойствами ряда возмущений $\{\varepsilon_t\}$. В этом случае можно воспользоваться авторегрессионным преобразованием. В линейной регрессионной модели либо в моделях, сводящихся к линейной, наиболее целесообразным и простым преобразованием является *авторегрессионная схема первого порядка AR(1)*.

Для простоты изложения AR(1) рассмотрим модель парной линейной регрессии

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (5.11)$$

Тогда наблюдениям t и $(t-1)$ соответствуют формулы:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t; \quad (5.12)$$

$$y_{t-1} = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1} \quad (5.13)$$

Пусть случайные возмущения подвержены воздействию авторегрессии первого порядка:

$$\varepsilon_t = \rho \varepsilon_{t-1} + \nu_t, \quad (5.14)$$

где ν_t ($t = 2, 3, \dots, n$) – случайные возмущения, удовлетворяющие всем предпосылкам МНК;

ρ – коэффициент автокорреляции, который известен.

Данная схема оказывается авторегрессионной, поскольку ε определяется значениями этой же величины с запаздыванием, и схемой первого порядка, потому что в этом случае максимальное запаздывание равно единице.

Вычтем из (5.12) соотношение (5.13), умноженное на ρ :

$$y_t - \rho y_{t-1} = \beta_0(1 - \rho) + \beta_1(x_t - \rho x_{t-1}) + (\varepsilon_t - \rho \varepsilon_{t-1}) \quad (5.15)$$

Положив

$$y'_t = y_t - \rho y_{t-1}; \quad x'_t = x_t - \rho x_{t-1}; \quad \beta'_0 = \beta_0(1 - \rho)$$

и с учетом (5.14) получим:

$$y'_t = \beta'_0 + \beta_1 x'_t + \nu_t \quad (5.16)$$

Преобразованное уравнение (5.16) не содержит автокорреляцию, и для оценки ее параметров (β'_0, β_1) используется обычный МНК. Оценка коэффициента β_1 из этой зависимости непосредственно используется и для исходного уравнения, а коэффициент β_0 рассчитывается по формуле $\beta_0 = \beta'_0 / (1 - \rho)$.

Способ вычисления x'_t, y'_t приводит к потере первого наблюдения. Число степеней свободы уменьшается на единицу, что при малых выборках может привести к потере эффективности. Эта проблема обычно преодолевается с помощью *поправки Прайса–Винстена*:

$$x'_1 = \sqrt{1 - \rho^2} \cdot x_1, \quad y'_1 = \sqrt{1 - \rho^2} \cdot y_1 \quad (5.17)$$

Отметим, что авторегрессионное преобразование может быть обобщено на произвольное число объясняющих переменных, т. е. использовано для уравнения множественной регрессии. Авторегрессионное преобразование первого порядка $AR(1)$ может быть обобщено на преобразования более высоких порядков $AR(2), AR(3)$ и т. д.:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \nu_t \quad (5.18)$$

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \rho_3 \varepsilon_{t-3} + \nu_t$$

На практике значение коэффициента ρ обычно неизвестно и его необходимо оценивать. Рассмотрим несколько методов оценивания.

Определение ρ на основе статистики Дарбина–Уотсона.

Согласно соотношению (5.10)

$$DW \approx 2(1 - r_{e_t e_{t-1}}).$$

Тогда в качестве оценки коэффициента ρ может быть взят коэффициент $r = r_{e_t e_{t-1}}$.

$$\text{Имеем: } r \approx 1 - \frac{DW}{2}.$$

В этом случае оценка r параметра ρ будет достаточно точной.

Метод Кохрана–Оркатта. Итерационный процесс, называемый методом Кохрана–Оркатта, включает следующие этапы (на примере парной регрессии):

1) применяя МНК к исходному уравнению регрессии, получают первоначальные оценки параметров β_0 и β_1 ;

2) вычисляют остатки e_t и в качестве оценки ρ используют коэффициент автокорреляции остатков первого порядка.

3) применяют МНК к преобразованному уравнению, получают новые оценки параметров β_0 и β_1 .

Процесс обычно заканчивается, когда очередное приближение ρ мало отличается от предыдущего.

Метод Хилдрета–Лу. По данному методу преобразованное уравнение оценивают для каждого значения ρ из интервала $(-1; 1)$ с заданным шагом внутри его (например, 0,001; 0,01 и т. д.). Выбирают то значение ρ , для которого сумма квадратов отклонений в преобразованном уравнении минимальна, а коэффициенты регрессии определяют при оценивании преобразованного уравнения с использованием этого значения.

► **Пример 5.4.** Изучается зависимость расходов на потребление Y от доходов X по данным за 20 лет. Требуется: а) оценить уравнение регрессии Y по X ; б) проверить наличие автокорреляции с помощью критерия Дарбина–Уотсона; в) при наличии автокорреляции переоценить уравнение регрессии, применяя авторегрессионную схему первого порядка.

Таблица 5.6 – Исходные данные

t	1	2	3	4	5	6	7	8	9	10
x_t	68	73	76	83	89	95	100	107	108	113
y_t	65	70	73	78	83	86	89	96	96	103

t	1	2	3	4	5	6	7	8	9	10
x_t	119	121	122	131	135	139	142	145	148	151
y_t	109	112	114	115	118	122	128	131	134	139

Решение. По МНК находим уравнение регрессии

$$\hat{y} = 6,6396 + 0,8513x.$$

Для расчета статистики Дарбина–Уотсона составим вспомогательную таблицу (табл. 5.7).

Таблица 5.7 – Расчет статистики Дарбина–Уотсона

t	x_t	y_t	$\hat{y}_t$	$e_t = y_t - \hat{y}_t$	e_{t-1}	$(e_t - e_{t-1})^2$	e_t^2
1	68	65	64,5280	0,4720	–	–	0,2228
2	73	70	68,7845	1,2155	0,4720	0,5528	1,4774
3	76	73	71,3384	1,6616	1,2155	0,1990	2,7609
4	83	78	77,2975	0,7025	1,6616	0,9199	0,4935
5	89	83	82,4053	0,5947	0,7025	0,0116	0,3537
6	95	86	87,5131	-1,5131	0,5947	4,4428	2,2895
7	100	89	91,7696	-2,7696	-1,5131	1,5788	7,6707
8	107	96	97,7287	-1,7287	-2,7696	1,0835	2,9884
9	108	96	98,5800	-2,5800	-1,7287	0,7247	6,6564
10	113	103	102,8365	0,1635	-2,5800	7,5268	0,0267
11	119	109	107,9443	1,0557	0,1635	0,7960	1,1145
12	121	112	109,6469	2,3531	1,0557	1,6832	5,5371
13	122	114	110,4982	3,5018	2,3531	1,3195	12,2626
14	131	115	118,1599	-3,1599	3,5018	44,3782	9,9850
15	135	118	121,5651	-3,5651	-3,1599	0,1642	12,7099
16	139	122	124,9703	-2,9703	-3,5651	0,3538	8,8227
17	142	128	127,5242	0,4758	-2,9703	11,8756	0,2264
18	145	131	130,0781	0,9219	0,4758	0,1990	0,8499
19	148	134	132,6320	1,3680	0,9219	0,1990	1,8714
20	151	139	135,1859	3,8141	1,3680	5,9834	14,5474
Сумма						83,9919	92,8668

По формуле (5.9) статистика Дарбина–Уотсона

$$DW = \frac{83,9919}{92,8668} = 0,90.$$

По таблице III приложений при $n = 20$, $p = 1$ критические значения $d_H = 1,20$; $d_B = 1,41$. Фактически найденное значение $DW = 0,90$ находится в пределах от 0 до d_H ($0 < 0,90 < 1,20$). Это свидетельствует о наличии положительной автокорреляции возмущений.

Для устранения автокорреляции используем авторегрессионную схему первого порядка $AR(1)$. Вместо переменных X , Y рассмотрим новые переменные X' , Y' , значения которых вычислим по следующим правилам:

$$x'_1 = \sqrt{1 - \rho^2} \cdot x_1; \quad y'_1 = \sqrt{1 - \rho^2} \cdot y_1;$$

$$x'_t = x_t - \rho x_{t-1}, \quad y'_t = y_t - \rho y_{t-1}, \quad t = 2, 3, \dots, 20.$$

Коэффициент ρ определим на основе статистики Дарбина–Уотсона:

$$\rho = 1 - \frac{DW}{2} = 1 - \frac{0,90}{2} = 0,55.$$

Результаты вычислений приведены в таблице 5.8.

Таблица 5.8 – Расчет значений переменных X' , Y'

t	x_t	y_t	x'_t	y'_t
1	68	65	56,79	54,29
2	73	70	35,60	34,25
3	76	73	35,85	34,50
4	83	78	41,20	37,85
5	89	83	43,35	40,10
6	95	86	46,05	40,35
7	100	89	47,75	41,70
8	107	96	52,00	47,05
9	108	96	49,15	43,20
10	113	103	53,60	50,20
11	119	109	56,85	52,35
12	121	112	55,55	52,05
13	122	114	55,45	52,40
14	131	115	63,90	52,30
15	135	118	62,95	54,75
16	139	122	64,75	57,10
17	142	128	65,55	60,90
18	145	131	66,90	60,60
19	148	134	68,25	61,95
20	151	139	69,60	65,30

С помощью МНК по переменным X' и Y' оценим коэффициенты уравнения $y' = b'_0 + b_1x'$. Получим $b_1 = 0,8649$, $b'_0 = 2,4728$.

Далее находим b_0 по формуле:

$$b_0 = \frac{b'_0}{1 - \rho} = \frac{2,4728}{1 - 0,55} = 5,4951.$$

Получим уравнение регрессии $\hat{y} = 5,4951 + 0,8649x$. ►

Задачи для самостоятельного решения

5.5. Постройте и протестируйте на отсутствие гетероскедастичности модель зависимости государственных расходов на образование от объема ВВП.

Таблица 5.9 – Государственные расходы на образование в зависимости от объема ВВП

i	x_i	y_i	i	x_i	y_i
1	5,67	0,34	18	211,78	4,79
2	10,13	0,22	19	186,33	5,46
3	66,97	1,60	20	261,41	18,90
4	115,97	6,40	21	40,15	0,75
5	11,34	0,32	22	51,62	2,80
6	18,88	1,23	23	57,71	4,90
7	20,94	1,81	24	66,32	4,45
8	23,83	1,27	25	76,88	4,26
9	27,56	0,67	26	63,03	3,50
10	22,16	1,02	27	101,65	5,31
11	24,67	1,07	28	249,72	8,92
12	27,57	1,25	29	395,52	15,95
13	119,49	7,15	30	534,97	29,90
14	124,15	11,22	31	655,29	33,59
15	140,98	8,66	32	815,00	38,62
16	153,85	5,56	33	1040,45	61,61
17	169,38	13,41	34	2586,40	181,30

5.6. Исследуется зависимость между доходом X домохозяйства и его расходом Y на продукты питания. Выборочные данные по 40 домохозяйствам представлены в таблице 5.10.

а) Постройте уравнение регрессии Y по X .

б) Проведите графический анализ отклонений и сделайте предположение о наличии (или об отсутствии) гетероскедастичности в модели.

в) Проведите анализ модели на гетероскедастичность по тесту ранговой корреляции Спирмена.

г) Постройте новое уравнение регрессии, используя взвешенный метод наименьших квадратов.

Таблица 5.10 – Зависимость между доходом X домохозяйства и
расходом Y на продукты питания

X	Y	X	Y	X	Y
25,5	14,5	45,5	13,8	75,8	26,1
26,5	11,3	48,3	16,0	76,9	20,0
27,2	14,7	49,5	18,2	79,2	19,8
29,6	10,2	52,3	19,1	81,5	21,2
35,7	13,5	55,7	16,3	82,4	29,0
38,6	9,9	59,0	17,5	82,8	17,3
39,0	12,4	61,0	10,9	83,0	23,5
39,3	8,6	61,7	16,1	85,9	22,0
40,0	10,3	62,5	10,5	86,4	18,3
41,9	13,9	64,7	10,6	86,9	13,7
42,5	14,9	69,7	29,0	88,3	14,5
44,2	11,6	71,2	8,2	89,0	27,3
44,8	21,5	73,8	14,3		
45,5	10,8	74,7	21,8		

5.7. Имеются статистические данные за 25 лет по темпам прироста заработной платы Y (%), производительности труда X_1 (%) и уровню инфляции X_2 (%).

Таблица 5.11 – Исходные данные к задаче 5.7

Год	X_1	X_2	Y	Год	X_1	X_2	Y
1	3,5	4,5	9,0	14	2,3	6,9	12,5
2	2,8	3,0	6,0	15	2,8	3,5	6,7
3	6,3	3,1	8,9	16	1,5	7,1	8,5
4	4,5	3,8	9,0	17	6,0	3,1	5,9
5	3,1	3,8	7,1	18	2,9	3,7	6,8
6	1,5	1,1	3,2	19	2,8	3,9	5,6
7	7,6	2,3	6,5	20	2,6	4,0	4,8
8	6,7	3,6	9,1	21	1,5	4,8	4,5
9	4,2	7,5	14,6	22	0,9	4,8	6,7
10	2,7	8,0	11,9	23	0,6	4,2	5,5
11	4,5	3,9	9,2	24	0,7	4,9	4,0
12	3,5	4,7	8,8	25	3,1	3,2	3,3
13	5,0	6,1	12,0				

Оцените по МНК уравнение регрессии Y по X_1 и X_2 . Оцените качество построенного уравнения, выполнив проверку на наличие гетероскедастичности и автокорреляции.

5.8. По статистическим данным за 30 лет, приведенным в таблице 5.12, необходимо:

- а) оценить по МНК уравнение регрессии Y по X_1 и X_2 ;
- б) оценить статистическую значимость коэффициентов и общее качество уравнения регрессии;
- в) используя статистику Дарбина–Уотсона, определить наличие автокорреляции остатков для построенного уравнения;
- г) при наличии автокорреляции переоценить уравнение регрессии, используя для этого авторегрессионную схему первого порядка $AR(1)$.

Таблица 5.12 – Исходные данные к задаче 5.8

X_1	X_2	Y	X_1	X_2	Y
18,12	11,55	8,58	17,50	21,95	23,75
11,05	13,25	10,45	16,43	23,10	25,75
9,00	10,90	8,35	7,40	25,65	24,20
17,00	10,45	10,65	15,45	26,15	25,20
16,25	15,10	9,70	19,15	25,55	26,20
13,80	17,50	12,00	5,45	28,10	28,60
19,95	17,77	13,45	9,52	24,20	30,60
18,74	16,10	14,20	7,95	32,30	31,32
13,80	10,59	14,45	7,45	21,50	26,00
9,55	10,65	13,85	19,90	22,95	26,85
19,30	9,32	16,55	8,65	30,45	32,10
15,20	11,00	18,00	21,35	24,60	32,95
12,40	15,05	18,40	11,11	32,50	33,30
16,50	15,10	20,40	15,82	31,20	33,85
5,95	22,70	21,00	21,67	29,50	35,60

Контрольные вопросы

1. Сформулируйте основные предпосылки классической регрессионной модели.
2. В чем суть гетероскедастичности?
3. Каковы последствия гетероскедастичности?
4. Как графически обнаружить гетероскедастичность?
5. Приведите схему теста Спирмена.
6. Приведите схему теста Голдфелда–Квандта.
7. В чем суть метода взвешенных наименьших квадратов?
8. Что такое автокорреляция?
9. Какая предпосылка МНК нарушается при автокорреляции?
10. Назовите основные причины автокорреляции в регрессионных моделях.
11. Каковы последствия автокорреляции?
12. Перечислите основные методы обнаружения автокорреляции?
13. Опишите схему использования теста Дарбина–Уотсона.
14. Опишите авторегрессионную схему первого порядка.
15. В чем смысл поправки Прайса–Винстена?
16. Опишите способы определения коэффициента автокорреляции ρ в авторегрессионной схеме первого порядка $AR(1)$.

Набор факторов в каждом уравнении может варьировать. Отсутствие того или иного фактора в уравнении системы может быть следствием как экономической нецелесообразности его включения в модель, так и несущественности его воздействия на зависимую переменную. Примером такой модели может служить модель экономической эффективности сельскохозяйственного производства, где в качестве зависимых переменных выступают показатели, характеризующие эффективность сельскохозяйственного производства, – продуктивность коров, себестоимость 1 ц молока, а в качестве факторов – специализация хозяйства, количество голов на 100 га пашни, затраты труда и т. п.

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

[illegible]

Для оценки параметров системы (6.3) применяют МНК.

3. Система взаимозависимых уравнений – когда одни и те же переменные одновременно рассматриваются как зависимые в одних уравнениях и как независимые в других.

[illegible]

Система взаимозависимых уравнений получила название *системы совместных, одновременных уравнений*.

Для нахождения параметров системы (6.4) МНК не применим.

6.2 Структурная и приведенная формы модели

Система совместных, одновременных уравнений (6.4) называется *структурной формой модели*.

Параметры структурной формы модели называют *структурными коэффициентами*.

Уравнения, составляющие исходную модель, называют *структурными уравнениями* модели. Обычно их подразделяют на *поведенческие уравнения* и *уравнения – тождества*. В поведенческих уравнениях описываются взаимодействия между переменными. Тождества представляют собой соотношения, которые должны выполняться во всех случаях. Они не содержат подлежащие оценке параметры и случайные составляющие.

При рассмотрении систем одновременных уравнений переменные делятся на два больших класса – *экзогенные* и *эндогенные* переменные.

Экзогенные переменные – это независимые переменные, значения которых определяются вне модели. *Эндогенные переменные* – это взаимозависимые переменные, значения которых определяются внутри модели, т. е. они формируют свои значения, подчиняясь уравнениям системы. Например, в системе (6.1) функции спроса и предложения, условие равновесия определяют величины спроса Q_t^D , предложения Q_t^S и цену P_t . Все эти переменные являются эндогенными, так как они определяются внутри системы. Значения переменной Y определяются вне модели. Из модели нельзя понять, как формируются ее значения, следовательно, она является экзогенной.

В качестве экзогенных переменных могут рассматриваться значения эндогенных переменных за предыдущий период времени, т. е. *лаговые* эндогенные переменные. Экзогенные и лаговые эндогенные переменные, значения которых определены до рассмотрения соотношения, называются *предопределенными переменными*. Например, в уравнении предложения (6.1₁) переменная P_{t-1} (цена товара в предыдущий момент времени) является предопределенной переменной.

Предполагается, что в каждом уравнении экзогенные переменные некоррелированы со случайным членом. В общем случае эндогенные переменные коррелированы со случайным членом, поэтому применение МНК к структурной форме модели приводит к смещенным и несостоятельным оценкам структурных коэффициентов. Для

определения структурных коэффициентов структурная форма модели преобразуется в приведенную форму.

Приведенная форма модели представляет собой систему уравнений, в каждом из которых эндогенные переменные выражены только через экзогенные переменные и случайные составляющие:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + \dots + \delta_{1m}x_m + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + \dots + \delta_{2m}x_m + u_2, \\ \\ y_n = \delta_{n1}x_1 + \delta_{n2}x_2 + \dots + \delta_{nm}x_m + u_n \end{cases} \quad (6.5)$$

Параметры δ_{ij} ($i=1,2,\dots,n$; $j=1,2,\dots,m$) приведенной формы модели называются *коэффициентами приведенной формы*. Они оцениваются обычным МНК, поскольку экзогенные переменные некоррелированы со случайным членом.

6.3 Проблема идентифицируемости

При переходе от приведенной формы модели к структурной форме возникает проблема идентифицируемости.

Под *проблемой идентифицируемости* понимается возможность численной оценки параметров структурных уравнений по оценкам коэффициентов приведенных уравнений.

С позиции идентифицируемости структурные модели можно подразделить на три вида:

- идентифицируемые,
- неидентифицируемые,
- сверхидентифицируемые.

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

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

Модель *сверхидентифицируема*, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае

на основе коэффициентов приведенной формы можно получить два и более значений одного структурного коэффициента.

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

Для быстрого формального определения идентифицируемости структурных уравнений применяются необходимые и достаточные условия.

Необходимое условие может быть записано в виде следующего счётного правила:

$D+1 = H$ – уравнение идентифицируемо;

$D+1 < H$ – уравнение неидентифицируемо;

$D+1 > H$ – уравнение сверхидентифицируемо;

где D – число экзогенных переменных, которые содержатся в системе, но не входят в данное уравнение;

H – число эндогенных переменных в уравнении.

Достаточное условие: Уравнение идентифицируемо, если по отсутствующим в нем переменным (эндогенным и экзогенным) можно из коэффициентов при них в других уравнениях системы получить матрицу, определитель которой не равен нулю, а ранг матрицы не меньше, чем число эндогенных переменных в системе без одного.

6.4 Оценивание параметров структурной модели

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

- косвенный метод наименьших квадратов,
- двухшаговый метод наименьших квадратов,
- метод инструментальных переменных.

Косвенный метод наименьших квадратов применяется в случае точно идентифицируемой структурной модели. Процедура применения этого метода предполагает выполнение следующих этапов:

1. Структурная модель преобразовывается в приведенную форму модели.

2. Для каждого уравнения приведенной формы модели обычным МНК оцениваются приведенные коэффициенты (δ_{ij}).

3. Путем алгебраических преобразований переходят от приведенной формы к уравнениям структурной формы модели, получая тем самым численные оценки структурных параметров.

Двухшаговый метод наименьших квадратов используется для оценки структурных коэффициентов сверхидентифицируемой модели и заключается в следующем:

1. Составляют приведенную форму модели и определяют численные значения параметров каждого ее уравнения обычным МНК.

2. Выявляют эндогенные переменные, находящиеся в правой части структурного уравнения, параметры которого определяют двухшаговым методом наименьших квадратов, и находят расчетные значения этих переменных по соответствующим уравнениям приведенной формы модели.

3. Обычным МНК определяют параметры структурного уравнения, используя в качестве исходных данных фактические значения экзогенных переменных и расчетные значения эндогенных переменных, стоящих в правой части данного структурного уравнения.

Сверхидентифицируемая структурная модель может быть двух типов:

- все уравнения системы сверхидентифицируемы;
- система содержит наряду со сверхидентифицируемыми точно идентифицируемые уравнения.

Если все уравнения системы сверхидентифицируемы, то для оценки структурных коэффициентов каждого уравнения используется двухшаговый метод наименьших квадратов. Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

Контрольные вопросы

1. Назовите возможные способы построения систем уравнений. Чем они отличаются друг от друга?
2. Как связаны между собой структурная и приведенная формы модели?
3. Какие переменные системы называются эндогенными? экзогенными?
4. В чем суть проблемы идентифицируемости?
5. Сформулируйте необходимые и достаточные условия идентифицируемости.
6. Раскройте суть косвенного метода наименьших квадратов.
7. В каких случаях используется двухшаговый метод наименьших квадратов? Раскройте его содержание.

Тестовые задания

1. Особенность эконометрики как прикладной науки заключается в _____ существующих взаимосвязей социально-экономических показателей и систем.

- а) количественном измерении;
- б) формулировании теорий;
- в) качественном описании;
- г) схематическом описании.

2. По типу функциональной зависимости между переменными различают _____ уравнения регрессии.

- а) линейные и нелинейные;
- б) множественные и парные;
- в) стохастические и вероятностные;
- г) линейные и парные.

3. Уравнение регрессии может включать одну или несколько независимых переменных. По данному классификационному признаку различают _____ регрессии.

- а) простую и парную;
- б) линейную и нелинейную;
- в) множественную и многофакторную;
- г) простую и множественную.

4. Простая линейная регрессия предполагает ...

- а) наличие двух и более факторов и нелинейность уравнения регрессии;
- б) наличие одного фактора и линейность уравнения регрессии;
- в) наличие одного фактора и нелинейность уравнения регрессии;
- г) наличие двух и более факторов и линейность уравнения регрессии.

5. В модели вида $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$ количество объясняющих переменных равно ...

- а) 4;
- б) 3;
- в) 2;
- г) 1.

6. Модель вида $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$ является _____ моделью _____ регрессии.

- а) линейной ... множественной;
- б) полиномиальной ... парной;
- в) множественной ... полиномиальной;
- г) полиномиальной ... множественной.

7. Под верификацией модели понимается ...

- а) спецификация модели;
- б) оценка параметров модели;
- в) сбор статистической информации об объеме исследования;
- г) проверка адекватности модели.

8. К основным задачам спецификации модели относится ...

- а) сбор необходимой статистической информации;
- б) выбор вида уравнения регрессии;
- в) прогнозирование значений зависимой переменной;
- г) проверка адекватности модели.

9. Для регрессионной модели несмещенность оценки параметра означает, что ее выборочное математическое ожидание равно ...

- а) свободному члену уравнения регрессии;
- б) оцениваемому параметру, рассчитанному по генеральной совокупности;
- в) коэффициенту парной корреляции между зависимой переменной и соответствующей независимой переменной;
- г) математическому ожиданию остатков модели.

10. Для регрессионной модели состоятельность оценки параметра означает, что при увеличении выборки значение оценки параметра стремиться к ...

- а) оцениваемому параметру, рассчитанному по другой выборке, объем которой значительно меньше исходной совокупности данных;
- б) коэффициенту парной корреляции между зависимой переменной и соответствующей независимой переменной;
- в) свободному члену уравнения регрессии;
- г) истинному значению параметра, вычисленному для генеральной совокупности.

11. Предпосылкой регрессионного анализа является ...

- а) автокорреляция в остатках;
- б) функциональная связь между зависимой и независимой переменными;
- в) гомоскедастичность остатков;
- г) присутствие в эконометрической модели более двух факторов.

12. Пусть t – рассчитанная для коэффициента регрессии статистика Стьюдента, а $t_{\text{табл.}}$ – табличное значение этой статистики. Коэффициент регрессии считается статистически значимым, если ...

- а) $|t| > t_{\text{табл.}}$;
- б) $|t| < t_{\text{табл.}}$;
- в) $t < t_{\text{табл.}}$;
- г) $t < |t_{\text{табл.}}|$.

13. При линейаризации нелинейных регрессионных моделей как один из видов преобразований используется замена переменных. Указанным способом может быть линейаризовано уравнение ...

- а) $y = b_0 x^{b_1}$;
- б) $y = b_0 + b_1 \cdot x^{-1}$;
- в) $y = e^{b_0 + b_1 x}$;
- г) $y = b_0 + b_1 x_1 + b_2 x_2$.

14. Для нелинейного уравнения регрессии значение индекса детерминации $R^2 = 0,9$. Следовательно, доля остаточной дисперсии в общей дисперсии зависимой переменной составляет ...

- а) 90%;
- б) 90;
- в) 10;
- г) 10%.

15. Если параметр эконометрической модели не является статистически значимым, то его значение признается ...

- а) равным коэффициенту парной корреляции;
- б) равным 0;
- в) отличным от 0;
- г) равным 1.

16. При МНК параметры уравнения парной линейной регрессии $\hat{y} = b_0 + b_1 x$ определяются из условия _____ остатков.

- а) равенства нулю суммы квадратов;
- б) минимизации суммы квадратов;
- в) минимизации модулей;
- г) равенства нулю.

17. Если общая сумма квадратов отклонений $\sum (y - \bar{y})^2 = 120$, остаточная сумма квадратов отклонений $\sum (y - \hat{y})^2 = 30$, то сумма квадратов отклонений, объясненная регрессией, будет равна ...

- а) 0,25;
- б) 4;
- в) 150;
- г) 90.

18. Из перечисленного условием выполнения предпосылок регрессионного анализа не является _____ остатков.

- а) нулевая средняя величина;
- б) случайный характер;
- в) гетероскедатичность;
- г) отсутствие автокорреляции.

19. Если параметр эконометрической модели не является статистически значимым, то соответствующая независимая переменная ...

- а) оказывает статистически значимое влияние на зависимую переменную;
- б) не оказывает влияния на зависимую переменную;
- в) тесно связана с зависимой переменной;
- г) оказывает основное (доминирующее) влияние на зависимую переменную.

20. Если зависимость объема спроса от цены характеризуется постоянной эластичностью, то моделирование целесообразно проводить на основе ...

- а) равносторонней гиперболы;
- б) экспоненциальной функции;
- в) степенной функции;
- г) параболы второй степени.

21. При выполнении основных предпосылок регрессионного анализа оценки параметров модели, рассчитанные с помощью МНК, обладают свойствами ...

- а) состоятельности, смещенности и неэффективности;
- б) состоятельности, смещенности и эффективности;
- в) несостоятельности, смещенности и эффективности;
- г) состоятельности, несмещенности и эффективности.

22. Для модели парной регрессии коэффициент детерминации $R^2=0,831$. На дисперсию зависимой переменной, объясненную уравнением, приходится _____ общей дисперсии зависимой переменной.

- а) 0,169 %;
- б) 16,9 %;
- в) 83,1 %;
- г) 0,831 %.

23. F -статистика рассчитывается как отношение _____ дисперсии к _____ дисперсии, рассчитанных на одну степень свободы.

- а) факторной ... остаточной;
- б) остаточной ... факторной;
- в) факторной ... к общей;
- г) остаточной ... общей.

24. Если общая сумма квадратов отклонений $\sum (y - \bar{y})^2 = 150$, остаточная сумма квадратов отклонений $\sum (y - \hat{y})^2 = 30$, то значение коэффициента детерминации равно ...

- а) 0,2;
- б) $\sqrt{0,8}$;
- в) 0,8;
- г) $\sqrt{0,2}$.

25. Для линеаризации модели $y = b_0 \cdot x^{b_1} \cdot \varepsilon$ используется ...

- а) замена переменных;
- б) приведение уравнения к виду $1/y$;
- в) логарифмирование;
- г) потенцирование.

26.Основной предпосылкой регрессионного анализа не является:
а)математическое ожидание случайного отклонения равно 0 для всех наблюдений;
б)случайные отклонения являются независимыми друг от друга;
в)дисперсия случайных отклонений постоянна для всех наблюдений;
г)случайные отклонения коррелируют друг с другом.

27.Гомоскедастичность остатков подразумевает выполнение следующей предпосылки регрессионного анализа ...

а)дисперсия случайных отклонений постоянна для всех наблюдений;
б)математическое ожидание случайного отклонения равно 0 для всех наблюдений;
в)случайные отклонения должны быть независимы от объясняющих переменных;
г)случайные отклонения являются независимыми друг от друга.

28.Оценено уравнение регрессии $\hat{y} = -70,85 + 0,88x$. t -статистика для коэффициентов b_0 и b_1 равна $-5,89$ и $5,9$ соответственно. Какое утверждение будет верным, если для 5%-го уровня значимости $t_{\text{табл.}} = 2,4$?

а)коэффициенты b_0 и b_1 незначимы;
б)коэффициенты b_0 и b_1 значимы;
в)коэффициент b_0 – значим, коэффициент b_1 – незначим;
г)коэффициент b_0 – незначим, коэффициент b_1 – значим.

29.Дано уравнение регрессии: $\hat{y} = 5 - 1,2x$. Известны среднеквадратичные отклонения для переменных и количество наблюдений: $\sigma_y = 0,64$, $\sigma_x = 0,36$, $n = 16$. Чему равен коэффициент корреляции r_{xy} ?

а) $r_{xy} = 0,9$;
б) $r_{xy} = 0,675$;
в) $r_{xy} = -0,675$;
г) $r_{xy} = -0,9$.

30.Коэффициент регрессии в уравнении $\hat{y} = 9,2 + 1,5x$, характеризующем связь между X (млн. руб.) и Y (млн. руб.) означает, что при увеличении X на 1 млн. руб. Y увеличивается в среднем на ...

- а) 1,5 %;
- б) 0,5 млн. руб.;
- в) 1,5 тыс. руб.;
- г) 1,5 млн. руб.

31. Известно, что между величинами X и Y существует обратная связь. В каких пределах будет находиться парный коэффициент корреляции?

- а) от -1 до 0 ;
- б) от 0 до 1 ;
- в) от -1 до 1 ;
- г) от $-0,5$ до $0,5$.

32. Величина индекса корреляции, равная 1,587, свидетельствует...

- а) о слабой зависимости;
- б) о сильной зависимости;
- в) об ошибках в вычислениях;
- г) об умеренной зависимости.

33. Если доверительный интервал для параметра проходит через точку ноль, следовательно ...

- а) параметр может принимать как отрицательные, так и положительные значения;
- б) параметр является значимым;
- в) параметр является существенным;
- г) параметр признается статистически незначимым.

34. Линеаризация подразумевает процедуру ...

- а) приведения уравнения множественной регрессии к парной;
- б) приведения нелинейного уравнения к линейному виду;
- в) приведения линейного уравнения к нелинейному виду;
- г) приведения уравнения парной регрессии к множественной.

35. Значение коэффициента корреляции равно 0,9. Следовательно, значение коэффициента детерминации составит ...

- а) 0,3;
- б) 0,81;
- в) 0,95;
- г) 0,1.

36. Если справедливо равенство $r_{xy} > 0$, где r_{xy} – коэффициент корреляции, то ...

- а) при возрастании переменной X значения переменной Y в среднем положительны;
- б) переменные X и Y положительны;
- в) переменные X и Y имеют одинаковые знаки;
- г) при возрастании переменной X значение переменной Y в среднем возрастает.

37. При оценке параметров парной линейной регрессии получены коэффициенты: $b_0 = 160$, $b_1 = 5$ и значения t -статистик: $t_{b_0} = 8$, $t_{b_1} = 10$. Чему будут равны стандартные ошибки коэффициентов b_0 и b_1 ?

- а) 0,05 и 2;
- б) 20 и 0,5;
- в) 0,02 и 5;
- г) 50 и 0,2.

38. В каком случае уравнение парной регрессии признается статистически значимым?

- а) $F_{табл} < F_{факт}$;
- б) $F_{табл} > F_{факт}$;
- в) $F_{факт} > 0$;
- г) $F_{факт} < 0$.

39. В рамках модели $y = \beta_0 + \beta_1 x + \varepsilon$ по формуле $cov(x, y) / \sigma_x^2$ вычисляется

- а) оценка параметра β_0 ;
- б) оценка случайного возмущения ε ;
- в) оценка среднего значения y ;
- г) оценка параметра β_1 .

40. Если зависимость спроса Y (усл. ед.) на товар от цены X (усл. ед.) товара характеризуется уравнением $\hat{y} = 105,2 x^{-1,12}$, то ...

- а) с увеличением цены на 1 усл. ед. спрос снизится в среднем на 1,12 усл. ед.;
- б) с увеличением цены на 1 усл. ед. спрос возрастет в среднем на 105,2 усл. ед.;
- в) с увеличением цены на 1 %. спрос снизится в среднем на 1,12 %;

г) с увеличением цены на 1%. спрос в среднем повысится на 1,12%.

41. Если в рамках модели $y = \beta_0 + \beta_1 x + \varepsilon$ при уровне значимости 0,05 гипотеза $H_0: \beta_1 = 0$ отвергается, то можно утверждать, что ...

- а) с вероятностью 95% связь между X и Y существует;
- б) с вероятностью 95% связь между X и Y отсутствует;
- в) с вероятностью 5% связь между X и Y существует;
- г) с вероятностью 5% связь между X и Y отсутствует.

42. Отбор факторов в эконометрическую модель множественной регрессии может быть осуществлен на основе ...

- а) матрицы парных коэффициентов корреляции;
- б) значений коэффициентов автокорреляции уровней ряда различных порядков;
- в) сравнения коэффициентов «чистой» регрессии;
- г) сравнения коэффициентов эластичности.

43. Фиктивные переменные могут принимать значения ...

- а) 1 и 0;
- б) 2;
- в) -1 и 1;
- г) любые значения.

44. Для оценки параметров регрессионной модели с гетероскедастичными остатками используется _____ МНК.

- а) двухшаговый;
- б) косвенный;
- в) взвешенный;
- г) традиционный.

45. В линейной регрессии $\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_jx_j + \dots + b_kx_k$ коэффициентом, характеризующим среднее изменение зависимой переменной при изменении независимой переменной на 1 единицу измерения, является ...

- а) b_0 ;
- б) b_j ;
- в) y ;
- г) x_j .

46. Ошибкой спецификации эконометрической модели уравнения регрессии является ...

- а) учет случайных факторов;
- б) расчет показателей качества модели;
- в) оценка параметров при помощи МНК;
- г) использование парной регрессии вместо множественной.

47. Значение коэффициента множественной корреляции находится в интервале ...

- а) $[0; 1]$;
- б) $[-1; 1]$;
- в) $[0; +\infty]$;
- г) $[-1; 0]$.

48. По результатам 50 наблюдений построено уравнение множественной регрессии $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$. Число степеней свободы остаточной суммы квадратов отклонений для этого уравнения равно ...

- а) 49;
- б) 48;
- в) 46;
- г) 47.

49. При моделировании линейного уравнения множественной регрессией вида $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$ необходимо, чтобы выполнялось требование отсутствия взаимосвязи между ...

- а) b_1 и b_2 ;
- б) x_1 и x_2 ;
- в) b_0 и $\{b_1; b_2\}$;
- г) y и $\{x_1; x_2\}$.

50. Примерами фиктивных переменных в эконометрической модели зависимости дохода работника предприятия от ряда факторов могут выступать ...

- а) стаж работы (количество лет, месяцев);
- б) уровень образования (начальное, среднее, высшее);
- в) величина среднемесячной заработной платы;
- г) производительность труда.

51. Укажите требование к факторам, включаемым в модель множественной линейной регрессии:

- а) факторы должны иметь одинаковую размерность;
- б) факторы должны представлять временные ряды;
- в) между факторами не должна существовать высокая корреляция;
- г) между факторами должна существовать высокая корреляция.

52. Под мультиколлинеарностью понимается...

- а) высокая взаимная коррелированность объясняющих переменных;
- б) высокая взаимная коррелированность объясняемых переменных;
- в) высокая коррелированность между объясняющими и объясняемыми переменными;
- г) корреляционная зависимость между последовательными уровнями временного ряда.

53. Если справедлива гипотеза $H_0: \beta_1 = \beta_2 = 0$ относительно коэффициентов β_1 и β_2 модели $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, то переменные X_1 и X_2 являются ...

- а) незначимыми;
- б) необходимыми;
- в) значимыми;
- г) желанными.

54. Если справедлива гипотеза $H_0: \beta_1 = 0$ относительно коэффициента β_1 модели множественной регрессии $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, то целесообразно ...

- а) удалить переменную X_2 из спецификации модели;
- б) сохранить переменную X_1 ;
- в) удалить переменную ε из спецификации модели;
- г) удалить переменную X_1 .

55. Какой из перечисленных методов не может быть применен для обнаружения автокорреляции?

- а) метод рядов;
- б) тест Глейзера;
- в) критерий Дарбина-Уотсона;
- г) графический метод.

56.Статистика Дарбина-Уотсона изменяется в пределах ...

- а)от нуля до четырех;
- б)от нуля до двух;
- в)меньше или равна двум;
- г)от нуля до единицы.

57.Получена производственная функция $Y=2,7 K^{0,8} L^{0,2}$, если объем капитала K увеличить на 1%, то объем производства в среднем

- а) увеличится на 0,8%;
- б) увеличится на 2,7%;
- в) увеличится на 0,2%;
- г) уменьшится на 0,8%.

58.В рамках инвестиционной модели $I_t = 200 + 0,18Y_t - 19R_t$, $R^2=0,82$ текущий доход Y_t и текущая реальная ставка процента R_t объясняют ...

- а) 18% величины I_t ;
- б) 4% величины I_t ;
- в) 50% величины I_t ;
- г) 82% величины I_t .

59.Коэффициент корреляции, характеризующий тесноту связи между двумя переменными при устранении влияния других переменных, называется ...

- а) множественным коэффициентом корреляции;
- б) коэффициентом линейной корреляции;
- в) парным коэффициентом корреляции;
- г) частным коэффициентом корреляции.

60.Факторы, описывающие трендовую компоненту временного ряда, характеризуются ...

- а) периодическим воздействием на величину экономического показателя;
- б) случайным воздействием на уровень временного ряда;
- в) долговременным воздействием на экономический показатель;
- г) сезонным воздействием на экономический показатель.

61.Уровнем временного ряда является ...

- а) значение временного ряда в конкретный момент времени;

- б) среднее значение временного ряда;
- в) совокупность значений временного ряда;
- г) значение конкретного момента времени.

62. Совокупность значений экономического показателя за несколько последовательных моментов (периодов) времени называется ...

- а) коррелограммой;
- б) автокорреляционной функцией;
- в) временным рядом;
- г) тенденцией.

63. Какие значения компонент T , S и ε формируют мультипликативную модель временного ряда ...

- а) $y_t = 7$; $T = 6,5$; $S = 0$; $\varepsilon = 0,5$;
- б) $y_t = 7$; $T = 3,5$; $S = 2$; $\varepsilon = 1$;
- в) $y_t = 7$; $T = 3,5$; $S = -2$; $\varepsilon = 1$;
- г) $y_t = 7$; $T = -3,5$; $S = -2$; $\varepsilon = -1$.

64. Для аддитивной модели временного ряда получено уравнение тренда $T = 3,14 + 2,07 t$. Для уровня y_3 значения сезонной и случайной компонент: $S_3 = 1,6$; $\varepsilon_3 = -0,3$. Значение y_3 будет равно ...

- а) 9,35;
- б) 1,3;
- в) 10,65;
- г) 6,51.

65. Под трендом временного ряда понимают ...

- а) изменение, определяющее общее направление развития;
- б) влияние случайной составляющей на уровень временного ряда;
- в) действия исследователя по приведению исходного временного ряда к стационарному виду;
- г) влияние циклических колебаний на уровень временного ряда.

Список литературы

1. Бородич, С.А. Эконометрика: учебное пособие / С.А. Бородич. – 2-е изд., испр. – Минск: Новое знание, 2004.
2. Бывшев, В.А. Эконометрика: учебное пособие / В.А. Бывшев. – Москва: Финансы и статистика, 2008.
3. Галочкин, В. Т. Эконометрика: учебник и практикум для бакалавриата и специалитета / В.Т. Галочкин. – Москва: Издательство Юрайт, 2019.
4. Евсеев, Е.А. Эконометрика: учебное пособие для вузов / Е.А. Евсеев, В.М. Буре. – 2-е изд., испр. и доп. – Москва: Издательство Юрайт, 2020.
5. Елисеева, И.И. Общая теория статистики: учебник / И.И. Елисеева, М.М. Юзбашев; под ред. И.И. Елисеевой. – 5-е изд., перераб. и доп. – Москва: Финансы и статистика, 2004.
6. Костюнин, В.И. Эконометрика: учебник и практикум для вузов / В.И. Костюнин. – Москва: Издательство Юрайт, 2020.
7. Кремер, Н.Ш. Эконометрика: учебник и практикум для вузов / Н.Ш. Кремер, Б.А. Путко; под редакцией Н.Ш. Кремера. – 4-е изд., испр. и доп. – Москва: Издательство Юрайт, 2020.
8. Новиков, А.И. Эконометрика: учебное пособие / А. И. Новиков. – Москва: ИНФРА-М, 2003.
9. Практикум по эконометрике: учебное пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко [и др.]; под редакцией И.И. Елисеевой. – Москва: Финансы и статистика, 2002.
10. Тимофеев, В.С. Эконометрика: учебник для академического бакалавриата / В.С. Тимофеев, А.В. Фаддеенков, В.Ю. Щеколдин. – 2-е изд., перераб. и доп. – Москва: Издательство Юрайт, 2019.
11. Эконометрика: учебник для вузов / И.И. Елисеева [и др.]; под редакцией И.И. Елисеевой. – Москва: Издательство Юрайт, 2020.

ПРИЛОЖЕНИЯ

Приложение I

Таблица значений F-критерия Фишера при уровне значимости $\alpha = 0,05$

k_2 k_1	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

**Критические значения t-критерия Стьюдента
при уровне значимости 0,10, 0,05, 0,01(двухсторонний)**

Число степеней свободы d.f.	α				Число степеней свободы d.f.	α		
	0,10	0,05	0,01			0,10	0,05	0,01
1	6,3138	12,706	63,657		18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248		19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409		20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041		21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321		22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074		23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995		24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554		25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498		26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693		27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058		28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545		29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123		30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768		40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467		60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208		120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982		∞	1,6449	1,9600	2,5758

Значения d_H и d_B критерия Дарбина – Уотсона на уровне значимости $\alpha = 0,05$

(n – число наблюдений, p – число объясняющих переменных)

n	$p=1$		$p=2$		$p=3$		$p=4$		$p=5$	
	d_H	d_B	d_H	d_B	d_H	d_B	d_H	d_B	d_H	d_B
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,56	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,34	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,58	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

Оксана Викторовна Ментюкова

ЭКОНОМЕТРИКА

Учебное пособие
для студентов, обучающихся
по направлению подготовки 38.03.01 Экономика

Компьютерная верстка *О.В. Ментюковой*

Подписано в печать
Бумага SvetoCopy
Тираж ____ экз.

Формат 60×84 1/16
Усл. печ. л. ____
Заказ № ____

РИО ПГАУ
440014, г. Пенза, ул. Ботаническая, 30