

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
Белорусский национальный технический университет

Кафедра «Маркетинг»

И. Е. Ругалёва

ЭКОНОМЕТРИКА

Пособие
для студентов специальности
1-26 02 03 «Маркетинг»

*Рекомендовано учебно-методическим объединением
по экономическому образованию*

Минск
БНТУ
2023

УДК 330.43(075.8)

ББК 65в6я7

P82

Р е ц е н з е н т ы:

зав. кафедрой экономического развития и менеджмента
Академии управления при Президенте Республики Беларусь,
канд. экон. наук, доцент *Д. Ф. Рутко*;
кафедра информационного менеджмента, маркетинга и учета БГАТУ
(зав. кафедрой, канд. пед. наук, доцент *О. Л. Сапун*)

Ругалёва, И. Е.

P82 Эконометрика : пособие для студентов специальности 1-26 02 03
«Маркетинг» / И. Е. Ругалёва. – Минск : БНТУ, 2023. – 103 с.
ISBN 978-985-583-903-4.

Настоящее пособие предназначено для студентов учреждений высшего образования специальности 1-26 02 03 «Маркетинг». В нем представлены общие положения, методические указания основ эконометрики парной и множественной регрессии, корреляции, типовые задачи, варианты индивидуальных заданий, контрольные вопросы и работы, необходимые статистические таблицы.

УДК 330.43(075.8)

ББК 65в6я7

ISBN 978-985-583-903-4

© Ругалёва И. Е., 2023

© Белорусский национальный
технический университет, 2023

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
1. ОПРЕДЕЛЕНИЕ ЭКОНОМЕТРИКИ, ЕЕ ПРЕДМЕТ И ОБЛАСТЬ ПРИМЕНЕНИЯ	6
1.1. Основные понятия эконометрики	6
1.2. Алгоритм эконометрической модели	8
1.3. Переменные эконометрической модели	10
1.4. Статистическое наблюдение	12
Контрольные вопросы	16
2. ПАРНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ	17
2.1. Парная корреляция	18
2.2. Множественная корреляция	23
2.3. Коэффициент корреляции	25
2.4. Парная регрессия	27
2.5. Множественная регрессия	28
2.6. Статистические свойства параметров	31
2.7. Уравнение регрессии	33
2.8. Показатели качества регрессии	35
2.9. Интервальное оценивание функции регрессии	37
2.10. Примеры решения задач	38
Контрольные вопросы	47
3. ВРЕМЕННЫЕ РЯДЫ	49
3.1. Автокорреляционная функция	51
3.2. Примеры решения задач	55
Контрольные вопросы	62
4. СИСТЕМЫ ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ	63
4.1. Понятие систем эконометрических уравнений	63
4.2. Идентификация систем одновременных уравнений	65
4.3. Типовые задания	70
Контрольные вопросы	72
5. ИНДИВИДУАЛЬНЫЕ ЗАДАНИЯ ПО ТЕМАМ	73
5.1. Множественная регрессия	73
5.2. Временные ряды	75
5.3. Системы одновременных уравнений	75
ТЕСТ 1	77
ТЕСТ 2	81
ЛИТЕРАТУРА	87
ПРИЛОЖЕНИЕ А	88
ПРИЛОЖЕНИЕ Б	89
ПРИЛОЖЕНИЕ В	92

ВВЕДЕНИЕ

В 1926 году норвежский экономист, статист Рагнар Фриш ввел термин «эконометрика», что означает экономическая метрика, занимающая в мировой науке важное место.

Эконометрика как научная дисциплина развивалась, интегрируя математическую экономику и статистику с экономической теорией.

Основным инструментом экономических исследований является аппарат теории вероятностей и математической статистики, реализованный в статистических пакетах прикладных программ. Эконометрика, в отличие от экономической теории, рассматривает количественные, а не качественные аспекты этих явлений. Дисциплина «Эконометрика» относится к государственному компоненту учебного плана, модулю «Статистика и эконометрика». Для ее успешного освоения студент должен обладать знаниями и умениями, полученными при изучении дисциплин «Высшая математика», «Статистика», «Теория вероятностей», «Экономика предприятия (организации)».

Целями дисциплины «Эконометрика» являются:

- обучение студентов основам эконометрического моделирования;
- использование количественных данных наблюдений для выявления закономерностей функционирования экономических систем разного уровня;
- изучение методов оценок и прогнозирования экономических показателей, характеризующих состояние и развитие анализируемых экономических систем;
- освоение современных компьютерных технологий эконометрического анализа и возможностей их применения для решения прикладных экономических задач;
- применение статистических методов, методов моделирования и прогнозирования, социально-экономических процессов для принятия обоснованных управленческих решений.

Задачами изучения дисциплины «Эконометрика» являются:

- изучение основных типов эконометрических моделей, методологии их разработки и практического использования в экономических приложениях;
- изучение теоретических основ и практическое применение методов эконометрического анализа;

- освоение методики подготовки статистических данных для проведения эконометрического анализа, практический опыт их применения для решения типовых задач эконометрики;
- изложение основ эконометрики, представленных регрессионными уравнениями, моделями временных рядов и системами эконометрических уравнений;
- рассмотрение использования экономических методов при изучении взаимосвязей экономических показателей с привлечением положений экономической теории.

1. ОПРЕДЕЛЕНИЕ ЭКОНОМЕТРИКИ, ЕЕ ПРЕДМЕТ И ОБЛАСТЬ ПРИМЕНЕНИЯ

1.1. Основные понятия эконометрики

Эконометрика, возникшая в итоге интеграции статистических методов, экономической теории и математических методов, нацелена на количественную оценку конкретных взаимосвязей, на замену коэффициентов, указываемых в них, конкретными постоянными, т. е. она взаимодействует с конкретными экономическими данными.

Исходя из теории микроэкономики, известно, что она устанавливает связь между ценой на товар и спросом на него, хотя не отвечает на вопрос: на сколько изменится спрос на данный товар в результате изменения его цены на определенную величину? Не дает количественных оценок данной связи.

Эконометрика с помощью математико-статистических методов придает конкретное количественное выражение общим закономерностям, устанавливаемым экономической теорией, связывает между собой экономическую статистику и экономическую теорию.

Аппаратом эконометрики для причинно-следственных взаимосвязей между признаками является система математических соотношений между переменными изучаемого экономического явления, которая объясняет отличия значений текущих переменных от значений определенных ранее других переменных и составляет описание процесса, который рассматривается.

Объект эконометрики – экономика на микро- и макроуровнях, различные экономические явления и взаимосвязи. *Предмет* – методы и модели измерения зависимостей между экономическими явлениями. Эконометрические методы базируются на использовании аппарата теории вероятностей и математической статистики в приложениях к изучению причинно-следственных зависимостей между различными экономическими показателями, представленными статистическими данными.

Эконометрическая модель, которая является объектом исследования, имеет следующий общий вид:

$$Y = f(X_1, X_2, \dots, X_k) + \varepsilon,$$

где $f(X_1, X_2, \dots, X_k)$ – объясненная часть;

Y – объясняемая, зависимая переменная, значение наблюдений признака;

X_1, X_2, \dots, X_k – объясняющие переменные, независимые;

ε – составляющая, случайная величина:

$$\varepsilon = Y - f(X_1, X_2, \dots, X_k).$$

Пример 1. Охарактеризовать уравнение эконометрической модели зависимости спроса на определенный товар от его цены, если известно, что X – цена товара, Y – спрос товара, а уравнение имеет вид

$$Y = 4500 - 3X + \varepsilon.$$

Решение. При увеличении цены товара на единицу спрос уменьшается на три единицы.

Выделяют три вида эконометрических моделей:

1. *Регрессионную модель с одним уравнением.* В таких моделях результативный признак (зависимая переменная) представляется в виде функции факторных признаков (независимых переменных). В уравнении регрессии может быть введен свободный член, и коэффициенты регрессии могут быть найдены методом наименьших квадратов (МНК). Отдельные коэффициенты при переменных могут быть равны нулю.

2. *Модель временных данных.* Результативный признак в таких моделях является функцией переменной времени или переменных, относящихся к другим моментам времени. В уравнении регрессии при выполнении исходных предпосылок коэффициенты регрессии могут быть найдены МНК.

3. *Системы одновременных уравнений.* Эти модели описываются системами взаимосвязанных регрессионных уравнений. Система объясняет, а также прогнозирует столько результативных признаков, сколько поведенческих уравнений входит в систему. Система наиболее распространенная. В ней одни и те же переменные одновременно являются зависимыми в одних уравнениях и независимыми в других. В отличие от предыдущих двух видов, для нахождения параметров этой модели a_{ij}, b_{ij} ($i = 1, 2, \dots, n, j = 1, 2, \dots, m$), называемых структурными коэффициентами, простой МНК неприменим.

В качестве независимых переменных поведенческие уравнения, которые требуют оценивания параметров, включают как факторные, так и результативные признаки уравнений системы.

С помощью эконометрической модели решаются разнообразные задачи:

- прогнозирование социальных, экономических показателей;
- вопросы ценообразования;
- имитация социально-экономического развития системы, выявление изменений различных параметров выходных характеристик.

1.2. Алгоритм эконометрической модели

Алгоритм экономической модели:

1. Определить конечные цели модели, набора участвующих факторных и результативных признаков.

2. Провести качественный (теоретический) анализ сущности изучаемого явления. Формализация априорной (известной до начала моделирования) информации.

3. Выбрать общий вид модели, состав и формы, входящих в нее связей.

4. Собрать необходимую информацию, провести анализ ее качества.

5. Рассчитать оценки параметров модели, т. е. провести идентификацию модели.

6. Оценить качество модели, ее достоверность и надежность. Если качество модели низкое, то вернуться к п. 2 и далее по алгоритму.

Постановка проблемы осуществляется в контексте положений экономической теории, с позиций последних осуществляется спецификация модели, т. е. выбор переменных модели, видов зависимостей между ними и ограничений на коэффициенты уравнений, задающих эти зависимости. Базой для спецификации модели может служить содержательный экономический анализ. В этом случае выбор факторов исследователем осуществляется произвольно в рамках знаний экономического объекта или с помощью статистических методов отбора. Метод исключения предполагает построение регрессионного уравнения, включающего всю совокупность факторов с последующим сокращением их числа до тех пор, пока не выполнится некоторое условие, а именно, не достигнутся удовлетвори-

тельный коэффициент детерминации, статистическая значимость регрессоров, отсутствие мультиколлинеарности. Суть метода включения состоит в последовательном включении факторов в модель до тех пор, пока регрессионная модель не будет отвечать заранее установленному критерию качества.

Фундаментом, основной базой данных, для эконометрических исследований служит официальная статистика или данные бухгалтерского учета. Для информационного обеспечения формируются ряды наблюдений, что предполагает сбор статистических данных по выделенному кругу показателей или в пространстве (предприятия, цеха, отрасли), или во времени (год, месяц, квартал). На этом этапе главным является обеспечение сопоставимости данных в пределах ряда, что требует знания особенностей формирования статистической отчетности.

Например, при прогнозировании спроса на конечную продукцию корректировка статистического ряда требуется при возможной потере или приобретении клиента, появлении товаров-заменителей.

При исследовании зависимости поступлений в бюджет и ВВП как налогооблагаемой базы методологическая сопоставимость ряда поступлений будет обеспечена в том случае, если в рамках наблюдаемого динамического ряда ставка налоговых поступлений остается неизменной, в противном случае ряд поступлений должен быть скорректирован, например, методом пропорциональной корректировки, при условии, что ставка налога сохраняется неизменной на уровне последнего отчетного периода, протяжении всего наблюдения.

В случае нестационарных данных могут потребоваться преобразования, приводящие ряды к стационарному виду. Чаще встречается во временных рядах, которые содержат тренд и могут содержать сезонную компоненту. Чтобы обеспечить требование стационарности, проводят корректировку ряда: удаляют сезонную компоненту и тренд. Система Mathcad, а также современные статистические пакеты программных продуктов дают возможность выполнить указанные преобразования данных в автоматизированном режиме.

Этап *идентификации модели* предусматривает построение оценок коэффициентов (параметров) модели на основе статистических данных и описание свойств этих оценок. Задача этапа состоит в получении статистических оценок этих коэффициентов, на основе ко-

торых определяются регрессионные уравнения, для описания зависимости ожидаемых значений (математических ожиданий) эндогенных переменных через значения предопределенных переменных. Свойства оценок коэффициентов во многом определяются свойствами случайных составляющих в уравнениях, а также свойствами значений предопределенных переменных. От свойств оценок коэффициентов зависит качество модели.

На этапе *верификации модели* проводится проверка ее качества и адекватности модели с помощью специальных тестов или по качеству прогноза. Если тесты указывают на неадекватность спецификации модели, то ее следует пересмотреть с целью устранения этой неадекватности. Для проверки качества модели часто применяют прием, в котором используется некоторая часть известных статистических данных, сравниваемые с предсказанными данными, полученными на основании остальной части статистических данных. При использовании модели в качестве прогноза находится ошибка прогноза, которая определяется с заданной вероятностью.

Если модель была построена на основе содержательного экономического анализа, то она должна быть подвергнута анализу с позиции экономической теории, а это осуществляется на этапе *экономической интерпретации результатов*, где особое значение придается экономической интерпретации коэффициентов регрессионной модели.

1.3. Переменные эконометрической модели

При моделировании экономических процессов используют два типа данных: пространственные данные (cross – sectional data) и временные данные (time – series data).

Пространственными данными является набор сведений по разным объектам, взятым за один и тот же период или момент времени. Например, набор сведений по разным фирмам (объем производства, численность работников, размер основных производственных фондов и пр.) или данные об объеме, ценах потребления некоторого товара по потребителям.

Временными данными являются набор сведений, характеризующий один и тот же объект, но за разные периоды или моменты времени: ежеквартальные данные о средней заработной плате, индексе

потребительских цен, числе занятых за последние годы; ежедневный курс доллара США или Евро на ММВБ. Отличительной особенностью временных данных является то, что они естественным образом упорядочены по времени.

Набор сведений представляет собой множество признаков, характеризующих объект исследования. Признаки являются взаимосвязанными, причем они выступают в одной из двух ролей:

- в роли результативного признака или объясняемой переменной Y ;
- в роли факторного признака или объясняющей переменной X , значение которой определяет результативный признак.

Переменные, участвующие в эконометрической модели любого типа, разделяются на:

а) *экзогенные* (независимые) X , значения которых задаются извне, автономно; в определенной степени они являются управляемыми (планируемыми);

б) *эндогенные* (зависимые) Y , значения которых определяются внутри модели и являются взаимозависимыми;

в) *лаговые* экзогенные или эндогенные переменные – это переменные эконометрической модели, датируемые предыдущими моментами времени и находящиеся в уравнении с текущими переменными; примером лаговых переменных являются:

Y_{t-1} – лаговая эндогенная переменная;

X_{t-1} – лаговая экзогенная переменная;

г) *предопределенные* переменные (объясняющие переменные); к ним относятся лаговые X_{t-1} и текущие X_t экзогенные переменные, а также лаговые эндогенные Y_{t-1} , Y_{t-2} переменные; любая эконометрическая модель предназначена для объяснения значений текущих эндогенных переменных Y_t , в зависимости от значений предопределенных переменных.

Основным элементом эконометрической модели выступает *случайная величина* (*случайное возмущение*, *случайный остаток*) – величина, входящая в уравнение модели, значения которой неизвестны и которая отражает влияние неучтенных факторов, неточность выбора переменных и вида зависимостей между ними, погрешности нахождения значений переменных зависимостей между ними. По поведению таких составляющих делают вывод об адекватности построенной эконометрической модели.

1.4. Статистическое наблюдение

Статистические данные – совокупность количественных характеристик социально-экономических явлений и процессов, полученных в результате статистического наблюдения, их обработки и соответствующих расчетов. Основными свойствами статистической информации являются ее массовость, стабильность и способность устаревать, поэтому выводы, сделанные на основе информации многолетней давности, могут быть неполными и даже неверными.

Статистическое исследование состоит из 3 стадий:

1. Статистическое наблюдение.
2. Первичная обработка, сводка и группировка результатов, наблюдений.
3. Анализ полученных сводных материалов.

Статистическое наблюдение – это научно-организованный сбор сведений об изучаемых социально-экономических процессах или явлениях (формирует цели и задачи).

Статистическое наблюдение – это массовое, планомерное, научно организованное наблюдение за явлениями социальной жизни, которое заключается в регистрации отобранных признаков у каждой единицы совокупности.

Процесс проведения статистического наблюдения включает следующие этапы:

- 1) подготовка наблюдения;
- 2) проведение массового сбора данных;
- 3) подготовка данных к автоматизированной обработке;
- 4) разработка предложений по совершенствованию статистического наблюдения.

При подготовке статистического наблюдения необходимо решить методологические вопросы, важнейшими из которых являются:

- определение цели и объекта наблюдения;
- определение состава признаков, подлежащих регистрации;
- разработка документов для сбора данных;
- выбор отчетной единицы и единицы, относительно которой будет проводиться наблюдение, а также методов и средств получения данных.

Целостность означает, что дальнейшее разложение индивидуальных явлений не повлечет за собой потери их качественной основы.

Вариация – это количественные изменения значений признака при переходе от одной единицы к другой.

Выбор времени наблюдения заключается в установлении критического момента (даты) или интервала времени и в определении срока (периода) наблюдения.

Критический момент – это момент времени, к которому приурочен сбор информации о каждой единице наблюдения. Собранные документы на этапе их подготовки к автоматизированной обработке подвергаются арифметическому и логическому контролю. Оба этих контроля основываются на знании взаимосвязей между показателями и качественными признаками.

Счетный контроль заключается в проверке точности арифметических расчетов, применявшихся при составлении отчетности или заполнении формуляров обследования.

Логический контроль заключается в проверке ответов на вопросы программы наблюдения путем их логического осмысления или путем сравнения полученных данных с другими источниками по этому же вопросу. Обычно, для исправления ошибок, выявленных в ходе логического контроля, требуется повторно обратиться к источнику сведений.

В процессе контроля может быть выявлено два типа ошибок:

– ошибка регистрации – это различие между значением показателя, полученного в ходе статистического наблюдения, и фактическим, действительным его значением;

– ошибка репрезентативности – отклонение значения показателя обследованной совокупности от его величины по исходной совокупности.

Объект наблюдения – некоторая совокупность, в которой протекает социально-экономические явления и процессы. Объектами могут быть совокупность физических лиц, физические единицы (дома), юридические лица. Чтобы определить объект необходимо четко определить границы изучаемой совокупности. Указать важнейшие признаки, отличающие его от других схожих объектов. Всякий объект состоит из единиц статистического наблюдения, являющихся носителем признаков, подлежащих регистрации.

Отчетная единица – субъект, от которого поступают данные о единице наблюдения.

Программа наблюдения – перечень признаков, подлежащих регистрации в процессе наблюдения. Обычно программа выражается в перечне вопросов прописного или опросного листа. Она должна содержать существенные признаки, характеризующие явления, его тип, основные черты и свойства. Вопросы должны быть точными, не двусмысленными, легкими для понимания, также должны быть вопросы контрольного характера для проверки и уточнения собираемых данных.

Для обеспечения единообразия получаемых сведений от каждого респондента программы оформляется документ, называемый статистическим бланком, т. е. документом единого образца, содержащего программу и результаты статистического наблюдения. Статистический бланк должен содержать титульную и адресную части. Первая содержит наименование статистического органа, проводящего наблюдение, информацию о том, кто и когда утвердил этот бланк. Вторая включает адрес отчетной единицы, ее подчиненность.

Различают две системы статистического бланка: индивидуальную (карточную) и списочную. Индивидуальный бланк предусматривает запись на нем ответов только об одной единице наблюдения, списочный – о нескольких.

Кроме бланка разрабатывается инструкция, определяющая порядок проведения наблюдения и заполнения формы отчетности.

На заключительном этапе исследования проводится анализ причин, которые привели к неверному заполнению статистических бланков, и даются предложения по совершенствованию наблюдения.

Основные формы статистического наблюдения представлены в табл. 1.1.

Статистическая отчетность – основная форма статистического наблюдения, с помощью которой статистические органы получают от предприятий необходимые данные в виде установленных в законном порядке отчетных документов, скрепляемых подписями лиц, юридически ответственных за предоставление сведений.

Отчетность как форма статистического наблюдения основана на первичном учете и является его обобщением. *Первичный учет* представляет собой регистрацию различных фактов, событий по мере их совершения. По срокам представления отчетность делят на ежедневную, месячную и годовую.

Таблица 1.1

Основные формы статистического наблюдения

Организационные формы статистического наблюдения	Виды статистического наблюдения		Способы статистического наблюдения
	по времени регистрации фактов	по охвату совокупности	
1. Статистическая отчетность 2. Специализированное наблюдение	1. Непрерывное или текущее. 2. Прерывное: – периодическое; – единовременное	1. Сплошное. 2. Несплошное: – выборочное; – основного массива; – монографическое	1. Непосредственное. 2. Документальное. 3. Опрос: – экспедиционный; – регистрация; – корреспондентский; – анкетный; – явочный

Специализированное наблюдение (специально организованное или перепись) проводится с целью получения сведений, отсутствующих в отчетности, или проверки ее данных.

Сплошное наблюдение – это вид статистического наблюдения, при котором обследованию подвергаются все единицы изучаемой совокупности. Основной предпосылкой подобного наблюдения, от которой зависит качество полученной статистической информации, является определение подлежащих регистрации признаков.

Несплошное наблюдение – это вид статистического наблюдения, при котором обследованию подвергается лишь часть единиц изучаемой совокупности. Основой подобного наблюдения является методологически точный выбор тех единиц, которые будут подвергнуты обследованию. Существует несколько видов несплошного наблюдения:

1. *Выборочное наблюдение* – это вид несплошного наблюдения, основанный на принципе случайного отбора тех единиц исследуемой совокупности, которые должны быть подвергнуты статистическому наблюдению; в структуре выборочной совокупности должны быть представлены все характерные типы единиц, составляющие исследуемую совокупность; в противном случае выборка не будет точно воспроизводить пропорции и зависимости, характерные для генеральной совокупности.

2. Разновидностью выборочного наблюдения является *метод моментных наблюдений*, при котором информация собирается путем регистрации признаков у единиц выборочной совокупности в заранее определенные моменты времени, поэтому данный метод требует выбора не только единиц исследуемой совокупности, но и моментов времени, в которые проводится регистрация. Он применяется, например, при проведении обследований доходов населения.

3. *Метод основного массива* – это вид сплошного наблюдения, при котором обследованию подвергаются те единицы совокупности, которые имеют наибольший удельный вес изучаемого признака. Так, наблюдение за динамикой цен осуществляется, как правило, на примере наиболее крупных городов страны. Методика подобного обследования исходит из предположения, что исключение из наблюдаемой совокупности «малозначимых» единиц не отразится существенным образом на результатах исследования.

4. *Монографическое обследование* – это вид сплошного наблюдения, при котором глубокому детальному обследованию подвергаются единичные объекты изучаемой совокупности, как правило, представители характерных либо новых типов общественных явлений и процессов. Соответственно, целью монографического обследования является характеристика уже имеющихся существенных тенденций в развитии изучаемого явления или определение наступающих изменений.

Контрольные вопросы

1. Связь эконометрики с другими областями знаний.
2. Понятие и виды эконометрических моделей и их переменных.
3. Алгоритм эконометрического моделирования.
4. Типы данных, используемых в эконометрических исследованиях.
5. Этапы построения регрессионной модели.
6. Статистическое наблюдение: основные понятия, этапы его проведения.
7. Программа, виды и формы статистического наблюдения.

2. ПАРНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

При статистической зависимости величины не связаны функционально, но как случайные величины заданы совместным распределением вероятностей. Исследование взаимозависимости случайных величин приводит к теории корреляции.

Основная задача корреляционного анализа состоит в установлении и оценивании тесноты линейной связи между изучаемым признаком Y и влияющих на него факторов X_1, X_2, \dots, X_k .

В статистике принято различать следующие варианты зависимостей:

1. *Парная корреляция* – связь между двумя признаками (результативным и факторным или двумя факторными).

2. *Частная корреляция* – зависимость между результативным и факторным признаками при фиксированном значении других факторных признаков.

3. *Множественная корреляция* – зависимость результативного признака от двух или более факторных признаков, включенных в исследование.

Теснота линейной связи количественно выражается величиной коэффициентов корреляции.

Парный (частный) коэффициент корреляции характеризует тесноту линейной зависимости между двумя переменными при исключении влияния всех остальных показателей, входящих в модель. Он изменяется в пределах от -1 до $+1$, причем, чем ближе коэффициент корреляции $k \pm 1$, тем сильнее зависимость между переменными. Если коэффициент корреляции больше 0 , то корреляция положительная, а если меньше 0 – отрицательная.

Положительная корреляция между случайными величинами характеризует такую зависимость между ними, когда при возрастании одной из них другая также в среднем возрастает.

Отрицательная корреляция характеризует такую зависимость, когда при возрастании одной случайной величины, другая в среднем убывает.

Близкое к нулю значение коэффициента корреляции свидетельствует об отсутствии прямолинейной статистической связи.

Очевидно, что для независимых случайных величин парный коэффициент корреляции равен нулю.

Множественный коэффициент корреляции характеризует тесноту линейной связи между резульативной переменной Y и остальными факторными переменными X_1, X_2, \dots, X_k , входящими в модель. Его значение изменяется в пределах от 0 до 1. Квадрат множественного коэффициента корреляции называется множественным коэффициентом детерминации. Он характеризует долю дисперсии резульативной переменной, обусловленной влиянием всех остальных аргументов, входящих в модель.

2.1. Парная корреляция

Коэффициент корреляции, его свойства

Пусть имеется набор значений двух переменных:

y_i – объясняемая переменная, или результат;

x_i – объясняющая переменная, или фактор, где $i = 1, 2, \dots, n$.

Эти величины имеют совместный двумерный нормальный закон распределения. Сопряженную изменчивость двух признаков в статистике можно количественно оценить с помощью *выборочного корреляционного момента*:

$$K_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{\sum_{i=1}^n x_i y_i}{n} - \bar{x} \bar{y}. \quad (2.1)$$

Хотя этот показатель непригоден в тех случаях, когда признаки выражены в разных единицах измерения. Этот недостаток устраняется с помощью известного приема нормирования: деления усредненной суммы произведения $(x_i - \bar{x})(y_i - \bar{y})$ на произведение средних квадратических отклонений $S_x S_y$. Результатом этой операции является показатель, который называют *выборочным коэффициентом корреляции*:

$$r_{xy} = \frac{K_{xy}}{S_x S_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n S_x S_y}. \quad (2.2)$$

Равенство (2.2) является оценкой *коэффициента корреляции генеральной совокупности* ρ_{xy} , который характеризует тесноту связи

между величинами X и Y в случае их *линейной зависимости*. Подставим в формулу (2.2) выражения для выборочной дисперсий:

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}; \quad (2.3)$$

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

Получим формулу

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (2.4)$$

Свойства коэффициента корреляции:

1. Коэффициент корреляции принимает значения на отрезке $[-1; 1]$: $|r_{xy}| \leq 1$. В случае *линейной* функциональной зависимости между X и Y коэффициент корреляции равен по модулю единице: $r_{xy} = \pm 1$.

2. Если $r > 0$, то связь между X и Y положительна и случайные величины одновременно возрастают (или одновременно убывают). Если $r_{xy} < 0$, то при увеличении X убывает Y .

3. Значение r_{xy} не зависит от выбора начала отсчета и единицы измерения. Значение r_{xy} не изменится в результате линейных преобразований: $r_{xy}(\lambda_1 X + C_1; \lambda_2 Y + C_2) = r_{xy}$; $\lambda_1 > 0$, $\lambda_2 > 0$.

4. Если X и Y независимы, то их коэффициент корреляции равен нулю: $r_{xy} = 0$; в этом случае $K_{xy} = 0$. Из независимости случайных величин следует их некоррелированность. Если $r_{xy} = 0$, то X и Y – независимы, но связь *нелинейная*.

5. Для случайных величин X и Y значения коэффициентов корреляции зависимостей Y по X и X по Y совпадают: $r_{xy} = r_{yx}$.

Чем больше по абсолютному значению коэффициент корреляции, тем теснее *линейная* связь между признаками. В табл. 2.1 при-

ведена условная градация коэффициента корреляции по силе линейной связи.

Таблица 2.1

Условная градация коэффициента корреляции по силе линейной связи

Диапазон измерений $ r_{xy} $	(0; 0,3]	(0,3; 0,5]	(0,5; 0,7]	(0,7; 0,9]	(0,9; 1)
Сила связи	слабая	умеренная	значительная	тесная	очень тесная, очень высокая

Коэффициент корреляции имеет реальный смысл только в случае линейной связи между величинами. Если связь нелинейная, то для описания корреляции используют другие показатели:

1) при небольшом объеме выборки ($n < 30$) можно воспользоваться формулой (2.4);

2) если значения x_i и y_i выражены в неудобных числах – дробных или многозначных – предпочтительна формула

$$r_{xy} = \frac{K_{xy}}{S_x S_y} = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \cdot \sqrt{\sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}}}; \quad (2.5)$$

3) если информация имеет вид корреляционной таблицы, то удобно пользоваться формулой

$$r_{xy} = \frac{K_{xy}}{S_x S_y} = \frac{\sum_{i=1}^n \sum_{j=1}^k x_i y_j n_{ij} - \frac{\sum_{i=1}^n x_i n_{xi} \cdot \sum_{j=1}^k y_j n_{yj}}{n}}{n S_x S_y}. \quad (2.6)$$

Пример 1. Приведены данные о величине основных производственных фондов X (млн руб.) и о суточной выработке продукции Y (т) на четырех предприятиях. Требуется выдать заключение о степени сопряженной изменчивости между этими величинами.

Решение. Данные промежуточные вычисления данных показаны в таблице (табл. 2.2).

Таблица 2.2

Расчетная таблица

№ предприятия	О.п.ф. X (млн руб.)	Продукция Y (т)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
1	70,60	32,67	21,90	7,97	479,61	63,52	174,54
2	35,65	17,91	-13,02	-6,79	169,52	46,10	88,41
3	40,48	22,45	-7,86	-2,25	61,78	5,06	17,68
4	47,68	25,79	-1,02	1,09	1,04	1,19	1,11
Σ	194,80	98,82	0,00	0,02	711,95	115,87	218,74
	$\bar{x} = 48,7$	$\bar{y} = 24,71$					

По формуле (2.4) вычисляем коэффициент корреляции:

$$r_{xy} = \frac{281,74}{\sqrt{711,95 \cdot 115,87}} = \frac{281,74}{287,22} = 0,98,$$

что говорит о почти функциональной линейной зависимости между величиной основных производственных фондов и выработкой продукции, причем прямой зависимости, т. е. большему значению X отвечает (в среднем) большее значение Y .

Пример 2. Найти коэффициент корреляции (в расчете на одного работающего) между производительностью труда Y (тыс. руб.) и энерговооруженностью труда X (кВт) для 14 предприятий региона.

Таблица 2.3

Расчетная таблица

x_i	2,8	2,2	3,0	3,5	3,2	3,7	4,0	4,8	6,0	5,4	5,2	5,4	6,0	9,0
y_i	6,7	6,9	7,2	7,3	8,4	8,8	9,1	9,8	10,6	10,7	11,1	11,8	12,1	12,4

Решение. Так как выборка мала (n), то воспользуемся формулой (2.5). Для этого вычислим необходимые суммы:

$$\sum_{i=1}^{14} x_i = 2,8 + 2,2 + 3 + 3,5 + 3,2 + 3,7 + 4 + 4,8 + \\ + 6 + 5,4 + 5,2 + 5,4 + 6 + 9 = 64,2;$$

$$\sum_{i=1}^{14} x_i^2 = 2,8^2 + 2,2^2 + 3^2 + 3,5^2 + 3,2^2 + 3,7^2 + 4^2 + 4,8^2 + \\ + 6^2 + 5,4^2 + 5,2^2 + 5,4^2 + 6^2 + 9^2 = 335,2;$$

$$\sum_{i=1}^{14} y_i = 6,7 + 6,9 + 7,2 + 7,3 + 8,4 + 8,8 + 9,1 + 9,8 + 10,6 + \\ + 10,7 + 11,1 + 11,8 + 12,1 + 12,4 = 132,9;$$

$$\sum_{i=1}^{14} y_i^2 = 6,7^2 + 6,9^2 + 7,2^2 + 7,3^2 + 8,4^2 + 8,8^2 + 9,1^2 + 9,8^2 + \\ + 10,6^2 + 10,7^2 + 11,1^2 + 11,8^2 + 12,1^2 + 12,4^2 = 1313,95;$$

$$\sum_{i=1}^{14} x_i y_i = 2,8 \cdot 6,7 + 2,2 \cdot 6,9 + 3 \cdot 7,2 + 3,5 \cdot 7,3 + 3,2 \cdot 8,4 + 3,7 \cdot 8,8 + \\ + 4 \cdot 9,1 + 4,8 \cdot 9,8 + 6 \cdot 10,6 + 5,4 \cdot 10,7 + 5,2 \cdot 11,1 + 5,4 \cdot 11,8 + \\ + 6 \cdot 12,1 + 9 \cdot 12,4 = 650,99.$$

Подставив найденные суммы в формулу (2.5), получим

$$r_{xy} = \frac{650,99 - \frac{64,2 \cdot 132,9}{14}}{\sqrt{335,26 - \frac{(64,2)^2}{14}} \cdot \sqrt{1313,95 - \frac{(132,9)^2}{14}}} = 0,898.$$

Это говорит о тесной линейной (прямой) связи между энерговооруженностью X и производительностью труда Y .

В некоторых случаях встречаются признаки, не поддающиеся количественной оценке, например, «способности» студентов к ма-

тематике X и «способности» к музыке Y ; в таком случае студентов располагают по порядку, в соответствии со степенью способностей, и присваивают каждому порядковый номер, который называется *рангом*. В литературе по статистике используют коэффициент корреляции рангов Кэнделла и коэффициент корреляции Спирмэна. Для нахождения достоверного показателя тесноты связи при любой форме зависимости по выборке вычисляется величина, называемая *эмпирическим корреляционным отношением X по Y* , или *эмпирическим корреляционным отношением Y по X* .

2.2. Множественная корреляция

В корреляционном анализе выборку объема n из $(k + 1)$ -мерной генеральной совокупности, подчиняющейся $(k + 1)$ -мерному нормальному закону распределения, представляют в матричном виде:

$$X = \begin{pmatrix} y_1 & x_{11} & \dots & x_{1k} \\ y_2 & x_{21} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ y_n & x_{n1} & \dots & x_{nk} \end{pmatrix}, \quad (2.7)$$

где i -строка $i = 1, 2, \dots, n$ характеризует наблюдение по всем $(k + 1)$ -показателям $j = 1, 2, \dots, k$.

По выборке определяют оценки параметров генеральной совокупности:

- вектор средних $\bar{X} = (\bar{Y}, \bar{X}_1, \dots, \bar{X}_k)$;
- вектор средних квадратических отклонений $\bar{S} = (S_y, S_{x_1}, \dots, S_{x_k})$;
- корреляционная матрица R порядка $k + 1$:

$$R = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_k} \\ r_{x_1y} & 1 & r_{x_1x_2} & \dots & r_{x_1x_k} \\ r_{x_2y} & r_{x_2x_1} & 1 & \dots & r_{x_2x_k} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_ky} & r_{x_kx_1} & r_{x_kx_2} & \dots & 1 \end{pmatrix}. \quad (2.8)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i; \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad j = 1, 2, \dots, k; \quad (2.9)$$

$$S_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}; \quad S_{x_j} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}, \quad j = 1, 2, \dots, k; \quad (2.10)$$

$$r_{y_i x_j} = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(x_{ij} - \bar{x}_j)}{S_y S_{x_j}}; \quad (2.11)$$

$$r_{x_l x_m} = \frac{\frac{1}{n} \sum_{i=1}^n (x_{il} - \bar{x}_l)(x_{im} - \bar{x}_m)}{S_{x_l} S_{x_m}}, \quad (2.12)$$

где x_{ij} – значение i -наблюдения j -фактора;

$r_{x_l x_m}$ – выборочный парный коэффициент корреляции, характеризующий тесноту линейной связи между факторами x_l и x_m .

Множественный коэффициент корреляции k -го порядка резуль- тативного признака Y определяется по формуле

$$r_{y/x_1, x_2, \dots, x_k} = \sqrt{1 - \frac{|R|}{R_{11}}}, \quad (2.13)$$

где $|R|$ – определитель матрицы R (2.8);

R_{11} – это алгебраическое дополнение элемента корреляционной матрицы R , находящегося в первой строке и первом столбце.

При этом

$$R_{ij} = (-1)^{i+j} M_{ij},$$

где M_{ij} – минор, определитель матрицы, получаемый из матрицы R путем вычеркивания i -строки и j -столбца.

2.3. Коэффициент корреляции

Если коэффициент корреляции равен нулю для генеральной совокупности, то это не означает, что он будет равен нулю для выборки. Наоборот, он будет отклоняться от истинного значения, но чем больше такое отклонение, тем менее оно вероятно при данном объеме выборки. При каждом конкретном значении коэффициента корреляции величин X и Y для генеральной совокупности выборочный коэффициент корреляции r_{xy} является случайной величиной. Следовательно, случайной величиной является так же любая его функция, и требуется указать такую функцию, которая имела бы одно из известных распределений, удобное для табличного анализа. Для выборочного коэффициента корреляции r_{xy} такой функцией является t -статистика, рассчитываемая по формулам

$$t = r_{xy} \frac{\sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \quad (\text{для двумерной модели}); \quad (2.14)$$

$$t = r \frac{\sqrt{n-l-2}}{\sqrt{1-r^2}} \quad (\text{для многомерной модели}), \quad (2.15)$$

где r – оценка частного или парного коэффициента корреляции;

l – число фиксируемых факторов, имеющее распределение Стьюдента с $(n-2)$ - или $(n-l-1)$ -степенями свободы (прил. Б).

Для коэффициента корреляции проверяется нулевая гипотеза о равенстве нулю коэффициента корреляции в генеральной совокупности. Эта гипотеза отвергается, если выборочный коэффициент корреляции слишком далеко отклоняется от нулевого значения, т. е. произошло событие, которое было бы маловероятным в случае равенства нулю коэффициента корреляции для генеральной совокупности. Для этого задается вероятность такого события, которая называется в статистике «уровень значимости». Чаще всего уровень значимости равен 1 % или 5 %. Так проверяемый коэффициент корреляции считается значимым, если значение t по модулю будет больше, чем $t_{кр}$, которое определяется по таблице t -распределения (прил. Б) для уровня значимости α и числа степеней свобод $\nu = n-2$ или $\nu = n-l-1$.

Пример 3. Данные о количестве выпускаемых деталей X (в тысячах) на 10 заводах и полных затратах Y (усл. ден. ед.), приведены в табл. 2.4.

Таблица 2.4

Данные о количестве деталей и полных затратах

X	50	55	65	65	70	85	90	90	100	100
Y	20	30	25	35	45	40	45	55	50	50

Вычислить выборочный коэффициент корреляции r_{xy} и проверить его значимость при $\alpha = 0,05$.

Решение. Вычисления выборочного коэффициента корреляции выполнить по формуле (2.4). Найти выборочные средние и суммы

$$\bar{x} = \frac{\sum x_i}{n} = \frac{770}{10} = 77,0; \quad \bar{y} = \frac{\sum y_i}{n} = \frac{405}{10} = 39,5.$$

$$\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y}) = 1685; \quad \sum_{i=1}^{10} (x_i - \bar{x})^2 = 3010; \quad \sum_{i=1}^{10} (y_i - \bar{y})^2 = 1222,5.$$

Выборочный коэффициент корреляции равен

$$r_{xy} = \frac{1685}{\sqrt{3010 \cdot 1222,5}} \approx 0,878.$$

Для оценки значимости r_{xy} и нулевой гипотезе $H_0(r_{xy} = 0)$ применить формулу

$$T = t_3 = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}} = \frac{0,878 \sqrt{8}}{\sqrt{1-(0,878)^2}} = \frac{2,4834}{0,4787} = 5,19.$$

По таблице (прил. Б) распределения Стьюдента при уровне значимости $\alpha = 0,05$ и числу степеней свободы $k = n - 2 = 8$ определить критическое значение $t_{кр} = t_{0,05;8} = 2,31$.

Сравнить значения t_3 и $t_{кр}$. Выборочные данные не согласуются с выдвинутой гипотезой H_0 , т. к. $5,19 > 2,31$. Значение r_{xy} намного отличается от нуля, с доверительной вероятностью $\gamma = 1 - \alpha = 0,95$ сделать вывод, что X и Y коррелированы.

2.4. Парная регрессия

Сформулируем регрессионную проблему для случая одного факторного признака.

Пусть имеется набор значений двух переменных: y_i (объясняемая переменная или результат) и x_i (объясняющая переменная или фактор); $i = 1, 2, 3, \dots, n$. Между этими переменными имеется объективная связь:

$$y = f(x). \quad (2.16)$$

Уравнение (2.16) будем называть истинным уравнением регрессии. Необходимо по данным наблюдений $((y_i, x_i); (i = 1, 2, 3, \dots, n))$ подобрать функцию $\hat{y} = \hat{f}(x)$ «наилучшим» образом описывающую «истинную» зависимость (2.17). Подобрать функцию – определить вид функциональной зависимости и значения параметров.

Нормальная (классическая) линейная парная (или однофакторная) регрессионная модель имеет вид

$$y = a_0 + a_1x + \varepsilon. \quad (2.17)$$

Величина y состоит из двух составляющих:

1. Неслучайной: $a_0 + a_1x$.

2. Случайной: ε .

Для модели (2.14) выполняются следующие предпосылки:

1) x – неслучайная величина;

2) $M(\varepsilon) = 0$ – математическое ожидание случайной составляющей, равное 0 в любом наблюдении;

3) $M(\varepsilon_i \varepsilon_j) = \sigma^2$ – теоретическая дисперсия случайной составляющей; постоянна для всех наблюдений;

4) $M(\varepsilon_i, \varepsilon_j) = 0$; ($i \neq j$) – отсутствие систематической связи между значениями случайной составляющей в любых двух наблюдениях (ковариация случайных составляющих в любых двух разных наблюдениях равна нулю);

5) ε – нормально распределенная случайная величина с параметрами 0 и σ^2 .

В матричной форме парная регрессионная модель (2.19) имеет вид

$$Y = AX + \varepsilon, \quad (2.18)$$

$$\text{где } Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}_{n \times 1};$$

$$A = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}_{2 \times 1};$$

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix}_{n \times 2};$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}_{n \times 1}.$$

2.5. Множественная регрессия

Традиционная линейная модель множественной регрессии (ЛММР) математически формулируется следующим образом:

$$\left\{ \begin{array}{l} y_i = a_0 + a_1 x_{i1} + \dots + a_k x_{ik} + \varepsilon_i; \quad i = 1, 2, 3, \dots, n; \\ M(\varepsilon) = 0; \\ M(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2, & i = j; \\ 0, & i \neq j; \end{cases} \\ x_{i1}, \dots, x_{ik}. \end{array} \right. \quad (2.19)$$

где x_{i1}, \dots, x_{ik} – неслучайные переменные ($k < n$);

$M(\varepsilon) = \sum_{i=1}^n \varepsilon_i \rho_i$ – математическое ожидание случайной составляющей ε ;

$$M(\varepsilon_i \varepsilon_j) = \sum_{i=1}^n \sum_{j=1}^n \varepsilon_i \varepsilon_j P_{ij};$$

σ – среднее квадратическое отклонение.

Модель линейной множественной регрессии (2.19), для которой выполняется условие нормального распределения случайной составляющей модели с параметрами 0 и σ^2 называется *нормальной линейной множественной регрессионной моделью*.

В дальнейшем считается неизменным взаимная некоррелируемость (независимость) случайных регрессионных остатков, т. к. для всех остатков $\varepsilon_1, \dots, \varepsilon_n$ выполняется соотношение $M(\varepsilon_i^2) = \sigma^2$. Это означает неизменность отклонений (дисперсий) регрессионных остатков. Это свойство называется *гомоскедастичностью регрессионных остатков*.

Запишем матричную формулу ЛММР.

Введем матрицы:

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}_{n \times (k+1)} \quad ; \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}_{n \times 1} \quad ;$$

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}_{n \times n} ; \quad A = \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_k \end{pmatrix}_{(k+1) \times 1} ; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}_{n \times 1} ;$$

$$X_1 = \begin{pmatrix} x_{11} \\ x_{21} \\ \dots \\ x_{n1} \end{pmatrix}, \dots, X_k = \begin{pmatrix} x_{1k} \\ x_{2k} \\ \dots \\ x_{nk} \end{pmatrix} ; \quad O_n = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}_{n \times 1} ;$$

$$M(\varepsilon\varepsilon^T) = \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & \dots & M(\varepsilon_1\varepsilon_n) \\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & \dots & M(\varepsilon_2\varepsilon_n) \\ \dots & \dots & \dots & \dots \\ M(\varepsilon_n\varepsilon_1) & M(\varepsilon_n\varepsilon_2) & \dots & M(\varepsilon_n^2) \end{pmatrix}_{n \times n} ,$$

где X_1, \dots, X_k – неслучайные переменные;

$M(\varepsilon\varepsilon^T)$ – ковариационная матрица «возмущений».

Матричная форма ЛММП будет иметь вид

$$\begin{cases} Y = XA + \varepsilon; \\ M(\varepsilon) = O_n; \\ M(\varepsilon\varepsilon^T) = \sigma^2 I_n; \\ X_1, \dots, X_k; \\ \text{rang} X = k + 1 < n. \end{cases} \quad (2.20)$$

Для получения оценок параметров ЛММП (2.19) применяется МНК, записанный в форме (2.18).

2.6. Статистические свойства параметров

При повторении выборок того же объема n из той же анализируемой генеральной совокупности и при тех же значениях объясняющих переменных X_1, X_2, \dots, X_k наблюдаемые значения результирующего показателя Y будут случайным образом варьироваться за счет случайного характера остатков ε . Значения Y будут варьироваться и зависеть от значения оценок \hat{A} и $\hat{\sigma}^2$. То есть оценки ведут себя как случайные величины. Поэтому, пытаясь определить их поведение, решаются следующие задачи:

1. К чему стремятся (по вероятности) оценки \hat{A} и $\hat{\sigma}^2$ при неограниченном объеме выборки ($n \rightarrow \infty$)?
2. Каковы средние значения этих оценок для выборок фиксированного объема n ?
3. Каковы основные характеристики случайного разброса в значениях этих оценок относительно истинных значений оцениваемых параметров?

Состоятельность оценок \hat{A} и $\hat{\sigma}^2$ определяется структурой матрицы X . Оценки \hat{A} и $\hat{\sigma}^2$ состоятельны только тогда, когда наименьшее собственное число матрицы $X^T \cdot X$ стремится к бесконечности, при $n \rightarrow \infty$. Наименьшее собственное значение матрицы $X^T \cdot X$ определяется как минимальный по величине корень уравнения

$$\left| X^T \cdot X - \lambda I_{k+1} \right| = 0,$$

где I_{k+1} – единичная матрица $(k + 1)$ -го порядка.

Для случая парной регрессии условие состоятельности \hat{A} и $\hat{\sigma}^2$

равносильно требованию $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \rightarrow \text{const}$, при $n \rightarrow \infty$.

Несмещенность оценок. Оценка \hat{A} параметра A называется несмещенной, если $M(\hat{A}) = A$. Можно показать, что

$$\begin{aligned}
 M(\hat{A}) &= M\left(\left(X^T \cdot X\right)^{-1} X^T Y\right) = M\left(\left(X^T \cdot X\right)^{-1} \cdot X^T (XA + \varepsilon)\right) = \\
 &= \left(X^T \cdot X\right)^{-1} X^T \cdot x M(A) + \left(X^T \cdot X\right)^{-1} X^T M(\varepsilon) = M(A) = A.
 \end{aligned}$$

Ковариационная матрица **МНК-оценок** \hat{A} определяется следующим образом:

$$C_{\hat{A}} = M\left[\left(\hat{A} - A\right) \cdot \left(\hat{A} - A\right)^T\right].$$

Диагональные элементы этой матрицы $M(\hat{a}_0 - a_0)^2$, $M(\hat{a}_1 - a_1)^2, \dots, M(\hat{a}_k - a_k)^2$ задают средние квадраты отклонений соответствующих оценок.

Ковариационная матрица $C_{\hat{A}}$ содержит важную информацию о качестве оценок \hat{A} . В терминах параметров ЛММР и исходных статистических данных $C_{\hat{A}}$ выражается следующим образом:

$$C_{\hat{A}} = \sigma^2 \left(X^T \cdot X\right)^{-1}.$$

На практике обычно используют матрицу:

$$C_{\hat{A}} = S^2 \left(x^T \cdot x\right)^{-1}, \quad (2.21)$$

$$\text{где } S^2 = \frac{1}{n - k - 1} (Y - \hat{A}X)^T \cdot (Y - \hat{A}X).$$

Чаще достаточно вычислить только среднеквадратичные ошибки:

$$S_j = S \sqrt{a_{jj}},$$

где S определяется из равенства (2.21).

МНК-оценки параметров и дисперсии случайной составляющей являются «наилучшими» (состоятельными, несмещенными и эффективными) оценками. Эффективность МНК-оценок определяется с помощью теоремы Гаусса–Маркова.

Теорема Маркова–Гаусса. Метод наименьших квадратов дает оценки, имеющие наименьшую дисперсию в классе всех линейных несмещенных оценок, если выполняются предпосылки нормальной линейной регрессионной модели:

- 1) $y = a_0 + a_1x_1 + \dots + a_kx_k + \varepsilon$;

- 2) x_1, x_2, \dots, x_k – неслучайные величины;

- 3) $M(\varepsilon) = 0$;

- 4) $M(\varepsilon_i\varepsilon_j) = \begin{cases} \sigma^2, & i = j, \\ 0, & i \neq j, \quad i = 1, \dots, n; \end{cases}$

- 5) $\varepsilon \in N(0, \sigma^2)$ (ε – нормально распределенная случайная величина).

2.7. Уравнение регрессии

С помощью МНК были получены лишь оценки параметров уравнения регрессии $\hat{A} = (\hat{a}_0, \hat{a}_1, \dots, \hat{a}_k)^T$.

Чтобы проверить значимы ли параметры уравнения регрессии, т. е. значимо ли они отличаются от нуля в «истинном» уравнении регрессии $y = a_0 + a_1x_1 + \dots + a_kx_k + \varepsilon$, применим статистические методы проверки гипотез. В качестве основной гипотезы H_0 выдвигают гипотезу о незначимом отличии от нуля «истинного» параметра регрессии. Альтернативной гипотезой H_1 при этом является гипотеза о неравенстве нулю «истинного» параметра, т. е. $H_0: a_i = 0$, $H_1: a_i \neq 0$. Для проверки гипотезы H_0 используется t -статистика критерия проверки гипотезы, имеющая распределение Стьюдента:

$$t = \frac{|\hat{a}_i|}{S_{\hat{a}_i}},$$

где $S_{\hat{a}_i} = S \cdot \sqrt{[(X^T X)^{-1}]_{ii}}$ – стандартная ошибка параметра регрессии \hat{a}_i .

Найденное по данным наблюдений значение t -статистики сравнивается с критическим значением t -статистики, определяемым по таблицам распределения Стьюдента (прил. Б). Критическое значение t -статистики определяется в зависимости от уровня значимости α и числа степеней свободы $n - (k + 1)$ и равно

$$t_{\text{кр}} = t_{1-\alpha; n-(k+1)},$$

где n – число наблюдений;

$k + 1$ – число оцениваемых параметров в уравнении регрессии.

В случае линейной парной регрессии $k + 1 = 2$, а число степеней свободы равно $n - 2$. Если фактическое значение t -статистики больше критического ($t > t_{\text{кр}}$), то основную гипотезу H_0 отвергают и считают, что с вероятностью $1 - \alpha$ «истинный» параметр регрессии значимо отличается от нуля.

Если фактическое значение t -статистики не больше критического ($t \leq t_{\text{кр}}$), то нет оснований отвергать основную гипотезу H_0 и значит «истинный» параметр регрессии незначимо отличается от нуля при уровне значимости α . Для линейного уравнения парной регрессии стандартные ошибки коэффициентов \hat{a}_1, \hat{a}_0 вычисляется по следующим формулам:

$$S_{\hat{a}_0} = \sqrt{\frac{S \cdot \frac{1}{n} \sum x_i^2}{\sum (x_i - \bar{x})^2}} = \sqrt{\frac{\sum (y_i - \hat{y}_i) \cdot \sum x_i^2}{n(n-2) \cdot \sum (x_i - \bar{x})^2}} \quad (2.22)$$

$$S_{\hat{a}_1} = \sqrt{\frac{S}{\sum (x_i - \bar{x})^2}} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{(n-2) \sum (x_i - \bar{x})^2}}.$$

2.8. Показатели качества регрессии

Качество модели регрессии связывают с адекватностью (соответствием) модели наблюдаемым данным. Проверка адекватности модели регрессии наблюдаемым данным проводится на основе анализа остатков $\varepsilon_i = y_i - \hat{y}_i$ ($i = 1, \dots, n$) (остаток представляет собой отклонение фактического значения переменной от теоретического значения).

При анализе качества модели регрессии используется теорема о разложении дисперсии, согласно которой общая дисперсия результирующего признака может быть разложена на две составляющие: объясненную и необъясненную уравнением регрессии дисперсии:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (2.23)$$

На основании равенства (2.23) рассчитываются показатели качества модели регрессии:

1. Теоретический коэффициент детерминации:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (2.24)$$

Коэффициент детерминации характеризует долю вариации результирующего признака, объясняемую регрессией изменчивостью объясняющих переменных.

При парной линейной регрессии $y = a_0 + a_1x + \varepsilon$ коэффициент детерминации равен квадрату парного линейного коэффициента корреляции

$$R^2 = r_{xy}^2.$$

2. Корень из коэффициента детерминации R^2 – коэффициент множественной корреляции. Если все точки корреляционного поля

лежат на теоретической линии регрессии, то коэффициент множественной корреляции равен 1. Следовательно, связь между величинами Y и X функциональная и уравнение регрессии точно описывает фактические данные.

3. Средняя квадратическая ошибка уравнения регрессии представляет собой среднее квадратическое отклонение наблюдаемых значений результативного признака от теоретических значений, рассчитанных по модели (по уравнению регрессии), т. е.:

$$S = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1}}, \quad (2.25)$$

где k – число факторов модели.

Величину (2.25) можно сравнить со средним квадратическим отклонением результативного признака $S_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}}$. Если S окажется меньше S_y , то использование модели регрессии является целесообразным.

4. Средняя ошибка аппроксимации:

$$A = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \quad (2.26)$$

Чем меньше рассеяние эмпирических точек вокруг теоретической линии регрессии, тем меньше средняя ошибка аппроксимации (ошибка аппроксимации меньше 7 % свидетельствует о хорошем качестве модели).

Оценка значимости уравнения регрессии производится для того, чтобы узнать пригодно ли оценочное уравнение регрессии для практического использования (прогноза). При этом выдвигают основную гипотезу о незначимости уравнения в целом, которая сводится к гипотезе о равенстве нулю параметров регрессии или, что то же самое, о равенстве нулю коэффициента детерминации $R^2 = 0$.

Альтернативная ей гипотеза – это гипотеза о значимости уравнения регрессии, т. е. гипотеза о неравенстве нулю параметров регрессии или неравенстве нулю коэффициента детерминации: $R^2 \neq 0$. Для ее проверки используют F -статистику критерия проверки гипотезы:

$$F = \frac{R^2}{1 - R^2} \frac{n - (k + 1)}{k}, \quad (2.27)$$

где n – число наблюдений;

k – число факторов модели.

Данная статистика имеет распределение Фишера.

По таблицам распределения Фишера находят критическое значение F -критерия, в зависимости от уровня значимости α и двух степеней свободы: $k_1 = k$; $k_2 = n - (k + 1)$.

Сравниваются значения F -критерия, рассчитанные по данным выборки, с критическим значением F_{α, k_1, k_2} . Если $F < F_{\alpha, k_1, k_2}$, то гипотезу отвергают и принимают альтернативную гипотезу о статистической значимости уравнения регрессии.

Для случая парной линейной модели оба способа проверки значимости с использованием F - и t -критериев равносильны, т. к. $F = t^2$.

2.9. Интервальное оценивание функции регрессии

Доверительный интервал (интервальная оценка) для значимого параметра регрессионной модели имеет вид

$$\hat{a}_i - t_{1-\alpha; n-(k+1)} \cdot S_{\hat{a}_i} \leq a_i \leq \hat{a}_i + t_{1-\alpha; n-(k+1)} \cdot S_{\hat{a}_i}. \quad (2.28)$$

Наиболее важным для оценки точности определения зависимой переменной (прогноза) является построение доверительного интервала для функции регрессии или для условного математического ожидания зависимой переменной $M_x(Y)$. Для парной и множественной регрессий доверительные интервалы для $M_x(Y)$ соответственно имеют виды (2.29) и (2.30):

$$\hat{y} - t_{1-\alpha; n-2} \cdot S_{\hat{y}} \leq M_x(Y) \leq \hat{y} + t_{1-\alpha; n-2} \cdot S_{\hat{y}}, \quad (2.29)$$

$$\text{где } S_{\hat{y}} = S \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}};$$

$$S = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}};$$

$$\hat{y} - t_{1-2; n-(k+1)} \cdot S_{\hat{y}} \leq M_x(Y) \leq \hat{y} + t_{1-\alpha; n-(k+1)} \cdot S_{\hat{y}}, \quad (2.30)$$

$$\text{где } S_{\hat{y}} = S \sqrt{(X_0^*)^T (X^T \cdot X)^{-1} X_0^*};$$

$$S = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-(k+1)}}.$$

Доверительный интервал для индивидуальных значений зависимой переменной Y в случае парной регрессии имеет вид

$$\hat{y}_0 - t_{1-\alpha; n-2} \cdot S_{\hat{y}_0} \leq y_0 \leq \hat{y}_0 + t_{1-\alpha; n-2} \cdot S_{\hat{y}_0}, \quad (2.31)$$

$$\text{где } S_{\hat{y}_0} = S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}};$$

при множественной регрессии

$$\hat{y}_0 - t_{1-\alpha; n-(k+1)} \cdot S_{\hat{y}_0} \leq y_0 \leq \hat{y}_0 + t_{1-\alpha; n-(k+1)} \cdot S_{\hat{y}_0}, \quad (2.32)$$

$$\text{где } S_{\hat{y}_0} = S \sqrt{1 + X_0^T (X^T \cdot X)^{-1} X_0}.$$

2.10. Примеры решения задач

Пример 4. По территориям 12 регионов приводятся данные (табл. 2.5) за год по среднему душевому прожиточному минимуму в день на одного трудоспособного человека (X , руб.) и средней заработной платы (Y , руб.).

Таблица 2.5

Исходные данные

<i>N</i>	1	2	3	4	5	6	7	8	9	10	11	12
<i>X</i>	67	73	76	78	79	82	87	87	88	89	106	115
<i>Y</i>	139	152	159	133	154	148	134	162	158	162	195	173

Требуется:

1. Построить линейное уравнение парной регрессии *Y* по *X*.
2. Проверить значимость уравнения регрессии.
3. В случае значимости уравнения регрессии выполнить прогноз заработной платы *Y* при прогнозном значении среднедушевого прожиточного минимума, составляющего 107 % от среднего уровня.
4. Оценить статистическую точность прогноза, рассчитав ошибку прогноза и его доверительный интервал с вероятностью 0,95.

Решение:

1. Для нахождения оценочного уравнения регрессии *Y* по *X*, предполагая, что все предпосылки регрессионного анализа выполнены, применим метод наименьших квадратов.

Решим систему линейных уравнений относительно параметров оценочного уравнения регрессии:

$$\begin{cases} a_0 n + a_1 \sum x_i = \sum y_i; \\ a_0 \sum x_i + a_1 \sum x_i^2 = \sum x_i y_i. \end{cases}$$

Для определения значений сумм в системе воспользуемся расчетной табл. 2.6.

Таблица 2.6

Расчетные данные

№	x_i	y_i	$x_i y_i$	x_i^2
1	67	139	9313	4489
2	73	152	11096	5329
3	76	159	12084	5776
4	78	133	10374	6084

№	x_i	y_i	$x_i y_i$	x_i^2
5	79	154	12166	6241
6	82	148	12136	6724
7	87	134	11658	7569
8	87	162	14094	7569
9	88	158	13904	7744
10	89	162	14418	7921
11	106	195	20670	11236
12	115	173	19895	13225
Сумма	1027	1869	161808	89907

Тогда система запишется в виде

$$\begin{cases} 12a_0 + 1027a_1 = 1869; \\ 1027a_0 + 89907a_1 = 161808. \end{cases}$$

Методом последовательного исключения неизвестных получаем следующее решение:

$$a_0 \approx 76,9765; \quad a_1 \approx 0,9204.$$

В итоге оценочное уравнение регрессии будет иметь вид

$$\hat{y} = 76,9765 + 0,9204x.$$

Построим линейное уравнение парной регрессии Y по X и исходные статистические данные (рис. 2.1).

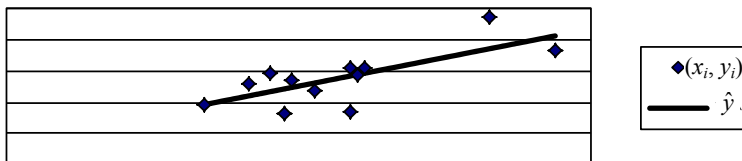


Рис. 2.1. Линия регрессии и совокупности

2. Выдвигаем гипотезы: $H_0: a_0 = 0, a_1 = 0$; $H_1: a_0 \neq 0, a_1 \neq 0$.

Определим расчетное значение статистики t -распределения Стьюдента с $(n - 2)$ -степенями свободы по формулам

$$t_{a_0} = \frac{|\hat{a}_0| \cdot n \cdot \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{S \sqrt{\sum_{i=1}^n x_i^2}},$$

$$t_{a_1} = \frac{|\hat{a}_1|}{S} \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2},$$

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

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}.$$

Получим следующие результаты:

$$S = 12,6; \quad \bar{x} = 85,6; \quad t_{a_0} = \frac{77}{24,3} \approx 3,2; \quad t_{a_1} = \frac{0,92}{0,281} \approx 3,3.$$

Сравним расчетные значения t_{a_0}, t_{a_1} с критическим значением $t_{кр}$ при уровне значимости $\alpha = 1 - 0,95 = 0,05$. Для этого по таблице значений распределения Стьюдента (прил. Б) найдем значение $t_{кр} = t_{1-\alpha; n-2} = t_{0,95; 10} = 2,23$. Имеем, что $t_{a_0} > t_{кр}, t_{a_1} > t_{кр}$. Это значит, что на уровне значимости 0,05 гипотезу $H_0: a_0 = 0, a_1 = 0$ отклоняем и признаем статистическую значимость уравнения регрессии в целом.

3. Выполним прогноз заработной платы Y при прогнозном значении среднедушевого прожиточного минимума X , составляющем 107 % от среднего уровня, т. к. установлена статистическая значи-

мость уравнения регрессии (п. 2). Найдем значение прожиточного минимума, составляющего 107 % от его среднего значения:

$$x = \bar{x} \cdot 1,07 = 85,6 \cdot 1,07 = 91,6 \text{ руб.}$$

Тогда по полученному уравнению регрессии в п. 1 решения имеем

$$\hat{y}(91,6) = 0,92 \cdot 91,6 + 76,98 \approx 161,25 \text{ руб.}$$

Окончательно имеем, что прогноз заработной платы в день составит 161,25 рублей при прожиточном минимуме в 91,6 рублей.

4. Оценим точность прогноза, полученного в п. 3, рассчитав ошибку прогноза и его доверительный интервал. Применяем формулу:

$$S_{\hat{y}(91,6)} = S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 12,6 \cdot \sqrt{1 + \frac{1}{12} + \frac{(91,6 - 85,6)^2}{2012,92}} \approx 13,2.$$

Тогда предельная ошибка прогноза, которая в 95 % случаев не будет превышена, составит

$$\Delta y = t_{\alpha; n-m} \cdot S_{\hat{y}(91,6)} = 2,23 \cdot 13,2 = 29,4$$

и доверительный интервал прогноза будет

$$(161 - 29,4; 161 + 29,4) = (131,6; 190,4).$$

Пример 5. По территориям 12 регионов приводятся данные (табл. 2.7) за год по среднедушевому прожиточному минимуму в день на одного трудоспособного человека (X , руб.) и среднедневной заработной плате (Y , руб.).

Таблица 2.7

Исходные данные

№	1	2	3	4	5	6	7	8	9	10	11	12
X	67	73	76	78	79	82	87	87	88	89	106	115
Y	139	152	159	133	154	148	134	162	158	162	195	173

Требуется оценить тесноту линейной связи между среднедушевым прожиточным минимумом в день на одного трудоспособного человека и среднедневной заработной платой. Оценить статистическую значимость коэффициента корреляции всей генеральной совокупности с вероятностью 0,95.

Решение. Тесноту линейной связи между двумя признаками оценивает коэффициент корреляции. Его точечной оценкой является выборочный коэффициент корреляции, который вычисляется по формуле

$$r_{xy} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{S_x \cdot S_y} = \frac{\frac{1}{n} \sum x_i y_i - \frac{1}{n^2} \sum x_i \sum y_i}{\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n} \sum (y_i - \bar{y})^2}}.$$

Расчеты значений сумм удобно проводить в табл. 2.8.

Таблица 2.8

Расчетные данные

№	x_i	y_i	$x_i y_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	67	139	9313	-18,6	345,3403	-16,8	280,5625
2	73	152	11096	-12,6	158,3403	-3,8	14,0625
3	76	159	12084	-9,6	91,84028	3,3	10,5625
4	78	133	10374	-7,6	57,50694	-22,8	517,5625
5	79	154	12166	-6,6	43,34028	-1,8	3,0625
6	82	148	12136	-3,6	12,84028	-7,8	60,0625
7	87	134	11658	1,4	2,006944	-21,8	473,0625
8	87	162	14094	1,4	2,006944	6,3	39,0625
9	88	158	13904	2,4	5,840278	2,3	5,0625
10	89	162	14418	3,4	11,67361	6,3	39,0625
11	106	195	20670	20,4	416,8403	39,3	1540,563
12	115	173	19895	29,4	865,3403	17,3	297,5625
Сумма	1027	1869	161808		2012,917		3280,25

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{1027}{12} = 85,6; \quad \bar{y} = \frac{1}{n} \sum y_i = \frac{1869}{12} = 155,8;$$

$$\overline{xy} = \frac{1}{n} \sum x_i y_i = \frac{161808}{12} = 13484.$$

При условии, что $S_x^2 = 167,7$, $S_y^2 = 273,4$, следует

$$S_x = \sqrt{S_x^2} = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2} = \sqrt{\frac{2012,919}{12}} \approx 12,952;$$

$$S_y = \sqrt{S_y^2} = \sqrt{\frac{1}{n} \sum (y_i - \bar{y})^2} = \sqrt{\frac{3280,25}{12}} \approx 16,533.$$

Итоговое значение выборочного коэффициента корреляции

$$r_{xy} = \frac{85,6 \cdot 155,8 - 13484}{12,952 \cdot 16,533} \approx 0,721.$$

Статистическую значимость коэффициента корреляции генеральной совокупности проверим с помощью критерия Стьюдента. Выдвигаем гипотезы: $H_0: \rho_{xy} = 0$, $H_1: \rho_{xy} \neq 0$. Рассчитаем значение статистики t -распределения Стьюдента с $(n - 2)$ -степенями свободы по формуле:

$$t = \frac{|r_{xy}| \cdot \sqrt{n-2}}{\sqrt{1-r_{xy}^2}} = \frac{0,721 \cdot \sqrt{10}}{\sqrt{1-0,721^2}} \approx 3,291.$$

Сравним расчетное значение t с критическим значением $t_{кр}$ при уровне значимости $\alpha = 1 - 0,95 = 0,05$. Для этого по таблице значений распределения Стьюдента найдем значение

$$t_{кр} = t_{1-\alpha; n-2} = t_{0,95; 10} = 2,23.$$

Имеем $t > t_{кр}$. Это значит, что на уровне значимости 0,05 гипотезу $H_0: \rho_{xy} = 0$ отклоняем и признаем статистическую значимость коэффициента корреляции генеральной совокупности.

Пример 6. Торговое предприятие имеет сеть, состоящую из 12 магазинов, информация о деятельности которых представлена в табл. 2.9.

Таблица 2.9

Исходные данные

№ магазина	Годовой товарооборот Y , млн руб.	Торговая площадь X_1 , тыс. м ²	Среднее число посетителей в день X_2 , тыс. чел.
1	19,76	0,24	8,25
2	38,09	0,31	10,24
3	40,95	0,55	9,31
4	41,08	0,48	11,01
5	56,29	0,78	8,54
6	68,51	0,98	7,51
7	75,01	0,94	12,36
8	89,05	1,21	10,81
9	91,13	1,29	9,89
10	91,26	1,12	13,72
11	99,84	1,29	12,27
12	108,55	1,49	13,92

Требуется рассчитать линейные коэффициенты частной корреляции и коэффициент множественной корреляции. Сравнить их с линейными коэффициентами парной корреляции, пояснить различия между ними.

Решение. Коэффициенты парной корреляции рассчитаем по формулам

$$r_{yx_1} = \frac{\overline{yx_1} - \bar{y} \cdot \bar{x}_1}{S_y \cdot S_{x_1}}; \quad r_{yx_2} = \frac{\overline{yx_2} - \bar{y} \cdot \bar{x}_2}{S_y \cdot S_{x_2}}; \quad r_{x_1x_2} = \frac{\overline{x_1x_2} - \bar{x}_1 \cdot \bar{x}_2}{S_{x_1} \cdot S_{x_2}}.$$

Необходимые расчеты занесем в табл. 2.10.

Таблица 2.10

Расчетные данные

№	y	x ₁	x ₂	(y _i - \bar{y}) ²	(x _{1i} - x ₁) ²	(x _{2i} - x ₂) ²	yx ₁	yx ₂	x ₁ x ₂
1	19,76	0,24	8,25	2355,484	0,423	5,772	4,742	163,020	1,980
2	38,09	0,31	10,24	912,241	0,336	0,170	11,808	390,042	3,174
3	40,95	0,55	9,31	747,658	0,116	1,802	22,523	381,245	5,121
4	41,08	0,48	11,01	740,566	0,168	0,128	19,718	452,291	5,285
5	56,29	0,78	8,54	144,080	0,012	4,463	43,906	480,717	6,661
6	68,51	0,98	7,51	0,047	0,008	9,875	67,140	514,510	7,360
7	75,01	0,94	12,36	45,114	0,002	2,916	70,509	927,124	11,618
8	89,05	1,21	10,81	430,839	0,102	0,025	107,751	962,631	13,080
9	91,13	1,29	9,89	521,513	0,160	0,581	117,558	901,276	12,758
10	91,26	1,12	13,72	527,468	0,053	9,410	102,211	1252,087	15,366
11	99,84	1,29	12,27	995,192	0,160	2,616	128,794	1225,037	15,828
12	108,55	1,49	13,92	1620,599	0,360	10,677	161,740	1511,016	20,741
Среднее значение	68,293	0,89	10,653	753,400	0,158	4,036	71,533	763,416	9,914

Отметим, что в последней строке расчетной табл. 2.10 приведены средние значения величин каждого столбца. Имеем следующие результаты:

$$S_y = 27,448; \quad S_{x_1} = 0,398; \quad S_{x_2} = 2,009;$$

$$r_{yx_1} = 0,984; \quad r_{yx_2} = 0,651; \quad r_{x_1x_2} = 0,542.$$

Линейные коэффициенты частной корреляции рассчитываются по формулам:

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}} = \frac{0,984 - 0,651 \cdot 0,542}{\sqrt{(1 - 0,651^2) \cdot (1 - 0,542^2)}} = 0,9899;$$

$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1x_2}^2)}} = \frac{0,651 - 0,984 \cdot 0,542}{\sqrt{(1 - 0,984^2) \cdot (1 - 0,542^2)}} = 0,796;$$

$$r_{x_1x_2/y} = \frac{r_{x_1x_2} - r_{yx_1} r_{yx_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{yx_2}^2)}} = \frac{0,542 - 0,984 \cdot 0,651}{\sqrt{(1 - 0,984^2) \cdot (1 - 0,651^2)}} = -0,738.$$

Если сравнивать значения коэффициентов парной и частной корреляции, то из-за значительной межфакторной связи и существует различие между этими коэффициентами.

Коэффициент множественной корреляции определим через матрицу парных коэффициентов корреляции по формуле

$$r_{yx_1x_2} = \sqrt{1 - \frac{|R|}{R_{11}}},$$

$$\text{где } |R| = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1x_2} \\ r_{yx_2} & r_{x_1x_2} & 1 \end{vmatrix} = 0,0082;$$

$$R_{11} = \begin{vmatrix} 1 & r_{x_1x_2} \\ r_{x_1x_2} & 1 \end{vmatrix} = 0,7058;$$

$$r_{yx_1x_2} = \sqrt{1 - \frac{0,0082}{0,7058}} \approx 0,994.$$

Вывод: зависимость y от x_1 и x_2 очень тесная.

Контрольные вопросы

1. Сформулируйте основную задачу корреляционного анализа.
2. Перечислите виды корреляционных зависимостей.
3. Раскройте сущность парной корреляции.
4. Укажите свойства коэффициента парной корреляции.
5. Дайте определение понятия множественной корреляции.
6. Перечислите свойства множественной корреляции.
7. Опишите этапы построения регрессионной модели.
8. Каков экономический смысл параметров моделей парной линейной, множественной линейной регрессий? Назовите отличия в интерпретации параметров этих моделей.
9. Опишите алгоритм метода наименьших квадратов для оценки параметров парной и множественной регрессий.
10. Дайте понятие линейной модели множественной регрессии.

11. Перечислите предпосылки нормальной линейной регрессионной модели.

12. Дайте определения понятиям некоррелированности, гомоскедастичности, гетероскедастичности остатков, мультиколлениарности факторов, классической нормальной регрессионной модели.

13. Сформулируйте теорему Маркова–Гаусса

14. Раскройте сущность коэффициента детерминации.

15. Расскажите о роли интервального оценивания функции регрессии и ее параметров в анализе регрессионной модели.

16. Дайте определение понятию нелинейной регрессионной модели. Перечислите методы оценок параметров НРМ.

3. ВРЕМЕННЫЕ РЯДЫ

Для построения эконометрической модели используют два типа исходных данных:

- характеризующие совокупность различных объектов в определенный момент (период времени);
- характеризующие один объект за ряд последовательных моментов времени.

Пространственными называются модели, построенные по данным первого типа; *моделями временных рядов* называются модели, построенные по данным второго типа.

Временным рядом в экономике называется 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.

Используя характер образования значений временного ряда, следует иметь в виду четыре типа факторов: долговременные, сезонные, циклические и случайные. При этом необязательно участие первых трех компонент:

$$y_t = I(u) \cdot u_t + I(v) \cdot v_t + I(c) \cdot c_t + \varepsilon_t \rightarrow (t = 1, n), \quad (3.1)$$

где $I(k) = \begin{cases} 1, & \text{если } k\text{-составляющая участвует в равенстве (2.16),} \\ 0, & \text{если } k\text{-составляющая отсутствует в равенстве (2.16);} \end{cases}$

u_t – функция трендов, описывающая влияние долговременных факторов на изменение значений Y_t ;

v_t – функция, описывающая воздействие сезонных факторов на значение Y_t ;

c_t – функция циклических факторов, отражающая повторяемость экономических процессов в течение длительных периодов;

ε_t – случайная компонента, или «случайные остатки», отражающая влияние не поддающихся учету факторов.

Первые три компонента в равенстве (3.1) являются закономерными, неслучайными.

Основная задача эконометрического исследования отдельного временного ряда состоит в определении количественного выраже-

ния каждой из перечисленных компонент в равенстве (2.16), для того, чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении модели-взаимосвязи двух и более временных рядов.

Совокупность существующих методов и моделей исследования временных рядов (рядов зависимых наблюдений) называется *анализом временных рядов*.

Основными этапами анализа временных рядов являются:

- 1) графическое представление и описание поведения временного ряда;
- 2) выделение и удаление закономерных (неслучайных) составляющих временного ряда (тренда, сезонных и циклических сопоставляющих);
- 3) сглаживание и фильтрация (удаление низко- или высокочастотных составляющих временного ряда);
- 4) исследование случайной составляющей временного ряда, построение и проверка адекватности математической модели для ее описания;
- 5) прогнозирование развития изучаемого процесса на основе имеющегося временного ряда;
- 6) исследование взаимосвязи между различными временными рядами.

Среди наиболее распространенных методов анализа временных рядов являются корреляционный и спектральный анализ, модели авторегрессии и скользящей средней. Члены временного ряда не являются статистически независимыми, не являются одинаково распределенными.

Если в результате выполнения п. 2 в списке этапов анализа временных рядов выяснилось, что имеющиеся наблюдения взаимно зависимы и неодинаково распределены (т. е. действительно образуют временной ряд), то приступают к подбору подходящей модели этого ряда. Множество моделей ограничивается обычно двумя классами линейных моделей: классом *стационарных временных рядов* (которые используются для описания поведения случайных остатков) и классом *нестационарных временных рядов*.

Стационарные (в широком смысле) временные ряды характеризуются тем, что их средние значения, дисперсии, ковариации не зависят от времени.

Временной ряд называется *нестационарным однородным*, когда существует такой порядок k , при котором последовательные разности $\Delta^k y(t)$ этого ряда порядка k образуют стационарный временный ряд, т. е. когда ε_t (случайный остаток) – стационарный временной ряд. Реальные временные ряды, встречающиеся в экономике, финансах, торговле, маркетинге, обычно являются нестационарными. Их нестационарность чаще всего проявляется в наличии зависящей от времени t неслучайной составляющей. В дальнейшем будем рассматривать только стационарные ряды.

Тренд характеризует устойчивую динамику показателя в течение предыдущего периода времени. Сезонная составляющая характеризует внутригодичные колебания значений показателя, обычно связанные с сезонностью (зима, весна, лето, осень) экономической конъюнктуры. Циклическая компонента рассматривается только для временных рядов, охватывающих периоды более десяти лет, и характеризует долгосрочные колебания экономической конъюнктуры, связанные с обновлением основных средств, изменениями технологий и т. д. Эти колебания рассматриваются обычно в макроэкономических исследованиях.

При прогнозировании временных рядов случайную составляющую не рассматривают, ограничиваясь оценкой ее дисперсии, для получения доверительного интервала прогнозируемых значений рассматриваемого показателя. Методы экстраполяции динамики временных рядов основаны на предположении, что закономерности изменения показателя в прошлом сохранятся и на период прогнозирования.

3.1. Автокорреляционная функция

Стационарные временные ряды – это ряды, вероятные свойства которых не изменяются во времени. Стационарные временные ряды применяются при описании случайных составляющих анализируемых рядов.

Временной ряд $y_t (t = 1, 2, \dots, n)$ называется *строго стационарным*, когда совместное распределение вероятностей n наблюдений y_1, y_2, \dots, y_n такое же, как и n наблюдений $y_{1+\tau}, y_{2+\tau}, \dots, y_{n+\tau}$ при любых n, τ, t . Свойства строго стационарных рядов y_t не зависят от момента t : и закон распределения, и его числовые характеристики

не зависят от t . Значит, математическое ожидание $a_y(t) = a$ с квадратичным отклонением $\sigma_y(t) = \sigma$ оценивается по наблюдениям $y_t (t = 1, 2, \dots, n)$ по формулам

$$\bar{y}_t = \frac{\sum_{i=1}^n y_i}{n};$$

$$S_t^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_t)^2}{n}.$$

Примером стационарного временного ряда служит и «белый шум», для которого $M(y_t) = a$, ε_t – некоррелированы.

Можно сказать, что $\rho(\tau)$ в классической линейной регрессионной модели образуют «белый шум».

Степень тесноты связи между последовательностями наблюдений временного ряда y_1, y_2, \dots, y_n и $y_{1+\tau}, y_{2+\tau}, \dots, y_{n+\tau}$ может быть определена с помощью коэффициента корреляции $\rho(\tau)$:

$$\rho(\tau) = \frac{M((y_t - a) \cdot (y_{t+\tau} - a))}{\sigma_y(t) \cdot \sigma_y(t + \tau)} = \frac{M((y_t - a) \cdot (y_{t+\tau} - a))}{\sigma^2}.$$

Коэффициент $\rho(\tau)$, измеряющий корреляцию между членами одного и того же ряда, называют *коэффициентом автокорреляции*, а зависимость $\rho(\tau)$ – *автокорреляционной функцией*.

Автокорреляционная функция зависит только от лага τ , где $\rho(-\tau) = \rho(\tau)$, учитывая стационарность временного ряда. При изучении $\rho(\tau)$ можно ограничиться только $\tau > 0$.

Статистической оценкой $\rho(\tau)$ является выборочный коэффициент корреляции

$$r(t) = \frac{(n - \tau) \cdot \sum_{t=1}^{n-\tau} y_t \cdot y_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \cdot \sum_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{(n - \tau) \cdot \sum_{t=1}^{n-\tau} y_t^2 - (\sum_{t=1}^{n-\tau} y_t)^2} \cdot \sqrt{(n - \tau) \sum_{t=1}^{n-\tau} y_{t+\tau}^2 - (\sum_{t=1}^{n-\tau} y_{t+\tau})^2}}. \quad (3.2)$$

С увеличением τ число пар $y_t, y_{t+\tau} - (n - \tau)$ уменьшается, поэтому лаг должен быть таким, чтобы число $(n - \tau)$ было достаточным для определения $r(\tau)$: $\tau \leq \frac{n}{4}$.

Для стационарных временных рядов с увеличением τ лага взаимосвязь членов временного ряда y_t и $y_{t+\tau}$ ослабевает и $r(\tau)$ должен убывать по абсолютной величине. В то же время для $r(\tau)$ при небольшом числе пар наблюдений $n - \tau$, свойство монотонного убывания при росте τ может нарушаться.

Одновременно с автокорреляционной функцией при исследовании стационарных временных рядов рассматривается частная автокорреляционная функция, имеющая $\rho_{\text{част}}(\tau)$ – частный коэффициент автокорреляции.

Коэффициентом автокорреляции называется коэффициент корреляции между t_τ и $t_{t+\tau}$ при устранении влияния промежуточных членов.

Статистической оценкой $\rho_{\text{част}}(\tau)$ является выборочная частная автокорреляционная функция, где $r_{\text{част}}(\tau)$ – выборочный частный коэффициент корреляции. Выборочный частный коэффициент автокорреляции первого порядка между членами временного ряда Y_t и Y_{t+2} может быть вычислен по формуле

$$r_{\text{част}}(2) = \frac{r(2) - r(1) \cdot r(1,2)}{\sqrt{(1 - r^2(1)) \cdot (1 - r^2(1,2))}}, \quad (3.3)$$

где $r(1)$, $r(2)$, $r(1,2)$ – выборочные коэффициенты автокорреляции между y_t и y_{t+1} , y_t и y_{t+2} , y_{t+1} и y_{t+2} , $t = 1, 2, \dots, n$, соответственно.

Знание выборочных автокорреляционных функций $r(\tau)$ и $r_{\text{част}}(\tau)$ может оказать существенную помощь при подборе и идентификации модели анализируемого временного ряда и статистической оценке его параметров.

Отсутствие корреляции между соседними членами дает возможность считать, что корреляция отсутствует в целом. Метод наименьших квадратов дает адекватные и эффективные результаты.

Рассмотрим два способа обнаружения автокорреляции остатков $e_t = y_t - \bar{y}_t$ (а, следовательно, и случайных составляющих):

1. Графический. С помощью метода наименьших квадратов оценивается регрессия. Рассчитываются остатки e_t . Строится график зависимости остатков e_t от номера наблюдения – $t \cdot (t = 1, 2, \dots, n)$.

2. Основан на применении критерия Дарбина–Уотсона. Тест Дарбина–Уотсона определяет наличие автокорреляции между соседними членами. Если корреляция ошибок регрессии не равна нулю, то она присутствует и в остатках регрессии e_t , получающихся в результате применения обычного метода наименьших квадратов.

В тесте Дарбина–Уотсона для оценки корреляции используется статистика вида

$$D = \frac{\sum_{t=1}^{n-1} (e_t - e_{t+1})^2}{\sum_{t=1}^n e_t^2}.$$

Для вычислений рекомендуется пользоваться формулой

$$D \approx 2 \cdot \left(1 - \frac{\sum_{t=1}^{n-1} e_t \cdot e_{t+1}}{\sum_{t=1}^n e_t^2}\right).$$

Критерий Дарбина–Уотсона не является статистическим, т. к. распределение статистики D зависит не только от числа наблюдений, но и от значений факторов (регрессоров). Поэтому нельзя указать критическую область, которая позволяла бы отвергнуть гипотезу об отсутствии корреляции, если бы оказалось, что в эту область попало наблюдаемое значение статистики D , но существуют два пороговых значения d_1 и d_2 , зависящие только от числа наблюдений, числа регрессоров и уровня значимости, такие, что выполняются условия:

1. Если $d_2 < D < 4 - d_2$, то принимается гипотеза об отсутствии автокорреляции.

2. Если $d_1 < D < d_2$ или $4 - d_2 < D < 4 - d_1$, то вопрос о присутствии автокорреляции открыт (зоны неопределенности).

3. Если $0 < D < d_1$, то принимается гипотеза о положительной автокорреляции.

4. Если $4 - d_2 < D < 4$, то принимается гипотеза об отрицательной автокорреляции. Применение теста Дарбина–Уотсона показано в *примере 1*.

Одной из важнейших задач исследования экономического временного ряда является выделение основной тенденции изучаемого процесса, выраженной неслучайной составляющей ($U(t), V(t), C(t)$). Для этого на начальном этапе анализа ВР проверяют гипотезу о неизменности среднего значения ВР. Проверку гипотезы $M(Y_t) = \text{const}$ можно проводить с помощью критерия серий. Подробнее критерий серий можно изучить по книге.

Для выявления основной тенденции чаще всего используется МНК. 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. Полученный таким образом ряд скользящих средних ведет себя более гладко, чем исходный ряд.

Если индивидуальный разброс значений членов временного ряда Y_t около своего среднего значения a характеризуется дисперсией σ^2 , то разброс средней из m членов временного ряда $(y_1 + y_2 + \dots + y_m) / m$ около того же значения a будет характеризоваться дисперсией σ^2 / m .

3.2. Примеры решения задач

Пример 1. Имеются данные (табл. 3.1) о величине доходов на одного члена семьи за шесть лет и расходы на товар. Требуется:

1. Построить линейную модель спроса на товар, включив в нее фактор времени. Интерпретировать полученные параметры.
2. Выявить на уровне значимости 0,05 наличие автокорреляции возмущений для временного ряда по критерию Дарбина–Уотсона. Оценить полученный результат.
3. Указать, пригоден ли уравнение для прогноза.

Таблица 3.1

Исходные данные

Показатель	2017	2018	2019	2020	2021	2022
Условное время x_1	1	2	3	4	5	6
Расход на товар x_2 , руб.	27	33	37	42	49	52
Доход на одного человека семьи y , % к 2010 г.	100	102	106	108	113	118

Решение:

1. Построить линейную модель спроса на товар, включив в нее фактор времени и интерпретировать полученные параметры.

Составить модель множественной регрессии для случая числа факторов $m = 2$, имея линейную модель множественной регрессии:

$$y = a_0 + a_1x_1 + a_2x_2,$$

где a_0, a_1, a_2 – оценки параметров, найденные по МНК.

Для нахождения коэффициентов регрессии составить таблицу на примере табл. 3.2.

Таблица 3.2

Расчетные данные

$n = 6$	x_1	x_2	y	$x_1 \cdot x_1$	$x_1 \cdot x_2$	$x_1 \cdot y$	$x_2 \cdot x_2$	$x_2 \cdot y$	$y_i^{(m)}$	ε_i	$(y_i - y_i^{(m)})^2$	$(y_i - \bar{y})^2$
1	1	27	100	1	27	100	729	2700	98,98		1,05	61,36
2	2	33	102	4	66	204	1089	3366	102,32	-0,316	0,1	34,03
3	3	37	106	9	111	318	1369	3922	106,16	-0,164	0,03	3,36
4	4	42	108	16	168	432	1764	4536	109,76	0,158	3,09	0,03
5	5	49	113	25	245	565	2401	5537	112,84	0	0,03	26,69
6	6	52	118	36	312	708	2704	6136	116,94		1,12	103,36
Сумма	21	240	647	91	929	2327	10056	26197	647		5,4	228,83
Среднее значение	3,5	40	107,83	15,17	154,83	387,83	1676	4366,17	107,83		0,9	38,14
	\bar{x}_1	\bar{x}_2	\bar{y}	\bar{x}_1^2	$\bar{x}_1 \cdot \bar{x}_2$	$\bar{x}_1 \cdot \bar{y}$	\bar{x}_2^2	$\bar{x}_2 \cdot \bar{y}$	$\bar{y}_i^{(m)}$	$\bar{\varepsilon}_i$	$\overline{(y_i - y_i^{(m)})^2}$	$\overline{(y_i - \bar{y})^2}$

Составить нормальную систему уравнений для определения параметров принятой модели:

$$\begin{cases} a_0 + a_1 \cdot \bar{x}_1 + a_2 \cdot \bar{x}_2 = y; \\ a_0 \cdot \bar{x}_1 + a_1 \cdot \bar{x}_1^2 + a_2 \cdot \overline{x_1 \cdot x_2} = \overline{x_1 y}; \\ a_0 \cdot \bar{x}_2 + a_1 \cdot \overline{x_1 x_2} + a_2 \cdot \bar{x}_2^2 = \overline{x_2 y}. \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} a_0 + a_1 \cdot 3,5 + a_2 \cdot 40 = 107,83; \\ a_0 \cdot 3,5 + a_1 \cdot 15,17 + a_2 \cdot 154,83 = 387,83; \\ a_0 \cdot 40 + a_1 \cdot 154,83 + a_2 \cdot 1676 = 4366,17. \end{cases}$$

Решаем систему и находим коэффициенты:

$$a_0 = 100,98; \quad a_1 = 4,86; \quad a_2 = -0,25.$$

Таким образом,

$$y = a_0 + a_1 x_1 + a_2 x_2 \Rightarrow y = 100,98 + 4,86 x_1 - 0,25 x_2.$$

Используем матричный способ решения:

$$A = (X^T \cdot X)^{-1} \cdot X^T \cdot Y,$$

$$\text{где } X = \begin{pmatrix} 1 & 1 & 27 \\ 1 & 2 & 33 \\ 1 & 3 & 37 \\ 1 & 4 & 42 \\ 1 & 5 & 49 \\ 1 & 6 & 52 \end{pmatrix}; \quad Y = \begin{pmatrix} 100 \\ 102 \\ 106 \\ 108 \\ 113 \\ 118 \end{pmatrix}; \quad A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}.$$

$$\text{В результате получено решение: } A = \begin{pmatrix} 100,98 \\ 4,86 \\ -0,25 \end{pmatrix}.$$

Объясненная часть модели, т. е. уравнение регрессии имеет вид:

$$y = 100,98 + 4,86x_1 - 0,25x_2.$$

Значения объясненной части приведены в табл. 3.2, их можно сравнить с наблюдаемыми значениями Y .

В теореме Гаусса–Маркова требуется, чтобы остатки ε_t , ε_j были независимыми при наличии зависимости между остатками при различных номерах наблюдений. Явление автокорреляции состоит в отсутствии этого условия, которое возникает при наличии зависимости между остатками при различных номерах наблюдений. Автокорреляция означает корреляционную зависимость последующих уровней временного ряда от его предыдущих уровней.

Коэффициент автокорреляции, показывающий зависимость между соседними уровнями t , $t + k$ временного ряда с разницей номеров, равной k (т. е. лаг равен t), находится по формуле

$$r_k = \frac{\overline{\varepsilon_t \cdot \varepsilon_{t+k}} - \overline{\varepsilon_t} \cdot \overline{\varepsilon_{t+k}}}{S_1 \cdot S_2},$$

где средние вычисляются как $\overline{\varepsilon_t \cdot \varepsilon_{t+k}} = \frac{1}{n-k} \cdot \sum_{t=1}^{n-k} \varepsilon_t \cdot \varepsilon_{t+k}$.

Составить расчетную таблицу на примере табл. 3.3.

Таблица 3.3

Расчетные данные

t	ε_t	ε_{t+1}	$\varepsilon_t \cdot \varepsilon_{t+1}$	$(\varepsilon_t)^2$	$(\varepsilon_t - \overline{\varepsilon_t})^2$	$(\varepsilon_{t+1} - \overline{\varepsilon_{t+1}})^2$	$(\varepsilon_t - \varepsilon_{t+1})^2$
1	1,02	-0,32	-0,3264	1,0404	1,517824	0,013456	1,7956
2	-0,32	-0,16	0,0512	0,1024	0,011664	0,001936	0,0256
3	-0,16	-1,76	0,2816	0,0256	0,002704	2,421136	2,56
4	-1,76	0,16	-0,2816	3,0976	2,396304	0,132496	3,6864
5	0,16	1,06	0,1696	0,0256	0,138384	1,597696	0,81
Сумма	-1,06	-1,02	-0,1056	4,2916	4,06688	4,16672	8,8776
Среднее значение	0,212	0,204	-0,02112	0,85832			

Средние квадратические отклонения $S_1 \cdot S_2$, при $k = 1$,

$$S_1 = \sqrt{\frac{1}{n-2} \sum_{t=1}^{n-1} (\varepsilon_t - \bar{\varepsilon}_t)^2} = 1,1643;$$

$$S_2 = \sqrt{\frac{1}{n-2} \sum_{t=1}^{n-1} (\varepsilon_{t+1} - \bar{\varepsilon}_{t+1})^2} = 1,1785.$$

Выборочный коэффициент автокорреляции

$$r_1 = \frac{\overline{\varepsilon_t \cdot \varepsilon_{t+1}} - \bar{\varepsilon}_t \cdot \bar{\varepsilon}_{t+1}}{S_1 \cdot S_2} = -0,0469.$$

2. Гипотеза H_0 об отсутствии автокорреляции в ряде остатков проверяется по статистике (критерий Дарбина–Уотсона):

$$D = \frac{\sum_{t=1}^{n-1} (\varepsilon_t - \varepsilon_{t+1})^2}{\sum_{t=1}^n \varepsilon_t^2} = 1,6394,$$

где $\sum_{t=1}^{n-1} (\varepsilon_t - \varepsilon_{t+1})^2 = 8,8776$;

$$\sum_{t=1}^n \varepsilon_t^2 = 5,4152.$$

Значение статистики D удовлетворяет условиям $0 \leq D \leq 4$.

Для заданного уровня значимости α , объема выборки n и числа регрессоров (факторов) m определяют две величины: d_1 (нижнюю границу значения статистики), d_2 (верхнюю границу значения статистики). Если выполнено $d_2 \leq D \leq 4 - d_2$, то гипотеза H_0 об отсутствии автокорреляции остатков принимается. Если же верно $0 \leq D \leq d_1$ или $4 - d_1 \leq D \leq 4$, то эта гипотеза отвергается. В осталь-

ных случаях решение о принятии или отклонении гипотезы остается неопределенным. В нашем случае имеем число объясняющих переменных $m = 3$, уровень значимости $\alpha = 0,05$, число уровней ряда $n = 6$. Значения нижней и верхней границ взяты из прил. Б: $d_1 = 0,61$ и $d_1 = 1,4$. Следовательно, условие $d_2 \leq D \leq 4 - d_2$ выполнено, поэтому гипотеза об отсутствии автокорреляции в ряде остатков принимается.

3. Данная модель применима для прогноза, т. к. выполнены условия теоремы Гаусса–Маркова, отсутствует автокорреляция. Пригодность модели для прогноза также подтверждается дисперсией вектора остатков:

$$S_0^2 = \frac{1}{n - m - 1} \cdot \sum_{t=1}^n \varepsilon_t^2 = \frac{1}{3} \cdot \sum_{t=1}^6 \varepsilon_t^2 = 1,8051.$$

Дисперсия достаточно мала, что свидетельствует о хорошем соответствии модели.

Значимость уравнения регрессии можно установить с помощью критерия Фишера.

Пример 2. Имеются данные о величине доходов на одного члена семьи за шесть лет и расходы на товар A (табл. 3.4).

Таблица 3.4

Исходные данные

Показатель	2017	2018	2019	2020	2021	2022
Условное время x_1	1	2	3	4	5	6
Расход на товар A x_2 , руб.	27	33	37	42	49	52
Доход на одного человека семьи y , % к 2014 г.	100	102	106	108	113	118

Провести сглаживание временного ряда методом скользящих средних с интервалом сглаживания $m = 3$ года.

Решение. Используя простую среднюю арифметическую с интервалом $m = 3$, рассчитаем средние по формуле

$$\tilde{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i}{m},$$

где $p = \frac{m-1}{2}$;

m – нечетное.

Получим

$$\tilde{y}_2 = \frac{100 + 102 + 106}{3} = 102,67;$$

$$\tilde{y}_3 = \frac{102 + 106 + 108}{3} = 105,33;$$

$$\tilde{y}_4 = \frac{100 + 108 + 113}{3} = 109;$$

$$\tilde{y}_5 = \frac{108 + 113 + 118}{3} = 113.$$

Для удобства составим табл. 3.5.

Таблица 3.5

Расчетные данные

t	1	2	3	4	5	6
\tilde{y}_t	–	102,67	105,33	109	113	–

В результате применения метода «скользящих средних» получили сглаженный ряд, который характеризует поведение неслучайной составляющей.

Контрольные вопросы

1. Дайте определение эконометрического временного ряда, стационарных и нестационарных временных рядов, автокорреляционной функции.
2. Назовите составляющие, участвующие в формировании уровней временного ряда.
3. Перечислите основные этапы анализа временных рядов.
4. Раскройте сущность автокорреляции остатков временного ряда.
5. Охарактеризуйте методы выделения неслучайной составляющей временного ряда.
6. Каков смысл прогнозирования на основе моделей временных рядов.

4. СИСТЕМЫ ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ

4.1. Понятие систем эконометрических уравнений

Одной из причин коррелированности объясняющих переменных со случайными членами являются факторы, действующие одновременно и на объясняющие переменные, и на объясняемые переменные при фиксированных значениях объясняющих переменных, т. е. в рассматриваемой экономической ситуации значения объясняемых переменных и объясняющих формируются одновременно под воздействием некоторых внешних факторов. Значит, рассматриваемая модель не полна и ее следует дополнять уравнениями, в которых объясняемыми переменными должны выступать объясняющие переменные. Следовательно, появляется необходимость рассматривать системы одновременных уравнений.

Системой одновременных уравнений (СОУ) называется набор взаимосвязанных регрессионных моделей, в которых одни и те же переменные могут одновременно быть в различных уравнениях системы объясняемыми и объясняющими переменными.

Классическим примером СОУ является одновременное формирование спроса Q^d и предложения Q^s , товара в зависимости от его цены P :

$$\begin{cases} Q^d = \beta_1 + \beta_2 P + \beta_3 I + E_1; \\ Q^s = \beta_4 + \beta_5 P + E_2, \end{cases} \quad (4.1)$$

где I – доход на душу населения.

Если предположить, что рынок находится в состоянии равновесия, то в равенствах (4.1) $Q^d = Q^s = Q$, а P – цена равновесия, формируемая одновременно спросом и предложением. Очевидно, что P и Q считаются объясняемыми переменными, а величина дохода I – объясняющей. Величины P и Q являются эндогенными, т. к. формируются внутри модели, а переменная I – экзогенной (задающейся вне модели).

Пусть y_1, \dots, y_m – эндогенные переменные; x_1, \dots, x_e – экзогенные переменные.

Введем матрицы, имеющие вид

$$B = \begin{pmatrix} \beta_{11} & \dots & \beta_{1m} \\ \dots & \dots & \dots \\ \beta_{m1} & \dots & \beta_{mm} \end{pmatrix};$$

$$\Gamma = \begin{pmatrix} \gamma_{11} & \dots & \gamma_{1e} \\ \dots & \dots & \dots \\ \gamma_{m1} & \dots & \gamma_{me} \end{pmatrix},$$

следовательно, общий вид СОУ представляется в матричной форме:

$$\begin{aligned} BY + \Gamma X &= E, \\ E &= (E_1, \dots, E_m)^T. \end{aligned} \tag{4.2}$$

Помимо регрессионных одновременных уравнений (4.2) модель может содержать тождества, которые представляют собой алгебраические соотношения между эндогенными переменными ($N: Q^d = Q^s$).

Тождества позволяют исключать некоторые эндогенные переменные и рассматривать систему регрессионных уравнений меньшей размерности.

$$Q^s = Q^d = Q;$$

$$Y = \begin{pmatrix} Q \\ P \end{pmatrix}; \quad X = \begin{pmatrix} 1 \\ I \end{pmatrix}; \quad E = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix};$$

$$B = \begin{pmatrix} 1 & -\beta_2 \\ 1 & -\beta_5 \end{pmatrix}; \quad \Gamma = \begin{pmatrix} -\beta_1 & -\beta_3 \\ -\beta_4 & 0 \end{pmatrix};$$

$$\begin{pmatrix} 1 & -\beta_2 \\ 1 & -\beta_5 \end{pmatrix} \begin{pmatrix} Q \\ P \end{pmatrix} + \begin{pmatrix} -\beta_1 & -\beta_3 \\ -\beta_4 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ I \end{pmatrix} = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix};$$

$$E_1 = Q - \beta_1 - \beta_2 P - \beta_3 I; \quad E_2 = Q - \beta_4 - \beta_5 P.$$

Рассмотрим случай двух уравнений с двумя эндогенными переменными:

$$\begin{cases} y_1 = \alpha_1 + \beta_1 x_1 + \gamma_1 y_2 + E_1; \\ y_2 = \alpha_2 + \beta_2 x_2 + \gamma_2 y_1 + E_2. \end{cases} \quad (4.3)$$

Если применить к уравнениям (4.3) обычный МНК, то получатся несостоятельные оценки параметров $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$. Оценивание СОУ требует специальных методов.

4.2. Идентификация систем одновременных уравнений

В основе *косвенного метода наименьших квадратов* (КМНК) лежит следующий алгоритм: для устранения коррелированности эндогенных переменных со случайными членами следует разрешить СОУ относительно эндогенных переменных Y так, чтобы в правых частях уравнений вида (4.3) остались только экзогенные переменные X . Применить обычный МНК к полученным уравнениям и получить оценки выражений от исходных параметров, из которых найти оценки параметров.

Применяя процедуру КМНК к системе (4.3), т. к. $y_2 = \alpha_2 + \beta_2 x_2 + \gamma_2 y_1 + E_2$, имеем

$$y_1 = \alpha_1 + \beta_1 x_1 + \gamma_1 (\alpha_2 + \beta_2 x_2 + \gamma_2 y_1 E_2) + E_1;$$

$$y_1 (1 - \gamma_1 \gamma_2) = (\alpha_1 + \gamma_1 \alpha_2) + \beta_1 x_1 + \gamma_1 \beta_2 x_2 + E_1 + \gamma_1 E_2;$$

$$y_1 = \underbrace{\frac{\alpha_1 + \gamma_1 \alpha_2}{1 - \gamma_1 \gamma_2}}_{\alpha_1} + \underbrace{\frac{\beta_1}{1 - \gamma_1 \gamma_2}}_{b_1} x_1 + \underbrace{\frac{\gamma_1 \beta_2}{1 - \gamma_1 \gamma_2}}_{c_1} x_2 + \underbrace{\frac{E_1 + \gamma_1 E_2}{1 - \gamma_1 \gamma_2}}_{v_1},$$

или

$$y_1 = \alpha_1 + b_1 x_1 + c_1 x_2 + v_1;$$

$$y_1 = \alpha_2 + \beta_2 x_2 + \gamma_2 y_1 + E_2 = \alpha_2 + \beta_2 x_2 + \gamma_2 \times \\ \times \left(\frac{\alpha_1 + \gamma_1 \alpha_2}{1 - \gamma_1 \gamma_2} + \frac{\beta_1}{1 - \gamma_1 \gamma_2} x_1 + \frac{\gamma_1 \beta_2}{1 - \gamma_1 \gamma_2} x_2 + \frac{E_1 + \gamma_1 E_2}{1 - \gamma_1 \gamma_2} \right) + E_2.$$

Преобразуя, имеем

$$\left[\alpha_2 + \frac{\gamma_2 \alpha_1 + \gamma_1 \gamma_2 \alpha_2}{1 - \gamma_1 \gamma_2} \right] + \left[\beta_2 + \frac{\gamma_2 \gamma_1 \beta_2}{1 - \gamma_1 \gamma_2} \right] x_2 + \frac{\gamma_2 \beta_1}{1 - \gamma_1 \gamma_2} x_1 + \\ + \left[\frac{E_1 + \gamma_1 \gamma_2 E_2}{1 - \gamma_1 \gamma_2} + E_2 \right] = \left[\frac{\alpha_2 - \gamma_1 \gamma_2 \alpha_2 + \gamma_2 \alpha_1 + \gamma_1 \gamma_2 \alpha_2}{1 - \gamma_1 \gamma_2} \right] + \frac{\gamma_2 \beta_1}{1 - \gamma_1 \gamma_2} x_1 + \\ + \left[\frac{\beta_2 - \gamma_1 \gamma_2 \beta_2 + \gamma_1 \gamma_2 \beta_2}{1 - \gamma_1 \gamma_2} \right] x_2 + \left[\frac{E_1 \gamma_2 + \gamma_1 \gamma_2 E_2 + E_2 - \gamma_1 \gamma_2 E_2}{1 - \gamma_1 \gamma_2} \right] = \\ = \frac{\alpha_2 + \gamma_2 \alpha_1}{1 - \gamma_1 \gamma_2} + \frac{\gamma_2 \beta_1}{1 - \gamma_1 \gamma_2} x_1 + \frac{\beta_2}{1 - \gamma_1 \gamma_2} + \frac{E_1 \gamma_2 + E_2}{1 - \gamma_1 \gamma_2};$$

$$y_2 = a_2 + b_2 x_1 + c_2 x_2 + v_2.$$

Получим систему одновременных уравнений:

$$\begin{cases} y_1 = a_1 + b_1 x_1 + c_1 x_2 + v_1; \\ y_2 = a_2 + b_2 x_1 + c_2 x_2 + v_2. \end{cases} \quad (4.4)$$

Равенства позволяют однозначно выразить исходные параметры $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$ через \bar{a} , \bar{b} , \bar{c} .

$$a_1 = \frac{\alpha_1 + \gamma_1 \alpha_2}{1 - \gamma_1 \gamma_2}, \quad b_1 = \frac{\beta_1}{1 - \gamma_1 \gamma_2}; \quad c_1 = \frac{\gamma \beta_2}{1 - \gamma_1 \gamma_2}; \quad v_1 = \frac{E_1 + \gamma_1 \gamma_2}{1 - \gamma_1 \gamma_2}. \\ a_2 = \frac{\alpha_2 + \gamma_2 \alpha_1}{1 - \gamma_1 \gamma_2}; \quad b_2 = \frac{\gamma_2 \beta_1}{1 - \gamma_1 \gamma_2}; \quad c_2 = \frac{\beta_2}{1 - \gamma_1 \gamma_2}; \quad v_2 = \frac{E_1 \gamma_2 + E_2}{1 - \gamma_1 \gamma_2}. \quad (4.5)$$

Форма (4.2) называется структурной формулой системы уравнений, а форма (4.4) называется приведенной формой системы. Параметры приведенной формы оцениваются с помощью метода наименьших квадратов. Однако экономический смысл и интерес для анализа представляют параметры структурной формы (4.2), т. е. структурная форма раскрывает экономический механизм формирования эндогенных переменных.

Структурный параметр называется *идентификационным*, если он может быть однозначно оценен с помощью КМНК. Уравнение идентифицируемо, если идентифицируемы все входящие в него структурные параметры.

Структурный параметр называется *неидентифицируемым*, если его значение невозможно получить, зная точные значения параметров приведенной формы.

Параметр называется *сверхидентифицируемым*, если КМНК дает несколько различных его оценок, т. е. несколько способов выражения через параметры приведенной формы.

Проблема сверхидентифицируемости – это проблемы структуры модели и количества наблюдений. С увеличением объема выборки все различные самостоятельные оценки параметра стремятся к одному и тому же истинному значению. Неидентифицируемость не исчезает с ростом количества наблюдений и означает, что существует бесконечное число структурных моделей, имеющих одну и ту же приведенную форму. Для идентифицируемости необходимо, чтобы количество оцениваемых структурных параметров было равно количеству оцененных параметров приведенной формы.

Неидентифицируемость модели означает, что КМНК неприменим. Введем следующие обозначения:

M – число предопределенных переменных в модели;

m – число предопределенных переменных в данном уравнении;

K – число эндогенных переменных в модели;

k – число эндогенных переменных в данном уравнении;

A – матрица коэффициентов при переменных, не входящих в данное уравнение.

Необходимое условие идентификации уравнения модели:

Для того чтобы уравнение модели было идентифицируемо, необходимо, чтобы число предопределенных переменных, не входящих

в уравнение, было не меньше числа эндогенных переменных, входящих в уравнение, минус единица, т. е.

$$M - m \geq k - 1.$$

Если $M - m = k - 1$, уравнение точно идентифицировано.

Если $M - m > k - 1$, уравнение сверхидентифицировано.

Достаточное условие идентификации уравнения модели:

Для того чтобы уравнение было идентифицируемым, достаточно, чтобы ранг матрицы A был равен $K - 1$.

Правила проверки необходимого и достаточного условий идентификации уравнения модели (применяются только к структурной форме СОУ):

1. Если $M - m > k - 1$ и ранг матрицы A равен $K - 1$, то уравнение сверхидентифицировано.

2. Если $M - m = k - 1$ и ранг матрицы A равен $K - 1$, то уравнение точно идентифицировано.

3. Если $M - m \geq k - 1$ и ранг матрицы A меньше $K - 1$, то уравнение неидентифицировано.

4. Если $M - m < k - 1$ и ранг матрицы A меньше $K - 1$, то уравнение неидентифицировано.

Для решения идентифицируемого уравнения применяется КМНК, для решения сверхидентифицированных – *двухшаговый метод наименьших квадратов* (ДМНК). Суть его заключается в следующем: составляют приведенную форму модели, определяют численные значения параметров каждого ее уравнения обычным МНК; выявляют эндогенные переменные, находящиеся в правой части структурного уравнения, параметры которого определяют двухшаговым МНК, находят расчетные значения по соответствующим уравнениям приведенной формы модели; обычным МНК определяют параметры структурного уравнения, используя в качестве исходных данных фактические значения предопределенных переменных и расчетные значения эндогенных переменных, стоящих в правой части данного структурного уравнения.

Методы получения оценок параметров структурной формы СОУ:

- метод инструментальных переменных;
- метод одновременного оценивания регрессионных уравнений;
- трехшаговый метод наименьших квадратов.

Метод инструментальных переменных (МИП) является одним из самых распространенных методов оценивания уравнений, в которых регрессоры (факторы) коррелируют со свободными членами. Данное явление характерно для систем одновременных уравнений.

Суть МИП состоит в подборе новых переменных $z_j, j = 1, 2, \dots, l$, тесно скоррелированных с x_j и нескоррелированных с ε в модели $Y = XA + \varepsilon$. Набор $\{z_j\}$ содержит большее число переменных, называемых инструментальными переменными, предназначенными для построения состоятельной оценки вектор-параметров A .

Метод одновременного оценивания регрессионных уравнений подробно рассматривается в учебниках. КМНК сводится к оцениванию по отдельности уравнений приведенной формы. Эффективность оценивания можно повысить, если объединить уравнения системы

$$\begin{cases} y_1 = a_1 + b_1x_1 + v_1; \\ y_2 = a_2 + b_2x_2 + v_2, \end{cases}$$

или

$$Y = X \cdot A + v,$$

где $X = \begin{pmatrix} 1 & x_1 & 0 \\ 1 & 0 & x_2 \end{pmatrix};$

$$Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix};$$

$$A = \begin{pmatrix} a_1 & a_2 \\ b_1 & 0 \\ 0 & b_2 \end{pmatrix};$$

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix},$$

и применить к полученному уравнению МНК

$$A = (X^T \cdot \Sigma^{-1} \cdot X)^{-1} X^T \cdot \Sigma \cdot Y,$$

где Σ – ковариационная матрица случайных остатков.

Трехшаговый МНК – более эффективная процедура оценивания систем регрессионных уравнений, сочетающая метод одновременного оценивания и метод инструментальных переменных. Метод заключается в том, что на первом шаге к исходной модели применяется обобщенный МНК с целью устранения корреляции случайных членов, далее к полученным уравнениям применяется двухшаговый метод наименьших квадратов.

4.3. Типовые задания

Задание. Составить приведенную форму модели, проверить каждое уравнение структурной модели заданной системы одновременных уравнений на идентификацию

$$\begin{cases} Y_1 = b_{12}Y_2 + b_{13}Y_3 + a_{11}X_1 + a_{12}X_2 + u_1; \\ Y_2 = b_{21}Y_1 + a_{21}X_3 + u_2; \\ Y_3 = b_{32}Y_2 + a_{31}X_1 + a_{33}X_3 + u_3. \end{cases}$$

Решение. В данной системе:

Y_1, Y_2, Y_3 – эндогенные переменные ($K = 3$);

X_1, X_2, X_3 – predetermined переменные ($M = 3$).

Составим приведенную форму модели:

$$\begin{cases} Y_1 = A_{11}X_1 + A_{12}X_2 + A_{13}X_3 + U_1; \\ Y_2 = A_{21}X_1 + A_{22}X_2 + A_{23}X_3 + U_2; \\ Y_3 = A_{31}X_1 + A_{32}X_2 + A_{33}X_3 + U_3. \end{cases}$$

Проверка выполнения необходимого условия идентификации для каждого уравнения:

1-е уравнение:

$$\begin{aligned} M - m < k - 1 &\Rightarrow \text{неидентифицировано;} \\ k = 3; m = 2; M - m = 1; k - 1 = 2. \end{aligned}$$

2-е уравнение:

$$M - m = 2 > k - 1 = 1 \Rightarrow \text{сверхидентифицировано;} \\ k = 2; m = 1.$$

3-е уравнение:

$$M - m = 1 < k - 1 = 1 \Rightarrow \text{идентифицировано;} \\ k = 2; m = 2.$$

Для того чтобы выполнялось достаточное условие идентификации для каждого уравнения системы, необходимо, чтобы определитель матрицы A (матрицы коэффициентов при переменных, не входящих в это уравнение) был равен $k - 1 = 2$.

Составим матрицу A для первого уравнения системы. В первом уравнении отсутствует лишь одна переменная системы – X_3 . По-

этому матрица A будет иметь следующий вид: $A = \begin{pmatrix} 0 \\ a_{22} \end{pmatrix}$. Ранг этой

матрицы равен единице, что меньше $k - 1 = 2$, следовательно, первое уравнение модели неидентифицировано.

Составим матрицу A для второго уравнения системы. Во втором уравнении отсутствуют переменные X_1, X_2, X_3 . Матрица A имеет

следующий вид: $A = \begin{pmatrix} b_{13} & a_{13} & 0 \\ 1 & a_{32} & a_{33} \end{pmatrix}$. Ранг этой матрицы равен двум,

что равно $k - 1 = 2$, следовательно, второе уравнение модели точно идентифицировано.

Составим матрицу A для третьего уравнения системы. В третьем уравнении отсутствуют переменные Y_1, X_2 , тогда матрица имеет

следующий вид: $A = \begin{pmatrix} 1 & a_{12} \\ b_{21} & 0 \end{pmatrix}$. Ранг этой матрицы равен двум, что

равно $k - 1 = 2$, следовательно, третье уравнение идентифицировано.

Вывод: первое уравнение системы неидентифицировано, т. к. не выполняются достаточное и необходимое условия идентификации. Второе уравнение сверхидентифицировано. Очевидно, система в целом является неидентифицируемой.

Контрольные вопросы

1. Дайте понятие системы одновременных уравнений.
2. Укажите формы СОУ и способы перехода.
3. Раскройте сущность проблемы идентифицируемости СОУ.
4. Опишите необходимое и достаточные условия идентифицируемости уравнения.
5. Сформулируйте косвенный метод наименьших квадратов.
6. Опишите двухшаговый метод наименьших квадратов.
7. Раскройте сущность метода инструментальных переменных.
8. Укажите метод одновременного оценивания регрессионных уравнений.

5. ИНДИВИДУАЛЬНЫЕ ЗАДАНИЯ ПО ТЕМАМ

Методические указания по выполнению индивидуальных заданий приведены в прил. А.

5.1. Множественная регрессия

Задание. По 10 предприятиям региона (табл. 5.1) изучается зависимость выработки продукции на одного работника Y (тыс. руб.) от ввода в действие новых основных фондов X_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих X_2 (%).

Таблица 5.1

Исходные данные

Вариант	Параметры	№ предприятия									
		1	2	3	4	5	6	7	8	9	10
B1	y	5	6	7	7	8	8	9	10	11	12
	x_1	4,8	5,9	6,8	6,9	7,1	7,8	8,3	8,5	8,8	9,0
	x_2	10	14	20	21	21	22	22	23	30	31
B2	y	11	11	11	11	12	12	13	13	14	14
	x_1	8,5	8,6	8,8	8,9	9,1	9,2	9,3	9,5	9,7	9,8
	x_2	25	27	29	30	31	32	32	33	34	35
B3	y	7	7	7	8	8	8	9	9	10	10
	x_1	4,5	4,5	4,6	5,1	5,4	5,7	6,1	6,3	6,4	6,9
	x_2	15	16	16	17	17	19	19	21	21	22
B4	y	11	11	11	11	11	12	12	13	13	14
	x_1	6,1	6,2	6,3	6,5	6,7	7,1	7,3	7,7	8	8,5
	x_2	17	17	18	18	19	20	21	21	22	22
B5	y	8	8	8	8	9	9	10	10	10	10
	x_1	4	4,1	4,2	4,2	4,3	4,4	4,5	5	5,1	5,3
	x_2	15	16	16	17	19	19	20	20	21	21
B6	y	10	11	12	12	12	13	13	14		14
	x_1	7	7,1	7,2	7,3	7,3	7,5	7,6	8	8,1	8,3
	x_2	25	25	26	26	27	28	29	29	30	30
B7	y	7	7	8	8	9	9	9	9	10	11
	x_1	4,2	4,2	4,3	5	5	5,1	5,2	6	6,1	6,2
	x_2		9	9	10	10	11	12	14	15	15
B8	y	11	11	12	13	13	14	14	15	16	17
	x_1	7,8	7,9	8	8,1	8,3	8,5	8,6	8,8	9,1	9,4
	x_2		11	12	13	13	15	16	17	18	19

Вариант	Параметры	№ предприятия									
		1	2	3	4	5	6	7	8	9	10
В9	y	7	7	7	7	8	8	9	9	10	10
	x_1	3,6	3,7	3,9	4	4	4,1	4,3	4,4	5,3	5,5
	x_2	11	12	13	14	15	15	18	18	20	21
В10	y	9	9	10	10	11	11	12	13	13	14
	x_1	5,5	5,6	6,4	6,5	6,6	7,6	7,9	8,2	8,5	8,7
	x_2	18	18	19	19	20	20	21	21	22	22

Требуется:

1. Построить диаграммы рассеяния выработки продукции на одного работника Y (тыс. руб.), в зависимости от ввода в действие новых основных фондов X_1 и от удельного веса рабочих высокой квалификации в общей численности рабочих X_2 , определить форму связи между результирующим показателем (Y) и каждым из факторов (X_1 и X_2).

2. Оценить с помощью МНК параметры линейного регрессионного уравнения Y от X_1 и X_2 .

3. Рассчитать значение коэффициента детерминации и интерпретировать его.

4. Определить скорректированный коэффициент детерминации и сравнить его со значением обычного коэффициента детерминации.

5. Проверить значимость полученного уравнения регрессии на уровне 0,05.

6. Найти 95%-доверительные интервалы для значимых коэффициентов регрессии.

5.2. Временные ряды

Задание. Найти среднее значение, среднее квадратическое отклонение и коэффициенты автокорреляции (для лагов $\tau = 1; 2$) временного ряда, уравнение тренда временного линейного ряда y_t , при уровне значимости 0,05. Провести сглаживание временного ряда y_t методом скользящих средних, используя простую среднюю арифметическую с интервалом сглаживания $m = 3$. Дать точечный и, с надежностью 0,95, интервальный прогнозы среднего и индивидуального значений доходов на пятый год.

Для вариантов 1–10 известны данные об урожайности озимой пшеницы y_t (ц/га) за 7 лет, отражающие динамику роста доходов на душу населения y_t (ден. ед.) за семилетний период (табл. 5.2).

Таблица 5.2

Исходные данные

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

5.3. Системы эконометрических уравнений

Задание:

1. Применив необходимое и достаточное условие идентификации, определить, идентифицировано ли каждое из уравнений модели, согласно варианту (табл. 5.3).
2. Определите метод оценки параметров модели.
3. Записать приведенную форму модели.

Таблица 5.3

Исходные данные

Вариант 1	Вариант 2
<p>Модель денежного и товарного рынков</p> $R_t = a_1 + b_{12} \cdot Y_t + b_{14} \cdot M_t + e_1;$ $Y_t = a_2 + b_{21} \cdot R_t + b_{25} \cdot G_t + e_2;$ $I_t = a_3 + b_{31} \cdot R_t + e_3;$ <p>где R – расходы на потребление; Y – реальный ВВП; M – денежная масса; I – внутренние инвестиции; t – текущий период; G – реальные государственные расходы</p>	<p>Макроэкономическая модель</p> $C_t = a_1 + b_{11} \cdot Y_t + b_{13} \cdot T_t + e_1;$ $I_t = a_2 + b_{21} \cdot Y_t + b_{24} \cdot K_{t-1} + e_2;$ $Y_t = C_t + I_t$ <p>где C – потребление; Y – доход; I – инвестиции; T – налоги; K – запас капитала; t – текущий период</p>

Продолжение табл. 5.3

Вариант 3	Вариант 4
<p>Одна из версий модифицированной модели Кейнса имеет вид</p> $C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot Y_{t-1} + e_1;$ $I_t = a_2 + b_{21} \cdot Y_t + b_{22} \cdot Y_{t-1} + e_2;$ $Y_t = C_t + I_t + G_t,$ <p>где C – расходы на потребление; Y – доход; G – государственные расходы; I – инвестиции; t – текущий период; $t - 1$ – предыдущий период</p>	<p>Модель мультипликатора-акселератора</p> $C_t = a_1 + b_{11} \cdot R_t + b_{12} \cdot C_{t-1} + e_1;$ $I_t = a_2 + b_{21} \cdot (R_t - R_{t-1}) + e_2;$ $R_t = C_t + I_t,$ <p>где C – расходы на потребление; I – инвестиции; R – доход; t – текущий период; $t - 1$ – предыдущий период</p>
Вариант 5	Вариант 6
<p>Конъюнктурная модель имеет вид</p> $C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + e_1;$ $I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + e_2;$ $r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + b_{33} \cdot P_t + e_3;$ $Y_t = C_t + I_t + G_t,$ <p>где C – расходы на потребление; Y – ВВП; r – процентная ставка; G – государственные расходы; M – денежная масса; I – инвестиции; t – текущий период; $t - 1$ – предыдущий период</p>	<p>Модель спроса и предложения кейнсианского типа</p> $Q_t^s = a_1 + a_2 P_t + a_3 P_{t-1} + e_1;$ $Q_t^d = b_1 + b_2 P_t + b_3 Y_t + e_2;$ $Q_t^s = Q_t^d,$ <p>где Q_t^d – спрос на товар; Q_t^s – предложение товара; P – цена товара; Y – доход; t – текущий период; $t - 1$ – предыдущий период</p>
Вариант 7	Вариант 8
<p>Модель Кейнса</p> $C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot Y_{t-1} + e_1;$ $I_t = a_2 + b_{21} \cdot Y_t + e_2;$ $Y_t = C_t + I_t + G_t,$ <p>где C – потребление; Y – ВВП; G – государственные расходы; I – инвестиции; t – текущий период; $t - 1$ – предыдущий период</p>	<p>Модель денежного рынка</p> $R_t = a_1 + b_{11} \cdot M_t + b_{12} \cdot Y_t + e_1,$ $Y_t = a_2 + b_{21} \cdot R_t + b_{22} \cdot I_t + e_2,$ <p>где R – процентная ставка; Y – ВВП; M – денежная масса; I – внутренние инвестиции; t – текущий период</p>

Вариант 9	Вариант 10
<p>Модель мультипликатора-акселератора</p> $C_t = a_1 + b_{11} \cdot R_t + b_{12} \cdot C_{t-1} + e_1;$ $I_t = a_2 + b_{21} \cdot (R_t - R_{t-1}) + e_2;$ $R_t = C_t + I,$ <p>где C – расходы на потребление; I – инвестиции; R – доход; t – текущий период; $t - 1$ – предыдущий период</p>	<p>Модель спроса и предложения на деньги</p> $R_t = a_1 + b_{11} \cdot M_t + b_{12} \cdot Y_t + e_1;$ $Y_t = a_2 + b_{21} \cdot R_t + e_2,$ <p>где R – процентные ставки; Y – ВВП; M – денежная масса; t – текущий момент</p>

ТЕСТ 1

Укажите номер (-а) правильного (-ых) ответа (-ов).

1. Эконометрика – это наука, которая...

1) имеет дело с массовыми явлениями любой природы, повторяющимися в пространстве и времени;

2) определяет количественное выражение взаимосвязей экономических явлений и процессов;

3) формулирует статистические гипотезы и проверяет их с помощью критериев значимости;

4) устанавливает взаимосвязь экономических явлений и процессов.

2. Инструментом эконометрических исследований являются...

1) статистические методы;

2) задачи экономической теории;

3) эконометрические модели;

4) математические методы.

3. Выберите основные классы эконометрических моделей:

1) линейные модели;

2) регрессионные модели с одним уравнением;

3) модели временных данных;

4) нелинейные модели;

5) системы одновременных уравнений.

4. Укажите последовательность этапов эконометрического моделирования:

- 1) идентификация модели;
- 2) параметризация модели;
- 3) постановочный этап;
- 4) априорный этап;
- 5) информационный этап;
- 6) проверка адекватности модели.

5. Перечислите типы данных, используемые при моделировании экономических процессов:

- 1) статистические данные;
- 2) пространственные данные;
- 3) временные данные;
- 4) априорные данные.

6. Укажите типы переменных, участвующие в эконометрической модели:

- 1) постоянные;
- 2) эндогенные;
- 3) лаговые;
- 4) predetermined;
- 5) экзогенные;
- 6) мнимые.

7. Количество predetermined переменных в эконометрической модели...

- 1) равно числу уравнений;
- 2) равно количеству эндогенных переменных;
- 3) может быть любым;
- 4) равно числу случайных возмущений.

8. Охарактеризуйте линейную регрессионную модель вида $y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$:

- 1) линейная по переменным;
- 2) трехфакторная;
- 3) двухпараметрическая;
- 4) линейная по параметрам.

9. Укажите обязательное условие для нормальной линейной регрессионной модели с одним уравнением $y = a_0 + a_1x + \varepsilon$:

- 1) $M(\varepsilon_i \cdot \varepsilon_j) = 0, i \neq j$, где i, j – различные серии наблюдений;
- 2) $\varepsilon \in N(1, \sigma^2)$;
- 3) $\varepsilon \in N(0, \sigma^2)$.

10. Значение условия $M(\varepsilon_i \varepsilon_j) = 0, i \neq j$, где i, j – различные серии наблюдений случайной составляющей ε в линейной регрессионной модели:

1) некоррелированность значений случайной составляющей для различных наблюдений;

2) имеется автокорреляция значений случайной составляющей;

3) отсутствие регрессионной связи;

4) отсутствие корреляции между значениями изучаемого признака и факторов.

11. По условию теоремы Гаусса–Маркова, матрица значений факторных величин X обязательно должна иметь:

1) нулевой столбец;

2) количество строк больше количества столбцов;

3) столбец из единиц;

4) количество строк, равное количеству столбцов.

12. По какой формуле находят оценки параметров уравнения регрессии в методе наименьших квадратов?

1) $A = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$;

2) $A = (X^{-1} \cdot X)^T \cdot (X^T \cdot Y)$;

3) $A = (X^T \cdot X)^{-1} \cdot Y \cdot X^T$;

4) $A = (X \cdot X^T)^{-1} \cdot (X^T \cdot Y)$.

13. По формуле $S_e^2 = \frac{1}{n-2} \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2$ вычисляют...

1) смещенную оценку дисперсии случайной составляющей;

2) несмещенную оценку дисперсии случайной составляющей;

3) эффективную оценку дисперсии случайной составляющей;

4) составляющую оценку дисперсии случайной составляющей.

14. По формуле $R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$ вычисляют...

- 1) значение критерия Фишера;
- 2) статистику критерия Дарбина–Уотсона;
- 3) коэффициент детерминации;
- 4) коэффициент корреляции.

15. Для построения интервального прогноза значения эндогенной переменной Y нужно знать...

- 1) величину $F_{кр}$;
- 2) коэффициент детерминации;
- 3) точечный прогноз эндогенной переменной;
- 4) вид функции регрессии.

16. Для проверки адекватности линейной эконометрической модели обязательно нужно знать...

- 1) коэффициент корреляции;
- 2) величину $F_{кр}$;
- 3) коэффициент детерминации;
- 4) оценки параметров уравнения регрессии.

17. Если множественный коэффициент детерминации равен 0,728, то это значит, что...

- 1) 27,2 % вариации зависимой переменной обусловлено влиянием не учтенных в модели факторов;
- 2) между изучаемым признаком и включенными в модель факторами существует значительная линейная связь;
- 3) между результирующим признаком и факторами существует прямая связь;
- 4) между результирующим признаком и факторами существует обратная связь.

18. Временной ряд, имеющий постоянное среднее значение и колеблющиеся вокруг него значения с постоянной дисперсией, называется...

- 1) стационарным;
- 2) нестационарным;
- 3) динамичным;
- 4) постоянным.

19. Перечислите составляющие, из которых складывается уровень временного ряда:

- 1) не основная тенденция;
- 2) основная тенденция;

- 3) временная составляющая;
- 4) циклическая составляющая;
- 5) сезонная составляющая;
- 6) случайная составляющая;
- 7) неслучайная составляющая.

Среди утверждений из второго столбца, подберите соответствующее термину из первого.

20. Если ... , то

- | | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> а) уравнение идентифицировано; б) уравнение неидентифицировано; в) уравнение сверхидентифицировано. | <ul style="list-style-type: none"> 1) оценки структурных параметров нельзя найти по коэффициентам приведенной формы; 2) для некоторых структурных параметров можно получить более одного численного значения; 3) оценки структурных параметров можно однозначно найти по коэффициентам приведенной формы; 4) приведенная форма системы одновременных уравнений не существует. |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

ТЕСТ 2

Укажите номер (-а) правильного (-ых) ответа (-ов).

1. Проблема спецификации регрессионной модели включает в себя...

- а) отбор факторов, включаемых в уравнение регрессии;
- б) оценку параметров уравнения регрессии;
- с) оценку надежности результатов регрессионного анализа;
- д) выбор вида уравнения регрессии.

2. Требования к факторам, включаемым в модель линейной множественной регрессии:

- а) число факторов должно быть в 6 раз меньше объема совокупности;
- б) факторы должны представлять временные ряды;
- с) факторы должны иметь одинаковую размерность;
- д) между факторами не должно быть высокой корреляции.

3. Отбор факторов в эконометрическую модель линейного уравнения множественной регрессии можно проводить на основе...

- a) исключения одного из пары коллинеарных факторов из модели;
 - b) включения коллинеарных факторов в одно и то же уравнение;
 - c) отбора более высоких значений коэффициентов регрессии модели в естественном масштабе переменных;
 - d) сравнения величины остаточной дисперсии до и после.
4. Дана матрица парных коэффициентов корреляции (табл. 5.4).
Найти наибольшее значение межфакторной корреляции.

Таблица 5.4

Матрица парных коэффициентов корреляции

	y	x_1	x_2	x_3
y	1			
x_1	0,72	1		
x_2	0,48	-0,02	1	
x_3	0,13	0,69	0,51	1

5. Дана матрица парных коэффициентов корреляции (табл. 5.4).
Какой фактор наименее тесно связан с результативной переменной?

- a) y ;
- b) x_1 ;
- c) x_2 ;
- d) x_3 .

6. Дана матрица парных коэффициентов корреляции (табл. 5.4).
Какие факторы целесообразно включать в модель, чтобы обеспечить отсутствие коллинеарности факторов?

- a) x_1, x_2, x_3 ;
- b) x_1, x_2 ;
- c) x_2, x_3 ;
- d) x_1, x_3 .

7. Для уравнения регрессии отклонение фактического значения результативной переменной от расчетного для точки с координатами (2;50) равно...

- a) 4;
- b) 44;
- c) 58;
- d) 6.

8. Изучается зависимость себестоимости единицы изделия (y , тыс. руб.) от величины выпуска продукции (x , тыс. шт.) по группам предприятий за отчетный период. Экономист исследовал $n = 5$ предприятий и получил результаты, представленные в табл. 5.5.

Таблица 5.5

Результаты исследования

n	x	y	x^2	xy
1	2	1,9	4	3,8
2	3	1,7	9	5,1
3	4	1,8	16	7,2
4	5	1,6	25	8
5	6	1,4	36	8,4
Сумма	20	8,4	90	32,5

Найти параметр a в уравнении линейной парной регрессии $y = a - 0,11x + \varepsilon$.

9. Зависимость объема продаж y (д.е.) от расходов на рекламу x (д.е.) характеризуется по 12 предприятиям следующим образом: $y = 10,6 + 0,6x + \varepsilon$, $r_{yx} = 0,83$. Интерпретация коэффициента регрессии будет следующей:

а) 60 % вариации объема продаж объясняется вариацией расходов на рекламу;

б) 83 % вариации объема продаж объясняется вариацией расходов на рекламу;

с) при увеличении расходов на рекламу на 1 д.е. объем продаж увеличивается на 0,83 д.е.;

д) при увеличении расходов на рекламу на 1 д.е. объем продаж увеличивается в среднем на 0,6 д.е.;

е) при увеличении объемов продаж на 1 д.е. расходы на рекламу увеличиваются в среднем на 10,6 д.е.

10. Верные утверждения относительно мультиколлинеарности факторов:

а) в модель линейной множественной регрессии рекомендуется включать мультиколлинеарные факторы;

б) мультиколлинеарность факторов приводит к снижению надежности оценок параметров уравнения регрессии;

с) мультиколлинеарность факторов проявляется в наличии парных коэффициентов межфакторной корреляции со значениями, большими 0,7;

д) мультиколлинеарность факторов проявляется в наличии парных коэффициентов межфакторной корреляции со значениями, меньшими 0,3.

11. При построении модели множественной регрессии методом пошагового включения переменных на первом этапе рассматривается модель с...

а) одной объясняющей переменной, которая имеет с зависимой переменной наименьший коэффициент корреляции;

б) одной объясняющей переменной, которая имеет с зависимой переменной наибольший коэффициент корреляции;

с) несколькими объясняющими переменными, которые имеют с зависимой переменной коэффициенты корреляции по модулю больше 0,5;

д) полным перечнем объясняющих переменных.

12. Даны значения количественного показателя: 5; -6; 7; 1; 2. Среднее квадратическое отклонение равно 4,45. Первое стандартизированное значение показателя равно...

а) 0,81;

б) 1,12;

с) 1,4;

д) 3,6.

13. К свойствам уравнения регрессии в стандартизированном виде относятся следующие утверждения:

а) коэффициенты регрессии при объясняющих переменных равны между собой;

б) постоянный параметр (свободный член уравнения) регрессии отсутствует;

с) стандартизированные коэффициенты регрессии несравнимы между собой;

д) входящие в состав уравнения переменные являются безразмерными.

14. Коэффициент множественной детерминации характеризует...

а) тесноту совместного влияния факторов на результат в уравнении линейной множественной регрессии;

б) тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в модель;

с) долю дисперсии результативного признака, объясненную регрессией в его общей дисперсии;

д) среднее изменение результативной переменной с изменением соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне.

15. Для устранения систематической ошибки остаточной дисперсии для оценки качества модели линейной множественной регрессии используется...

а) коэффициент множественной детерминации;

б) коэффициент множественной корреляции;

с) скорректированный коэффициент множественной детерминации;

д) скорректированный коэффициент частной корреляции.

16. Оценка статистической значимости уравнения линейной множественной регрессии в целом осуществляется с помощью...

а) критерия Стьюдента;

б) критерия Фишера;

с) критерия Дарбина–Уотсона;

д) критерия Фостера–Стюарта.

17. Если коэффициент регрессии является существенным, то для него выполняются следующие условия:

а) фактическое значение t -критерия Стьюдента меньше критического;

б) фактическое значение t -критерия Стьюдента больше критического;

с) доверительный интервал проходит через ноль;

д) стандартная ошибка не превышает половины значения параметра.

18. Если коэффициент регрессии является существенным, то фактическое значение t -критерия...

а) больше критического;

б) меньше критического;

с) близко к единице;

д) близко к нулю.

19. Автокорреляционная функция – ...

a) зависимость коэффициента автокорреляции от первых разностей уровней временного ряда;

b) зависимость уровня временного ряда от коэффициента корреляции с его номером;

c) последовательность коэффициентов автокорреляции, расположенных по возрастанию их порядка;

d) последовательность коэффициентов автокорреляции, расположенных по возрастанию их значений.

ЛИТЕРАТУРА

1. Ливинская, В. А. Эконометрика : методические рекомендации к лабораторным работам для студентов направления 27.03.05 «Инноватика» / В. А. Ливинская. – Могилев : Белорусско-Российский университет, 2018. – 31 с.

2. Ленькова, Р. К. Эконометрика и экономико-математические методы и модели: курс лекций : учебно-методическое пособие «Экономика и организация производства в отраслях агропромышленного комплекса» : в 3 ч. / Р. К. Ленькова. – Горки : БГСХА, 2017. – Ч. 1 : Эконометрика. – 60 с.

3. Заяц, Т. А. Эконометрика и экономико-математические методы и модели: пособие для реализации содержания образовательных программ высшего образования 1-й ступени и переподготовки руководящих работников и специалистов : в 3 ч. / Т. А. Заяц, О. И. Еськова, М. А. Грибовская. – Гомель : Белорусский торгово-экономический университет потребительской кооперации, 2021. – Ч. 1. – 95 с.

4. Ленькова, Р. К. Эконометрика и экономико-математические методы и модели : учебно-методическое пособие для студентов специальности 1-74 01 01 «Экономика и организация производства в отраслях агропромышленного комплекса» : в 2 ч. / Р. К. Ленькова, Е. В. Карачевская. – Горки : БГСХА, 2020. – Ч. 1. – 60 с.

5. Голда, О. А. Эконометрика и экономико-математические методы и модели : учебно-методическое пособие для студентов специальности 1-28 01 01 «Экономика электронного бизнеса» : в 2 ч. / О. А. Голда, Н. М. Матвейчук. – Минск : БГУИР, 2019. – Ч. 1 : Эконометрика. – 153 с.

6. Кулаков, В. Н. Эконометрика и экономико-математические методы и модели. Компьютерное моделирование в среде Excel : методические указания к лабораторной работе для студентов экономических специальностей / В. Н. Кулаков, М. Н. Марченко, Д. В. Редько. – Горки : БГСХА, 2017. – 20 с.

ПРИЛОЖЕНИЕ А

Методические указания по выполнению индивидуальных заданий

Вариант индивидуального задания выбирает студент в соответствии с порядковым номером его зачетной книжки.

Индивидуальное задание выполняется студентами самостоятельно, после изучения теоретического материала по соответствующей теме. Работа над индивидуальным заданием оформляется студентами самостоятельно в виде отчета, содержащего выполненные расчеты с необходимыми пояснениями в системе Mathcad (прил. В). В начале отчета дается заглавный (титульный) лист, после которого следует лист с индивидуальным заданием согласно варианту. В процессе выполнения студенту необходимо дать подробное описание порядка выполнения им индивидуального задания. Все расчеты и пояснения по отдельным вопросам даются в соответствии с методическими указаниями по выполнению индивидуального задания.

В тексте не допускается:

- применять произвольные словообразования;
- применять сокращения слов, кроме установленных правилами русской орфографии, стандартами или в данном документе;
- сокращать обозначения единиц физических величин, если они употребляются без цифр, за исключением единиц физических величин в таблицах и расшифровках буквенных обозначений, входящих в формулы и рисунки.

Числовые значения величин одного наименования следует указывать с одинаковым округлением. Рисунки, схемы, графики следует располагать в тексте на странице, где они упоминаются, или на следующей странице. Таблицы нумеруются сквозной нумерацией или по разделам. По тексту необходимо указать ссылки на каждую таблицу. При ссылке на таблицу необходимо писать слово «Таблица» с указанием ее номера. Таблицы должны быть расположены ниже текста, где они упоминаются, либо на следующей странице. Уравнения и формулы выделяются из текста свободными строками. Выше и ниже формулы должно быть оставлено не менее одной строки. Пояснения символов и числовых коэффициентов формул должны быть приведены непосредственно под формулой, если они не пояснялись ранее в тексте.

ПРИЛОЖЕНИЕ Б

Таблица Б.1

Критические значения t -критерия Стьюдента при уровне значимости $\alpha = 0,10; 0,05; 0,01$ (двухсторонний)

Число степеней свободы	Уровень значимости α			Число степеней свободы	Уровень значимости 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				

**Таблица значений F -критерия Фишера
при уровне значимости $\alpha = 0,05$**

$k_1 \backslash k_2$	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	234,52
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

Окончание табл. Б.2

$k_2 \backslash k_1$	1	2	3	4	5	6	8	12	24	∞
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	1,65	1,31
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

ОСНОВНЫЕ ЭЛЕМЕНТЫ ИНТЕРФЕЙСА MATHCAD

Возможность использования формул, функций, графиков и диаграмм является одним из важнейших свойств программы системы Mathcad, позволяющей проводить статистический анализ числовых значений.

При запуске Mathcad отображается пустой документ, в который можно ввести текст, уравнения, графики и изображения.

Главное меню содержит основные вкладки, в которых главными являются *Вид / Панели инструментов / Математика* и *Символьные операции*.

Документ Mathcad представляет собой рабочий лист с размещенными блоками: математические выражения, фрагменты текста и графические области. Текстовые блоки размещаются в любом месте документа произвольно, используются для комментариев, а математические выражение и графические области располагают слева направо и снизу вверх.

Математика в документе Mathcad отображается в привычной математической нотации. При вводе буквы или цифры отображаются в окружении строк редактирования формул, чтобы показать, какие символы выбраны. Операторы вставляются при нажатии соответствующей кнопки на панелях инструментов оператора или клавиши.

Редактор уравнений Mathcad предназначен для работы в структуре математического выражения, поэтому выражения не столько типизируются, сколько создаются. Mathcad собирает детали, составляющие выражение, используя правила приоритета плюс некоторые дополнительные правила для упрощения ввода знаменателей, степеней и выражений в радикалах.

Панель инструментов *Форматирование* содержит списки прокрутки и кнопки для указания характеристик шрифта в уравнениях и тексте. Если курсор находится в математической области, в этом окне указывается математический стиль выделенных символов или области. Чтобы узнать, что делает кнопка, курсор необходимо навести на кнопку, пока в строке состояния не появится подсказка и описание. Для создания текстовой области и ввода текста после

ввода пробела Mathcad начинает текстовую область. Появится текстовое поле с красным текстовым курсором.

Рабочий лист, на котором располагаются блоки, – это документ программы MathCad. Блоки делятся на:

1. Математические выражения и графические области (располагаются слева направо, сверху вниз).

2. Фрагменты текста (размещаются произвольно, предназначены для комментариев).

Место ввода каждого блока указывается курсором, имеющим вне блоков вид красного крестика. Блоки можно копировать в буфер обмена, перемещать по рабочему листу, выделять.

Если текст введен без кавычек, то система воспринимает его как математическое выражение.

Ввод текста начинают с кавычек, курсор принимает вид вертикальной красной линии.

Математические выражения содержат переменные, константы, операторы, функции, управляющие структуры. Переменные разделяют на локальные, глобальные, системные (встроенные).

Переменным до их использования следует присвоить значения. Действие локальных переменных осуществляют оператором присваивания «:=». Значения глобальных переменных известны в любом месте документа, в том числе и до точки их объявления. Для глобального присваивания служит знак «≡», расположенный на панели инструментов *Вычисление*. Системные, или встроенные, переменные определены в самой системе с точностью *TOL*. Например, $\pi = 3,142$, $g = 9,807 \text{ м/с}^2$.

Для вычисления и вывода значения выражения предназначен знак «=».

Для ввода математических выражений применяются шаблоны и специализированные панели инструментов *Вид / Панели инструментов / Математика*, содержащие часто используемые функции, конструкции и отсутствующие на клавиатуре символы операций. Чтобы ввести математические выражения, необходимо просто расположить строки редактирования математики в документе и начать вводить числа или операторы, такие как «+» или «-».

Для получения результата функции в числовом виде переменную функции определяют как результат численного расчета в одной строке и применяют аналитическое преобразование.

Команда *Формат / Результат* применима для форматирования численных результатов для дробных или смешанных чисел, а для форматирования отдельного результата, не изменяя остальные, следует дважды щелкнуть нужный результат.

Чтобы расширить область в выражении, следует нажать [*Пробел*].

Для выделения пунктирного прямоугольника выделения, нужно щелкнуть эту область, удерживая нажатой клавишу [*Ctrl*].

Формулами в системе Mathcad являются математические выражения и функции, ввод которых осуществляют с помощью панелей инструментов (рис. В.1).



Рис. В.1. Панели инструментов программы Mathcad

В математике *функцией* называют соответствие (закон), по которому каждому значению x ставится в соответствие определенное значение y . Основным способом задания функции является формула.

Встроенные функции, имеющиеся в Mathcad, вызывают командой *Вставка / Функция* или кнопкой $f(x)$ на панели инструментов

Стандартная. Имеется возможность создания *собственных функций (функции пользователя)* следующим образом:

имя(аргументы) := выражение .

Слева (после имени) в скобках перечисляются аргументы, а справа (после знака присваивания) записывается выражение, которое может содержать переменные, константы, формулы и функции.

В то время как обычные переменные содержат одно значение, массивы (общий термин для матриц и векторов) содержат много значений. Они представлены в виде таблиц чисел с заданным количеством строк и столбцов. Можно создавать и использовать массивы в расчетах так же, как числовые или скалярные переменные. Многие операторы и функции общего назначения могут использоваться с массивами, но также существуют некоторые специальные операторы и функции для манипулирования массивами.

Символьные вычисления

Система Mathcad позволяет производить не только разнообразные вычисления, но и выполнять символьные операции, результатами которых являются не числа, а формулы.

Некоторые символы, которые можно использовать для имен переменных и функций, трудно ввести в Mathcad. Панель инструментов *Вид / Панели инструментов / Дополнительные символы* позволяет вводить их в математические области, нажимая кнопки. В настоящее время доступны следующие символы: F, C, /F, /C, ±, ≈.

Функции температуры и функции обратной температуры определены в Mathcad для назначения и преобразования значений с единицами измерения температуры. Другие символы не определены, они доступны для назначения собственных функций и имен переменных.

Символьные операции

Для символьного преобразования и вывода выражений предназначен знак «→» (стрелка), который вводится с панелей инструментов или комбинацией клавиш [Ctrl]+[.]

Строковая переменная устанавливается в строке с помощью клавиши [стрелка вправо] вертикальным курсором выделения справа от последней пары кавычек без пробелов.

При вычислении выражения с символическим знаком равенства Mathcad упрощает результат, выполняя арифметику и комбинируя сходные переменные. Например, $2x + 5 + 4x - 3 \rightarrow 6x + 2$.

Для выполнения более сложных символьных операций можно вставить ключевое слово, указывающее операцию перед символьным знаком равенства.

Вывод символьных данных

Символьные операции полезны, хотя не всегда вид полученных формул является ожидаемым (вместо упрощения может получиться усложнение). Поэтому следует указывать тип преобразования, выбирая подходящие варианты на панели *Символические операции*.

Преобразование *Simplify* (*Упростить*) предназначено для приведения к общему знаменателю, подобных слагаемых, применению тригонометрических тождеств.

Преобразование *Expend* (*Расширить*) применяется для разложения по степеням, раскрытия скобок.

Преобразование *Collect* (*Собрать*) позволяет группировать подобные.

Преобразование *Solve* (*Решить*) выражает указанную переменную через другую в виде формулы, т. е. аналитически решает уравнения или неравенства.

Операция *Solve* может быть полезна для получения и анализа формул при решении уравнений и неравенств и для оценки начальных приближений.

При использовании преобразований *Expend*, *Collect*, *Solve* необходимо указывать переменную, относительно которой производится операция.

Команда *Символьные операции* предназначена для символьных вычислений.

В то время как обычные переменные содержат одно значение, массивы (общий термин для матриц и векторов) содержат много значений. Они представлены в виде таблиц чисел с заданным количеством строк и столбцов. Можно создавать и использовать массивы в расчетах так же, как числовые или скалярные переменные.

Многие операторы и функции общего назначения могут использоваться с массивами, но также существуют некоторые специальные операторы и функции для манипулирования массивами.

Для доступа к определенному элементу массива используйте оператор подстрочного индекса. Для доступа к целому столбцу используйте оператор столбца.

Любую матрицу можно вывести как изображение в оттенках серого командой *Вставка / Изображение*. Имеется возможность масштабирования отображаемых значений пикселей соответствующих уровням серого от 0 до 255.

Если элемент матрицы с наибольшим индексом стоит первым. Быстрота вычислений с матрицами увеличивается.

Команда *Вставка / Элемент управления* служит для ставки кнопок, ползунков, текстовых полей и других элементов управления в Mathcad.

1.2. Решение уравнений и систем в Mathcad

Mathcad содержит мощные средства решения систем уравнений, собранные в категории *Solving (Решить)* и вызываются нажатием кнопки $f(x)$ на панели инструментов *Стандартная*. Рассмотрим некоторые способы решения уравнений в Mathcad.

Решение линейных уравнений

Mathcad поддерживает множество функций для решения одного уравнения в одном неизвестном с помощью больших систем линейных, нелинейных и дифференциальных уравнений с несколькими неизвестными. Существует также множество процедур подгонки и интерполяции для генерации функций и аппроксимаций к данным.

Блок решения относится к группе команд Mathcad, используемых для решения *системы* линейных, дифференциальных уравнений или дифференциальных уравнений в частных производных, задачи оптимизации или задачи линейного программирования.

Блоки решения позволяют настроить эти задачи в натуральной математической нотации, поэтому ограничения, функции и начальные значения проще идентифицировать.

Каждый блок решения включает ключевое слово *Given*, набор ограничений, набор уравнений и функцию решения.

Структура блока решения:

1. Ключевое слово *Given*, вводимое как математическая область.

2. Набор уравнений и / или неравенств, представляющих ограничения. Для создания этих ограничений на логической панели инструментов предназначены следующие операторы: логические значения равны; больше, чем; больше или равно; менее, чем; меньше или равно. Равенство обозначается логическими равными, [Ctrl] + [=].

Функции решают линейные и нелинейные уравнения и системы уравнений. Функции *Найти (Find)*, *Максимизировать (Maximize)*, *Минимизировать (Minimize)* и *Минер (Minerr)* используют, главным образом, внутри блоков решения, что позволяет получать естественное представление. Остальные функции используют в программах для итеративных вычислений или самостоятельно в одном выражении.

В Mathcad численное значение корня быстро и точно определяется с помощью функции *Root()*. Имеются некоторые задачи, в которых возможности Mathcad позволяют находить решения в символьном (аналитическом) виде.

Множество значений приближений можно задавать с помощью ранжированной переменной. При удачном выборе ее шага функция *Root(f(x),x)* возвращает сразу все корни, причем некоторые из них могут повторяться.

Точность вычислений в MathCad задается системной переменной *TOL*. По умолчанию $TOL = 0,001$. При необходимости требуемое значение может быть присвоено непосредственно на рабочем листе. Формат отображения результата должен соответствовать устанавливаемой точности (по умолчанию: общий, 3 десятичных знака). Настройки можно поменять в окне вызываемого меню *Формат / Результат*.

Чтобы решить уравнение символически, нужно ввести уравнение, используя логический знак равенства [Ctrl] + [=], или нажать [Ctrl] + [Shift] + [.] и ввести ключевое слово *solve*, или щелкнуть решение на символьной панели инструментов, затем нажать [Enter].

Если функция содержит более одной переменной, введите запятую после *Solve*, а затем ввести разделенный запятыми список переменных, для которых требуется найти решение. По возможности Mathcad возвращает символьные решения уравнения, иначе возвращается числовое решение.

Если ранее переменной, для которой выполняется решение, было назначено числовое значение, то Mathcad возвращает ошибку. Чтобы решить эту проблему, нужно или очистить значение переменной, введя $x := x$, или применить ключевое слово *explicit*.

Если уравнение имеет несколько решений, Mathcad возвращает решения в векторе, если решение не является периодическим.

Чтобы решить уравнение, правая сторона которого равна 0, необходимо ввести только левую сторону уравнения. Например, можно решить $x - 9 = 0$ следующим образом: $x - 9 \text{ solve} \rightarrow 9$.

Если уравнение содержит более одной переменной, необходимо указать переменную, которую требуется решить после *solve*.

Если уравнение включает числа с десятичными точками, Mathcad возвращает десятичный ответ. Например, $5x - 4,5 = 0 \text{ solve} \rightarrow 0,9$.

Для решения уравнения, подверженного ограничениям в области переменной (например, операции над вещественными числами) используют ключевое слово *предположить* с *решить*. Если уравнение имеет периодическое решение, то *solve* возвращает единственное значение из набора решений. Например, $\sin(x) = 0 \text{ solve} \rightarrow 0$.

Чтобы увидеть более подробное решение, добавляют модификатор *fully* после *solve*.

Mathcad возвращает решение в виде новой генерируемой переменной, представляющей произвольное целое число. Созданной переменной предшествует подчеркивание, чтобы избежать конфликтов с другими переменными, которые вы могли определить в другом месте листа.

Использование векторных и матричных операций упрощает решение многих практических задач, данные в которых можно интерпретировать как векторы или матрицы, а также лежит в основе многих методов решения уравнений и систем.

Работа с векторами и матрицами обеспечивается набором встроенных функций на панели инструментов *Матрица*.

Матрицей называют поименованную область элементов, которые размещаются в таблице и упорядочены по индексам (номерам строк и столбцов). Индексы могут иметь только целочисленные значения и начинаться с нуля или единицы (как строк, так и столбцов). Это значение следует установить в начале работы на вкладке *Сервис / Параметры* или присвоить *ORIGIN := 1* непосредственно на рабочем листе.

Обратной называют матрицу, умножение которой на исходную дает единичную матрицу (диагональные элементы равны 1, остальные – 0).

Векторы и матрицы задаются или заполнением шаблона вручную, или с использованием в качестве индексов ранжированной переменной, если возможен способ вычисления значений элементов через их индексы.

Системы линейных и нелинейных уравнений

Простейшим методом решения систем линейных алгебраических уравнений является метод обратной матрицы, в котором вектор V решения системы равен произведению обратной матрицы коэффициентов при неизвестных M^{-1} и вектора – столбца правых частей V :

$$X = M^{-1} \cdot V.$$

Данный метод реализуется встроенной функцией $Isolve(M,V)$.

Пример. Решить систему уравнений

$$\begin{cases} 12x - y + z = 6; \\ z + x - 2y = 2; \\ 5x + 3y - 2z = 4. \end{cases}$$

Решение:

1. Ввести матрицу коэффициентов при неизвестных M и вектор-столбец V правых частей уравнений:

$$M := \begin{pmatrix} 12 & -5 & 1 \\ 1 & -2 & -7 \\ 5 & 3 & -2 \end{pmatrix}; \quad V := \begin{pmatrix} 6 \\ 2 \\ 4 \end{pmatrix}.$$

2. Вычислить определитель системы $|M| = 478$, проверив условие о единственности решения.

3. Записать произведение $M^{-1} \cdot V$ и вывести результат:

$$M^{-1}V = \begin{pmatrix} 0,594 \\ 0,176 \\ -0,251 \end{pmatrix}.$$

Решение с функцией $I\text{solve}(M,V)$ отличается последним этапом, на котором она записывается. Матрица коэффициентов и вектор-столбец являются ее аргументами.

I\text{solve} решает линейную систему уравнений, заданных в матричной записи.

Найти решает линейные или нелинейные системы уравнений.

Максимизация (минимизация) максимизирует (минимизирует) значение функции, включая линейное программирование.

Minerr минимизирует ошибку наименьших квадратов при решении системы уравнений, решая систему приблизительно.

Используют следующие функции для нахождения решения:

– *Найти (Find)* или *Минер (Minerr)* для решения системы уравнений;

– *Максимизировать (Maximize)* или *Минимизировать (Minimize)* для оптимизации функции;

– *Odesolve* для решения систем обыкновенных дифференциальных уравнений;

– *Pdesolve* для решения систем дифференциальных уравнений в частных производных.

Методы и алгоритмы решения

Блоки решения могут быть решены численно или, в случае поиска, символически. Символьная оценка недоступна для дифференциальных уравнений или задач \max / \min . Функции решения и оптимизации системы могут использовать множество алгоритмов, в зависимости от класса задачи в блоке решения. Существует несколько обычных дифференциальных уравнений и алгоритмов дифференциальных уравнений в частных производных.

Блоки решения сначала оценивают свои ограничения, используя начальные приближения для проверки ошибок и определения области решения, что может диктовать, возвращаются ли реальные или

сложные решения. Если в проблеме рассматривается, как реальные и сложные результаты встречаются во время решения, то принимаются абсолютные значения. Если тип является сложным, то комплексные результаты выполняются как есть.

Во всех методах численного решения нелинейные задачи очень чувствительны к начальным приближениям, к которым применяются примечания по выбору и изменению начальных приближений для корневой функции. Для решения систем уравнений служит функция *Haytm* (*var1*, *var2*, ...). Данная функция возвращает значения *var1*, *var2*, ..., которые удовлетворяют уравнениям и неравенствам в блоке решения. Если решается для *n* переменных, блок решения должен иметь *n* уравнений. Допускается матричная нотация, как и решение для матричных переменных. Функция *Haytm* выбирает подходящий метод из группы доступных методов, в зависимости от того, является ли задача линейной или нелинейной, и других атрибутов.

Аргументы: *var1*, *var2*, ... – скалярные или матричные переменные, найденные в системе уравнений. Начальные приближения для каждой переменной должны быть определены выше ключевого слова *Given* или в теле блока решения. Если ожидается, что решения будут сложными, необходимо использовать комплексные начальные приближения.

Уравнения, которые необходимо решить, должны быть определены с помощью булевых функций. Значения в уравнениях и начальные приближения могут быть определены в теле блока с использованием знака « \Rightarrow ». Вывод функции *Haytm* может быть назначен одной переменной, вектору явных имен переменных или функции других имен переменных в блоке решения (включая переменные приближения), параметризируя блок решения. Если переменные имеют различные единицы измерения, они могут быть назначены только явному вектору имен или явному вектору имен функций, чтобы избежать смешанных единиц измерения в матрицах.

Если существует одна неизвестная скалярная переменная, решение является скалярным. В противном случае решением является вектор, первый элемент которого *var1*, второй – *var2* и т. д. Невозможно выполнить решение для одного элемента вектора, используемого в блоке. Все векторные значения корректируются одновременно для минимизации ошибки.

TOL и $CTOL$ могут влиять на решение нелинейных систем. Установка этих значений слишком маленькими может привести к тому, что решатель не сойдется. Если настройка этих параметров не помогает, попробуйте использовать другие начальные приближения или добавьте ограничение неравенства. Если система не сходится и необходимо знать найденное решение, даже если оно не соответствует критериям сходимости, то используют конструкцию *Minerr*.

Для решения систем уравнений и неравенств с n неизвестными применяется конструкция *Given – Find* (*Дано – Найти*).

Наибольшая трудность при решении систем нелинейных уравнений состоит в определении начальных приближений, которые должны задаваться до блока решений, начинающегося ключевым словом *Given* (*Дано*). Внутри блока помещаются уравнения и неравенства. Завершает блок решения функция от n аргументов $find(x1, x2, \dots, xn)$. Они перечисляются в том порядке, в котором будут выведены в вектор-столбец решения системы.

Учебное издание

РУГАЛЁВА Ирина Евгеньевна

ЭКОНОМЕТРИКА

Пособие

для студентов специальности
1-26 02 03 «Маркетинг»

Редактор *Е. И. Бенищевич*

Компьютерная верстка *Н. А. Школьниковой*

Подписано в печать 30.06.2023. Формат 60×84 ¹/₁₆. Бумага офсетная. Ризография.

Усл. печ. л. 6,05. Уч.-изд. л. 4,73. Тираж 100. Заказ 387.

Издатель и полиграфическое исполнение: Белорусский национальный технический университет.

Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий № 1/173 от 12.02.2014. Пр. Независимости, 65. 220013, г. Минск.