

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА  
И ПРОДОВОЛЬСТВИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования  
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ  
АГРАРНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

## **КОЛИЧЕСТВЕННЫЕ МЕТОДЫ АНАЛИЗА В АПК**

*Допущено Министерством образования Республики Беларусь  
в качестве учебного пособия для студентов второй ступени  
(магистратура) учреждений высшего образования  
по специальности «Экономика (магистратура)»*

Минск  
БГАТУ  
2020

УДК 330.43(075)  
ББК 65я7  
К60

Авторы:

доктор физико-математических наук, профессор *И. В. Белько*,  
кандидат педагогических наук, доцент *О. Л. Сапун*,  
старший преподаватель *Е. И. Подашевская*,  
старший преподаватель *Е. А. Криштанович*

Рецензенты:

кафедра общей математики и информатики БГУ  
(доктор экономических наук, профессор,  
заведующий кафедрой *С. А. Самаль*);  
доктор физико-математических наук, профессор,  
заведующий отделом НИЭИ  
Министерства экономики Республики Беларусь *М. К. Кравцов*

**Белько, И. В.**  
К60      Количественные методы анализа в АПК : учебное пособие /  
И. В. Белько [и др.]. – Минск : БГАТУ, 2020. – 232 с.  
ISBN 978-985-25-0027-2.

Излагаются количественные методы анализа эконометрических моделей и условия их применения для решения прикладных задач АПК; рассматривается теория вероятностей и математической статистики в эконометрическом анализе; исследуются парный регрессионный анализ как инструмент исследования экономических зависимостей, построение и использование модели множественной линейной регрессии при анализе и прогнозе в АПК. Приводятся решения конкретных экономико-математических задач с использованием программы MS Excel. Даются оценки параметров моделей, проверяется их значимость.

Для студентов второй ступени (магистратура) учреждений высшего образования, аспирантов и преподавателей экономических специальностей.

УДК 330.43(075)  
ББК 65я7

ISBN 978-985-25-0027-2

© БГАТУ, 2020

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	5
Тема 1. Применение количественного анализа для решения прикладных задач.....	6
Тема 2. Использование теории вероятностей и математической статистики в количественном анализе.....	18
Практическое занятие № 1. Основные характеристики случайных величин и их оценки .....	41
Тема 3. Парный регрессионный анализ как инструмент исследования экономических зависимостей .....	49
Практическое занятие № 2. Оценка тесноты и значимости стохастической зависимости между экономическими показателями.....	71
Практическое занятие № 3. Использование регрессионных моделей в анализе и прогнозе .....	73
Тема 4. Построение и использование модели множественной линейной регрессии при анализе и прогнозе .....	78
Практическое занятие № 4. Матричные вычисления оценок модели множественной линейной регрессии .....	89
Практическое занятие № 5. Компьютерное построение модели множественной линейной регрессии и прогнозирование .....	93
Тема 5. Устранение нарушений классических модельных предположений в процессе эконометрического анализа.....	98
Практическое занятие № 6. Выявление мультиколлинеарности и методика ее устранения .....	118
Практическое занятие № 7. Тестирование модели на наличие гетероскедастичности и автокорреляции остатков .....	121
Тема 6. Нелинейные модели регрессии. Использование фиктивных переменных в количественном анализе. ....	125
Практическое занятие № 8. Оценивание параметров LOGIT- и PROBIT-моделей методом максимального правдоподобия .....	139

Тема 7. Модели анализа и прогнозирования экономических процессов на основе временных рядов .....	142
Практическое занятие № 9. Моделирование компонент временных рядов для построения трендовых и сезонных моделей .....	167
Практическое занятие № 10. Анализ развития отраслей материального производства с использованием МОБ.....	174
Тема 8. Системы одновременных уравнений .....	179
Практическое занятие № 11. Применение счетного правила, условия ранга и оценка структурных коэффициентов .....	198
Тема 9. Прикладные модели в экономике.....	210
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....	226
ПРИЛОЖЕНИЕ 1 .....	227
ПРИЛОЖЕНИЕ 2 .....	228
ПРИЛОЖЕНИЕ 3 .....	229
ПРИЛОЖЕНИЕ 4 .....	231

## ВВЕДЕНИЕ

Количественные методы анализа в АПК – учебная дисциплина, включает те же разделы эконометрики, в которых изучаются и оцениваются связи между показателями производственной деятельности в АПК, указываются практические возможности их использования. Основу исследования в количественном анализе составляют положения экономической теории, которые используют количественные статистические методы.

Изучение предмета предполагает приобретение студентами теоретических знаний и практических навыков, необходимых для решения таких задач, как:

- построение модели (спецификация, оценивание параметров, экономическая интерпретация);
- проверка статистической адекватности построенной модели на основе статистических тестов и тестовых статистик;
- оценка прогностической способности модели;
- экономическое обоснование полученных на основе моделей результатов анализа и прогнозирования.

Особенности преподавания курса «Количественные методы анализа в АПК» определяются уровнем базовой математической подготовки студентов, степенью адаптации лекционного курса и практикума к наиболее актуальным для данной специальности задачам исследования. Использование авторами пакета Excel мотивируется следующими причинами: этот пакет обладает достаточно широкими возможностями решения задач, при этом он способствует более глубокому усвоению теории, так как требует введения самостоятельных корректировок и этапов построения моделей, основанных на положениях теории.

Сложившимся стандартом в организации учебно-методического обеспечения курса является наличие двух составляющих: базового учебного пособия, в котором излагаются теоретические разделы курса и практикума, включающего задачи по основным разделам учебного пособия, методические указания по решению типовых задач на компьютере, а также примеры решения типовых задач.

Особенностями учебного пособия являются:

- соединение теоретических разделов и практикума в виде типовых примеров и задач эконометрического анализа и моделирования;
- использование для иллюстрации решения задач эконометрического анализа и моделирования на основе статистических данных приложения Microsoft Office Excel.

## **Тема 1. ПРИМЕНЕНИЕ КОЛИЧЕСТВЕННОГО АНАЛИЗА ДЛЯ РЕШЕНИЯ ПРИКЛАДНЫХ ЗАДАЧ**

### *Вопросы*

- 1.1. Количественные методы анализа экономики АПК как наука и учебная дисциплина.
- 1.2. Типы данных в количественном анализе.
- 1.3. Системы эконометрических уравнений.

Даются определение предмета «Количественные методы анализа экономики в АПК» как науки и учебной дисциплины, основные принципы построения эконометрических моделей. Приводятся классические примеры систем эконометрических уравнений и их особенности. Анализируются основные типы данных.

### **1.1. Количественные методы анализа экономики АПК как наука и учебная дисциплина**

Количественные методы анализа экономики АПК (КМА) зародились и получили развитие на основе слияния экономической теории, математической экономики, экономической и математической статистики.

Основной задачей курса является обучение будущих экономистов методам исследования экономических процессов, основанных на использовании реальных статистических данных и положениях экономической теории.

Наиболее часто используют определение эконометрики, которое предложил известный российский ученый С. А. Айвазян: «Эконометрика – самостоятельная научная дисциплина, объединяющая совокупность теоретических результатов, приемов, методов и моделей, предназначенная для того, чтобы на базе экономической теории, экономической статистики, математико-статистического инструментария придавать конкретное количественное выражение общим закономерностям, обусловленным экономической теорией взаимосвязей экономических явлений и процессов».

Эконометрическая модель – система математических соотношений между входными и выходными переменными изучаемого

экономического явления, предназначенная для объяснения значений текущих переменных в зависимости от значений заранее определенных других переменных и позволяющая составить описание рассматриваемого процесса.

При исследовании причинно-следственных взаимосвязей приходится рассматривать не функциональные, а *стохастические зависимости*. Пусть изучается зависимость объема продаж некоторого товара от показателя – расходов на рекламу. Из экономической теории известно, что помимо последнего на объемы продаж влияют и другие факторы: денежные доходы покупателей, уровень инфляции, динамика экономической конъюнктуры и т. д. В данном контексте невозможно функционально описать влияние на объем продаж расходов на рекламу. В отличие от функциональной зависимости, где каждому набору значений факторов соответствует единственное значение результативного показателя, в данном случае значению фактора «расходы на рекламу» соответствует не одно конкретное значение объема продаж, а несколько, поскольку помимо набора исследуемых факторов на динамику объема продаж оказывают влияние и другие факторы. Поэтому в экономических исследованиях выделяют набор основных факторов и группу неучтенных факторов. В этом случае возможна ситуация, когда одинаковым значениям основного фактора  $x$  соответствует множество различных значений показателя  $y$ .

*Пример 1.1.* Изучалась зависимость объема продаж некоторого товара ( $y$ ) на предприятиях торговли от расходов на рекламу ( $x$ ). Были собраны данные по 12 однородным предприятиям (таблица). Требуется оценить наличие связи между показателями.

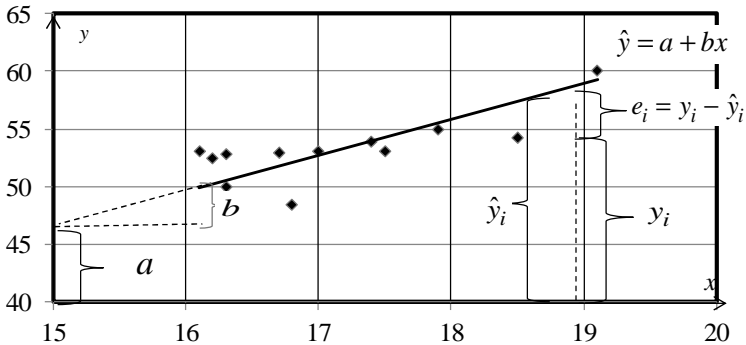
Таблица

Исходные данные об объеме продаж товара  
и расходах на рекламу на предприятиях торговли

Предприятие $i$	Расходы на рекламу (тыс. усл. ден. ед.) $x_i$	Объем продаж (тыс. шт.) $y_i$
1	16,3	52,8
2	16,8	48,4
3	18,5	54,2

Предприятие $i$	Расходы на рекламу (тыс. усл. ден. ед.) $x_i$	Объем продаж (тыс. шт.) $y_i$
4	16,3	50
5	17,9	54,9
6	17,4	53,9
7	16,1	53,1
8	16,2	52,4
9	17	53
10	16,7	52,9
11	17,5	53,1
12	19,1	60,1

*Решение.* Представим данную информацию с помощью графика рассеяния (рисунок).



*Рисунок.* График рассеяния, соответствующий данным примера 1.1, и общий вид регрессионной прямой

По скоплению точек на графике видно, что зависимость между показателями существует, но не такая жесткая, как функциональная. Она – прямая: с ростом  $x$  увеличивается  $y$  (см. рисунок).

Зависимость между показателями (пример 1.1) является стохастической и может быть формально описана регрессионной моделью. Линия, которая аппроксимирует зависимость, называется



линией регрессии, описывающее ее уравнение  $\hat{y} = a + bx_i$ ,  $i = \overline{1,12}$  – уравнением регрессии. Величины  $e_i = y_i - \hat{y}$ ,  $i = \overline{1,12}$  называют случайными остатками уравнения регрессии.

На современном этапе своего развития эконометрика как наука является следствием междисциплинарного подхода к изучению экономики и представляет собой сочетание экономической теории, математической экономики, экономической и математической статистики, информатики с использованием статистических пакетов прикладных программ. Принципиальное отличие эконометрики от других наук заключается в том, что цель эконометрики – количественная характеристика экономических закономерностей, выявляемых экономической теорией в общих чертах. В отличие от экономической теории, эконометрика делает упор на количественные, а не качественные аспекты этих явлений. Основным инструментом эконометрических исследований является аппарат теории вероятностей и математической статистики, реализованный в статистических пакетах прикладных программ.

## 1.2. Типы данных в количественном анализе

В эконометрической модели выделяют следующие переменные.

*Результирующая (зависимая, эндогенная) переменная*  $Y$  характеризует результат или эффективность функционирования экономической системы и соответствует, чаще всего, интегральному экономическому показателю.

*Объясняющие (независимые, экзогенные) переменные*  $X$  поддаются регистрации и описывают условия функционирования реальной экономической системы. Они в значительной мере определяют значения результирующих переменных. Обычно независимые переменные поддаются управлению. Значения этих переменных могут задаваться вне анализируемой системы (поэтому их и называют экзогенными; другое название – факторные признаки). В регрессионном анализе объясняющие переменные – это аргументы результирующей функции  $Y$ . По своей природе они могут быть как случайными, так и неслучайными. В примере 1.1 объясняющая переменная  $X$  – расходы на рекламу.

*Лаговые переменные* – это те переменные (экзогенные или эндогенные), значения которых зависят от показателя времени. Значения эндогенных переменных в предыдущие моменты времени могут заранее задаваться или вычисляться по уравнениям модели. Тогда эти эндогенные переменные играют роль экзогенных.

*Предопределенные переменные* – это экзогенные переменные вместе с их лаговыми значениями и лаговые значения эндогенных переменных в предыдущие моменты времени, которые служат для нахождения значений эндогенных переменных в данный момент времени.

Важным элементом эконометрической модели выступает случайная составляющая (случайное возмущение, случайный остаток)  $E$  – величина, входящая в уравнение модели, значения которой неизвестны, отражающая влияние неучтенных факторов, неточность выбора переменных и вида зависимостей между ними, погрешности нахождения значений переменных и др. Именно по поведению таких составляющих можно сделать заключение об адекватности построенной эконометрической модели. В примере 1.1 также имеется случайная составляющая  $e_i = y_i - \hat{y}_i$ ,  $i = \overline{1,12}$ .

В эконометрике помимо выявления причинно-следственных зависимостей решаются задачи прогнозирования экономических и социально-экономических показателей анализируемой экономической системы и имитации различных возможных сценариев ее развития.

В эконометрических исследованиях используют следующие типы экономических данных:

- *пространственные* – характеризуют ситуацию по конкретной переменной (или набору переменных), относящейся к пространственно разделенным сходным объектам в один и тот же момент или период времени. Таковы данные примера 1.1;

- *временные ряды* – отражают изменения (динамику) какой-либо переменной на промежутках времени, имеющих одинаковую продолжительность. В качестве примера временных рядов можно привести ежегодные данные объема продаж и расходов на рекламу торгового предприятия за несколько лет;

Современные эконометрические модели используют еще один тип данных – *панельные данные*, которые сочетают в себе как





параметров, при условии выполнения исходных предпосылок, может быть использован метод наименьших квадратов.

*Пример 1.2.* В упрощенной модели спроса-предложения (см. пример 1.1), которая описывает зависимость величины спроса  $Y^d$ , предложения  $Y^s$  и рыночной цены  $P$  некоторого блага от величины дохода  $X$ :

$$\begin{cases} Y^d = a_0 + a_1 P + a_2 X + \varepsilon^d, \\ Y^s = b_0 + b_1 P + \varepsilon^s, \\ Y^d = Y^s, \end{cases}$$

требуется определить экзогенные и эндогенные переменные модели и представить модель в приведенной форме.

*Решение.* В этой модели экзогенной переменной служит переменная  $X$ , эндогенные переменные –  $Y^d$ ,  $Y^s$  и  $P$ . Слагаемые  $\varepsilon^d$  и  $\varepsilon^s$  являются случайными составляющими. Последнее равенство является тождеством, отражающим суть рыночной цены. Нетрудно установить, что  $Y^d$ ,  $Y^s$  и  $P$  выражаются через  $X$  в виде:

$$\begin{cases} Y^d = \frac{b_1 a_0 - a_1 b_0}{b_1 - a_1} + \frac{b_1 a_2}{b_1 - a_1} X + (\varepsilon^d - \varepsilon^s), \\ Y^s = \frac{b_1 a_0 - a_1 b_0}{b_1 - a_1} + \frac{b_1 a_2}{b_1 - a_1} X + (\varepsilon^d - \varepsilon^s), \\ P = \frac{a_0 - b_0}{b_1 - a_1} + \frac{a_2}{b_1 - a_1} X + \frac{1}{b_1} (\varepsilon^d - 2\varepsilon^s). \end{cases}$$

Это и есть приведенная форма модели спроса-предложения. Слагаемые  $\varepsilon^d - \varepsilon^s$  и  $\frac{1}{b_1} (\varepsilon^d - 2\varepsilon^s)$  являются случайными остатками.

*Пример 1.3.* Пусть состояние экономики в заданном периоде времени описывается следующими переменными:  $Y_t$  – валовой внутренний продукт,  $C_t$  – уровень потребления,  $I$  – величина инвестиций,  $G_t$  – государственные расходы. В данных переменных модель Самуэльсона–Хикса имеет вид:

$$\begin{cases} C_t = a_0 + a_1 Y_{t-1} + u_t, \\ I_t = b(Y_{t-1} - Y_{t-2}) + v_t, \\ G_t = gG_{t-1} + w_t, \\ Y_t = C_t + I_t + G_t. \end{cases} \\ 0 < a_1 < 1, b > 0, g > 1.$$

Представить в приведенной форме и установить предопределенные переменные.

*Решение.* В модели (формула (1.2)) величины  $u_t$ ,  $v_t$ ,  $w_t$  являются случайными составляющими. Последнее равенство не содержит случайной составляющей и называется тождеством. Эндогенные переменные – величины  $C_t$ ,  $I_t$ ,  $Y_t$ ,  $G_t$ , экзогенные переменные:

$Y_{t-1}$  – ВВП за предшествующий период,

$(Y_{t-1} - Y_{t-2})$  – прирост ВВП за предшествующий период,

$G_{t-1}$  – государственные расходы в предшествующий период.

Приведенная форма модели получается подстановкой первых трех равенств:

$$\begin{cases} C_t = a_0 + a_1 Y_{t-1} + u_t, \\ I_t = b(Y_{t-1} - Y_{t-2}) + v_t, \\ G_t = gG_{t-1} + w_t, \\ Y_t = a_0 + a_1 Y_{t-1} + b(Y_{t-1} - Y_{t-2}) + gG_{t-1} + (u_t + v_t + w_t). \end{cases} \\ 0 < a_1 < 1, b > 0, g > 1.$$

Предопределенными переменными этой модели служат лаговые значения  $Y_{t-1}$ ,  $(Y_{t-1} - Y_{t-2})$ ,  $G_{t-1}$ .

Процесс построения и использования эконометрических моделей является достаточно сложным и включает в себя следующие основные этапы:

- постановка проблемы и выбор спецификации модели;
- получение данных, анализ их качества;
- идентификация модели;
- верификация модели;
- экономическая интерпретация результатов.

*В рамках первого этапа* экономическая проблема рассматривается в контексте положений экономической теории, с позиций последних осуществляется спецификация модели. Спецификацией

модели называется выбор переменных модели, вида зависимостей между ними и ограничений на коэффициенты уравнений, задающих эти зависимости. Спецификация модели обусловлена поставленной задачей исследования объекта. Ее базой служит экономическая теория или содержательный экономический анализ. В том случае, если в основу построения эконометрической модели положен содержательный экономический анализ, то выбор факторов исследователем осуществляется произвольно в рамках знаний экономического объекта. При отборе влияющих факторов могут быть использованы также статистические методы отбора. Метод исключения предполагает построение регрессионного уравнения, включающего всю совокупность факторов с дальнейшим последовательным сокращением числа факторов в модели до тех пор, пока не выполнится некоторое условие (удовлетворительный коэффициент детерминации, статистическая значимость регрессоров, отсутствие мультиколлинеарности). Суть метода включения – в последовательном включении факторов в модель до тех пор, пока регрессионная модель не будет отвечать заранее установленному критерию качества.

*Основной базой данных* для эконометрических исследований служат данные официальной статистики или данные бухгалтерского учета. В рамках информационного обеспечения формируются ряды наблюдений, что предполагает сбор статистических данных по выделенному кругу показателей или в пространстве (предприятия, цеха или отрасли), или во времени (год, месяц, квартал). На данном этапе важным является обеспечение методологической сопоставимости данных в пределах ряда, что требует знания особенностей формирования статистической отчетности. Так, при прогнозировании спроса на конечную продукцию корректировка статистического ряда требуется при возможной потере или приобретении клиента, появлении товаров-заменителей, исследовании зависимости поступлений в бюджет и ВВП.

Методологическая сопоставимость ряда поступлений в бюджет будет обеспечена в том случае, если в рамках наблюдаемого динамического ряда ставка налоговых поступлений оставалась неизменной; в противном случае ряд поступлений должен быть скорректирован, например, методом пропорциональной корректировки при условии, что ставка налога сохранялась неизменной

(на уровне последнего отчетного периода) на протяжении всего периода наблюдений. Или, в случае нестационарных данных, могут потребоваться преобразования, приводящие ряды к стационарному виду. Это, прежде всего, относится к временным рядам, которые часто содержат тренд и могут содержать сезонную компоненту. С тем, чтобы обеспечить требование стационарности, проводят определенную корректировку ряда, например, удаляют сезонную компоненту и тренд. Современные статистические пакеты программных продуктов позволяют осуществлять указанные преобразования данных в автоматизированном режиме.

Второй этап – *идентификация (параметризация) модели*, что предусматривает построение оценок коэффициентов (параметров) модели на основе статистических данных и описание свойств этих оценок. Задачей этапа является получение статистических оценок этих коэффициентов, на основе которых определяются регрессионные уравнения для описания зависимости средних ожидаемых значений (математических ожиданий) эндогенных переменных через значения предопределенных переменных. Свойства оценок коэффициентов, которым в эконометрике отводится особая роль, во многом определяются свойствами случайных составляющих в уравнениях и, конечно же, свойствами значений предопределенных переменных. Качество модели зависит от свойств оценок коэффициентов.

На этапе *верификации* модели проводится проверка качества и адекватности модели, что осуществляется с помощью специальных тестов или по качеству прогноза. Если тесты указывают на неадекватность спецификации модели, последняя должна быть пересмотрена с целью устранения этой неадекватности. Часто для проверки качества модели используется некоторая часть известных статистических данных, которые сравниваются с предсказанными данными, полученными на основании остальной части статистических данных. При использовании модели в качестве прогноза находится ошибка прогноза, которая определяется с заданной вероятностью.

В том случае, если модель была построена на основе содержательного экономического анализа, она дополнительно должна быть подвергнута анализу с позиции экономической теории, что осуществляется в рамках этапа *экономической интерпретации*



*результатов.* В этой части особое значение придается экономической интерпретации коэффициентов регрессионной модели.

### **Вопросы для контроля знаний**

1. Дайте определение эконометрики, эконометрической модели. Что является объектом и предметом изучения эконометрики?
2. Что такое «стохастическая зависимость»?
3. Какие типы переменных выделяют при построении регрессионной модели?
4. Что является информационной базой эконометрической модели?
5. Перечислите основные классы эконометрических моделей. В чем особенность каждого из них?
6. Какие виды систем эконометрических уравнений принято выделять? Приведите примеры.
7. Каковы этапы построения регрессионной модели?

## **Тема 2. ИСПОЛЬЗОВАНИЕ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В КОЛИЧЕСТВЕННОМ АНАЛИЗЕ**

### *Вопросы*

- 2.1. Случайные величины и функции их распределения.
- 2.2. Примеры дискретных и непрерывных случайных величин.
- 2.3. Основные понятия математической статистики.
- 2.4. Проверка статистических гипотез.

Поскольку объектом эконометрики выступают стохастические зависимости между показателями, для математически корректного построения эконометрических моделей разработчик модели должен владеть следующими понятиями: «случайная величина и функция ее распределения»; «параметр генеральной совокупности», «выборочная оценка параметра»; «эффективность, состоятельность и несмещенность оценок»; уметь формулировать статистические гипотезы.

### **2.1. Случайные величины и функции их распределения**

При использовании понятий теории вероятностей в анализе экономических объектов и систем вероятностное пространство трудно поддается (а скорее не поддается) описанию из-за сложности экономических систем. На первый план выходит понятие случайной величины (СВ), которая формально является функцией на пространстве событий. Например, такая СВ, как объем продаж, зависит от множества факторов, не все из которых можно учесть и оценить. Поэтому СВ будет пониматься как величина, значения которой зависят от исходов эксперимента (опыта). При использовании СВ будет учитываться множество принимаемых ею значений вместе с соответствующими им вероятностями. Случайная величина называется дискретной, если множество принимаемых ею значений конечно или счетно. Значения непрерывной СВ заполняют числовой промежуток или всю числовую прямую.

Законом распределения дискретной СВ называется совокупность ее значений и их вероятностей. Закон распределения задается в виде таблицы.

Таблица

Распределение вероятностей  
дискретной случайной величины

$x_i$	$x_1$	$x_2$	...	$x_n$
$p_i$	$p_1$	$p_2$	...	$p_n$

Поскольку в таблице указываются все значения случайной величины, сумма их вероятностей  $\sum_{i=1}^n p_i = 1$ .

Особенностью непрерывной случайной величины  $X$  является тот факт, что вероятность принятия ею любого конкретного значения  $x$  равна нулю:

$$P(X = x) = 0,$$

поэтому табличный способ не годится для задания закона распределения непрерывной СВ.

Другой способ задания закона распределения случайной величины  $X$ , который пригоден также и для непрерывных случайных величин, основан на использовании функции распределения.

Функцией распределения случайной величины  $X$  называется числовая функция  $F(x)$ , значение которой в точке  $x \in \mathbf{R}$  равно вероятности того, что величина  $X$  принимает значения меньшие, чем  $x$ :

$$F(X = x) = P(X < x).$$

Например, для дискретной случайной величины (см. таблицу):

$$F(x_2) = P(X < x_2) = P(X = x_1) = p_1.$$

С помощью функции распределения  $F(x)$  можно находить вероятности попадания СВ на произвольные промежутки:

$$P(a \leq X < b) = F(b) - F(a).$$

Для непрерывных СВ функция распределения является непрерывной.

Другой способ задания закона распределения непрерывной СВ использует понятие функции плотности вероятностей.

Функцией плотности распределения вероятностей непрерывной случайной величины  $X$  называется неотрицательная функция  $p(x)$ , такая, что

$$F(x) = \int_{-\infty}^x p(t) dt.$$

Функция плотности также позволяет находить вероятности попадания случайной величины  $X$  на произвольные промежутки:

$$P(a \leq X \leq b) = \int_a^b p(t) dt.$$

По известной теореме о производной от интеграла с переменным верхним пределом вытекает связь между функцией распределения и функцией плотности:  $F'(x) = p(x)$ .

Это равенство может также служить определением функции плотности в случае дифференцируемой функции распределения.

Таблица распределения и функция распределения для дискретной случайной величины, функция распределения и функция плотности для непрерывной случайной величины задают лишь вероятности значений случайной величины. В то же время сама случайная величина является функцией на пространстве событий.

Часто для применения случайных величин достаточно знать их основные характеристики: математическое ожидание (среднее или ожидаемое значение) и дисперсию.

Математическим ожиданием дискретной случайной величины  $X$  (например, заданной в таблице) называется число

$$MX = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i.$$

Число  $MX$  является взвешенным средним значений  $\{x_i\}$  величины  $X$  с коэффициентами  $\{p_i\}$ .

Для непрерывной случайной величины, имеющей функцию плотности вероятностей  $p(x)$ , математическое ожидание определяется формулой

$$MX = \int_{-\infty}^{\infty} xp(x) dx.$$

Второй характеристикой случайной величины  $X$ , которая учитывает разброс значений  $X$  относительно ее среднего  $MX$ , служит дисперсия (или вариация).

Дисперсией (вариацией) дискретной случайной величины  $X$ , заданной в табл.2.1, называется число

$$DX = \sum (x_i - MX)^2 p_i,$$

которое также записывается в виде

$$DX = M(X - MX)^2.$$

Это число является математическим ожиданием квадрата отклонения  $(X-MX)^2$  величины  $X$  от ее математического ожидания  $MX$ . Можно показать, что дисперсия также задается формулой

$$DX = M(X^2) - (MX)^2,$$

которая более удобна при вычислениях.

Для непрерывной случайной величины  $X$  с функцией плотности  $p(x)$  дисперсия определяется формулой

$$DX = \int_{-\infty}^{\infty} (x - MX)^2 p(x) dx,$$

или

$$DX = \int_{-\infty}^{\infty} x^2 p(x) dx - (MX)^2.$$

Наряду с дисперсией, часто используется другая характеристика случайной величины – стандартное отклонение. Стандартным отклонением случайной величины  $X$  называется число

$$\sigma_X = \sqrt{DX}.$$

В эконометрических моделях значения экономических показателей появляются в виде значений случайных величин. Как правило, при этом используются совокупности различных показателей, которые будут соответствовать наборам значений случайных величин.

Для непрерывных случайных величин независимость компонент определяется через функцию распределения. Непрерывные СВ  $X$  и  $Y$  называются независимыми, если их совместная функция распределения  $F(x, y) = P(X < x, Y < y)$  равна произведению функций распределения компонент:

$$F_{(X, Y)}(X < x, Y < y) = F_X(X < x) F_Y(Y < y).$$

В регрессионном анализе часто учитывается поведение одной из переменных в зависимости от поведения другой переменной. Эта величина характеризуется условным распределением вероятностей. Функция  $F(x / Y = y) = P(X < x / Y = y)$  называется функцией условного распределения вероятностей. Она задает вероятность того, что величина  $X$  принимает значение меньше  $x$  при условии, что  $Y = y$ .

Для двух случайных величин  $X$  и  $Y$  мерой их линейной зависимости служит ковариация. Ковариацией дискретных случайных величин  $X$  и  $Y$  называется число

$$\begin{aligned} \text{cov}(X, Y) &= M((X - MX)(Y - MY)) = M(XY) - MX \cdot MY = \\ &= \sum_i \sum_j (x_i - MX)(y_j - MY) p_{ij} = \sum_i \sum_j x_i \cdot y_j \cdot p_{ij} - MX \cdot MY. \end{aligned}$$

Обычно наряду с ковариацией используется безразмерная величина – коэффициент корреляции

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y},$$

где  $\sigma_X = \sqrt{DX}$ ,  $\sigma_Y = \sqrt{DY}$  – стандартные отклонения.

Коэффициент корреляции принимает значения от  $-1$  до  $+1$ . Теснота линейной связи между переменными тем выше, чем  $|\rho|$  ближе к единице.

Наряду со средним значением и дисперсией для характеристики распределений случайных величин используются моменты высших порядков. Моментом порядка  $k$  случайной величины  $X$  относительно ее среднего значения  $\mu = MX$  называется число  $\mu_k = M(X - \mu)^k$ .

С помощью моментов второго, третьего и четвертого порядков определяются асимметрия и эксцесс закона распределения случайной величины.

## 2.2. Примеры дискретных и непрерывных случайных величин

Рассмотрим основные примеры распределений вероятностей случайных величин.

*Биномиальное распределение.* Пусть имеется  $n$  независимых испытаний Бернулли с вероятностью успеха в одном испытании, равной  $p$ , и вероятностью неуспеха –  $q$ ,  $p + q = 1$ . Дискретная случайная величина  $X$ , значения которой равны числу успехов, имеет распределение  $X = k$ :

$$p_k = P(X = k) = C_n^k \cdot p^k \cdot q^{n-k}, k = \overline{0, n}.$$

Это распределение называется *биномиальным* с параметрами  $p$  и  $q$ . Заметим, что сумма вероятностей, вычисляемых по формуле Бернулли, равна

$$\sum_{k=0}^n p_k = \sum_{k=0}^n C_n^k p^k q^{n-k} = (p + q)^n = 1.$$

Математическое ожидание и дисперсия биномиальной случайной величины выражаются как

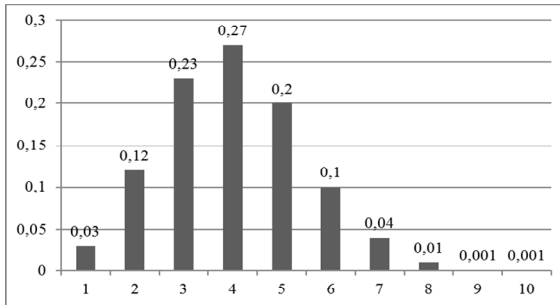
$$MX = n \cdot p, DX = n \cdot p \cdot q.$$

*Пример 2.1.* Пусть  $A$  – некоторое событие,  $X$  – СВ, равная числу свершений события  $A$ . СВ  $X$  распределена по биномиальному закону. Требуется построить графики распределения СВ  $X$ , если:

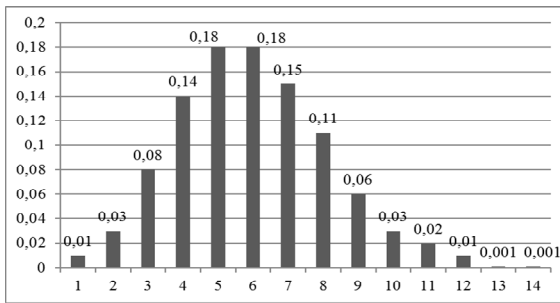
а) вероятность свершения события  $A$  равна 0,3, количество испытаний равно 10;

б) вероятность успеха – 0,05, число испытаний – 100.

*Решение.* Диаграммы биномиального распределения для параметров  $p = 0,3, n = 10$  и  $p = 0,05, n = 100$  приведены на рис. 2.1, а, б.



а



б

Рис. 2.1. Диаграммы биномиального распределения с параметрами  $p = 0,3, n = 10$  и  $p = 0,05, n = 100$

*Геометрическое распределение.* СВ  $X$ , имеющая *геометрическое распределение*, представляет собой число испытаний Бернулли до первого успеха.

Эта случайная величина принимает значения  $k = 1, 2, 3, \dots$  (счетное множество значений) с вероятностями

$$p_k = P(X = k) = p \cdot q^{k-1}, 0 < p < 1, q = 1 - p.$$

Определение является корректным, так как

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} p \cdot q^{k-1} = p \frac{1}{1-q} = 1.$$



Математическое ожидание и дисперсия  $X$  следующие:

$$MX = \frac{1}{p}, \quad DX = \frac{q}{p^2}.$$

*Пример 2.2.* В большой партии изделий вероятность брака равна  $p$ . Контроль качества проводится до первого появления бракованного изделия. В результате серии проверок обнаружилось, что бракованное изделие появлялось, в среднем, при десятом испытании. Требуется оценить вероятность брака  $p$ .

*Решение.* Пусть  $X$  – число испытаний до первого появления бракованного изделия. Эта случайная величина имеет геометрическое распределение. По условию ее среднее значение равно  $MX = 10$ .

Так как  $MX = \frac{1}{p}$ , то  $p = \frac{1}{10}$ .

*Гипергеометрическое распределение.* Дискретная случайная величина  $X$  имеет *гипергеометрическое распределение*, если она принимает значения  $m$  с вероятностями

$$p_m = P(X = m) = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n},$$

где  $m = 0, 1, \dots, k$ ;  $k = \min(n, M)$ ;

$$M \leq N;$$

$$n \leq N.$$

Вероятность  $p_m$  является вероятностью появления  $m$  объектов, обладающих заданным свойством, в множестве  $n$  объектов, случайно извлеченных (без возврата) из совокупности  $N$  объектов, среди которых  $M$  объектов обладают заданным свойством.

Математическое ожидание и дисперсия случайной величины, имеющей гипергеометрическое распределение с параметрами  $n, M, N$ :

$$MX = n \frac{M}{N}, \quad DX = n \frac{M}{N-1} \left(1 - \frac{M}{N}\right) \left(1 - \frac{n}{N}\right). \quad (2.1)$$

*Пример 2.3.* Среди продукции цеха сельхозоборудования 10 изделий из партии 100 штук не удовлетворяют стандарту. При приемке

продукции проверяют 10 единиц продукции. Требуется определить, какое среднее количество нестандартных единиц продукции будет обнаружено.

*Решение.* Количество нестандартных плат имеет гипергеометрическое распределение с параметрами  $n = 10$ ,  $M = 10$ ,  $N = 10$ , значит, согласно формуле (2.1):

$$MX = n \frac{M}{N} = 10 \frac{10}{100} = 1.$$

*Закон Пуассона.* Принято считать, что случайная величина  $X$  распределена по закону Пуассона, если она принимает целые значения  $k = 1, 2, 3, \dots$  с вероятностями

$$p_k = P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!},$$

где  $\lambda > 0$  – параметр распределения.

При этом

$$\sum_{k=0}^{\infty} p_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} p_k = e^{-\lambda} e^{\lambda} = 1.$$

Значения вероятностей  $p_k$  приводятся в таблицах распределения Пуассона. Распределение Пуассона иногда называют распределением редких событий. Примерами величин, распределенных по закону Пуассона, могут служить: число несчастных случаев, число дефектов в производстве и т. п.

Математическое ожидание и дисперсия пуассоновской случайной величины равны параметру распределения:

$$MX = \lambda, \quad DX = \lambda.$$

Распределение Пуассона используется для приближенных вычислений. Если в испытаниях Бернулли  $n$  принимает большие значения, а вероятность успеха  $p$  в одном испытании относительно мала (число  $\lambda = np$ ), то распределение Бернулли с достаточно высокой точностью можно заменить распределением Пуассона (рис. 2.1, б).

*Пример 2.4.* В ходе проверки аудитор случайным образом отбирает 10 счетов. Требуется найти вероятность того, что он обнаружит один счет с ошибкой, если в среднем 3 % счетов содержат ошибки.

*Решение.* Требуемую вероятность найдем по формуле Пуассона, в которой  $n = 10$ ,  $p = 0,03$ ,  $\lambda = np = 0,3$ :

$$p_1 \approx \frac{e^{-0,3} \cdot 0,3^1}{1!} \approx 0,2222.$$

Это же значение можно найти по таблицам значений распределения Пуассона:  $p_1 \approx 0,2222$ .

*Нормальное распределение.* Непрерывная случайная величина  $x$  имеет *нормальный закон распределения (закон Гаусса)* с параметрами  $a$  и  $\sigma$ , если ее плотность вероятностей имеет вид:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

где  $\sigma > 0$ .

Математическое ожидание случайной величины, распределенной по нормальному закону, равно параметру  $a$ , дисперсия равна  $\sigma^2$ , то есть

$$MX = a, \quad DX = \sigma^2.$$

Графики плотностей вероятностей нормально распределенных случайных величин при различных значениях параметров  $a$  и  $\sigma$  приведены на рис. 2.2 и 2.3. Параметр  $a$  характеризует положение,  $\sigma$  – форму кривой нормального распределения.

При  $a = 0$  и  $\sigma = 1$  распределение называется *стандартным нормальным* (рис. 2.4). Нормальное распределение возникает всегда при систематических отклонениях случайной величины от своего среднего, например, при измерении физических величин, отклонениях заданных координат при стрельбе (и т. д.), т. е. в тех случаях, когда на величину влияет большое количество случайных факторов, ни один из которых не является определяющим.

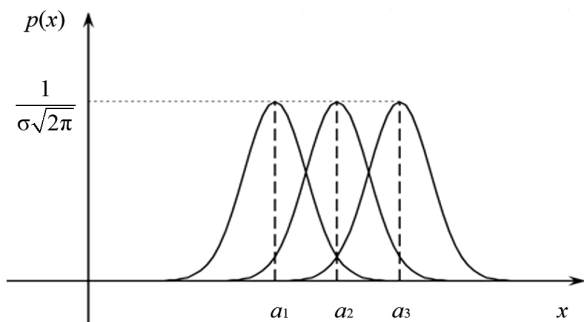


Рис. 2.2. Графики функций плотности вероятности нормального распределения с различными значениями параметра  $a$

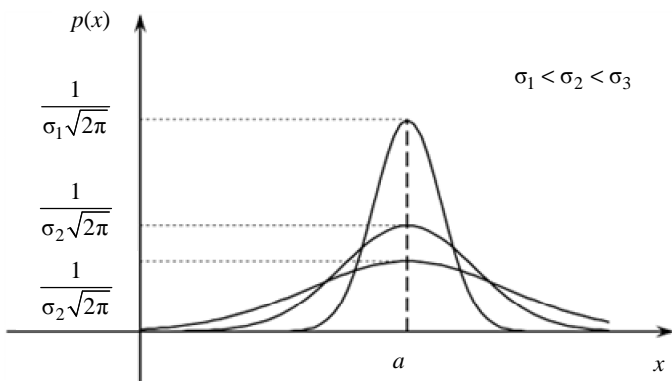


Рис. 2.3. Графики функций плотности вероятности нормального распределения

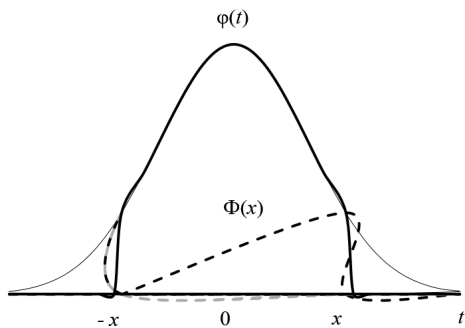


Рис. 2.4. Кривая плотности вероятностей стандартного нормального распределения

Геометрически  $\Phi(x)$  представляет собой часть площади под кривой  $\varphi(x)$  (см. рис. 2.4).

Пусть  $X \in N(a; \sigma)$ , тогда вероятность того, что  $X$  принимает значения из интервала  $(x_1, x_2)$ , можно вычислить, используя функцию  $\Phi(x)$ , по формуле

$$P(x_1 < X < x_2) = \frac{1}{2} \Phi\left(\frac{x_2 - a}{\sigma}\right) - \frac{1}{2} \Phi\left(\frac{x_1 - a}{\sigma}\right).$$

При решении задач статистики используется ряд специальных распределений:  $\chi$ -квадрат распределение, распределение Стьюдента, распределение Фишера, которые строятся на основе нормальных случайных величин.

### 2.3. Основные понятия математической статистики

Предметом математической статистики является изучение случайных событий и случайных величин по результатам наблюдений. Статистической совокупностью называется совокупность предметов или явлений, объединенных каким-либо общим признаком, который сохраняет свои свойства при многократных наблюдениях. Результатом наблюдений над статистической совокупностью являются статистические данные – сведения о том, какие значения принял в итоге наблюдений интересующий признак (случайная величина  $X$ ).

Обработка статистических данных методами математической статистики приводит к установлению определенных закономерностей, присущих массовым явлениям. При этом точность статистических выводов повышается с ростом числа наблюдений.

Статистические данные, как правило, представляют собой ряд значений  $\{x_1, x_2, \dots, x_n\}$  некоторой случайной величины  $X$ . Исследование случайной величины начинается с обработки этого ряда значений. Затем строятся функции, характеризующие случайную величину, и эти функции называются статистиками.

Таким образом, статистика – функция  $T:(x_1, x_2, \dots, x_n) \rightarrow T(x_1, x_2, \dots, x_n)$  от реализации случайной величины. Поскольку аргументы

функции  $T$  являются значениями случайной величины, значения самой функции  $T$  также становятся реализацией случайной величины.

В теоретических исследованиях удобно рассматривать статистику  $T$  как функцию от случайных величин  $X_1, X_2, \dots, X_n$ , каждая из которых имеет такое же распределение вероятностей, как и сама случайная величина  $X$  :

$$T = T(X_1, X_2, \dots, X_n).$$

Таким образом, изучая случайную величину  $X$ , мы рассматриваем набор одинаково распределенных, независимых в совокупности случайных величин  $\{X_1, X_2, \dots, X_n\}$ . В такой трактовке статистика становится случайной величиной, и изучение ее распределения приводит к выводам о распределении самой случайной величины  $X$ .

*Генеральная и выборочная совокупности.* Генеральной совокупностью называется совокупность объектов или наблюдений, а также их результатов, все элементы которой подлежат изучению при статистическом анализе.

В математической статистике генеральная совокупность часто понимается как совокупность всех мыслимых результатов наблюдений, которые могли быть получены при выполнении некоторых условий. Понятие генеральной совокупности аналогично понятию случайной величины (закону распределения вероятностей), так как обе они полностью определяются заданными условиями.

Генеральная совокупность может быть конечной или бесконечной. Объемом генеральной совокупности называется число ее объектов (наблюдений). Выборочной совокупностью называется часть объектов генеральной совокупности, используемая для исследования.

Например, число единиц какого-либо товара, произведенного фирмой в течение года, есть конечная генеральная совокупность. Для исследования качества продукции на практике рассматривается выборка, состоящая из  $n$  единиц товара. Признаком, или случайной величиной, может быть число единиц товара, удовлетворяющих сертификатным требованиям.

Сущность выборочного метода в математической статистике заключается в том, чтобы по определенной части генеральной совокупности (выборке) судить о ее свойствах в целом.

Выборочный метод является единственно возможным в случае бесконечной генеральной совокупности или тогда, когда исследование связано с уничтожением (гибелью) наблюдаемых объектов (например, исследование предельных режимов работы приборов, исследование действия вирусов на подопытных животных и т. д.). Чтобы по выборке можно было адекватно судить о случайной величине, она должна быть представительной (репрезентативной). Репрезентативность выборки обеспечивается объемом выборки и случайностью отбора ее элементов, при котором все элементы генеральной совокупности должны иметь одинаковую вероятность попадания в выборку.

Имеются два способа образования выборки:

1) повторная выборка, когда каждый элемент, случайно отобранный и исследованный, возвращается в общую совокупность и может быть отобран повторно;

2) бесповторная выборка, когда отобранный элемент не возвращается в общую совокупность.

*Точечные оценки параметров.* Пусть распределение случайной величины  $X$  (генеральной совокупности) задается вероятностями  $P(x_i, \theta)$  – для дискретной случайной величины или плотностью вероятностей  $P(x, \theta)$  – для непрерывной случайной величины, которые зависят от неизвестного параметра  $\theta$ . Этим параметром может быть, например, параметр  $\lambda$  закона Пуассона или параметры  $a$  и  $\sigma$  нормального распределения. На практике о величине параметра  $\theta$  можно судить по выборке  $\{x_1, x_2, \dots, x_n\}$  объема  $n$  из генеральной совокупности.

Оценкой  $\bar{\theta}_n$  параметра  $\theta$  называется любая функция от значений выборки  $\bar{\theta}_n = \bar{\theta}_n(x_1, x_2, \dots, x_n)$ , т. е. статистика. Неожиданным в этом определении является отсутствие связи оценки с самим параметром  $\theta$ . Эта связь появляется при введении свойств оценки.

Заметим, что сам параметр  $\theta$  является некоторым постоянным (неслучайным) числом, которое представляет истинное значение параметра генеральной совокупности. Статистику  $\bar{\theta}_n$  можно рассматривать как функцию от случайных величин  $X_1, X_2, \dots, X_n$ , когда  $x_i$  есть реализация случайной величины  $X_i$ :

$$\bar{\theta}_n = \bar{\theta}_n(X_1, X_2, \dots, X_n).$$

Таким образом, оценка  $\bar{\theta}_n$  является случайной величиной, и ее реализации заменяют неизвестные значения параметра  $\theta$ . Следовательно, оценку  $\bar{\theta}_n$  следует выбирать так, чтобы ее значения как можно точнее оценивали неизвестный параметр  $\theta$ . Например, можно потребовать, чтобы математическое ожидание оценки равнялось значению оцениваемого параметра  $\theta$ . Оценки с таким свойством имеют название «несмещенные». Однако на практике иногда приходится использовать оценки, не обладающие этим свойством.

Оценка  $\bar{\theta}_n$  называется несмещенной, если

$$M\bar{\theta}_n = \theta. \quad (2.2)$$

Если это требование не выполняется, то в среднем оценка  $\bar{\theta}_n$  будет всегда давать значение  $\theta$  с некоторым отклонением. Для несмещенных оценок устраняется возможность появления систематической ошибки при оценке параметра  $\theta$ . Статистика  $\bar{\theta}_n$  зависит от объема выборки  $n$ , что отмечено в ее обозначении, и при удачном построении оценки естественно ожидать, что при больших  $n$  значение статистики приближается к истинному значению параметра  $\theta$ . Это свойство оценки описывается с помощью понятия предела по вероятности: для любого  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\{|\bar{\theta}_n - \theta| < \varepsilon\} = 1.$$

Такая оценка  $\bar{\theta}_n$  называется состоятельной. Условие ( $p \lim$ ) означает, что  $p \lim \bar{\theta}_n = \theta$  ( $\bar{\theta}_n$  стремится к  $\theta$  по вероятности), поэтому при больших  $n$  с вероятностью, близкой к единице, отклонение  $\bar{\theta}_n$  от  $\theta$  становится сколь угодно малым.

Оценкой качества несмещенной оценки является ее дисперсия. Несмещенная оценка  $\bar{\theta}_n$  называется эффективной, если ее дисперсия

$$D\bar{\theta}_n = M(\bar{\theta}_n - \theta)^2,$$

является наименьшей среди дисперсий всех возможных оценок параметра  $\theta$ , вычисленных по одному и тому же объему выборки  $n$ .



При этом оценка  $\bar{\theta}_n$  предполагается выбранной из определенного класса функций.

Оценки  $\bar{\theta}_n$  называются точечными, так как они оценивают одно численное значение параметра  $\theta$  (точку).

Рассмотрим повторную выборку  $\{x_1, x_2, \dots, x_n\}$  значений генеральной совокупности  $X$ . Эта выборка может рассматриваться как реализация системы случайных величин  $\{X_1, X_2, \dots, X_n\}$ . При этом считается, что случайные величины  $X_1, X_2, \dots, X_n$ , имеющие такое же распределение вероятностей, как и сама  $X$ , являются независимыми в совокупности. Пусть  $MX = a, DX = \sigma^2$  – генеральные среднее и дисперсия случайной величины  $X$ . В качестве оценок для  $a$  и  $\sigma$  рассмотрим среднюю арифметическую выборки

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (2.3)$$

и выборочную дисперсию

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

Оценка  $\bar{x}$  для математического ожидания  $MX = a$  является несмещенной, состоятельной и эффективной. Ее дисперсия, в случае повторной выборки, равна

$$\sigma_x^2 = \frac{\sigma^2}{n}.$$

Оценка  $S^2$  для дисперсии  $DX = \sigma^2$  является состоятельной, но смещенной. Поэтому на практике часто пользуются несмещенной оценкой

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

Исправленной выборочной дисперсией называется оценка  $\overline{S^2}$  для дисперсии  $\sigma^2$ , определяемая формулой (2.3).

Для бесповторной выборки оценки (формулы (2.3), (2.4)) также являются несмещенными и состоятельными, а дисперсия оценки  $\bar{x}$  равна

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right), \quad (2.5)$$

где  $N$  – объем генеральной совокупности. При  $N \rightarrow \infty$  бесповторная выборка неотличима от повторной, и формула (2.4) переходит в формулу (2.4а). В случае, когда дисперсия  $\sigma^2$  неизвестна, она заменяется значением ее оценки  $S^2$ .

Пусть генеральная совокупность содержит  $M$  элементов, обладающих некоторым признаком  $A$ .

Генеральной долей признака называется величина

$$p = \frac{M}{N}. \quad (2.6)$$

Для доли  $p$  несмещенной и состоятельной оценкой будет выборочная доля

$$\bar{\omega} = \frac{m}{n},$$

где  $m$  – число элементов выборки, обладающих признаком  $A$ .

В случае повторной выборки дисперсия выборочной доли

$$\sigma_{\bar{\omega}}^2 = \frac{pq}{n},$$

в случае бесповторной выборки

$$\sigma_{\bar{\omega}}^2 = \frac{pq}{n} \left(1 - \frac{n}{N}\right),$$

где  $q = 1 - p$ .

Если  $n$  намного меньше  $N$ , то повторная выборка практически не отличается от бесповторной, и последние две формулы дают

приблизительно одинаковый результат. Если же  $n = N$ , то объем выборки равен объему генеральной совокупности, и выборочная доля равна генеральной, тогда  $\sigma_{\frac{z}{\omega}}^2 = 0$ .

В случае, когда  $p$  неизвестно, его заменяют выборочным значением  $\omega$ .

*Интервальные оценки параметров.* Точечная оценка  $\bar{\theta}_n$  параметра  $\theta$  дает лишь некоторое приближенное значение  $\theta$ . Чтобы получить представление о точности и надежности оценки, используют интервальную оценку параметра.

Интервальной оценкой параметра  $\theta$  называется случайный интервал  $(\alpha, \beta)$ , который с заданной вероятностью  $\gamma$  накрывает неизвестное значение параметра  $\theta$  (рис. 2.5).

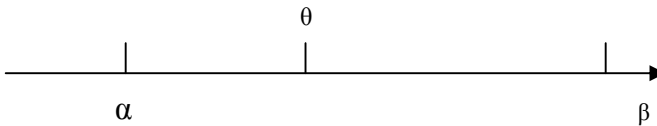


Рис. 2.5. Интервальная оценка параметра  $\theta$

Такой интервал  $(\alpha, \beta)$  называется доверительным интервалом, вероятность  $\gamma$  – доверительной вероятностью,  $\alpha = 1 - \gamma$  – уровнем надежности. Обычно доверительный интервал выбирается симметричным относительно оценки  $\bar{\theta}_n$ , тогда он определяется формулой

$$P\{|\bar{\theta}_n - \theta| < \Delta\} = \gamma$$

и имеет вид

$$(\bar{\theta}_n - \Delta, \bar{\theta}_n + \Delta),$$

т. е. неравенства  $\bar{\theta}_n - \Delta < \theta < \bar{\theta}_n + \Delta$  выполняются с вероятностью  $\gamma$ .

Наибольшее отклонение  $\Delta$  выборочного значения оценки  $\bar{\theta}_n$  параметра от истинного значения параметра  $\theta$  при заданном объеме выборки называется предельной ошибкой выборки.

Доверительные интервалы для генеральной средней и генеральной доли выборки. Пусть  $\{x_1, x_2, \dots, x_n\}$  – выборка из генеральной совокупности объема  $n$ ,  $\bar{x}$  – выборочное среднее,  $\overline{S^2}$  – выборочная дисперсия,  $\overline{S} = \sqrt{\overline{S^2}}$  – выборочное среднее квадратическое отклонение,  $\omega = \frac{m}{n}$  – выборочная доля признака.

Доверительный интервал с доверительной вероятностью  $\gamma$  для генеральной средней  $a$  имеет следующий вид:

$$\bar{x} - \Delta < a < \bar{x} + \Delta,$$

где  $\Delta$  – предельная ошибка выборки, зависящая от  $\gamma$  и  $n$ .

При  $n > 30$  для повторной выборки

$$\Delta = t \frac{\overline{S}}{\sqrt{n}},$$

для бесповторной выборки

$$\Delta = t \frac{\overline{S}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}.$$

В этих формулах  $t$  определяется из условия:

$$\Phi(t) = \gamma, \tag{2.7}$$

где  $\Phi(t)$  – функция Лапласа.

Если  $n \leq 30$ , то доверительный интервал для  $a$  строится только для нормальной генеральной совокупности. Для повторной выборки

$$\Delta = t_{n-1} \frac{\overline{S}}{\sqrt{n}}.$$

Здесь  $t_{n-1}$  определяется из условия:

$$P\{|\xi| < t_{n-1}\} = \gamma,$$

где  $\xi$  – случайная величина, имеющая распределение Стьюдента с  $n-1$  степенями свободы.

Для бесповторной выборки:

$$\Delta = t_{n-1} \frac{\bar{S}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}.$$

Доверительный интервал для генеральной доли  $p$  равен  $(\omega - \Delta, \omega + \Delta)$ , то есть

$$\omega - \Delta < p < \omega + \Delta,$$

Для повторной выборки (при  $n > 30$ )

$$\Delta = t \sqrt{\frac{\omega(1-\omega)}{n}}. \quad (2.8)$$

Для бесповторной выборки

$$\Delta = t \sqrt{\frac{\omega(1-\omega)}{n}} \sqrt{1 - \frac{n}{N}} \quad (2.9)$$

$t$  определяется равенством (2.7).

При  $n \leq 30$  рассматривается только выборка из нормальной совокупности. Для повторной выборки  $\Delta$  определяется по формуле (2.8), для бесповторной – по формуле (2.9).

## 2.4. Проверка статистических гипотез

Статистической гипотезой называется любое предположение о свойствах распределения вероятностей, лежащих в основе наблюдаемых явлений.

По своему содержанию статистические гипотезы можно подразделить на несколько основных типов:

1. Гипотезы о виде закона распределения исследуемой случайной величины.
2. Гипотезы о числовых значениях параметров случайной величины.
3. Гипотезы об общем виде модели, описывающей статистическую зависимость между признаками.

4. Гипотезы о принадлежности некоторого признака к тому или иному классу.

Гипотезы обозначаются прописными буквами латинского алфавита:  $H_0, H_1, H_2, \dots, H_k$ . Как правило, через  $H_0$  обозначается основная гипотеза, в справедливости которой важно убедиться ( $H_0$  – благополучная посадка самолета, выздоровление больного, успех в коммерческой деятельности и т. д.).

Гипотезе  $H_0$  противопоставляются гипотезы  $H_1, H_2, \dots, H_k$ , которые называются альтернативными. В дальнейшем будут рассматриваться только две гипотезы:  $H_0$  – основная и  $H_1$  – альтернативная (конкурирующая). Принятие гипотезы  $H_0$  или ее альтернативы основано на исследовании статистических данных, т. е. на выборке  $\{x_1, x_2, \dots, x_n\}$  из некоторой генеральной совокупности.

Гипотеза  $H$  называется простой, если она полностью определяет теоретическое распределение случайной величины по имеющейся выборке ее значений. В противном случае гипотеза называется сложной. Обычно основная гипотеза  $H_0$  является простой, в то время как ее альтернатива может быть простой или сложной.

Пусть имеется выборка значений случайной величины  $\{x_1, x_2, \dots, x_n\}$  из  $N(a, \sigma)$ , и  $\sigma$  известна. Тогда, если основная гипотеза  $H_0: a = a_1$ , альтернативная гипотеза  $H_1: a = a_2$ , где  $a_1$  и  $a_2$  – некоторые числа, то и  $H_0$  и  $H_1$  – простые гипотезы. Если  $H_0: a = a_0$ ,  $H_1: a \neq a_0$ , то  $H_0$  – простая гипотеза,  $H_1$  – сложная.

Правило  $K$ , по которому принимается или отвергается гипотеза, называется критерием. Вероятность отвергнуть гипотезу  $H_0$ , если она верна, и принять гипотезу  $H_1$ , называется ошибкой первого рода, или уровнем значимости, и обозначается  $\alpha$ :

$$P\{H_1 | H_0\} = \alpha.$$

Величина  $1 - \alpha$ , равная вероятности принять верную гипотезу  $H_0$ , называется уровнем доверия, или доверительной вероятностью:

$$P\{H_0 | H_0\} = 1 - \alpha = \gamma.$$

Вероятность принять основную гипотезу, если она неверна, называется ошибкой второго рода и обозначается  $\beta$ :

$$P\{H_0 | H_1\} = \beta.$$

Вероятность принять гипотезу  $H_1$ , если она верна, называется мощностью критерия:

$$P\{H_1 | H_0\} = 1 - \beta.$$

Процедура построения критерия осуществляется следующим образом. По имеющимся данным выборки составляется статистика  $\bar{\theta}_n$  так, чтобы в случае, если гипотеза  $H_0$  верна, распределение случайной величины  $\bar{\theta}_n = \bar{\theta}_n(X_1, \dots, X_n)$  было бы известным. Построение критерия, в зависимости от вида гипотезы  $H_0$ , заключается в выборе таких значений  $\theta_1$  и  $\theta_2$ , что если  $\theta_1 < \bar{\theta}_n < \theta_2$ , то гипотеза  $H_0$  принимается. При этом возможно, что  $\theta_1 = -\infty$  или  $\theta_2 = +\infty$ . Значения  $\theta_1$  и  $\theta_2$  называются критическими, а область  $T_D = \{\theta: \theta_1 < \theta < \theta_2\}$  называется областью допустимых значений. Если  $\bar{\theta}_n \in T_D$ , то гипотеза  $H_0$  принимается, если  $\bar{\theta}_n \notin T_D$  – отвергается, т. е. принимается альтернатива  $H_1$ . При этом область  $T_{кр} = \{\theta: \theta_1 \leq \theta, \theta \geq \theta_2\}$  называется критической.

*Критерии согласия.* Одной из главных задач математической статистики является установление истинного закона распределения случайной величины на основании экспериментальных данных. Вид закона распределения определяют из общих технических, финансовых и других соображений, учитывая схожесть условий эксперимента с исследованными ранее, или из теоретических предпосылок.

На практике о виде закона распределения можно судить по графику выборочной плотности распределения вероятностей. Параметры закона распределения обычно неизвестны, и их заменяют на выборочные значения. Однако как бы ни выбирали закон распределения и его параметры, полной уверенности в том, что в результате получится истинный закон распределения, к которому принадлежит имеющаяся выборка, не существует. Поэтому может ставиться вопрос лишь о том, что на определенном уровне доверия выбранный закон согласуется с данными выборки.

В соответствии с этим критерии, устанавливающие закон распределения, называются критериями согласия.

*Критерий Пирсона.* Пусть  $\{x_1, x_2, \dots, x_n\}$  – выборка из некоторой генеральной совокупности  $X$ ,  $F(X)$  – предполагаемая функция

теоретического распределения. На основании выборки построим интервальный ряд  $\{\Delta_i, n_i\}$ ,  $i = \overline{1, m}$ , где  $n_i$  – число элементов выборки, попавших в интервал  $\Delta_i = [a_i, a_{i+1})$ . Для каждого интервала  $\Delta_i$  вычислим теоретические вероятности  $p_i$  попадания случайной величины  $X$  в интервал  $\Delta_i$ :

$$p_i = P\{X \in \Delta_i\} = F(a_{i+1}) - F(a_i).$$

Числа  $n_i$  и  $np_i$  называются, соответственно, эмпирическими и теоретическими частотами. Доказано, что при  $n \rightarrow \infty$ , статистика

$$t = \sum_{i=1}^n \frac{(n_i - np_i)^2}{np_i},$$

имеет  $\chi^2$ -распределение с  $k = m - r - 1$  степенями свободы, в котором  $m$  – число интервалов вариационного ряда,  $r$  – число параметров теоретического распределения, вычисленных по экспериментальным данным.

Основная гипотеза  $H_0$  состоит в том, что функцией распределения случайной величины  $X$  является выбранная теоретическая функция  $F(X)$ .

Для заданного уровня доверия  $\gamma$  по таблицам распределения  $\chi_k^2$  находим критическое значение:

$$\chi_{k, \text{кр}}^2 : P\{\chi_k^2 < \chi_{k, \text{кр}}^2\} = \gamma.$$

Гипотеза  $H_0$  о согласии экспериментальных данных с распределением  $F(X)$  принимается, если  $t < \chi_{k, \text{кр}}^2$ .

Заметим, что статистика  $t$  имеет распределение  $\chi^2$  при  $n \rightarrow \infty$ , поэтому критерий Пирсона следует применять только при  $n > 30$ .

### Вопросы для контроля знаний

1. Что такое «математическое ожидание» и «дисперсия случайной величины»?
2. Как формулируется закон распределения дискретной случайной величины?
3. Когда две случайные величины считаются независимыми?



4. Что такое «ковариация» и «корреляция»?
5. Что такое «стандартное нормальное распределение»?
6. Дайте определения генеральной и выборочной совокупности.
7. Дайте определения свойствам несмещенности, состоятельности и эффективности оценок случайной величины.
8. Что такое «интервальная оценка параметров»?
9. Как проверить гипотезу о численной величине среднего значения?

## **Практическое занятие № 1**

### **Основные характеристики случайных величин и их оценки**

**Цель работы:** генерация выборок значений нормальной и равномерно распределенной случайных величин, вычисление их статистических характеристик, построение интервальных вариационных рядов и гистограмм.

#### **Методические рекомендации**

При выполнении практических работ необходима надстройка Excel **Анализ данных**. Актуализация этой команды проводится следующим образом.

• Проверьте доступ к пакету анализа. Для этого в главном меню последовательно выберите **Файл/Параметры/Надстройки/Надстройки Excel перейти**. Установите флажок **Пакет анализа**.

• В главном меню выберите **Данные/Анализ данных**. Затем выбираете требуемый инструмент анализа.

*Пример 2.1. Нормальная случайная величина*

Для генерации данных примера нормальной случайной величины используем инструмент **Генерация случайных чисел**. В качестве параметров ввода используем следующие. *Число переменных – 1. Число случайных чисел – 100. Распределение – нормальное. Параметры: Среднее=650 Стандартное отклонение=80. Параметры вывода – выходной интервал, ячейка A2. Ячейку A1 используем для указания метки **НормРасп**.*

Полученные данные для компактности записи представлены несколькими столбцами (табл. 1).

## Исходные данные

НормРасп	621,3	540,8	666,5	723,4	684,7
584,1	687,4	581,9	735,1	713,1	670,3
541,1	661,2	712,8	647,8	597,3	702,7
503,3	587,4	544,9	534,8	569,9	622,8
556,8	645,8	791,3	717,3	715,5	577,3
642,2	557,4	636,1	741,1	647,6	677,8
783,0	749,9	758,4	614,3	583,3	597,2
647,1	815,3	855,2	581,4	624,5	763,9
712,0	650,8	536,9	676,7	604,3	638,3
794,7	567,6	670,8	648,1	681,4	656,0
635,3	621,2	748,5	535,5	676,9	578,7
726,4	624,7	626,4	619,3	725,0	626,5
596,2	675,6	652,8	619,1	579,2	922,1
670,7	722,7	703,0	675,8	706,7	748,2
692,8	628,7	615,1	595,9	538,2	634,0
638,6	705,5	714,3	691,1	639,3	647,6
519,9	592,9	550,8	745,0	658,4	

Статистические характеристики для исходной выборки рассчитываются с помощью надстройки **Анализ данных**, инструмент **Описательная статистика**. Входные данные: *Входной интервал* – \$A\$1:\$A\$101. *Группирование* – по столбцам. *Метки в первой строке* – ставим «галочку». *Параметры вывода: выходной интервал* – указываем ячейку. *Итоговая статистика* – ставим «галочку».

Получены следующие результаты: *Среднее* равно 653,8; *Минимум* равен 503,3; *Максимум* равен 922,1; *Дисперсия выборки* равна 5865,0; *Стандартное отклонение* равно 76,6.

Построение интервального вариационного ряда. Для определения количества интервалов  $k$  воспользуемся формулой Стерджерса:

$k = 1 + 1,4 \ln n$ , где  $n$  – количество исследуемых данных. При подстановке  $n = 100$  получим  $k = 6,447$ . Возьмем  $k$  равным 7.

В качестве аргумента функции в обоих случаях выступает интервал A2:A101. Минимальное значение равно 503,3. Максимальное значение равно 922,1.

Определим длину интервала  $\Delta$ : размах (в Excel – интервал), равный 418,8, делим на  $k = 7$ . Получено значение 59,8, принимаем  $\Delta = 60$ . Разобьем данные на 7 интервалов (см. табл. 1). При этом в 1-й интервал попадут все значения, меньшие, чем 550, в 7-й интервал – все значения, большие, чем 850. При этом длина последнего интервала равна 100 (табл. 2).

Таблица 2

Разбиение данных на интервалы

Номер интервала	Интервал	4	[670,730)
1	[500,550)	5	[730,790)
2	[550,610)	6	[790,850)
3	[610,670)	7	[850,950)

Занесем полученные данные в Excel (рис. 1). *Обратите внимание, что последний интервал не заполняется, так как является открытым.*

	А	В
	Номер интервала	Граница интервала
1	1	550
2	2	610
3	3	670
4	4	730
5	5	790
6	6	850
7	7	
8		

Рис. 1. Исходные данные для определения частот попаданий в интервалы

Для подсчета попаданий значений в интервалы:

1) выделяем место для вывода E2:E8;

2) вызываем функцию **ЧАСТОТА (Мастер функций/Статистические)**. Аргументы функции: *Массив данных* – A2:A100. *Массив интервалов* – D2:D8. Нажимаем комбинацию клавиш: **Ctrl+Shift+Enter**.

Получим следующие частоты по интервалам: 1) 9; 2) 19; 3) 31; 4) 27; 5) 9; 6) 3; 7) 2.

Строим гистограмму распределения (**Вставка–Диаграммы–Гистограмма**) (рис. 2).

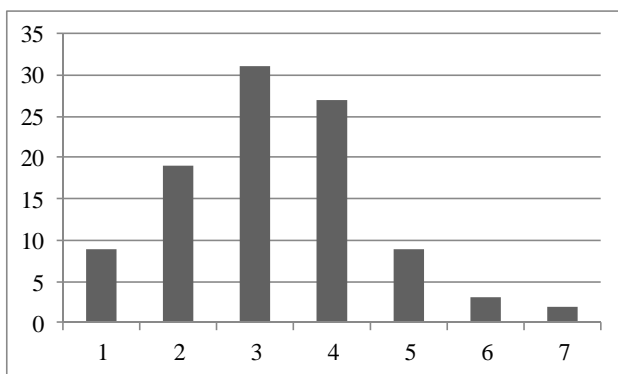


Рис. 2. Гистограмма распределения для примера 2.1

Визуальный анализ показывает соответствие графику функции плотности вероятности нормального распределения.

Для сравнительного анализа вычислим статистические характеристики интервального вариационного ряда.

Строим интервальный вариационный ряд, указывая середины интервалов и частоты попадания в них.

Середины интервалов вычисляются по формуле  $(a_i + b_i)/2$ , в которой  $a_i$  и  $b_i$  – начало и конец интервала. Для первого интервала  $a_i = 500$  и  $b_i = 550$ , для второго –  $a_i = 550$  и  $b_i = 610$ , для седьмого –  $a_i = 850$  и  $b_i = 950$ .

Вместо исходного ряда данных получаем ряд, в котором середины интервалов повторяются в соответствии с их частотами (табл. 3).

Расчет середин интервалов

Номер интервала	$a_i$	$b_i$	Середина интервала	Частота
1	500	550	525	9
2	550	610	580	19
3	610	670	640	31
4	670	730	700	27
5	730	790	760	9
6	790	850	820	3
7	850	950	900	2

Вычислим статистические характеристики интервального вариационного ряда.

Для расчета среднего значения умножим значения середин интервалов на их частоты попадания, просуммируем полученные значения и результат поделим на количество значений в исходной выборке (100). Получено среднее значение по интервальному ряду, равное 655,9. Оно незначительно отличается от среднего значения по выборке (653,8).

Для расчета исправленной выборочной дисперсии: определяем отклонения середин интервалов от среднего значения, возводим их в квадрат, суммируем и делим на  $n - 1$ , где  $n = 100$ . Получено значение 6277,8, которое более значимо отличается от дисперсии выборки (5864,9). Однако стандартное отклонение интервального ряда, рассчитываемое как корень квадратный из дисперсии, равно 78,0 и незначительно отличается от стандартного отклонения исходной выборки (76,6).

#### *Пример 2.2. Равномерная случайная величина*

Для генерации данных примера равномерной случайной величины используем инструмент **Генерация случайных чисел**. На новом листе Excel в ячейке A1 записываем название (метку) **Равномерн-Распр**. Вызываем надстройку **Анализ данных/Генерация случайных чисел**. В качестве параметров ввода используем следующие. *Число переменных* – 1. *Число случайных чисел* – 120. *Распределение* – равномерное. *Параметры* – между 5 и 30. *Параметры вывода* – выходной интервал, ячейка A2.

Для примера 2.2 были получены следующие значения выборки (табл. 4).

Таблица 4

Исходные данные

РавномернРасп	9,6	18,8	27,8	20,1	5,5
	19,4	29,0	16,8	26,2	13,9
	29,6	25,9	16,0	5,9	25,0
15,3	16,7	29,9	16,4	24,3	16,6
15,0	6,0	15,4	20,1	29,3	28,0
5,7	12,5	8,7	10,8	22,3	13,7
19,9	28,0	14,6	19,3	17,9	23,9
14,1	14,2	21,7	29,6	24,5	23,5
20,3	6,8	16,7	22,0	8,8	25,6
24,3	14,0	23,2	21,4	27,8	29,2
16,0	9,2	21,9	16,4	15,7	13,2
29,8	21,5	5,3	19,8	23,2	16,6
9,7	8,6	25,4	28,4	7,2	8,4
28,3	10,4	13,5	6,4	12,5	8,4
18,0	21,5	14,0	18,0	6,8	26,1
12,8	7,3	26,6	13,9	8,7	26,5
23,6	24,7	8,8	18,9	14,2	7,3
8,0	14,8	27,5	9,5	5,1	28,6
9,6	5,5	11,1	7,5	23,7	
12,8	23,5	20,9	6,1	16,6	
23,7	18,4	9,6	14,1	12,2	

Рассчитываем статистические характеристики аналогично примеру 2.1.

Получены следующие результаты: *Среднее* равно 16,7; *Минимум* равен 5,1; *Максимум* равен 29,8; *Дисперсия выборки* равна 54,7; *Стандартное отклонение* равно 7,4.

Количество интервалов  $k$  при  $n = 120$  равно 6,7. Возьмем  $k$  равным 7. Минимальное значение равно 5,1. Максимальное значение равно 29,8. Интервал  $\Delta$  равен 3,5. Разобьем данные на 7 интервалов (табл. 5).

Разбиение данных на интервалы

Номер интервала	Интервал		
		4	[15,5;19)
1	[5,8;5)	5	[19;22,5)
2	[8,5;12)	6	[22,5;26)
3	[12;15,5)	7	[26;30)

Занесем полученные данные в Excel.

Рассчитаем частоты попаданий значений в интервалы. Получим следующие частоты: 1) 18; 2) 14; 3) 21; 4) 17; 5) 15; 6) 16; 7) 19.

Строим гистограмму распределения (рис. 3).

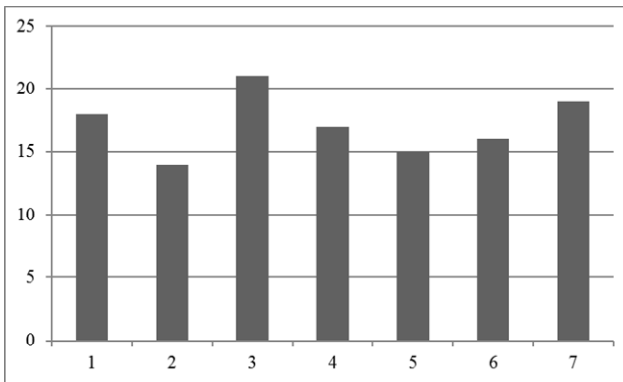


Рис. 3. Гистограмма распределения для примера 2.2

Визуальный анализ показывает соответствие графику функции плотности вероятности равномерного распределения. Равномерное распределение имеет теоретическую функцию плотности, равную константе на отрезке, где она отлична от нуля.

Для сравнительного анализа вычислим статистические характеристики интервального вариационного ряда.

Строим интервальный вариационный ряд, указывая середины интервалов (6,75; 10,25; 13,75; 17,25; 20,75; 24,25; 28,00) и частоты попадания в них: 18; 14; 21; 17; 15; 16; 19.

Вычислим статистические характеристики интервального вариационного ряда.

Получено среднее значение по интервальному ряду, равное 20,8 при среднем значении по выборке, равном 16,7.

Исправленная выборочная дисперсия равна 63,3 при дисперсии выборки, равной 54,4. Стандартное отклонение интервального ряда равно 7,9 при стандартном отклонении исходной выборки, равной 7,4. Следовательно, выборочные характеристики незначительно отличаются от исходных.

### Задания для самостоятельной работы

Провести генерации нормальной и равномерно распределенной случайных величин согласно варианту по номеру в списке группы. Исходные количества  $n$  выбираем из табл. 6.

Таблица 6

Исходные данные

№ варианта	Нормальная случайная величина	Равномерно распределенная случайная величина
1–3	$n = 90; x_{\text{ср}} = 500; \sigma_x = 40$	$n = 100;$ параметры – между 10 и 50
4–6	$n = 120; x_{\text{ср}} = 600; \sigma_x = 45$	$n = 80;$ параметры – между 5 и 25
7–9	$n = 80; x_{\text{ср}} = 400; \sigma_x = 25$	$n = 120;$ параметры – между 20 и 60
10–12	$n = 110; x_{\text{ср}} = 700; \sigma_x = 50$	$n = 90;$ параметры – между 20 и 50

Выполнить расчет в Excel. Оформить отчет.



## **Тема 3. ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ КАК ИНСТРУМЕНТ ИССЛЕДОВАНИЯ ЭКОНОМИЧЕСКИХ ЗАВИСИМОСТЕЙ**

### *Вопросы*

- 3.1. Стохастическая зависимость, функция регрессии.
- 3.2. Корреляция и линейная парная регрессия.
- 3.3. Коэффициент детерминации и средняя ошибка аппроксимации.
- 3.4. Использование функции регрессии в анализе и прогнозе.

В данной теме рассмотрены методы корреляционного и регрессионного анализа, которые используются для построения простейшего вида эконометрических моделей – регрессионных. Подробно рассмотрена методика парного регрессионного анализа, что обусловлено ее аналогией с методикой множественной линейной регрессии, а также возможностью линеаризации нелинейных моделей.

### **3.1. Стохастическая зависимость, функция регрессии**

На практике между экономическими показателями существует стохастическая (или вероятностная) зависимость. В отличие от функциональной зависимости, каждому из значений  $x$  независимой переменной  $X$  соответствует не одно конкретное значение другой переменной, а случайная величина  $Y(x)$ . Поскольку случайная величина характеризуется своим законом распределения вероятностей, стохастическая зависимость понимается как соответствие  $x$  **a**  $Y(x)$ , при котором величина  $Y(x)$  понимается как распределение вероятностей случайной величины. В дискретном случае распределение  $Y(x)$  можно задать таблицей значений и их вероятностей, а в непрерывном случае распределение задается функцией плотности вероятностей. Одной из основных характеристик случайной величины служит ее среднее значение (или математическое ожидание), которое является числом. Таким образом, при заданной стохастической зависимости  $x$  **a**  $Y(x)$  возникает обычная числовая функция  $x$  **a**  $MY(x)$ , значениями которой являются

осредненные значения случайной величины  $Y$ . В статистике часто вместо термина «случайная величина» используется термин «генеральная совокупность». При этом характеристики, относящиеся к случайным величинам, называют генеральными, чтобы отличать их от выборочных характеристик.

Функция, которая числу  $x$  ставит в соответствие условное математическое ожидание  $M_x Y = MY(x)$ , т. е. функция  $M_x(Y)$ , называется функцией регрессии, соответствующей стохастической зависимости  $x$  а  $Y(x)$ , в которой  $x$  – число (значение независимой переменной),  $Y(x)$  – случайная величина, соответствующая числу  $x$ .

На практике полная (генеральная) совокупность значений показателей отсутствует, имеем только их выборки.

Поле корреляции называется графическое изображение стохастической зависимости, при этом точки плоскости соответствуют значениям  $(x_i, y_i)$ . При исследовании стохастических зависимостей на практике оценку функции регрессии проводят по данным выборки.

Выборочной функцией регрессии называется оценка функции регрессии, построенная по выборке. Существуют различные методы построения такой оценки.

### 3.2. Корреляция и линейная парная регрессия

Исследование зависимости между экономическими показателями начинается с корреляционного анализа, основная задача которого заключается в выявлении взаимосвязи между случайными переменными путем точечной и интервальной оценок парных (частных) коэффициентов корреляции, вычисления и проверки значимости множественных коэффициентов корреляции и коэффициента детерминации. Кроме того, с помощью корреляционного анализа решаются следующие задачи: отбор факторов, оказывающих наиболее существенное влияние на результативный признак, на основании измерения тесноты связи между ними; обнаружение ранее неизвестных причинных связей. Корреляция непосредственно не выявляет причинных связей между показателями, но устанавливает численное значение этих связей и достоверность суждений об их наличии.

При проведении корреляционного анализа вся совокупность данных рассматривается как множество переменных (факторов), каждая из которых содержит  $n$  наблюдений.

При изучении взаимосвязи между факторами их обозначают следующим образом:  $X = (x_1, x_2, \dots, x_n)$  и  $Y = (y_1, y_2, \dots, y_n)$ .

Для заданной пары случайных величин  $X$  и  $Y$  рассматривается выборка их значений  $(x_i, y_i)$ , где  $i = 1, \dots, n$ . На основании этой выборки можно сделать заключение о взаимосвязи величин. Характеристикой такой взаимосвязи выступает коэффициент ковариации.

Ковариацией случайных величин  $X$  и  $Y$  называется число

$$\text{cov}(X, Y) = M((X - MX)(Y - MY)) = M(XY) - MX \cdot MY. \quad (3.1)$$

Ковариация – ожидаемое значение произведения отклонений величин от их ожидаемых значений. Оценкой ковариации по выборке значений величин служит выборочная ковариация.

Выборочной ковариацией случайных величин  $X$  и  $Y$  называется величина

$$\text{cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad (3.2)$$

где  $\bar{x}$ ,  $\bar{y}$  – выборочные средние, которые находятся по формуле:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (3.3)$$

При этом выборку значений случайной величины  $X$  мы понимаем двояко: либо это набор случайных величин  $\{X_1, X_2, \dots, X_n\}$ , каждая из которых имеет такое же распределение вероятностей, как и сама  $X$ , либо это набор чисел  $\{x_1, x_2, \dots, x_n\}$ , в котором  $x_i$  ( $i = 1, \dots, n$ ) является значением случайной величины  $X_i$ . Кроме того, предполагается, что пары  $(X_i, Y_j)$  для  $i \neq j$  являются независимыми случайными величинами.

Формулу (3.2) можно понимать как правило (процедуру) оценивания ковариации.

Ковариация имеет положительный знак, если при увеличении (уменьшении) одной из переменных вторая переменная тоже увеличивается (уменьшается). Отрицательный знак ковариации

означает, что переменные изменяются в противоположные стороны (рисунок).

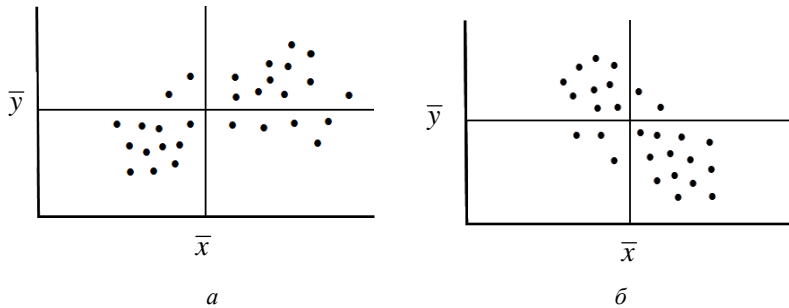


Рисунок. Графики прямой (а) и обратной (б) зависимости факторов

Если случайные величины  $X$  и  $Y$  независимы, то  $\text{cov}(X, Y) = 0$ . Если  $\text{cov}(X, Y) \neq 0$ , то  $X$  и  $Y$  – зависимые случайные величины. Для нормально распределенных случайных величин понятия независимости и некоррелированности ( $\text{cov}(X, Y) = 0$ ) совпадают.

Дисперсия  $\sigma_x^2$  случайной величины  $X$  является частным случаем ковариации:  $\sigma_x^2 = \text{cov}(X, X)$ . Ее несмещенной оценкой служит исправленная выборочная дисперсия

$$s_x^2 = \text{cov}(X, X) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Выборочная дисперсия  $s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  также является оценкой дисперсии  $\sigma_x^2$ , но смещенной.

Для измерения тесноты линейной стохастической зависимости между двумя случайными величинами  $X$  и  $Y$  используется безразмерная статистическая характеристика – коэффициент корреляции:

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y}.$$

Выборочным коэффициентом корреляции называется величина

$$r_{x,y} = \frac{\text{cov}(X, Y)}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x s_y}, \quad (3.4)$$

где  $s_x, s_y$  – (исправленные) выборочные дисперсии величин  $X$  и  $Y$  (или оценки дисперсий  $\sigma_x, \sigma_y$  этих величин).

Ковариация и корреляция представляют, по существу, одну и ту же информацию о взаимной вариации величин. Однако корреляция является более удобной безразмерной величиной.

Выборочный коэффициент корреляции обладает следующими свойствами:

- 1)  $-1 \leq r \leq 1$ ;
- 2) чем ближе  $|r|$  к единице, тем теснее линейная связь;
- 3) при  $|r| = 1$  между переменными существует функциональная линейная зависимость.

Выборочный коэффициент корреляции является оценкой генерального коэффициента корреляции.

Так как для построения оценки коэффициента корреляции пользуются выборкой, возникает вопрос о значимости коэффициента корреляции. Проверим нулевую гипотезу  $H_0$  об отсутствии линейной связи между величинами  $X$  и  $Y$  на основе  $t$ -статистики Стьюдента. Расчетное значение критерия определяем по формуле

$$t_p = \frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \sqrt{n - 2}. \quad (3.5)$$

С учетом заданного уровня значимости  $\alpha$  и числа степеней свободы, равного  $(n-2)$ , по таблице Стьюдента или в Excel находят критическое значение  $t_{кр}$ . Если  $|t_p| > t_{кр}$ , то коэффициент корреляции считается значимым. Нулевая гипотеза отвергается. Это означает, что выборочные данные на уровне значимости  $\alpha$  не противоречат наличию линейной статистической связи между  $X$  и  $Y$ .

Для оценки тесноты линейных связей между различными парами переменных из набора переменных  $\{X_1, X_2, \dots, X_m\}$  строится матрица коэффициентов парной корреляции  $\mathbf{R}$ . По выборке пар значений этих переменных  $(x_{i1}, x_{i2}, \dots, x_{im}), I = 1, \dots, n$  вычисляются парные коэффициенты корреляции и составляется матрица:

$$\mathbf{R} = \begin{pmatrix} 1 & r_{x_1x_2} & \mathbf{K} & r_{x_1x_m} \\ r_{x_1x_2} & 1 & \mathbf{K} & r_{x_2x_m} \\ \mathbf{K} & \mathbf{K} & \mathbf{K} & \mathbf{K} \\ r_{x_1x_m} & r_{x_2x_m} & \mathbf{K} & 1 \end{pmatrix}. \quad (3.6)$$

Эта матрица используется при построении модели множественной регрессии, при определении выборочных коэффициентов множественной корреляции и выборочных частных коэффициентов корреляции.

Теснота линейной связи одной из случайных величин с совокупностью остальных величин проверяется с помощью выборочного множественного коэффициента корреляции, который вычисляется по формуле

$$R_{j;12\dots\hat{j}\dots m} = \sqrt{1 - \frac{|\mathbf{R}|}{R_{jj}}}, \quad (3.7)$$

где  $|\mathbf{R}|$  – определитель матрицы  $\mathbf{R}$ ,

$R_j$  – определитель матрицы, полученной из матрицы  $\mathbf{R}$ , вычеркиванием  $j$ -го столбца и  $j$ -й строки. Надстрочный символ индекса в формуле (3.7) означает пропуск этого индекса.

Для вычисления выборочного коэффициента множественной корреляции  $R_{j;1,2,\mathbf{K},\mathbf{K}m}$  в Excel можно воспользоваться матрицей  $\mathbf{R}^{-1}$ , обратной к матрице  $\mathbf{R}$  парных коэффициентов корреляции (3.6). Величина  $\frac{|\mathbf{R}|}{R_{jj}}$  является обратной величиной  $j$ -го элемента

главной диагонали матрицы  $\mathbf{R}^{-1}$ .

Выборочным коэффициентом детерминации называется квадрат выборочного коэффициента множественной корреляции:

$$R_{j:12\dots j\dots m}^2 = 1 - \frac{|\mathbf{R}|}{R_{jj}}. \quad (3.8)$$

Этот коэффициент принимает значения от 0 до 1 и может только увеличиваться при добавлении в набор новых переменных, но не увеличивается при исключении каких-либо переменных.

Пусть между показателями  $X$  и  $Y$  существует стохастическая зависимость. Предполагая эту зависимость линейной, запишем ее в следующем виде:

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$

где  $X$  – детерминированная (объясняющая) переменная;  
 $\beta_0, \beta_1$  – коэффициенты;  
 $\varepsilon$  – случайная составляющая.

Вместе с  $\varepsilon$  переменная  $Y$  также является случайной величиной. На случайную величину  $\varepsilon$  накладывается условие обращения в нуль ее математического ожидания:  $M\varepsilon = 0$ . С учетом этого условия математическое ожидание  $Y$  принимает вид:

$$M_x Y = \beta_0 + \beta_1 X.$$

Это уравнение и является теоретическим уравнением регрессии.

Таким образом, уравнение регрессии задает зависимость среднего значения случайной переменной  $Y$  от значений детерминированной переменной  $X$ . Основной задачей регрессионного анализа является построение оценок коэффициентов  $\beta_0, \beta_1$  по данным выборки и описание качества оценок.

Выборочным уравнением регрессии называется выражение

$$\hat{y} = b_0 + b_1 x, \quad (3.9)$$

где  $\hat{y}$  – выборочная оценка  $M_x Y$ ;

$b_0, b_1$  – выборочные оценки коэффициентов  $\beta_0, \beta_1$ .

Одним из основных методов построения названных оценок является метод наименьших квадратов (МНК). Обозначим через  $(x_i, y_i)$  ( $i = 1, \dots, n$ ) выборку значений объема  $n$  показателей  $X$  и  $Y$ .

Тогда, согласно предположению, между этими значениями существует зависимость:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n.$$

Соответствующая выборочная оценка теоретического уравнения регрессии имеет вид:

$$\hat{y}_i = b_0 + b_1 x_i, \quad i = 1, \mathbf{K}, n.$$

В этих соотношениях известны только значения  $x_i$ ; величины  $b_0$ ,  $b_1$  и  $\hat{y}_i$  будут вычисляться по МНК.

Согласно МНК минимизируется сумма квадратов отклонений:

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2,$$

которая рассматривается как функция от  $b_0$ ,  $b_1$ . Это соотношение включает уже и переменные  $y_i$ .

Выборочным коэффициентом регрессии  $Y$  по  $X$  называется коэффициент  $b_1$  в уравнении (3.9). Он показывает, на сколько единиц в среднем изменяется переменная  $Y$  при изменении переменной  $X$  на одну единицу.

Коэффициент регрессии  $b_1$  можно также записать в следующем виде:

$$b_1 = \frac{\text{cov}(x, y)}{s_x^2},$$

где  $\text{cov}(x, y)$  – выборочная ковариация,  $\text{cov}(x, y) = \overline{xy} - \bar{x} \cdot \bar{y}$ ;

$s_x^2$  – исправленная выборочная дисперсия переменной  $X$ ,  
 $s_x^2 = \overline{x^2} - \bar{x}^2$ .

Коэффициент  $b_0$  не имеет экономической интерпретации, но если переменная  $X$  может принимать нулевое значение,  $b_0$  является соответствующим значением переменной  $\hat{y}$ , и может интерпретироваться как значение  $Y$  при  $x = 0$ .



Уравнение регрессии (выборочное) записывается также в виде

$$\hat{y} = \bar{y} + b_1(x - \bar{x}).$$

Коэффициенты регрессии  $b_0$ ,  $b_1$  называют также МНК-оценками коэффициентов регрессии, т. е. коэффициентами, полученными по методу наименьших квадратов.

### 3.3. Коэффициент детерминации и средняя ошибка аппроксимации

Пусть  $y = b_0 + b_1 + e$  – выборочное уравнение линейной регрессии, построенной по выборке  $\{x_i, y_i \mid i = 1, \dots, n\}$ . Ошибки регрессии находят по формуле

$$e_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_i.$$

Эти ошибки используют для проверки качества уравнения регрессии. Разброс наблюдаемых значений зависимой переменной  $y$  характеризуется дисперсией  $\sigma_y^2$ . Из равенства  $y_i = \hat{y}_i + e_i$  следует, что

$$\sigma_y^2 = \sigma_{\hat{y}}^2 + \sigma_e^2,$$

где  $\sigma_y^2 = \frac{1}{n} \sum (y_i - \bar{y})^2$ ,  $\sigma_{\hat{y}}^2 = \frac{1}{n} \sum (\hat{y}_i - \bar{y})^2$ ,  $\sigma_e^2 = \frac{1}{n} \sum (e_i - \bar{e})^2 = \frac{1}{n} \sum e_i^2$  – дисперсии переменных  $y_i$ ,  $\hat{y}_i$  и  $e_i$ .

Таким образом, дисперсия  $\sigma_y^2$  состоит из двух слагаемых:  $\sigma_{\hat{y}}^2$  – слагаемое, объясненное регрессией, и  $\sigma_e^2$  – необъясненная часть вариации  $y_i$ .

Коэффициент детерминации

$$R^2 = \frac{\sigma_{\hat{y}}^2}{\sigma_y^2} = 1 - \frac{\sigma_e^2}{\sigma_y^2},$$

характеризует долю вариации зависимой переменной  $y$ , объясненную уравнением регрессии. Необъясненная часть вариации равна  $\frac{\sigma_e^2}{\sigma_y^2}$ .

Если  $R^2 = 1$ , то  $\sigma_y^2 = \sigma_{\hat{y}}^2$ ,  $\sigma_e^2 = 0$ ,  $y_i = \hat{y}_i$ , и все точки выборки лежат на прямой регрессии.

Если  $R^2 = 0$ , то  $\sigma_y^2 = \sigma_e^2$ ,  $\sigma_{\hat{y}}^2 = 0$ ,  $\hat{y}_i = \bar{y}$ , и все точки выборки лежат на горизонтальной прямой  $y = \bar{y}$ .

В этом случае изменение переменной  $x$  не вызывает изменения  $y$ , зависимость между ними отсутствует.

Чем ближе  $R^2$  к единице, тем лучше регрессия  $\hat{y}$  объясняет  $y$ . В этом смысл коэффициента детерминации, который является одной из эффективных оценок адекватности регрессионной модели.

Для парной линейной регрессии коэффициент детерминации  $R^2$  равен квадрату коэффициента корреляции  $r_{xy}^2$  переменных  $x$  и  $y$ , то есть

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

Статистическую значимость коэффициента детерминации  $R^2$  можно проверить на основе гипотезы  $H_0: \rho_{x,y}^2 = 0$ , где  $\rho_{x,y}$  – генеральный коэффициент парной корреляции. Соответствующей статистикой для проверки этой гипотезы служит  $F$ -статистика Фишера.

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

Эта статистика имеет  $F$ -распределение Фишера с параметрами  $v_1 = 1$ ,  $v_2 = n - 2$ . Критическое значение  $F_{кр}$  при заданном уровне значимости  $\alpha$  и параметрах  $v_1 = 1$ ,  $v_2 = n - 2$  находят по таблице  $F$ -распределений Фишера или в Excel с помощью функции  $F_{кр} = \text{ФРАСПОБ}(\alpha; v_1; v_2)$ .

Если наблюдаемое значение  $F < F_{кр}$ , то гипотеза  $H_0$  не отвергается, т. е.  $R^2$  незначим на заданном уровне.

Если наблюдаемое значение  $F > F_{кр}$ , то  $H_0$  отклоняется, т. е.  $R^2$  значим.

Оценку качества модели регрессии можно также проводить с помощью средней ошибки аппроксимации.

Средней ошибкой аппроксимации называется среднее процентное относительное отклонение фактических значений зависимой переменной от ее расчетных значений:

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

Модель регрессии считается качественной, если ошибка аппроксимации не превышает 8 %–10 %.

Оценкой этой модели по МНК служат уравнения:

$$\hat{y}_i = b_0 + b_1 x_i; \quad i = 1, \mathbf{K}, n,$$

где  $b_0, b_1$  находят по формулам:

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}; \quad b_0 = \frac{1}{n} \sum_{i=1}^n y_i - \frac{1}{n} b_1 \sum_{i=1}^n x_i.$$

Как видно из последних формул относительно переменных  $y_i$ , эти величины являются линейными (что является важной особенностью оценок);  $b_0$  и  $b_1$  будут рассматриваться как (точечные) оценки параметров  $\beta_0$  и  $\beta_1$  модели регрессии. Основными, «хорошими», свойствами статистических оценок являются: несмещенность, эффективность и состоятельность. Для выполнения этих свойств оценок вводятся дополнительные условия на случайные остатки  $\varepsilon_i$  и переменные  $x_i, y_i$  в модели регрессии.

Условиями Гаусса–Маркова называются следующие условия.

1. Объясняющая переменная  $x_i$  является детерминированной (неслучайной) величиной с ненулевой дисперсией, остаток  $\varepsilon_i$  является случайной величиной (вместе с переменной  $y_i$ ) ( $i = 1, \dots, n$ ).

2. Математическое ожидание случайного остатка  $\varepsilon_i$  равно нулю, т. е.  $M\varepsilon_i = 0$  ( $i = 1, \dots, n$ ).

3. Дисперсия случайного остатка  $\varepsilon_i$  постоянна для всех наблюдений, т. е.  $D(\varepsilon_i) = D(\varepsilon_i^2) = \sigma^2$  ( $i = 1, \dots, n$ ) (гомоскедастичность остатков).

4. Случайные остатки некоррелированы между собой, т. е.  $M(\varepsilon_i \varepsilon_j) = 0$  при  $i \neq j$ .

5. Случайный остаток  $\varepsilon_i$  является нормально распределенной случайной величиной (с нулевым средним и постоянной дисперсией  $\sigma^2$ ). Вместе с  $\varepsilon_i$  зависимая переменная  $y_i$  также является нормальной случайной величиной.

Классической нормальной линейной регрессионной моделью называется модель регрессии, для которой выполнены условия 1–5.

Для такой модели некоррелированность случайных остатков (или зависимых переменных) равносильна их независимости.

#### *Качество оценок модели регрессии*

Теорема Гаусса–Маркова. При выполнении условий 1–4 оценки  $b_0, b_1$  параметров  $\beta_0, \beta_1$ , полученные по методу наименьших квадратов, обладают следующими свойствами:

1) несмещенность:  $Mb_0 = \beta_0, Mb_1 = \beta_1$  (отсутствие систематического смещения оценки от самого параметра);

2) эффективность: оценки имеют наименьшую дисперсию в классе всех линейных (относительно  $Y$ ) несмещенных оценок;

3) состоятельность: при  $n \rightarrow \infty$  оценки сходятся по вероятности к самим параметрам.

Одной из важных характеристик оценок служит их дисперсия. Для оценок  $b_0$  и  $b_1$  дисперсии определяются по формулам:

$$\sigma_{b_0}^2 = \frac{\overline{x^2} \sigma^2}{ns_x^2}, \quad \sigma_{b_1}^2 = \frac{\sigma^2}{ns_x^2}, \quad (3.10)$$

где  $\sigma^2 = \sigma_e^2$  – общая дисперсия остатков;

$s_x^2$  – выборочная дисперсия  $x$ .

Выборочная остаточная дисперсия

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

служит несмещенной оценкой дисперсии остатков  $\sigma^2$ . Стандартной ошибкой регрессии называется величина

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

Заменив в формулах (3.10) дисперсию остатков  $\sigma^2$  ее выборочной оценкой  $s^2$ , получим выборочные оценки для дисперсий  $\sigma_{b_0}^2$ ,  $\sigma_{b_1}^2$ :

$$s_{b_0}^2 = \frac{\overline{x^2} s^2}{n s_x^2}, \quad s_{b_1}^2 = \frac{s^2}{n s_x^2}.$$

Стандартными ошибками оценок коэффициентов регрессии называются величины:

$$s_{b_0} = s \sqrt{\frac{\overline{x^2}}{n s_x^2}}, \quad s_{b_1} = \frac{s}{\sqrt{n s_x^2}}. \quad (3.11)$$

Качество регрессионной модели можно проверить не только на основе коэффициента детерминации  $R^2$  и ошибки аппроксимации  $A$ , но и на основе выборочного коэффициента регрессии  $b_1$ . Пусть для модели  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$  по выборке значений  $X$  и  $Y$  получен коэффициент регрессии  $b_1$ . Если выполняется условие 5 о нормальности распределения случайного остатка  $\varepsilon$ , то статистика

$$t = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{(b_1 - \beta_1) \sqrt{n s_x^2}}{s}$$

имеет  $t$ -распределение Стьюдента с  $\nu = n - 2$  степенями свободы. (При подсчете степеней свободы от объема выборки  $n$  вычитается количество связей, равное количеству уравнений в системе нормальных уравнений, т. е. 2). По таблице критических точек Стьюдента по заданному уровню значимости  $\alpha$  (или доверительной вероятности  $\gamma = 1 - \alpha$ ) и числу степеней свободы  $\nu = n - 2$  находят критическое значение  $t_{кр}$ . Критическое значение  $t_{кр}$  при заданном  $\alpha$  и  $\nu = n - 2$  определяют также в Excel с помощью функции

$t_{кр} = \text{СТЮДРАСПОБР}(\alpha, \nu)$ . Для построения доверительного интервала коэффициента  $\beta_1$  используется двусторонний критерий, основанный на статистике  $t_{кр}$ . Соответствующая доверительная область задается равенством

$$P\left(\frac{b_1 - \beta_1}{s_{\beta_1}} < t_{кр}\right) = 1 - \alpha. \quad (3.12)$$

Доверительным интервалом на уровне значимости  $\alpha$  для коэффициента  $\beta_1$  служит следующем интервал:

$$b_1 - t_{кр}s_{b_1} < \beta_1 < b_1 + t_{кр}s_{b_1}, \quad (3.13)$$

который записывается таким образом:

$$b_1 - t_{кр} \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})}} < \beta_1 < b_1 + t_{кр} \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})}}.$$

Для проверки значимости коэффициента регрессии тестируется гипотеза  $H_0: \beta_1 = 0$ . Значимость коэффициента регрессии  $\beta_1$  означает, что доверительный интервал не содержит число ноль. Это равносильно условию

$\left| \frac{b_1}{s_{b_1}} \right| > t_{кр}$ , при котором гипотеза  $H_0$  отвергается.

Таким образом, если  $\left| \frac{b_1}{s_{b_1}} \right| < t_{кр}$ , то коэффициент регрессии незна-

чим ( $\beta_1 = 0$ ) на уровне  $\alpha$ .

В этом интервале находятся статистически незначимые значения коэффициента  $\beta_1$ .

Прогнозирование на основе функции регрессии основано на формуле

$$\hat{y}_p = b_0 + b_1 x_p.$$

Поскольку выборочная функция регрессии строится по выборке и является оценкой функции регрессии, для нее можно построить

доверительный интервал. Для этого требуется определить дисперсию величины  $\hat{y}$ , которая является выборочной оценкой для  $M_x Y$ . С этой целью представим уравнение регрессии в виде

$$\hat{y} = \bar{y} + b_1(x - \bar{x}).$$

Из этого уравнения дисперсия  $\sigma_{\hat{y}}^2$  представляется в виде

$$\sigma_{\hat{y}}^2 = \sigma_y^2 + \sigma_{b_1}^2(x - \bar{x})^2. \quad (3.14)$$

Дисперсия  $\sigma_{\hat{y}}^2$  равна  $\frac{\sigma^2}{n}$ , дисперсия  $\sigma_{b_1}^2$ , по формуле (3.10), равна  $\sigma_{b_1}^2 = \frac{\sigma^2}{ns_x^2}$ . Поэтому  $\sigma_{\hat{y}}^2 = \frac{\sigma^2}{n} + \frac{\sigma^2}{ns_x^2}(x - \bar{x})^2 = \sigma^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{ns_x^2} \right)$ . Заменяя  $\sigma^2$  ее выборочной оценкой  $s^2$ , получим оценку дисперсии  $\sigma_{\hat{y}}^2$ :

$$s_{\hat{y}}^2 = s^2 \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right). \quad (3.15)$$

Доверительный интервал для функции регрессии  $M_x Y$  строится на основе статистики

$$t = \frac{\hat{y} - M_x Y}{s_{\hat{y}}}, \quad (3.16)$$

которая имеет  $t$ -распределение Стьюдента с  $\hat{y}_k = n - 2$  степенями свободы. Таким образом, доверительный интервал на уровне  $\alpha$  для функции регрессии  $M_x Y$  имеет вид:

$$\hat{y} - t_{1-\alpha, k} s_{\hat{y}} \leq M_x Y \leq \hat{y} + t_{1-\alpha, k} s_{\hat{y}}, \quad (3.17)$$

где  $t_{1-\alpha, k}$  – критическое значение  $t$ -распределения Стьюдента;

$s_{\hat{y}}$  – находят из формулы (3.15).

Индивидуальные значения зависимой переменной  $y_p$  имеют относительно линии регрессии стандартную ошибку  $s$ . Поэтому доверительный интервал для индивидуальных значений  $y_p$  имеет следующий вид:

$$y_p - t_{1-\alpha, k} s_{y_p} \leq y_p \leq y_p + t_{1-\alpha, k} s_{y_p},$$

где

$$s_p = s \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}},$$

стандартная ошибка прогноза для индивидуального значения  $y_p$ ,  $y_p = b_0 + b_1 x_p$  – прогнозное значение  $y$  для  $x = x_p$ .

Из формул для стандартных ошибок  $s_{\hat{y}}$  и  $s_{y_p}$  следует, что при  $x = \bar{x}$  они минимальны, а при увеличении отклонения  $x_p$  от  $\bar{x}$  – возрастают.

### 3.4. Использование функции регрессии в анализе и прогнозе

В экономическом анализе и прогнозировании часто используют экономический смысл коэффициента парной регрессии  $b_1$ . Несложно показать, что коэффициент парной регрессии  $b_1$  указывает среднюю величину изменения переменной  $y$  при увеличении фактора  $x$  на одну единицу. Параметр  $b_1$  имеет единицу измерения – отношение размерности зависимой переменной к размерности фактора. В представленной интерпретации прослеживается тесная зависимость коэффициента парной регрессии и среднего за наблюдаемый период коэффициента эластичности. Последний характеризует, на сколько процентов изменится результативный показатель  $y$  при изменении фактора  $x$  на 1 %, и рассчитывается по формуле

$$\eta = \frac{\Delta y}{y} \div \frac{\Delta x}{x} = \frac{\Delta y}{\Delta x} \cdot \frac{\bar{x}}{\bar{y}} = b_1 \cdot \frac{\bar{x}}{\bar{y}}.$$



При построении регрессионной модели для целей прогнозирования особое значение имеет знак коэффициента парной регрессии: знак «+» указывает на прямую зависимость  $y$  и  $x$  (с увеличением значения фактора возрастает и значение резульативного показателя), знак «-» характеризует обратную зависимость показателей. Важным представляется обоснованность описанного уравнением характера связи с точки зрения экономической теории. Если полученная зависимость сомнительна, то от данного уравнения, даже несмотря на его удовлетворительные статистические характеристики, лучше отказаться: на прогнозный период имеет смысл распространять зависимости, не противоречащие основным положениям экономической теории.

*Пример 3.1.* По данным примера 1.1 построить уравнение регрессии. Оценить качество построенной модели. Выполнить на ее основе анализ и прогноз объема продаж, если расходы на рекламу в прогнозном периоде составят 20 тыс. усл. ден. ед., прочие условия не меняются. В отчетном периоде расходы на рекламу составляли 19 тыс. усл. ден. ед., объем продаж – 58 тыс. шт.

*Решение.* Входная информация представляет собой пространственные данные, поэтому предпосылка теоремы Гаусса–Маркова относительно отсутствия автокорреляции остатков выполняется автоматически.

Построим рабочую таблицу для вычисления оценок параметров регрессии в соответствии с формулами.

Таблица

Рабочая таблица для вычисления оценок параметров уравнения регрессии

$i$	$x_i$	$y_i$	$x_i^2$	$x_i y_i$	$y_i^2$	$\hat{y}_i$	$e_i$	$e_i^2$	$(x_i - \bar{x})$
1	16,3	52,8	265,7	860,6	2787,8	51,40	1,405	1,97	0,722
2	16,8	48,4	282,2	813,1	2342,6	52,48	-4,077	16,62	0,122
3	18,5	54,2	342,3	1002,7	2937,6	56,15	-1,952	3,812	1,823
4	16,3	50	265,7	815,0	2500,0	51,40	-1,395	1,95	0,722
5	17,9	54,9	320,41	982,7	3014,01	54,86	0,045	0,002	0,563
6	17,4	53,9	302,8	937,9	2905,21	53,77	0,126	0,02	0,063
7	16,1	53,1	259,2	854,9	2819,61	50,96	2,137	4,57	1,102

$i$	$x_i$	$y_i$	$x_i^2$	$x_i y_i$	$y_i^2$	$\hat{y}_i$	$e_i$	$e_i^2$	$(x_i - \bar{x})$
8	16,2	52,4	262,4	848,9	2745,76	51,18	1,221	1,49	0,902
9	17	53	289,0	901,0	2809,00	52,91	0,091	0,008	0,022
10	16,7	52,9	278,9	883,4	2798,41	52,26	0,640	0,41	0,202
11	17,5	53,1	306,3	929,3	2819,61	53,99	-0,890	0,79	0,123
12	19,1	60,1	364,8	1147,9	3612,01	57,45	2,650	7,02	3,803
Сумма	205,8	638,8	3539,6	10977	34091,7	638,8	0,00	38,66	10,17
Среднее	17,2	53,23	294,9	914,8	2840,9	53,23	0,00	3,22	0,848

Получим:  $b_1 = \frac{914,78 - 17,15 \cdot 53,23}{294,97 - 17,15^2} = 2,16,$

$$b_0 = 53,23 - 2,16 \cdot 17,15 = 16,15.$$

Запишем уравнение регрессии:

$$\hat{y} = 16,15 + 2,16x.$$

*Статистические характеристики уравнения регрессии.* Коэффициент корреляции рассчитан по формуле (3.4).

$$r_{yx} = \frac{914,78 - 17,15 \cdot 53,23}{\sqrt{294,97 - 17,15^2} \cdot \sqrt{2840,97 - 53,23^2}} = 0,74.$$

Для оценки качества подбора линейной функции рассчитывается коэффициент детерминации, который характеризует долю дисперсии результативного показателя  $y$ , объясняемую регрессией, в общей дисперсии результативного показателя. В соответствии с формулой (3.4):

$$R^2 = r_{yx}^2 = 0,74^2 = 0,55.$$

Это означает, что объем продаж на 55 % объясняется расходами на рекламу в представленной форме связи, что говорит об удовлетворительном качестве модели.

Проверим статистическую значимость параметров уравнения регрессии.

В соответствии с формулами (3.10), (3.11) рассчитаем величины стандартных ошибок для примера, которые, соответственно, составили:

$$s = \sqrt{\frac{38,659}{12-2}} = 1,966,$$

$$s_{b_1} = \frac{s}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} = \frac{1,966}{\sqrt{10,17}} = 0,617,$$

$$s_{b_0} = \frac{s \cdot \sqrt{x^2}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} = s_{b_1} \cdot \sqrt{x^2} = \frac{1,966 \cdot \sqrt{294,97}}{\sqrt{10,17}} = 10,589.$$

Величина стандартных ошибок совместно с  $t$ -распределением Стьюдента при  $n - 2 = 10$  степенях свободы применяется для проверки статистической значимости коэффициента регрессии. Проверим, принимается ли нулевая гипотеза о равенстве нулю истинного значения параметров построенного уравнения регрессии  $H_0 : \beta_1(\beta_0) = 0$ . Расчетные значения  $t$ -статистик Стьюдента:

$$t_{b_1} = \frac{b_1}{s_{b_1}} = \frac{2,16}{0,617} = 3,5;$$

$$t_{b_0} = \frac{b_0}{s_{b_0}} = \frac{16,15}{10,589} = 1,5.$$

При  $\alpha = 0,05$  (для двустороннего критерия) и числе степеней свободы 10 табличное значение  $t_{\frac{\alpha}{2}, n-2} = t_{0,025; 10} = 2,23$ . Так как

$|t_{b_1}| > 2,23$ , то гипотезу о несущественности коэффициента регрессии  $b_1$  можно отклонить, в то время как  $|t_{b_0}| < 2,23$ , что

не позволяет принять альтернативную гипотезу о существенном отличии от нуля истинного значения свободного коэффициента уравнения регрессии.

При оценке значимости коэффициента регрессии на начальном этапе можно использовать «грубое» правило, позволяющее не прибегать к таблицам: «Если  $|t_b| > 3$  или если отношение стандартной ошибки параметра и самого параметра меньше 30 %:  $\frac{s_b}{b} < 0,3$ , то это свидетельствует о статистической значимости параметра регрессии  $b$ . Для числа наблюдений больше 10 это правило практически всегда работает».

Уравнение регрессии с набором диагностических тестов удобно записывать в следующем виде:

$$y = 16,15 + 2,16x, R^2 = 0,55, s = 1,966.$$

с.о.(10,6) (0,6)  
 $t_{ст}$  (1,5) (3,5)

Параметр регрессии  $b_1 = 2,16$  показывает, что при увеличении расходов на рекламу на 1 тыс. усл. ден. ед. объем продаж в среднем возрастает на 2,16 тыс. шт., и, как указывает  $t$ -статистика параметра регрессии, это влияние на 5%-м уровне является статистически значимым. Отсюда можно сделать вывод, что введенный в модель фактор (расходы на рекламу) является статистически значимым.

В целом, в части анализа получили следующий вывод. Статистическая значимость параметра регрессии и коэффициент детерминации указывают на значимость влияния фактора на результат. Однако невысокий коэффициент детерминации указывает на необходимость дальнейшего изменения модели в направлении повышения коэффициента детерминации. В качестве этапов дальнейшего улучшения модели можно рассматривать: изменение формы связи модели, введение дополнительного фактора, оказывающего влияние на объем продаж.

Учитывая невысокий коэффициент детерминации, использование для прогноза построенной регрессионной модели нецелесообразно.

Самый простой способ прогноза основан на использовании уравнения регрессии и предполагает подстановку значений фактора на прогнозный период в уравнение регрессии.

$$y = 16,15 + 2,16 \cdot 20 = 59,4 \text{ (тыс. шт.)}.$$

Получен точечный прогноз. Для расчета интервального прогноза найдем ошибку прогноза:

$$s_{\%} = s \sqrt{1 + \frac{1}{n} + \frac{(x_{\text{пр}} - \bar{x})^2}{ns_x^2}} = 1,966 \cdot \sqrt{1 + \frac{1}{12} + \frac{(20-17)^2}{12 \cdot 0,848}} = 2,697.$$

Тогда предельная ошибка прогноза  $\Delta_{\%} = t_{\text{кр}} \cdot s_{\%} = 2,23 \cdot 2,697 = 6,01$ .

Доверительный интервал прогноза  $\gamma_{\%} = 59,4 \pm 6$ .

Если условия производства на предприятии не изменятся, то при объеме инвестиций 20 тыс. усл. ден. ед. объем производства с вероятностью 95 % будет находиться в пределах от 53,4 до 65,4 тыс. шт.

При получении прогноза была использована постоянная регрессии, уровень статистической значимости которой неудовлетворительный. Поэтому нужен способ прогнозирования, основанный не на постоянной регрессии, а сориентированный на экономический смысл параметра регрессии.

В другом способе используют экономический смысл параметра регрессии. Исходя из отчетных данных предприятия, можно найти ожидаемый прирост инвестиций в прогнозном периоде:  $\Delta x = 20 - 19 = 1$  (тыс. усл. ден. ед.). Ориентируясь на экономический смысл параметра регрессии: изменение объема производства при изменении инвестиции на одну единицу, можно найти прирост объема производства в прогнозном периоде:  $\Delta y = 2,16 \cdot 1 = 2,16$  (тыс. шт.). Поскольку в отчетном периоде объем производства составлял 58 тыс. шт., то в прогнозном периоде объем производства составит  $y = 58 + 2,16 = 60,16$  (тыс. шт.).

Следующий способ – использование коэффициента эластичности. Рассчитаем коэффициент эластичности по формуле (3.9):

$$\epsilon = 2,16 \cdot \frac{17,15}{53,23} = 0,7.$$

Ожидаемый темп прироста инвестиций в прогнозном периоде составит

$$\frac{\Delta x}{x} = \frac{20 - 19}{19} \cdot 100 \% = 5,3 \%$$

Используя коэффициент эластичности, рассчитаем ожидаемый темп прироста объема производства:

$$\frac{\Delta y}{y} = 0,7 \cdot 5,3\% = 3,7\%$$

Рассчитаем ожидаемый объем производства:

$$y = 58 \cdot 1,037 = 60,15 \text{ (тыс. шт.)}$$

Выбор способа прогнозирования в каждом конкретном случае определяется пользователем, исходя из его опыта прогнозирования. Если конкретные соображения по поводу способа прогнозирования отсутствуют, то можно взять среднюю по всем способам прогнозирования. Выбираем второй или третий способ прогнозирования в силу статистической незначимости параметра  $b_0$ , тем более что прогнозные значения по последним вариантам практически совпадают.

В практике прогнозирования наиболее часто предпочтение отдается линейной регрессии по причине экономической интерпретируемости параметров регрессии и возможности экспертной корректировки параметра при изменении в прогнозном периоде условий экономической конъюнктуры. Экономической интерпретации поддаются параметры некоторых нелинейных форм связи.

### **Вопросы для контроля знаний**

1. Что такое «уравнение регрессии»? В чем различие между уравнением регрессии и выборочным уравнением регрессии?
2. Что понимается под спецификацией модели?
3. Опишите алгоритм метода наименьших квадратов для оценки параметров парной и множественной регрессии.
4. Каковы предпосылки МНК?

5. В каких случаях МНК-оценки параметров регрессии эффективные, состоятельные и несмещенные?

6. Каков экономический смысл параметров модели парной линейной регрессии?

7. Как рассчитать коэффициент парной линейной корреляции? Каков смысл коэффициента парной линейной корреляции? пределы изменения?

## Практическое занятие № 2

### Оценка тесноты и значимости стохастической зависимости между экономическими показателями

**Цель работы:** отработать практические навыки определения величины и значимости стохастической зависимости между экономическими показателями.

#### Методические рекомендации

1. Для дальнейшего использования генерируем две выборки гипотетических показателей  $X$  и  $Y$ .

2. Организуем работу в Excel следующим образом. В ячейки A1:D1 вводим метки  $n$ ,  $e$ ,  $X$  и  $Y$ . В столбец A2:A41 вводим номера показателей:  $n_i = 1, 2, 3, \dots, 40$ .

3. В столбец B2:B41 генерируем случайные числа  $e$  по правилу: в меню **Сервис/Данные/Анализ данных** (при отсутствии надстройки **Анализ данных** подключаем ее). Генерация случайных чисел для примера представлена в табл. 1, столбец  $X$ .

4. Стандартное отклонение  $s$  выбираем равным 0,02.

5. В диалоговом окне выбираем **Генерация случайных чисел** и заполняем новое диалоговое окно: число переменных – 1. Число случайных чисел – 40, распределение – нормальное, со средним – 0, стандартным отклонением  $s = 0,02$ ; случайное рассеивание не задается, выходной интервал – B2, ОК. Столбец B2:B41 заполняется случайными числами  $e \in N(0;0,02)$ .

6. В столбец C2:C41 по формуле «= 40 + 0,1\*A2» вводим  $X$ .

7. В столбец D2:D41 по формуле «= 10 + 0,11\*C2 + B2» вводим  $Y$ . Полученные результаты представлены в табл. 1.

## Генерация данных

$n$	$e$	$X$	$Y$	$n$	$e$	$X$	$Y$
1	-0,022 54	40,1	14,39	21	-0,146 19	42,1	14,48
2	0,065 111	40,2	14,49	22	0,034 41	42,2	14,68
3	0,022 008	40,3	14,46	23	-0,025 28	42,3	14,63
4	-0,046 96	40,4	14,40	24	-0,047 47	42,4	14,62
5	0,037 738	40,5	14,49	25	0,085 66	42,5	14,76
6	0,106 577	40,6	14,57	26	0,016 545	42,6	14,70
7	0,089 078	40,7	14,57	27	-0,075 33	42,7	14,62
8	0,013 438	40,8	14,50	28	-0,000 45	42,8	14,71
9	0,071 461	40,9	14,57	29	0,010 386	42,9	14,73
10	0,054 353	41,0	14,56	30	-0,0095	43,0	14,72
11	-0,049 48	41,1	14,47	31	0,014 056	43,1	14,76
12	-0,1051	41,2	14,43	32	-0,029 66	43,2	14,72
13	-0,024 81	41,3	14,52	33	0,032 183	43,3	14,80
14	-0,065 89	41,4	14,49	34	-0,085 27	43,4	14,69
15	0,056 691	41,5	14,62	35	-0,049 65	43,5	14,74
16	0,001 168	41,6	14,58	36	0,063 618	43,6	14,86
17	0,033275	41,7	14,62	37	0,162 823	43,7	14,97
18	-0,022 92	41,8	14,58	38	-0,04534	43,8	14,77
19	-0,040 79	41,9	14,57	37	-0,037 17	43,9	14,79
20	-0,077 047	42,0	14,58	40	-0,0141	44,0	14,83

8. Для вычисления коэффициента корреляции  $r_{xy}$  используем **Анализ данных/Корреляция**. В диалоговом окне **Корреляция** задается входной интервал столбцов C и D с метками X и Y, группировка – по столбцам, метки в первой строке (ставим «галочку»), выходной интервал – E1 (указываем верхний угол интервала). В ячейке F3 появится значение  $r_{xy} = 0,954$  (для данных приведенного примера), которое свидетельствует о достаточно тесной положительной линейной связи между X и Y.

9. Проверяем значимость коэффициента корреляции. Для тестирования нулевой гипотезы  $H_0$  об отсутствии линейной связи между X и Y используется  $t$ -статистика Стьюдента. Расчетное значение критерия  $t_p$  находим по формуле (3.5), где число наблюдений  $n = 40$ ,



$r_{xy} = 0,954$ . Реализуем формулу в ячейке F5 и получаем значение  $t_p = 19,6$ .

10. Значение  $t_{кр}$  определяется по таблице Стьюдента (прилож. 1). Число степеней свободы

$$v = n - m - 1,$$

где  $n$  – число наблюдений,  $n = 40$ ;

$m = 1$  – количество  $X$ .

В данном случае  $v = 38$ , следовательно  $t_{кр} = 2,024$ .

11. Значение  $t_{кр}$  можно найти другим способом: рассчитаем его в ячейке F7 по формуле «=СТЮД.ОБР.2X(0,05;38).

12. Поскольку  $|t_p| > t_{кр}$ , гипотеза  $H_0$  отклоняется. Это свидетельствует о значимости  $r_{xy}$  и наличии линейной связи.

### Задания для самостоятельной работы

Генерируем вариант данных для парной регрессии по приведенным методическим рекомендациям.

Значение стандартного отклонения  $s$  выбираем по табл. 2.

Таблица 2

Значение стандартного отклонения

№ варианта	$s$	№ варианта	$s$
1–3	0,12	7–9	0,03
4–6	0,06	10–12	0,02

Количество наблюдений – 50.

Выполнить расчет в Excel. Проанализировать полученные результаты. Оформить отчет.

## Практическое занятие № 3 Использование регрессионных моделей в анализе и прогнозе

**Цель работы:** отработать практические навыки построения уравнения парной линейной регрессии.

## Методические рекомендации

1. По данным табл. 1 построить уравнение регрессии. Оценить качество построенной модели. Выполнить на ее основе анализ и прогноз настрига шерсти, если живая масса овцы составляет 57 кг, прочие условия не меняются.

Таблица 1

Исходные данные

№ наблюдения	Живая масса, кг $x$	Настриг шерсти, кг $y$	№ наблюдения	Живая масса, кг $x$	Настриг шерсти, кг $y$
1	50,2	5,3	11	52,9	6,3
2	50,3	6,2	12	53	6,7
3	50,4	5	13	53,2	7,8
4	51	5,7	14	53,5	6,8
5	51,2	6,3	15	53,6	7,2
6	51,4	5,4	16	54,3	7,7
7	51,6	6,8	17	54,7	8,5
8	51,9	7,3	18	55,2	7,8
9	52,3	6,6	19	55,5	8,4
10	52,4	6,2	20	56,5	8,8

2. Ввести данные из табл. 1, проверить подключение надстройки **Анализ данных** в Excel.

3. Уравнение регрессии на компьютере можно построить двумя способами. Первый способ – графический. Выделите курсором анализируемые ряды (вместе с названиями) и вызовите режим **Вставка**, тип диаграммы – *точечная*.

Подведя курсор к любой точке графика, щелкните правой кнопкой мыши – отобразится меню, в котором надо выбрать пункт **Добавить линию тренда**. В качестве параметров линии тренда выбираем линейную зависимость, и ставим «галочку» на пункте **Показывать уравнение на диаграмме**.

4. Для получения значения коэффициента детерминации надо поставить «галочку» на пункте **Поместить на диаграмму величину достоверности аппроксимации**.

5. На экране будет представлено уравнение регрессии:

$$y = -21,37 + 0,5347x, \text{ где } b_0 = -21,37; b_1 = 0,5347.$$

6. Второй способ построения регрессионной модели – использование специального инструмента **Анализ данных/Регрессия**.

7. В диалоговом окне инструмента **Регрессия** необходимо заполнить следующие данные: Входной интервал  $Y$  – диапазон, содержащий данные резульативного показателя; Входной интервал  $X$  – диапазон, содержащий данные факторов. В предъявленном меню указываем входные интервалы  $Y$  и  $X$  (вместе с названиями), ставим «галочку» на пункте **Метки (Метки – это названия)**. В качестве выходного интервала можно указать как новый рабочий лист, так зрительно удобную ячейку на рабочем листе (рис. 1).

	A	B	C	D	E	F	G	H	I
	№ наблюдения	Живая масса, кг $X$	Настриг шерсти, кг $Y$						
1	1	50,2	5,3						
2	2	50,3	6,2						
3	3	50,4	5						
4	4	51	5,7						
5	5	51,2	6,3						
6	6	51,4	5,4						
7	7	51,6	6,8						
8	8	51,9	7,3						
9	9	52,3	6,6						
10	10	52,4	6,2						
11	11	52,9	6,3						
12	12	53	6,7						
13	13	53,2	7,8						
14	14	53,5	6,8						
15	15	53,6	7,2						
16	16	54,3	7,7						
17	17	54,7	8,5						
18	18	55,2	7,8						
19	19	55,5	8,4						
20	20	56,5	8,8						

Регрессия

Входные данные

Входной интервал Y:

Входной интервал X:

Метки  Константа - ноль

Уровень надежности:  %

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Остатки

Остатки  График остатков

Стандартизированные остатки  График подбора

Нормальная вероятность

График нормальной вероятности

OK Отмена Справка

Рис. 1. Диалоговое окно ввода параметров инструмента **Регрессия**

8. На экран выводится стандартизированный вывод итогов. В окне **Множественный R** выводится коэффициент линейной

корреляции,  $b_0 = -21,369$  и  $b_1 = 0,535$  указаны в столбце **Коэффициенты** (рис. 2).

Вывод итогов			
<i>Регрессионная статистика</i>			
Множественный R	0,894063795		
R-квадрат	0,79935007		
Нормированный R-квадрат	0,788202852		
Стандартная ошибка	0,501567965		
Наблюдения	20		
<i>Дисперсионный анализ</i>			
	<i>df</i>	<i>SS</i>	<i>MS</i>
Регрессия	1	18,03973238	18,03973238
Остаток	18	4,52826762	0,251570423
Итого	19	22,568	
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>
Y-пересечение	-21,36980797	3,333196059	-6,41120642
живая масса, кг X	0,534732404	0,063146785	8,468085902

Рис. 2. Протокол регрессионного анализа

9. В протоколе регрессионного анализа (см. рис. 2) находим коэффициент детерминации  $R^2$  (*R*-квадрат), стандартную ошибку уравнения регрессии, параметры регрессии, их стандартные ошибки и *t*-статистики.

Рассчитаем статистические характеристики уравнения регрессии. Коэффициент корреляции  $r_{xy}$  (обозначен в протоколе как **Множественный R**) составляет 0,894, что по шкале Чеддока (прилож. 2) свидетельствует о высокой связи результативного признака с факторным показателем.

10. Коэффициент детерминации  $R^2$  составит 0,799. Это означает, что настриг шерсти на 79,9 % объясняется живой массой овцы в представленной форме связи, что свидетельствует об адекватности модели.

11. Параметр регрессии  $b_1 = 0,535$  показывает, что при увеличении живой массы овцы на 1 кг настриг шерсти в среднем возрастает на 0,535 кг.

12. Учитывая высокий коэффициент детерминации, использование для прогноза построенной регрессионной модели целесообразно. Самый простой способ прогноза предполагает подстановку значений фактора на прогнозный период в уравнение регрессии. Получен точечный прогноз:

$$\hat{y} = -21,369 + 0,535x = -21,369 + 0,535 \cdot 57 = 9,226 \text{ кг.}$$

13. Для расчета интервального прогноза выполняем следующие действия.

14. Определяем  $\bar{x} = 52,755$  (функция СРЗНАЧ).

15. Определяем выборочную дисперсию  $s_x^2 = 3,321$  (функция ДИСП.В).

16. Стандартную ошибку регрессии берем из протокола регрессии:  $s = 0,502$ .

17. Рассчитываем ошибку прогноза:

$$s_{\hat{y}} = s \sqrt{1 + \frac{1}{n} + \frac{(x_{\text{пр}} - \bar{x})^2}{20s_x^2}} = 0,502 \sqrt{1 + \frac{1}{20} + \frac{(57 - 52,75)^2}{20 \cdot 3,321}} = 0,577.$$

18. По формуле «СТЮД.ОБР.2Х(0,05;18)» или по таблице Стьюдента рассчитываем значение  $t_{\text{кр}} = 2,1$ .

19. Тогда предельная ошибка прогноза  $\Delta_{\hat{y}} = t_{\text{кр}} s_{\hat{y}} = 2,1 \cdot 0,577 = 1,211$ .

20. Доверительный интервал прогноза представляет собой отклонение точечного прогноза на величину предельной ошибки прогноза:  $9,226 \pm 1,211$ .

Если условия выращивания овец не изменятся, то овца массой 57 кг с вероятностью 95 % будет давать от 8,015 до 10,437 кг шерсти.

### Задания для самостоятельной работы

Используя данные индивидуального задания практического занятия № 2, построить уравнение регрессии графическим способом и с помощью инструмента **Регрессия**.

Проанализировать полученные результаты. Оформить отчет.

## Тема 4. ПОСТРОЕНИЕ И ИСПОЛЬЗОВАНИЕ МОДЕЛИ МНОЖЕСТВЕННОЙ ЛИНЕЙНОЙ РЕГРЕССИИ ПРИ АНАЛИЗЕ И ПРОГНОЗЕ

### Вопросы

4.1. Матричная запись модели множественной линейной регрессии.

4.2. Оценка параметров множественной линейной регрессии.

4.3. Оценка дисперсии случайных остатков и надежности результатов множественной регрессии.

4.4. Использование множественной регрессии в анализе и прогнозе и их построение в пакете прикладных программ.

### 4.1. Матричная запись модели множественной линейной регрессии

Модель множественной регрессии является естественным обобщением парной регрессии. Основное внимание будет уделено линейной регрессии. Множественная регрессия служит для изучения влияния многих факторов  $X_1, X_2, \dots, X_p$  на результирующий показатель  $X$ . Предполагая зависимость линейной, модель множественной регрессии можно записать в виде:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon. \quad (4.1)$$

Как и для парной модели, наличие случайной составляющей  $\varepsilon$  объясняется следующими причинами:

- а) неточностью спецификации модели;
- б) наличием неучтенных факторов;
- в) погрешностями измерений.

Для построения модели будем считать, что имеется выборка значений показателей  $Y, X_1, X_2, \dots, X_p$ :

$$\left\{ (y_i, x_{i1}, x_{i2}, \mathbf{K}, x_{ip}), i = 1, \mathbf{K}, n \right\}. \quad (4.2)$$

Тогда модель примет вид:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i. \quad (4.3)$$

Соотношение (4.3) запишем в матричном виде для удобства вычислений с применением матричных операций, для чего введем следующие матрицы:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} - \text{вектор-столбец значений зависимой переменной } Y ,$$

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} - (n \cdot (p+1)) - \text{матрица, } (j+1)\text{-й столбец}$$

которой содержит значения  $j$ -го фактора  $X_j (j=1, \mathbf{K}, p)$ ,

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_n \end{pmatrix} - \text{вектор-столбец параметров,}$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix} - \text{вектор-столбец случайных остатков (возмущений).}$$

По правилу матричного умножения (строка на столбец) соотношение (4.3) запишем в виде:

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix},$$

или

$$Y = X\beta + \varepsilon. \quad (4.4)$$

Формула (4.1) задает стохастическую зависимость  $(x_1, x_2, \dots, x_p)$  **a**  $Y(x_1, x_2, \dots, x_p)$ , формула (4.3) – наблюдаемую выборку значений этой зависимости. Для случайной величины  $Y(x_1, x_2, \dots, x_p)$  определяется условное математическое ожидание  $MY(x_1, x_2, \dots, x_p)$ .

Функция множественной линейной регрессии  $(x_1, x_2, \dots, x_p)$  **a**

**a**  $MY(x_1, x_2, \dots, x_p)$  записывается в виде:

$$MY(x_1, x_2, \dots, x_p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p. \quad (4.5)$$

Оценка этой функции регрессии по выборке (4.2) записывается в виде:

$$\hat{y}_i = b_0 + b_1 x_{i1} + \mathbf{K} + b_p x_{ip}, i = 1, \mathbf{K}, n, \quad (4.6)$$

где  $\hat{y}_i$  – выборочная групповая средняя, которая является оценкой групповой средней  $MY(x_{i1}, x_{i2}, \dots, x_{ip})$ , ( $i = 1, \dots, n$ );

$$b = \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_p \end{pmatrix} \text{ – вектор оценок;}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_n \end{pmatrix} \text{ – вектор параметров.}$$

Соотношение (4.6) запишем в матричном виде:

$$-\hat{Y} = Xb, \quad (4.7)$$

где  $\hat{Y} = \begin{pmatrix} \hat{y}_1 \\ \mathbf{K} \\ \hat{y}_n \end{pmatrix}$ .

Выборочным уравнением множественной линейной регрессии называется уравнение (4.6), а также его матричная запись (формула (4.7)).

Построение вектора оценок  $b$  будет проводиться методом наименьших квадратов.

## 4.2. Оценка параметров множественной линейной регрессии

Пусть для модели (формула (4.4)) множественной линейной регрессии  $Y = X\beta + \varepsilon$  с выборочной функцией регрессии (формула (4.7))  $\hat{Y} = X\beta$  имеется выборка наблюдений (формула (4.2)).



Рассмотрим ошибки регрессии:

$$e_i = y_i - \hat{y}_i, i = 1, \mathbf{K}, n. \quad (4.8)$$

Вектор ошибок регрессии записывается в матричном виде:

$$e = Y - \hat{Y} = Y - Xb.$$

Как и в случае парной регрессии, согласно методу наименьших квадратов для построения вектора оценок  $b$  минимизируется сумма квадратов ошибок регрессии:

$$S = \sum_{i=1}^n e_i^2 \rightarrow \min.$$

Запишем эту сумму в матричном виде:

$$S = (e_1, e_2, \dots, e_n) \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix} = e^T e,$$

где  $e^T$  – транспонирование матрицы-столбца  $e$ .

Тогда

$$S = e^T e = (Y - Xb)^T (Y - Xb) = Y^T Y - b^T X^T Y - Y^T Xb + b^T X^T Xb.$$

Поскольку матрицы  $b^T X^T Y$  и  $Y^T Xb$  имеют порядок  $1 \times 1$  и взаимно транспонированы ( $b^T X^T Y$ )<sup>T</sup> =  $Y^T Xb$ , то они совпадают. Следовательно:

$$S = Y^T Y - 2b^T X^T Y + b^T X^T Xb,$$

т. е. сумма квадратов ошибок как функция от переменных ( $b_0$ ,

$b_1, \dots, b_p$ ) состоит из трех слагаемых: первое слагаемое  $Y^T Y = \sum_{i=1}^n y_i^2$

не зависит от  $b$ , второе слагаемое  $2b^T X^T Y$  является линейной формой от переменных  $b^T$  с матрицей-столбцом коэффициентов  $2X^T Y$ , третье слагаемое  $b^T X^T Xb$  является квадратичной формой от переменных  $b^T$  с симметрической матрицей  $X^T X = (X^T X)^T$ .

Введем обозначения для элементов матриц:

$$X^T Y = \begin{pmatrix} c_0 \\ c_1 \\ \dots \\ c_p \end{pmatrix}, X^T X = \begin{pmatrix} a_{00} & a_{01} & \dots & a_{0p} \\ a_{10} & a_{11} & \dots & a_{1p} \\ \dots & \dots & \dots & \dots \\ a_{p0} & a_{p1} & \dots & a_{pp} \end{pmatrix}. \quad (4.9)$$

Тогда

$$S = \sum_{i=1}^n y_i^2 - 2 \sum_{m=0}^p c_m b_m + \sum_{m=0}^p \sum_{l=0}^p a_{ml} b_m b_l.$$

Необходимыми (и достаточными) условиями минимума функции  $S$  являются равенства:

$$\frac{\partial S}{\partial b_0} = 0, \frac{\partial S}{\partial b_1} = 0, \dots, \frac{\partial S}{\partial b_p} = 0,$$

которые можно записать в виде равенства нулю вектора-столбца:

$$\frac{\partial S}{\partial \mathbf{b}} = \begin{pmatrix} \frac{\partial S}{\partial b_0} \\ \dots \\ \frac{\partial S}{\partial b_p} \end{pmatrix} = 0.$$

Для вычисления частных производных перепишем  $S$ :

$$S = \sum_{i=1}^n y_i^2 - 2 \sum_{m=0}^p c_m b_m + \sum_{m=0}^p a_{mm} b_m^2 + \sum_{m=0}^p \sum_{l=0(l \neq m)}^p a_{ml} b_m b_l.$$

Теперь найдем частные производные:

$$k = 0, 1, \dots, p. \quad \frac{\partial S}{\partial b_k} = -2c_k + 2a_{kk} b_k + \sum_{l=0(l \neq k)}^p a_{kl} b_l + \sum_{m=0(m \neq k)}^p a_{mk} b_m$$

(два последних слагаемых получены при  $m = k$  и  $l = k$  соответственно).

Заменяя в последнем слагаемом индекс суммирования  $m$  на  $l$  и учитывая, что  $a_{kl} = a_{lk}$ , получим:

$$\frac{\partial S}{\partial b_k} = -2c_k + 2a_{kk}b_k + 2 \sum_{l=0(l \neq k)}^p a_{kl}b_l = -2c_k + 2 \sum_{l=0}^p a_{kl}b_l, k = 0, 1, \dots, p.$$

Последние равенства в матричном виде с учетом обозначений (формула (4.9)) имеют вид:

$$\frac{\partial S}{\partial b} = -2X^T Y + 2X^T X b.$$

Таким образом, условия минимума суммы квадратов остатков регрессии имеют вид:

$$-2X^T Y + 2X^T X b = 0 \text{ или } X^T X b = X^T Y. \quad (4.10)$$

Системой нормальных уравнений для нахождения вектора оценок  $b$  называется система (4.10).

Если матрица  $X^T X$  – невырожденная, система нормальных уравнений (4.10) имеет единственное решение:

$$b = (X^T X)^{-1} X^T Y,$$

которое получается умножением обеих частей равенства (4.10) на матрицу  $(X^T X)^{-1}$ .

Наряду с вектором параметров  $\beta$  в уравнение регрессии входит параметр  $\sigma^2$  – общая дисперсия случайных остатков. Для ее построения рассматривается вектор ошибок. Статистика  $s^2 = \frac{1}{n-p-1} \sum_{i=1}^n e_i^2$  является несмещенной оценкой общей дисперсии  $\sigma^2$  случайных остатков  $\varepsilon_i, i = 1, \dots, n$ .

### 4.3. Оценка дисперсии случайных остатков и надежности результатов множественной регрессии

Практическая значимость уравнения множественной регрессии оценивается с помощью показателей множественной корреляции и детерминации. Коэффициент множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком, т. е. оценивает тесноту совместного влияния факторов на результат.

Независимо от формы связи коэффициент множественной корреляции может быть найден по формуле

$$R = \sqrt{1 - \frac{\sigma_e^2}{\sigma_y^2}}, \quad (4.11)$$

где  $\sigma_e^2$  – остаточная дисперсия,  $\sigma_e^2 = \frac{\sum_{i=1}^n e_i^2}{n}$ ;  
 $\sigma_y^2$  – общая дисперсия фактора  $y$ .

Коэффициент множественной корреляции изменяется в пределах от 0 до 1. Чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов. Иногда показателям тесноты связи можно дать качественную оценку по шкале Чеддока (прилож. 2).

Качество модели множественной регрессии оценивают при помощи коэффициента детерминации, введенного в предыдущей теме.

Для оценки значимости уравнения множественной регрессии в целом рассчитывают  $F$ -критерий Фишера:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m},$$

где  $R^2$  – коэффициент множественной детерминации;

$n$  – количество наблюдений;

$m$  – число параметров при переменных  $x$  (в линейной регрессии совпадает с числом включенных в модель факторов).

Нулевую гипотезу о статистической незначимости уравнения регрессии (о равенстве нулю истинного значения коэффициента детерминации) отклоняют, если  $F > F_{кр}$  при уровне надежности  $\alpha = 0,05$  и степенях свободы  $Df_1 = m$ ,  $df_2 = n - m - 1$ . Для поиска критического значения  $F$ -статистики в Excel используют статистическую функцию  $F_{кр} = \text{ФРАСПОБ}(\alpha; df_1; df_2)$ .

#### 4.4. Использование множественной регрессии в анализе и прогнозе

Множественная регрессия охватывает совокупное одновременное влияние факторов  $x_1, x_2, \dots, x_m$ , поэтому коэффициенты регрессии  $b_1, b_2, \dots, b_m$  указывают соответствующие усредненные частные влияния переменных  $x_1, x_2, \dots, x_m$  в предположении, что остальные переменные сохраняются на постоянном уровне. Так,  $b_1$  указывает среднюю величину изменения  $y$  при увеличении  $x_1$  на одну единицу при условии, что другие переменные остаются без изменения;  $b_2$  показывает, на сколько единиц в среднем изменится  $y$ , если бы переменная  $x_2$  увеличилась на единицу при условии, что переменные  $x_i$  ( $i \neq 2$ ) остались бы без изменения, и т. д. С точки зрения статистической методологии, нет различия между множественной и парной регрессиями. Такая содержательная интерпретация коэффициентов регрессии могла бы привести к ошибочному заключению, что достаточно определить несколько простых линейных регрессий переменной  $y$  по отдельным переменным  $x_i$ . Однако следует иметь в виду, что коэффициент множественной регрессии  $b_i$  исключает влияние независимых переменных, включенных в регрессию, в то время как при простой линейной регрессии влияние прочих факторов частично отражается в коэффициенте регрессии. Следовательно, если располагать достаточной информацией и эмпирическим числовым материалом по нескольким причинам-факторам для переменной  $y$ , то целесообразно и теоретически обоснованно строить множественную регрессию.

*Пример 4.1* (расширение примера 1.1). Изучается зависимость объема продаж  $y$  от расходов на рекламу  $x_1$  и цены на товары-заменители  $x_2$ . Были собраны данные по 12 предприятиям (таблица).

Требуется построить уравнение множественной регрессии, оценить его качество, сделать экономический анализ на основе модели и выполнить интервальный прогноз объема продаж, если расходы на рекламу в прогнозном периоде составят 20 тыс. усл. ден. ед., цена на товары-заменители – 100,5 усл. ден. ед.

Исходные данные

Предприятие	Расходы на рекламу (тыс. усл. ден. ед.)	Цена на товары-заменители (усл. ден. ед.)	Объем продаж (тыс. шт.)
$i$	$x_1$	$x_2$	$y$
1	16,3	99,5	52,8
2	16,8	98,9	48,4
3	18,5	99,2	54,2
4	16,3	99,3	50,0
5	17,9	99,8	54,9
6	17,4	99,6	53,9
7	16,1	99,8	53,1
8	16,2	99,7	52,4
9	17,0	99,8	53,0
10	16,7	99,9	52,9
11	17,5	100,0	53,1
12	19,1	100,2	60,1

*Решение.* Построим модель с помощью табличного процессора Excel.

В случае множественной регрессии требуется проверка мультиколлинеарности факторов. В этом случае в главном меню выберем **Сервис/Анализ данных/Корреляция**. Заполним диалоговое окно ввода данных (рис. 4.1).

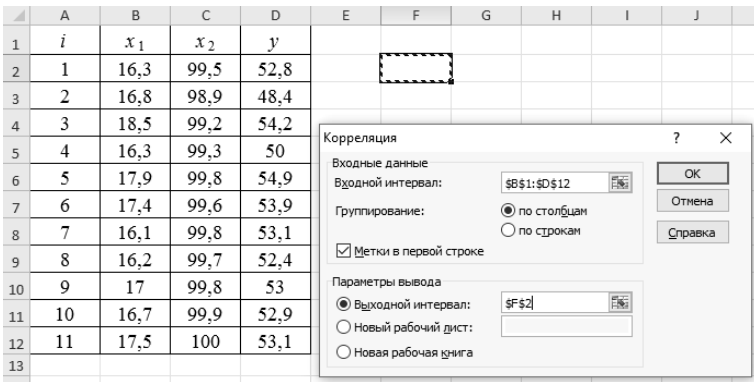


Рис. 4.1. Диалоговое окно ввода параметров инструмента **Корреляция** для поиска коэффициентов парной линейной корреляции (пример 4.1)

Анализируя полученную корреляционную матрицу (рис. 4.2), заметим, что факторы  $x_1$ ,  $x_2$  оказывают достаточно сильное влияние на результат  $y$  ( $r_{x_1,y} = 0,74$ ,  $r_{x_2,y} = 0,72$ ). Причем расходы на рекламу оказывают несколько большее влияние на объем продаж, нежели цена на товары-заменители.

	F	G	H	I
1				
2		$x_1$	$x_2$	$y$
3	$x_1$	1		
4	$x_2$	-0,05857	1	
5	$y$	0,510219	0,633766	1

Рис. 4.2. Корреляционная матрица для данных примера 4.1

Также отметим отсутствие явной коллинеарности между факторами, поскольку  $r_{x_1,x_2} = 0,27 < 0,3$ , что говорит о слабой линейной зависимости.

Для построения модели множественной линейной регрессии в Excel в меню команды **Анализ данных/Регрессия**, чтобы задать входной интервал X, выделим два столбца, содержащих факторы  $x_1$ ,  $x_2$ . Остальные действия аналогичны построению парной линейной регрессии (см. п. 3.2).

С помощью ППП построено следующее регрессионное уравнение:

$$y = -408,5 + 1,72x_1 + 4,34x_2, \quad R^2 = 0,85. \quad (4.12)$$

$(-4,0) \quad (4,4) \quad (4,2) \quad F = 25,3$

Сравнивая построенную модель с моделью примера (3.1), заключаем, что включение в регрессионную модель второго фактора значительно увеличило коэффициент детерминации – с 0,55 до 0,85. Это означает, что изменение объема производства на 85 % определяется вариацией двух факторов: расходы на рекламу, цена на товары-заменители. Модель имеет высокое качество. Основные статистические характеристики уравнения указывают на удовлетворительную спецификацию построенной модели: все  $t$ -статистики по абсолютной величине больше 3, т. е. все параметры статистически

значимы. Кроме того, модель в целом также является адекватной, поскольку  $F = 25,3$  превышает критический уровень:  $F = F_{0,05;2,3} = 4,26$ .

Если сравнить коэффициент парной регрессии для фактора «расходы на рекламу» в примере 3.1 ( $b = 2,16$ ) с аналогичным коэффициентом данного примера ( $b_1 = 1,72$ ), можно заключить, что уменьшение значения коэффициента во втором случае связано с элиминированием влияния из набора прочих факторов показателя цены на товары-заменители. Таким образом, при увеличении расходов на рекламу на 1 тыс. усл. ден. ед. объем продаж в среднем увеличится на 1,72 тыс. шт. при условии неизменности цены на товары-заменители. Если же цена на товары-заменители увеличится на 1 усл. ден. ед. при неизменности расходов на рекламу, то объем продаж возрастет на 4,34 тыс. шт.

Высокое качество модели (формула (4.12)), удовлетворительные статистические характеристики позволяют строить на ее основе прогноз. Точечный прогноз объема продаж

$$y_{\text{пр}} = -408,5 + 1,72 \cdot 20 + 4,34 \cdot 100,5 = 61,86 \text{ тыс. шт.}$$

Ошибка прогноза для множественной регрессии можно найти по формуле

$$s_y = s \sqrt{1 + X_{\text{пр}}^T (X^T X)^{-1} X_{\text{пр}}},$$

где  $X_{\text{пр}} = \begin{pmatrix} 1 \\ x_{1\text{пр}} \\ x_{2\text{пр}} \end{pmatrix}$  – вектор-столбец прогнозных значений независимых факторов;

$X = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} \end{pmatrix}$  – матрица наблюдаемых значений независимых переменных;

$$s = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n - m - 1}} \text{ – стандартная ошибка уравнения регрессии.}$$



Пример:

$$S_y = S \sqrt{1 + X_{\text{пр}}^T (X^T \cdot X)^{-1} X_{\text{пр}}} =$$
$$= 1,2 \sqrt{1 + (1 \ 20 \ 100,5) \begin{bmatrix} 1 & 1 & \dots & 1 \\ 16,3 & 16,8 & \dots & 19,1 \\ 99,5 & 98,9 & \dots & 100,2 \end{bmatrix}^{-1} \begin{pmatrix} 1 \\ 20 \\ 100,5 \end{pmatrix}} = 1,75.$$

Тогда предельная ошибка прогноза:  $\Delta y = t_{\text{кр}} s_{\hat{y}} = 2,26 \cdot 1,75 = 3,96$ ,  
интервал прогноза составит  $\gamma_y = y_{\text{пр}} \pm \Delta_y = 61,86 \pm 3,96$ .

Таким образом, при прогнозируемом уровне расходов на рекламу в объеме 20 тыс. усл. ден. ед. и ценах на товары-заменители 100,5 усл. ден. ед. объем продаж должен находиться в пределах от 57,9 до 65,8 тыс. шт.

### Вопросы для контроля знаний

1. Каков экономический смысл параметров модели парной линейной регрессии? множественной линейной регрессии? Каковы различия в интерпретации параметров этих моделей?
2. Каковы отличия коэффициентов парной и множественной корреляции?
3. Как рассчитать коэффициент детерминации, что он показывает?
4. Для чего применяют  $t$ -статистики Стьюдента? Какие выводы можно сделать на ее основании?
5. Как рассчитывают и для чего используют  $F$ -статистику?
6. Как связаны между собой  $t$ - и  $F$ -статистика в уравнении парной линейной регрессии?
7. Как рассчитать доверительный интервал для оценки прогноза парной линейной регрессии? множественной линейной регрессии?

### Практическое занятие № 4 Матричные вычисления оценок модели множественной линейной регрессии

**Цель работы:** отработать практические навыки матричных вычислений модели множественной линейной регрессии.

## Методические рекомендации

По данным выборочной совокупности сельскохозяйственных предприятий, приведенным в табл. 1, построить регрессионную модель урожайности гречихи.

Таблица 1

Исходные данные

$i$	Урожайность гречихи $y_i$ , ц/га	Число сельскохозяйственных машин (приведенной мощности) $x_{i1}$ , на 100 га	Количество химических средств обработки почвы $x_{i2}$ , ц/га	Балл плодородия почвы $x_{i3}$
1	19,4	3,64	0,46	51
2	16,8	0,80	1,25	36
3	18,0	4,99	0,61	48
4	19,8	11,07	1,02	51
5	19,2	4,32	0,55	47
6	17,2	4,85	0,49	38
7	25,0	1,41	0,65	55
8	15,2	0,77	0,29	42
9	13,8	1,01	0,28	41
10	27,0	6,44	2,10	57
11	19,4	4,97	0,90	43
12	21,4	5,70	0,39	47
13	24,2	20,87	0,77	53
14	19,4	3,98	0,99	50
15	14,0	1,19	0,48	36
16	14,4	0,58	0,24	39
17	16,4	3,08	0,28	41
18	16,8	0,14	0,63	39
19	26,2	0,11	0,93	54
20	17,4	2,53	1,41	42

1. Вводим данные в Excel в колонки А–Е, заголовок вводится в первую строку.

2. Строим матрицу парных коэффициентов корреляции (**Анализ данных/Корреляция**) и на ее основе оцениваем правомерность включения показателей в модель.

3. Оцениваем правомерность включения показателей в модель, используя шкалу Чеддока (прилож. 2). Для приведенного примера сила связи варьируется от умеренной до высокой, следовательно, все показатели включаем в модель.

4. Используя **Анализ данных/Регрессия**, строим многофакторную модель. Принцип построения аналогичен описанному в практическом занятии № 3 «Построение уравнения однофакторной регрессии», в качестве входного интервала  $X$  указываем прямоугольную таблицу показателей  $X$  с их названиями (ячейки В1:Е21).

5. Анализируем показатели итогового отчета. Коэффициент детерминации  $R$ -квадрат, равный 0,845, означает, что 84,5 % вариации  $Y$  объясняется изменением включенных в модель факторов.

6. Стандартная ошибка регрессии, равная 1,695, будет использоваться для оценки ошибок и при прогнозировании при последующей работе.

7. В разделе «Дисперсионный анализ» указано расчетное значение  $F$ -статистики Фишера, равное 29,088. Его значимость равна 1,029E-06, что свидетельствует о значимости уравнения регрессии в целом.

8. Выписываем оценку уравнения регрессии:

$$\hat{y} = -0,769 + 0,052x_1 + 3,034x_2 + 0,382x_3.$$

### **Задания для самостоятельной работы**

Выбираем вариант данных для регрессии по табл. 2 и 3.

Таблица 2

Исходные данные

Фактор						Результат $y$
A	B	C	D	E	F	
65	82	6	19	38	57	35
69	76	30	45	55	79	57
75	60	58	20	35	24	41
17	100	51	49	55	36	56

Окончание таблицы 2

Фактор						Результат у
A	B	C	D	E	F	
54	31	49	77	56	53	66
43	47	100	89	55	75	67
46	68	82	29	38	56	42
0	35	17	54	23	8	18
44	33	46	30	40	38	41
17	27	10	66	63	39	46
40	38	49	87	67	59	63
27	60	54	84	51	83	61
76	75	76	60	69	88	70
73	37	76	48	65	57	65
58	36	43	38	40	29	45
14	36	52	71	37	60	45
66	51	40	59	32	40	26
85	48	69	22	60	58	71
94	77	0	13	44	22	32
22	39	51	37	35	31	43
63	61	63	37	61	46	55
30	48	20	44	57	71	46
24	58	26	22	22	32	30
54	20	63	37	43	41	49
30	27	22	39	43	40	46
100	57	46	47	70	54	59
35	65	53	70	30	56	47
90	66	57	22	47	39	42
52	81	57	28	64	57	46
41	68	46	13	45	23	40
60	56	66	98	57	48	58

Таблица 3

## Включаемые факторы

№ варианта	Включаемые факторы	№ варианта	Включаемые факторы
1–3	A, C, E, F	7–9	C, D, E, F
4–6	B, D, E, F	10–12	B, C, D, F

Выполнить расчеты в Excel. Проанализировать полученные результаты. Оформить отчет.

## **Практическое занятие № 5** **Компьютерное построение модели** **множественной линейной регрессии и прогнозирование**

*Цель работы:* отработать практические навыки оценки качества многофакторной эконометрической модели и построения прогноза.

### **Методические рекомендации**

По данным выборочной совокупности сельскохозяйственных предприятий (практическое занятие № 4, табл. 1) провести оценку качества построенной регрессионной модели урожайности гречихи и построить модель прогноза.

1. Проанализируем значимость коэффициентов по  $t$ -статистике Стьюдента (см. практическое занятие № 2). В качестве  $m$  берем число факторов уравнения. Для примера  $v = 16$ , следовательно,  $t_{кр} = 2,1199$ . Для признания коэффициента уравнения значимым должно выполняться условие:  $|t_p| > |t_{кр}|$ .

2. Свободный член уравнения имеет  $t$ -статистику, равную – 0,241, следовательно, он незначим.  $t$ -статистика коэффициента при  $x_1$  равна 0,571. Анализируя экономический смысл фактора  $x_1$ , можно предположить, что важно не само число сельскохозяйственных машин, а качество и своевременность их использования, и при дальнейших исследованиях предлагается расчленить этот фактор на составляющие.

3.  $t$ -статистики коэффициентов при  $x_2$  и  $x_3$  равны, соответственно, 2,680 и 4,619; следовательно, коэффициенты значимы. Данные выборки свидетельствуют о том, что нулевая гипотеза о равенстве нулю этих коэффициентов отвергается (практическое занятие № 1).

4. Проверка значимости коэффициентов уравнения регрессии может быть осуществлена и другим способом: коэффициент будет значим, если интервал **Нижние–Верхние** не содержит нуля. Интервалы для коэффициентов  $x_2, x_3$  обладают этим свойством (таблица).

Интервал **Нижние–Верхние**

Y-пересечение	Нижние 95 %	Верхние 95 %
	-7,542	6,003
$x_{i1}$	-0,142	0,246
$x_{i2}$	0,634	5,434
$x_{i3}$	0,207	0,557

5. Для построения прогноза по значимым факторам исключим  $x_1$  и построим новую модель регрессии.

6. Оценка уравнения регрессии имеет вид:

$$\hat{y} = -1,410 + 2,956x_2 + 0,402x_3.$$

7. Высокое качество модели и удовлетворительные статистические характеристики позволяют строить на ее основе прогноза. Точечный прогноз урожайности гречихи при задаваемых прогнозных значениях факторов  $x_2$  и  $x_3$ :  $x_{2пр} = 1,3$  и  $x_{3пр} = 55$  будет равен

$$\hat{y} = -1,410 + 2,956 \cdot 1,3 + 0,402 \cdot 55 = 24,553.$$

8. Основным показателем для прогноза служит его ошибка, которая определяется по формуле

$$s_{\hat{y}} = s \sqrt{1 + X_{пр}^T (X^T \cdot X)^{-1} X_{пр}},$$

где  $s$  – стандартная ошибка уравнения регрессии. Ее значение уже рассчитано в отчете по регрессии:  $s = 1,661$ ;

$$X_{пр} = \begin{pmatrix} 1 \\ x_{1пр} \\ \dots \\ x_{nпр} \end{pmatrix} = \begin{pmatrix} 1 \\ x_{2пр} \\ x_{3пр} \end{pmatrix} = \begin{pmatrix} 1 \\ 1,3 \\ 55 \end{pmatrix} - \text{вектор-столбец прогнозных значений независимых факторов;}$$

$X_{пр}^T = (1 \ 1,3 \ 55)$  – вектор-строка прогнозных значений независимых факторов;

$X$  – матрица наблюдаемых значений по двум факторам:

$$X = \begin{pmatrix} 1 & 0,46 & 51 \\ 1 & 0,85 & 36 \\ 1 & 0,61 & 48 \\ \dots & \dots & \dots \\ 1 & 1,23 & 54 \\ 1 & 1,01 & 42 \end{pmatrix}.$$

9. Для удобства работы оформим матрицу  $X$  на отдельном листе Excel, без меток (используем ячейки A1:C20), затем транспонируем полученную матрицу. Последовательность работы следующая: копируем матрицу  $X$ , ставим курсор в ячейку E1 и переходим к команде **Специальная вставка/Транспонировать** (рис. 1).

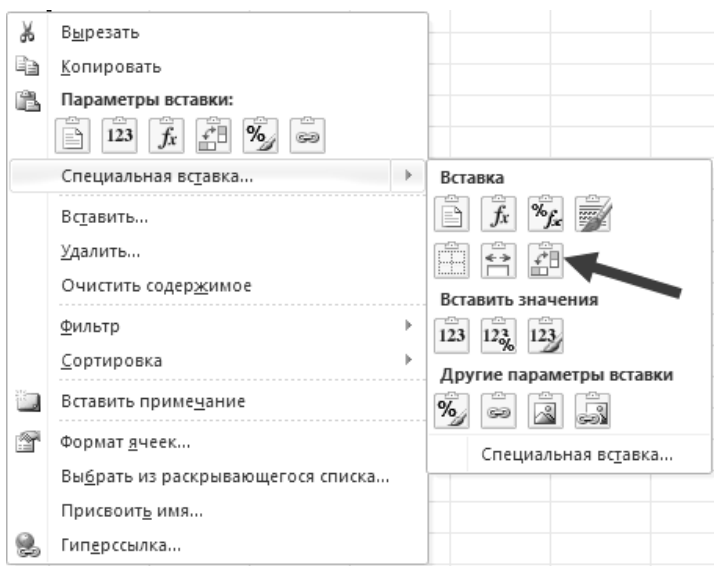


Рис. 1. Команда транспонирования

Результатом применения команды будет транспонированная матрица:

$$X^T = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 0,46 & 0,85 & 0,61 & \dots & 1,23 & 1,01 \\ 51 & 36 & 48 & \dots & 54 & 42 \end{pmatrix}.$$

10. Операцию  $X^T X$  выполняем следующим образом: выделяем диапазон для вывода (например E5:G7), вызываем команду **МУМНОЖ** (функции, категория **Математические**). В качестве аргументов записываем массивы транспонированной матрицы и матрицы наблюдаемых значений: =МУМНОЖ(E1:X3;A1:C20), затем нажимаем комбинацию клавиш Ctrl-Shift-Enter. Получены следующие результаты (рис. 2).

	E	F	G
4			
5	20,0	14,6	910,0
6	14,6	14,4	699,8
7	910,0	699,8	42240,0

Рис. 2. Результат умножения матриц

11. Для определения обратной матрицы выделяем диапазон для вывода (например E9:G11), вызываем команду **МОБР** (функции, категория **Математические**). В качестве аргумента записываем E5:G7. выделяем ячейки I5:G7, затем нажимаем комбинацию клавиш Ctrl-Shift-Enter. Получены следующие результаты (рис. 3).

	E	F	G
8			
9	3,115	0,508	-0,076
10	0,508	0,440	-0,018
11	-0,076	-0,018	0,002

Рис. 3. Значения обратной матрицы

12. Для примера (умножение матриц выполняем последовательно):

$$s_{\hat{y}} = s \sqrt{1 + X_{\text{пр}}^T (X^T \cdot X)^{-1} X_{\text{пр}}} =$$

$$= 1,661 \sqrt{1 + (1 \quad 1,13 \quad 55) \begin{pmatrix} 3,115 & 0,508 & -0,076 \\ 0,508 & 0,440 & -0,018 \\ -0,076 & -0,018 & 0,002 \end{pmatrix} \begin{pmatrix} 1 \\ 1,3 \\ 55 \end{pmatrix}} = 1,798.$$



13. Тогда предельная ошибка прогноза:  $\Delta y = t_{кр} s_{\hat{y}} = 2,1199 \times 1,798 = 3,811$ , интервал прогноза составит:

$$y_{пр} \pm \Delta_{\hat{y}} = (24,553 - 3,811; 24,553 + 3,811) = (20,742; 28,364).$$

14. Таким образом, при прогнозируемом количестве химических средств для удобрения и обработки почвы, расходуемых на гектар, равном 1,3, и балле плодородия почвы, равном 55, урожайность гречихи должна находиться в пределах от 20,74 до 28,36 ц/га.

### **Задания для самостоятельной работы**

Продолжить работу с индивидуальным заданием практического занятия № 4.

Выполнить расчеты в Excel. Проанализировать полученные результаты. Оформить отчет.

## Тема 5. УСТРАНЕНИЕ НАРУШЕНИЙ КЛАССИЧЕСКИХ МОДЕЛЬНЫХ ПРЕДПОЛОЖЕНИЙ В ПРОЦЕССЕ ЭКОНОМЕТРИЧЕСКОГО АНАЛИЗА

### *Вопросы*

- 5.1. Автокорреляция случайных остатков.
- 5.2. Гетероскедастичность остатков.
- 5.3. Понятие мультиколлинеарности, ее диагностика и способы устранения.

После нахождения МНК-оценок параметров модели необходимо удостовериться в их несмещенности, состоятельности и эффективности. В соответствии с теоремой Гаусса–Маркова, остатки должны обладать свойствами, близкими к свойствам случайных ошибок. В данной теме предлагаются на рассмотрение три вида нарушений этой теоремы: наличие автокорреляции случайных остатков, их гетероскедастичность, наличие мультиколлинеарности факторов.

### **5.1. Автокорреляция случайных остатков**

Теорема Гаусса–Маркова включает предположение о некоррелированности остатков:  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$ , (так как  $M(\varepsilon_i) = 0$ , то  $M(\varepsilon_i, \varepsilon_j) = 0$ ). При нарушении этого условия не выполняются нужные свойства оценок параметров модели. Для проверки некоррелированности остатков проверяется статистическая гипотеза:

$$H_0: \text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j.$$

Можно отметить, что коррелированность остатков может быть вызвана ошибкой в выборе спецификации модели, например, пропуском значимой экзогенной переменной. Причинами коррелированности остатков также может служить характер данных, их неточности (данные временного ряда часто являются коррелированными).

Ложной корреляцией называется корреляция остатков, вызванная ошибкой спецификации модели.

*Диагностика автокорреляции остатков.* Для проверки наличия корреляции в остатках (гипотезы  $H_0$ ) применяется тест Дарбина–Уотсона. Этот тест основан на том, что наличие автокорреляции остатков  $\varepsilon_i$  влечет коррелированность их оценок  $e_i$ , построенных по МНК. Другой особенностью теста Дарбина–Уотсона является то, что проверяется наличие автокорреляции между соседними остатками. Ясно, что при отсутствии автокорреляции остатков:  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$ , она отсутствует и между соседними остатками:  $\text{cov}(\varepsilon_i, \varepsilon_{i-1}) = 0$ . Обоснованием обратного утверждения может служить специфика связей между рядами экономических показателей. В таких рядах наибольшее влияние на последующие наблюдения оказывают предыдущие наблюдения. В случае сезонности данных можно составлять ряды их показателей с лагом, кратным циклу сезонности. Тогда наиболее коррелированными снова будут соседние наблюдения.

*Тест Дарбина–Уотсона.* По заданной выборке наблюдений находят оценки МНК для вектора остатков  $(\varepsilon_i)$ :  $(e_i) = (y - \hat{y})$  (в протоколе регрессионного анализа Excel оценки остатков находятся в столбце «Остатки» вывода остатков). Затем вычисляют статистику Дарбина–Уотсона:

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}. \quad (5.1)$$

Эта статистика принимает значения в интервале (0;4). Ее можно приблизительно вычислять по формуле

$$DW = 2(1 - r),$$

где  $r = \frac{\sum_{i=2}^n e_i \cdot e_{i-1}}{\sum_{i=1}^n e_i^2} = \frac{\text{cov}(e_i, e_{i-1})}{s_{e_i} \cdot s_{e_{i-1}}}$  – коэффициент корреляции оценок

соседних остатков. (Коэффициент корреляции можно вычислить средствами Excel с помощью функции **КОРРЕЛ**.)

В формуле (5.1) учитывается, что остатки гомоскедастичны и распределены по нормальному закону, выборочное среднее  $\bar{e}$  оценок остатков равно 0.

При отсутствии автокорреляции остатков значения статистики  $DW$  близки к числу 2.

Точнее, для применения теста  $DW$  по количеству наблюдений  $n$  числу объясняющих переменных  $k$  и уровню значимости  $\alpha$  из таблицы Дарбина–Уотсона находят нижнее и верхнее значения  $d_L$  и  $d_U$ . Затем отрезок  $[0,4]$  разбивают на пять частей. При попадании статистики  $DW$  в интервал  $(d_U; 4 - d_U)$  принимается гипотеза об отсутствии корреляции между соседними остатками:

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0. \quad \text{cov}(\varepsilon_i, \varepsilon_{i-1}) = 0.$$

Если статистика  $DW$  попадает в интервал  $(0; d_L)$  или в интервал  $(4 - d_L; 4)$ , то принимается альтернативная гипотеза о наличии корреляции между соседними остатками:  $\text{cov}(\varepsilon_i, \varepsilon_j) \neq 0$ . Если же значение  $DW$  принадлежит интервалу  $(d_L; d_U)$  или интервалу  $(4 - d_U; 4 - d_L)$ , то нельзя ни принять, ни отклонить нулевую гипотезу (рис. 5.1).

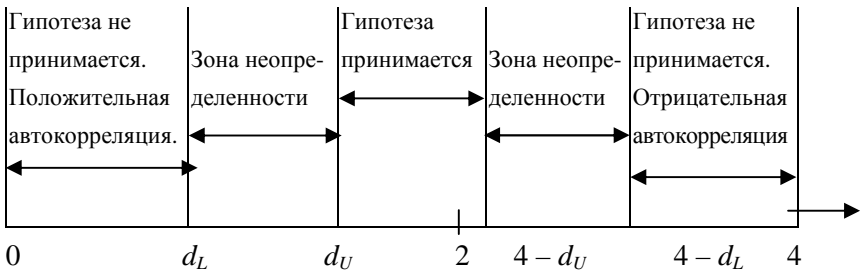


Рис. 5.1. Схема проведения теста Дарбина–Уотсона

Для устранения автокорреляции остатков применяется процедура Кохрейна–Оркатта, которая включает следующие этапы:

- построение оценок параметров уравнения регрессии;
- вычисление ее остатков;
- построение нового уравнения регрессии и нахождение новых оценок параметров.

## 5.2. Гетероскедастичность остатков

Одним из условий теоремы Гаусса–Маркова является условие гомоскедастичности случайных остатков:

$$M\varepsilon_i^2 = \sigma^2, i = 1, \dots, n.$$

Гетероскедастичными называются остатки, для которых не выполняется условие равенства дисперсий (гомоскедастичности), при этом предполагается, что выполняются условия некоррелированности остатков и нормальности их распределения. Гетероскедастичность означает, что анализируемые объекты не однородны, а дисперсии случайных остатков в разных наблюдениях различны.

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

Оценки, полученные по МНК, при наличии гетероскедастичности не будут эффективными (т. е. они не будут иметь наименьшую дисперсию по сравнению с другими оценками данного параметра). Стандартные ошибки оценок коэффициентов уравнения регрессии  $s_{b_i}$  будут занижены. Поэтому их  $t$ -статистики Стьюдента  $t_{b_i} = b_i/s_{b_i}$  будут завышены, что может привести к признанию статистически значимыми оценок коэффициентов, которые таковыми не являются. Доверительные интервалы  $b_i \pm t_{\alpha; n-m-1} s_{b_i}$  теоретических коэффициентов уравнения линейной регрессии получаются короче, чем на самом деле. Существуют два вида гетероскедастичности: истинная гетероскедастичность (вызывается непостоянством дисперсии случайного остатка, ее зависимостью от различных наблюдений) и ложная гетероскедастичность (вызывается ошибочной спецификацией модели регрессии).

Причины появления гетероскедастичности следующие:

- в пространственных выборках при зависимости масштаба измерений зависимой переменной от некоторой переменной, называемой фактором пропорциональности;
- во временных рядах, когда зависимая переменная имеет большой интервал качественно неоднородных значений или высокий темп изменения;

- в любой модели, в случае, если качество данных варьируется внутри выборки.

Наиболее распространенный случай гетероскедастичности – рост дисперсии одновременно с ростом одного из факторов (рис. 5.2, а и 5.2, б).

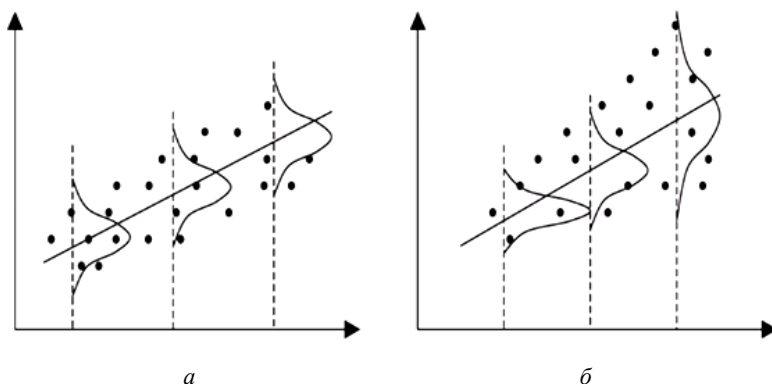


Рис. 5.2. Графики гомоскедастичности (а) и гетероскедастичности (б) остатков

*Диагностика гетероскедастичности остатков.* Наличие гетероскедастичности можно проверить визуально на основе графика остатков. Для этого нужно провести построение уравнения регрессии с выводом графика остатков (например, средствами Excel). О наличии гетероскедастичности свидетельствует величина разброса остатков относительно оси абсцисс. Если разброс примерно одинаков, то можно считать, что выполняется условие гомоскедастичности. Если же разброс увеличивается или уменьшается с ростом какого-либо из факторов, то остатки, вероятнее всего, гетероскедастичны. Здесь предполагается зависимость дисперсии остатков от значений факторов (или одного из них). Вид этой зависимости постулируется в основных используемых тестах.

Если факторов несколько, протокол расчета в Excel включает вывод графика остатков по всем факторам. Их сравнение позволяет выбрать тот фактор, который имеет наибольшее влияние на разброс остатков.

Выводы о величине дисперсии остатков, полученные визуально на графиках, можно уточнить и проверить с помощью тестов

ранговой корреляции Спирмена, Голдфелда–Квандта, Парка, Глейзера, Уайта и других.

*Тест ранговой корреляции Спирмена.* Предполагается, что дисперсии отклонений будут либо увеличиваться, либо уменьшаться с ростом значений  $x$ . Пусть  $n$  – число наблюдений. Значения переменной  $x_i$  и  $|e_i|$  ранжируются (упорядочиваются по величине). Рангами значений переменных  $x_i$  и  $|e_i|$  называются порядковые номера этих переменных в ранжированном ряду. Пусть  $d$  – разность между рангами значений переменной  $x_i$  и  $|e_i|$ . Коэффициент ранговой корреляции вычисляется по формуле

$$r_{x,e} = 1 - 6 \frac{\sum d^2}{n(n^2 - 1)}.$$

Зададим доверительную вероятность  $\gamma$ , пусть  $\alpha = (1 - \gamma)/2$  – уровень значимости. По  $t$ -таблицам Стьюдента находим граничную точку  $t_{\alpha;n-2}$ . Расчетное значение  $t$ -статистики  $t = \frac{r_{x,e} \sqrt{n-2}}{\sqrt{1-r_{x,e}^2}}$ . Если

$t < t_{\alpha;n-2}$ , то на уровне значимости  $\alpha$  принимается гипотеза об отсутствии гетероскедастичности. Иначе гипотеза об отсутствии гетероскедастичности отклоняется. В модели, содержащей несколько факторов, проверка гипотезы об отсутствии гетероскедастичности проводится с помощью статистики  $t$  для каждого из них отдельно.

*Пример 5.1.* Изучается зависимость себестоимости единицы изделия ( $y$ , тыс. руб.) от величины выпуска продукции ( $x$ , тыс. шт.) по группам предприятий за отчетный период (табл. 5.1).

Таблица 5.1

Исходные данные

№	1	2	3	4	5
$x$	2	3	4	5	6
$y$	1,9	1,7	1,8	1,6	1,4

Требуется проверить остатки уравнения парной линейной регрессии зависимости себестоимости от выпуска продукции на гетероскедастичность при помощи теста ранговой корреляции Спирмена.

*Решение.* Полагая, что между переменными  $x$ ,  $y$  имеет место линейная зависимость, определим выборочное уравнение линейной регрессии  $\hat{y} = 2,12 - 0,11x$ . Найдем остатки  $e_i$  и коэффициент детерминации (табл. 5.2).

Таблица 5.2

Результаты расчета

№	$x$	$y$	$\hat{y} = 2,12 - 0,11x$	$e = y - \hat{y}$
1	2	1,9	1,9	0,00
2	3	1,7	1,79	-0,09
3	4	1,8	1,68	0,12
4	5	1,6	1,57	0,03
5	6	1,4	1,46	-0,06
Сумма	20	8,4	8,4	0

Коэффициент детерминации  $r^2 = (-0,904)^2 = 0,817$ , т. е. 81,7 % общей вариации себестоимости  $y$  зависит от выпуска продукции  $x$ . Проверим гипотезу об отсутствии гетероскедастичности с помощью теста ранговой корреляции Спирмена. Зададим доверительную вероятность  $\gamma = 0,95$  %,  $\alpha = 0,025$ .

Таблица 5.3

Итоговый расчет

$x$	$e$	$ e_i $	$d_1$	$d_2$	$d = d_1 - d_2$	$d^2$
2	0,00	0,00	5	5	0	0
3	-0,09	0,09	4	2	2	4
4	0,12	0,12	3	1	2	4
5	0,03	0,03	2	4	-2	4
6	-0,06	0,06	1	3	-2	4
Сумма	0	0,3	15	15	0	16



Коэффициент ранговой корреляции равен  $r_{x,e} = 1 - 6 \frac{\sum d^2}{n(n^2 - 1)} =$   
 $= 1 - 6 \frac{16}{5(5^2 - 1)} = 0,2$  (табл. 5.3).

Находим граничную точку  $t_{\alpha;n-2} = t_{0,025;5-2} = 3,182$  по  $t$ -таблицам.

Статистика  $t = \frac{r_{x,e} \sqrt{n-2}}{\sqrt{1-r_{x,e}^2}} = \frac{0,2\sqrt{5-2}}{\sqrt{1-0,2^2}} = 0,354 < 3,182$ . Принимаем

гипотезу об отсутствии гетероскедастичности на уровне значимости 5 %.

*Тест Голдфелда–Квандта.* Одним из наиболее популярных тестов является тест Голдфелда–Квандта. Как правило, его применяют, если есть предположение о прямой зависимости дисперсии ошибки от величины некоторой независимой переменной в линейной модели. Предполагается, что стандартное отклонение  $\sigma_i = \sigma(\varepsilon_i)$  пропорционально значению переменной  $x$  в этом наблюдении:  $\sigma_i^2 = \sigma^2 x_i^2$ ,  $i = 1, \dots, n$ ,  $n$  – число наблюдений. Также предполагается, что  $\varepsilon_i$  имеет нормальное распределение и отсутствует автокорреляция. Все  $n$  наблюдений упорядочиваются по величине  $x$ . Эта упорядоченная выборка делится на три приблизительно равные части объемов  $k$ ,  $n - 2k$ ,  $k$  соответственно. При  $n = 30$   $k = 11$ , при  $n = 60$   $k = 22$ . Для каждой из выборок объема  $k$  оценивается свое уравнение регрессии, и находятся суммы квадратов отклонений  $S_1 = \sum_{i=1}^k e_i^2$

и  $S_3 = \sum_{i=n-k+1}^n e_i^2$  соответственно.

Зададим доверительную вероятность  $\gamma$ ,  $\alpha = 1 - \gamma$ . По  $F$ -таблицам находим граничную точку  $F_{\alpha;k-m-1;k-m-1}$ , где  $m$  – число факторов модели. Статистика  $F = S_3/S_1$  при  $S_3 > S_1$  и  $F = S_1/S_3$  при  $S_1 > S_3$ .

Если  $F < F_{\alpha;k-m-1;k-m-1}$ , то на уровне значимости  $\alpha$  принимается гипотеза об отсутствии гетероскедастичности. Иначе гипотеза об отсутствии гетероскедастичности отклоняется. Для множественной регрессии тест обычно проводится для того фактора, который

в максимальной степени связан с  $\sigma_i$ . При этом выбирают  $k > m + 1$ . Если нет уверенности относительно выбора фактора  $x_i$ , то данный тест можно осуществить для каждого фактора.

Существуют различные методы исправления модели в случае нарушений условий некоррелированности и гомоскедастичности случайных остатков.

Обобщенный метод наименьших квадратов (ОМНК) применяется в тех случаях, когда нарушаются условия Гаусса–Маркова о гомоскедастичности случайных остатков и их некоррелированности между собой.

Применение обычного МНК в указанных условиях ведет к тому, что найденные оценки коэффициентов уравнения регрессии не будут эффективными, а их дисперсии будут смещенными. Поэтому применение МНК ведет к ложным выводам при оценке качества модели. Для применения ОМНК необходима оценка ковариационной матрицы случайных остатков:

$$\sum_{\varepsilon} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma_n^2 \end{pmatrix}.$$

Эта матрица, как правило, неизвестна. Сделать оценку ее элементов по выборке объема  $n$  принципиально невозможно из-за количества элементов матрицы. Выделяются два важных специальных случая, когда количество элементов матрицы уменьшается и появляется возможность получения нужных оценок. Первый из них – это случай гетероскедастичности, когда матрица  $\sum_{\varepsilon}$  является диагональной и содержит не более  $n$  ненулевых элементов. Второй разновидностью реализуемого ОМНК служит модель с автотокоррелированными остатками при их постоянной дисперсии.

Для построения оценок вектора параметров матрица  $\sum_{\varepsilon}$  представляется в виде  $\sum_{\varepsilon} = \sigma_0^2 \Omega$ , где  $\Omega$  – симметрическая невырожденная матрица с безразмерными элементами,  $\sigma_0^2$  – некоторая положительная константа, размерность которой равна размерности  $y^2$ . Для классического случая, когда  $\sum_{\varepsilon} = \sigma^2 E_n$ ,  $\sigma_0^2 = \sigma^2$ ,  $\Omega = E_n$ .

Оценка вектора параметров линейной множественной регрессии методом ОМНК в матричном виде задается формулой

$$b = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y,$$

которая называется процедурой Эйткена. В этой формуле матрица  $\Omega = \frac{1}{\sigma_0^2} \sum_{\varepsilon}$ , а элементы матрицы  $\sum_{\varepsilon}$  заменяются их оценками, которые можно получить в практически реализуемых случаях.

В качестве примера рассмотрим случай, когда случайные остатки автокоррелированы и имеют одинаковую дисперсию. В этом случае ковариационная матрица

$$\sum_{\varepsilon} = \begin{pmatrix} \sigma^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma^2 & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma^2 \end{pmatrix}.$$

При дополнительном предположении относительно взаимосвязи элементов матрицы между собой количество неизвестных  $\sigma_{ij}$  можно уменьшить. Предполагается, что случайные остатки связаны автокорреляционной зависимостью первого порядка:

$$e_i = \rho e_{i-1} + v_i, \quad (5.2)$$

где  $v_i$  – случайные остатки, удовлетворяющие предпосылкам МНК,  $\rho$  – неизвестный параметр. В этом случае матрицу  $\Omega$  можно представить в виде

$$\Omega = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}, \quad \sum_{\varepsilon} = \sigma^2 \Omega,$$

где  $\sigma^2$  – общая дисперсия остатков,  $\sigma_0^2 = \sigma^2$ .

Оценку параметра  $\rho$  производят, исследуя взаимосвязи остатков, полученных после применения к модели регрессии обычного МНК.

### 5.3. Понятие мультиколлинеарности, ее диагностика и способы устранения

Часто при построении модели линейной множественной регрессии объясняющие переменные в той или иной мере связаны друг с другом. Линейная зависимость между объясняющими переменными приводит к негативному явлению эконометрики – мультиколлинеарности. При наличии мультиколлинеарности нарушается условие б (см. п. 3.2) о линейной независимости столбцов матрицы  $X$ . В этом случае матрица  $X^T X$  необратима, и вектор оценок параметров модели (формула (3.49)) определяется неоднозначно.

Причинами мультиколлинеарности часто являются следующие погрешности спецификации модели:

1. Включение в модель факторов, представляющих линейную комбинацию других факторов, или факторов, имеющих между собой сильную корреляционную зависимость.
2. Использование фиктивных переменных, количество которых равно периоду сезонности.
3. Наличие в модели большого числа лаговых переменных.

*Диагностика мультиколлинеарности.* О наличии коллинеарности в модели можно судить по следующим фактам:

1. Высокие значения коэффициентов парной корреляции между факторами (больше 0,8).
2. Значительные изменения коэффициентов регрессии при изменении набора факторов.
3. Чрезмерно высокие, экономически противоречивые значения коэффициентов модели.

*Методы устранения мультиколлинеарности.* Иногда мультиколлинеарность не является настолько серьезной проблемой, чтобы прилагать существенные усилия по ее выявлению и устранению. Если основная задача модели – прогноз будущих значений резуль- тативного признака, то при  $R^2 \geq 0,9$  наличие мультиколлинеарности обычно не сказывается на прогнозных качествах модели.

Единого метода устранения мультиколлинеарности не существует. Простейшим методом устранения мультиколлинеарности является исключение из модели ряда коррелированных переменных. В прикладных моделях лучше не сокращать число факторов до тех пор, пока мультиколлинеарность не станет серьезной проблемой.

Иногда для уменьшения мультиколлинеарности достаточно увеличить объем выборки, но при этом может усилиться автокорреляция. Иногда проблема мультиколлинеарности может быть решена с помощью изменения спецификации модели. Решению проблемы устранения мультиколлинеарности может помочь и переход к уравнениям приведенной формы.

Для устранения мультиколлинеарности может быть использован переход от исходных факторов к новым переменным, представляющим линейные комбинации исходных. При этом новые переменные должны быть слабо коррелированы. Например, векторы наблюдений над регрессорами могут быть ортогонализированы с использованием методов прикладного статистического анализа (методы факторного анализа).

Еще одним из возможных методов уменьшения мультиколлинеарности является использование пошаговых процедур отбора наиболее информативных переменных. Например, на первом шаге рассматривается лишь одна объясняющая переменная, имеющая с зависимой переменной  $Y$  наибольший коэффициент детерминации. На втором шаге выбирается новая переменная, которая вместе с первоначально отобранной имеют с  $Y$  наиболее высокий скорректированный коэффициент детерминации. На третьем шаге добавляется третья переменная, которая вместе с выбранными двумя имеет с  $Y$  наибольший скорректированный коэффициент детерминации, и т. д. Процедура введения новых переменных продолжается до тех пор, пока будет увеличиваться соответствующий скорректированный коэффициент детерминации.

Существует еще один метод устранения мультиколлинеарности без изменения состава регрессоров. Согласно этому методу для обеспечения нужных свойств матрицы  $X^T X$  к ней прибавляется с некоторым коэффициентом  $\tau$  диагональная матрица  $L$ , элементами которой являются диагональные элементы матрицы  $X^T X$ . Полученные после преобразования оценки коэффициентов модели получаются смещенными, поэтому полученную таким образом регрессию называют ридж-регрессией. Подбором величины коэффициента  $\tau$ , который, как правило, находится в интервале  $(0,1; 0,5)$ , можно добиться стабилизации коэффициентов уравнения модели.

В качестве примера построения регрессионной модели в условиях мультиколлинеарности приведены данные зависимости урожайности зерновых культур от пяти факторов.

*Пример 5.2.* По данным 20 сельскохозяйственных районов региона требуется построить регрессионную модель урожайности. Исходные данные приведены в табл. 5.4.

Таблица 5.4

Исходные данные

$i$	$y_i$	$x_{i1}$	$x_{i2}$	$x_{i3}$	$x_{i4}$	$x_{i5}$
1	9,70	1,59	0,26	2,05	0,32	0,14
2	8,40	0,34	0,28	0,46	0,59	0,66
3	9,00	2,53	0,31	2,46	0,30	0,31
4	9,90	4,63	0,40	6,44	0,43	0,59
5	9,60	2,16	0,26	2,16	0,39	0,16
6	8,60	2,16	0,30	2,69	0,32	0,17
7	12,50	0,68	0,29	0,73	0,42	0,23
8	7,60	0,35	0,26	0,42	0,21	0,08
9	6,90	0,52	0,24	0,49	0,20	0,08
10	13,50	3,42	0,31	3,02	1,37	0,73
11	9,70	1,78	0,30	3,19	0,73	0,17
12	10,70	2,40	0,32	3,30	0,25	0,14
13	12,10	9,36	0,40	11,51	0,39	0,38
14	9,70	1,72	0,28	2,26	0,82	0,17
15	7,00	0,59	0,29	0,60	0,13	0,35
16	7,20	0,28	0,26	0,30	0,09	0,15
17	8,20	1,64	0,29	1,44	0,20	0,08
18	8,40	0,09	0,22	0,05	0,43	0,20
19	13,10	0,08	0,25	0,03	0,73	0,20
20	8,70	1,36	0,26	1,17	0,99	0,42

Для анализа будут использованы следующие показатели:  
 $y$  – урожайность зерновых культур (ц/га);  
 $x_1$  – число колесных тракторов (приведенной мощности) на 100 га;  
 $x_2$  – число зерноуборочных комбайнов на 100 га;  
 $x_3$  – число орудий поверхностной обработки почвы на 100 га;  
 $x_4$  – количество удобрений, расходуемых на гектар, ц;  
 $x_5$  – количество расходуемых химических средств оздоровления растений (ц/га).

*Решение.* Предварительно, с целью анализа взаимосвязи показателей, построена таблица парных коэффициентов корреляции  $R$  (табл. 5.5).

Анализ матрицы парных коэффициентов корреляции показывает, что результативный показатель наиболее тесно связан с фактором  $x_4$  – количеством удобрений, расходуемых на 1 га ( $r_{yx_4} = 0,58$ ).

Таблица 5.5

Матрица парных коэффициентов корреляции

	$y$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
$y$	1,00	0,43	0,37	0,40	0,58	0,33
$x_1$	0,43	1,00	0,85	0,98	0,11	0,34
$x_2$	0,37	0,85	1,00	0,88	0,03	0,46
$x_3$	0,40	0,98	0,88	1,00	0,03	0,28
$x_4$	0,58	0,11	0,03	0,03	1,00	0,57
$x_5$	0,33	0,34	0,46	0,28	0,57	1,00

В то же время связь между признаками-аргументами достаточно тесная. Так, существует практически функциональная связь между числом колесных тракторов ( $x_1$ ) и числом орудий поверхностной обработки почвы ( $x_3$ ) –  $r_{x_1x_3} = 0,98$ .

О наличии мультиколлинеарности свидетельствуют также коэффициенты корреляции  $r_{x_1x_2} = 0,85$  и  $r_{x_2x_3} = 0,88$ . Учитывая тесную взаимосвязь показателей  $x_1$ ,  $x_2$  и  $x_3$ , в регрессионную модель урожайности следует включить один из них.

Чтобы продемонстрировать отрицательное влияние мультиколлинеарности, рассмотрим регрессионную модель урожайности, включив в нее все исходные показатели:

$$\hat{y} = 3,515 - 0,006x_1 + 15,542x_2 + 0,110x_3 + 4,475x_4 - 2,932x_5;$$

(0,65)      (-0,01)      (0,72)      (0,13)      (2,91)      (-0,95)

$$\hat{R}_y^2 = 0,517; \hat{s}^2 = 2,59; \bar{\delta} = 10,4\%; F_{\text{набл}} = 121.$$

В скобках указаны расчетные значения  $t$ -критерия для проверки значимости коэффициентов регрессии.

Под уравнением регрессии представлены следующие его параметры адекватности: множественный коэффициент детерминации  $\hat{R}_y^2$ , исправленная оценка остаточной дисперсии  $\hat{s}^2$ , средняя относительная ошибка аппроксимации  $\bar{\delta}$  и расчетное значение  $F$ -критерия  $F_{\text{набл}}$ .

Уравнение регрессии значимо, так как  $F_{\text{набл}} > F_{\text{кр}} = 2,85$ .  $F_{\text{кр}}$  определено по таблице  $F$ -распределения на уровне значимости  $\alpha = 0,05$  при числе степеней свободы числителя и знаменателя, соответственно,  $\nu_1 = 5$  и  $\nu_2 = 14$ . Из этого следует, что хотя бы один из коэффициентов уравнения  $\beta_j$  ( $i = 1, 2, \dots, 5$ ) не равен нулю.

Для проверки гипотезы о значимости отдельных коэффициентов регрессии  $H_0: \beta_j$  ( $i = 1, 2, \dots, 5$ ) сравним критическое значение  $t_{\text{кр}} = 2,14$ , найденное по таблице  $t$ -распределения на уровне значимости  $\alpha = 0,05$  при числе степеней свободы  $\nu = 14$ , с расчетными значениями

$t$ -статистик  $t_j = \frac{b_j}{\hat{s}_{b_j}}$ . Из уравнения следует, что статистически

значимым является коэффициент регрессии только при  $x_4$ , так как  $|t_4| = 2,91 > t_{\text{кр}} = 2,14$ .

Не поддаются экономической интерпретации отрицательные знаки коэффициентов регрессии при  $x_1$  и  $x_5$ , из чего следует, что повышение насыщенности сельского хозяйства колесными тракторами ( $x_1$ ) и средствами оздоровления растений ( $x_5$ ) отрицательно сказывается на урожайности. Таким образом, полученное уравнение регрессии неприемлемо.

Для получения уравнения регрессии со всеми значимыми коэффициентами используем пошаговый алгоритм регрессионного анализа.

Исключим из модели переменную  $x_1$ , коэффициенту при которой соответствует минимальное по абсолютной величине значение  $t$ -статистики  $|t_1| = 0,01$ . Для оставшихся переменных вновь построим уравнение регрессии:

$$\hat{y} = 3,507 + 15,574 x_2 + 0,105 x_3 + 4,473 x_4 - 2,935 x_5;$$

(0,65)    (0,77)            (0,35)            (2,91)            (-3,02)

$$\hat{R}_y^2 = 0,517; \hat{s}^2 = 2,39; \bar{\delta} = 10,4\%; F_{\text{набл}} = 151.$$

Полученное уравнение значимо, так как  $F_{\text{набл}} > F_{\text{кр}} = 2,90$ .  $F_{\text{кр}}$  определено по таблице  $F$ -распределения на уровне значимости



$\alpha = 0,05$  при числе степеней свободы числителя и знаменателя, соответственно,  $\nu_1 = 4$  и  $\nu_2 = 15$ . Однако в уравнении значим только коэффициент регрессии при  $x_4$ . Модули расчетных значений  $t$ -статистик для остальных коэффициентов меньше  $t_{кр} = 2,131$ , найденного по таблице  $t$ -распределения при  $\alpha = 0,05$  и  $\nu = 15$ .

Исключив из модели переменную  $x_3$ , которой соответствует минимальное значение  $t_3 = 0,35$ , получим уравнение регрессии:

$$\hat{y} = 1,980 + 21,801x_2 + 4,614x_4 - 3,338x_5;$$

(0,79)      (2,43)              (3,34)              (-1,30)

$$\hat{R}_y^2 = 0,513; \hat{s}^2 = 2,39; \bar{\delta} = 11,0\%; F_{набл} = 205.$$

В полученном уравнении статистически не значим и не интерпретируем коэффициент при  $x_5$ . Исключив переменную  $x_5$ , получим уравнение регрессии:

$$\hat{y} = 3,425 + 15,793x_2 + 3,486x_4;$$

(1,50)      (2,06)              (3,17)

$$\hat{R}_y^2 = 0,462; \hat{s}^2 = 2,35; \bar{\delta} = 11,7\%; F_{набл} = 263.$$

Получено значимое уравнение регрессии со значимыми и интерпретируемыми коэффициентами при регрессорах.

Однако данное уравнение является не самой лучшей моделью урожайности для приведенного примера.

Покажем, что в условиях мультиколлинеарности пошаговый алгоритм с включением переменных является более эффективным. На первом шаге в модель урожайности включим объясняющую переменную  $x_4$ , имеющую самый высокий коэффициент корреляции с зависимой переменной  $y$ , равный 0,58 (табл. 5.5). На втором шаге, включая поочередно в уравнение наряду с  $x_4$  переменные  $x_1$  и  $x_3$ , получаем модели, которые по экономическим соображениям и статистическим характеристикам превосходят полученные ранее:

$$\hat{y} = 7,291 + 0,282x_3 + 3,475x_4;$$

(11,50)      (2,17)              (3,25)

$$\hat{R}_y^2 = 0,482; \hat{s}^2 = 2,26; \bar{\delta} = 10,5\%; F_{набл} = 273.$$

$$\hat{y} = 7,342 + 0,345x_1 + 3,294x_4;$$

(11,12)      (2,03)              (3,02)

$$\hat{R}_y^2 = 0,469; \hat{s}^2 = 2,32; \bar{\delta} = 10,5\%; F_{набл} = 266.$$

Включение в эти уравнения дополнительно любой из трех оставшихся переменных ухудшает их свойства.

Таким образом, получены три приемлемых модели урожайности, из которых нужно по экономическим и статистическим соображениям выбрать наилучшую.

По статистическим критериям наиболее адекватна первая из двух последних моделей. Ей соответствует минимальное значение остаточной дисперсии и средней относительной ошибки аппроксимации, а также наибольшие значения коэффициента детерминации и  $F$ -статистики. Несколько худшими показателями обладает последняя модель, затем следует модель, полученная методом пошагового исключения переменных.

Сравним модели, полученные методом пошагового включения переменных. Они отличаются друг от друга одним регрессором. Наряду с общим для них фактором  $x_4$  второй объясняющей переменной в одной модели является  $x_1$ , а в другой –  $x_3$ . Однако в модели урожайности переменная  $x_1$  (число колесных тракторов на 100 га) более предпочтительна, чем переменная  $x_3$  (число орудий поверхностной обработки почвы на 100 га), которая является в некоторой степени вторичной по отношению к  $x_1$ . В связи с этим из экономических соображений предпочтение следует отдать модели с регрессорами  $x_1$  и  $x_4$ .

Таким образом, после реализации алгоритма пошагового регрессионного анализа выбираем окончательное уравнение регрессии:

$$\hat{y} = 7,342 + 0,345 x_1 + 3,294 x_4.$$

(11,12)    (2,03)            (3,02)

Уравнение значимо при  $\alpha = 0,05$ , так как  $F_{\text{набл}} = 266 > F_{\text{кр}} = 3,20$ .  $F_{\text{кр}}$  определено по таблице  $F$ -распределения при  $\alpha = 0,05$ ;  $\nu_1 = 2$  и  $\nu_2 = 17$ . Значимыми являются и все коэффициенты при регрессорах. Коэффициент регрессии  $\beta_1$  следует признать значимым ( $\beta_1 \neq 0$ ) из экономических соображений, при этом  $t_1 = 2,09$  лишь незначительно меньше  $t_{\text{кр}}(\alpha = 0,05; \nu = 17) = 2,11$ .

Из построенной регрессионной модели следует, что увеличение на единицу числа тракторов на 100 га пашни приводит (при фиксированном значении количества удобрений) к росту урожайности зерновых в среднем на 0,345 ц/га.

Расчет коэффициентов эластичности  $\mathcal{E}_1 \approx 0,068$  и  $\mathcal{E}_2 \approx 0,161$  показывает, что при увеличении факторов  $x_1$  или  $x_4$  на 1 % при фиксированном значении остальных факторов урожайность зерновых повышается в среднем на 0,068 % и 0,161 % соответственно.

Множественный коэффициент детерминации  $\hat{R}_y^2 = 0,469$  свидетельствует о том, что только 46,9 % вариации урожайности объясняется вошедшими в модель факторами  $x_1$  и  $x_4$ , т.е. насыщенностью растениеводства тракторами и удобрениями. Остальная часть вариации обусловлена действием неучтенных факторов ( $x_2$ ,  $x_3$ ,  $x_5$ , погодными условиями, особенностями почвы и др.). Средняя относительная ошибка аппроксимации  $\bar{\delta} = 10,5\%$  и величина остаточной дисперсии  $\hat{s}^2 = 1,97$  характеризуют адекватность модели.

При интерпретации уравнения регрессии интерес представляют значения относительных ошибок аппроксимации  $\delta_i = [(y_i - \hat{y}_i) / y_i] \cdot 100\%$ . Модельное значение  $\hat{y}_i$  резульативного показателя характеризует среднее для совокупности рассматриваемых районов значение урожайности при условии, что значения объясняющих переменных  $x_1$  и  $x_4$  зафиксированы на одном и том же уровне, а именно:  $x_1 = x_{i1}$  и  $x_4 = x_{i4}$ . Тогда по значениям  $\delta_i$  можно сопоставлять районы по урожайности. Районы, которым соответствуют значения  $\delta_i > 0$ , имеют урожайность выше средней,  $\delta_i < 0$  – ниже средней.

В примере растениеводство по урожайности наиболее эффективно ведется в седьмом районе, которому соответствует  $\delta_i = 28\%$ , т.е. урожайность в нем на 28 % выше средней по региону, наименее эффективно – в двадцатом районе с  $\delta_i = 27,3\%$ .

По экономическим соображениям часто бывает необходимо включение в модель тесно связанных между собой факторов. В ходе рассмотренных в данном примере процедур пошагового регрессионного анализа приходилось выбирать лучшую из двух практически равноценных моделей, отличающихся друг от друга тем, что в одной из них регрессором являлось число тракторов, а в другой вместо него присутствовало число прицепных орудий. Очевидно, что эти факторы важны для объяснения поведения зависимой переменной. Однако их одновременное присутствие в уравнении регрессии делает его некорректным вследствие близкого

к единице коэффициента корреляции регрессоров:  $r_{x_1x_3} = 0,98$  (см. табл. 5.5).

$$\hat{y} = 7,275 + 0,242x_1 + 0,467x_3 + 3,600x_4;$$

(10,68)    (-0,28)    (0,69)    (2,99)

$$\hat{R}_y^2 = 0,482; F_{\text{набл}} = 4,69.$$

Несмотря на значимость уравнения по  $F$ -критерию на уровне  $\alpha = 0,05$ , модель несостоятельна вследствие незначимости коэффициентов при двух ее регрессорах и по экономическим соображениям. Знаки коэффициентов показывают, что увеличение количества тракторов должно приводить к снижению средней урожайности.

Несколько изменим состав объектов, исключив из исходной выборки два последних наблюдения. Такая операция в обычных условиях может изменить значения оценок коэффициентов при сохранении общих закономерностей. Однако наблюдается радикальное изменение модели:

$$\hat{y} = 7,132 + 0,688x_1 + 0,230x_3 + 3,328x_4;$$

(12,79)    (0,85)    (-0,36)    (3,17)

$$\hat{R}_y^2 = 0,63; F_{\text{набл}} = 7,95.$$

Теперь закономерность изменения урожайности под влиянием обеспеченности факторами выглядит корректно, но ее зависимость от числа прицепных орудий изменилась на обратную. Такие изменения в модели объясняются тесной взаимосвязью включенных в модель факторов. Если их одновременное присутствие важно для анализа по экономическим соображениям, для построения модели урожайности можно применить ридж-регрессию.

Первоначально прибегнем к незначительному (на 10 %) усилению диагональных элементов матрицы  $(\mathbf{X}^T\mathbf{X})$ , при этом  $\tau = 0,1$ . Проанализируем изменения коэффициентов. В уравнениях со всеми объектами и без двух последних:

$$\hat{y} = 5,822 + 0,240x_1 + 0,236x_3 + 4,600x_4,$$

$$\hat{y} = 8,485 + 0,260x_1 + 0,348x_3 - 2,600x_4.$$

Если первое уравнение согласуется с представлением о влиянии факторов на урожайность, то второе уравнение непригодно из-за тенденции к снижению сбора зерновых культур при увеличении

количества удобрений. Вследствие очень тесной связи признаков небольшое усиление главной диагонали не приводит к приемлемым результатам. При более существенном ее усилении картина меняется. Так, при  $\tau = 0,2$  уравнения приобретают следующий вид:

$$\begin{aligned}\hat{y} &= 5,050 + 0,320x_1 + 0,260x_3 + 4,899x_4, \\ \hat{y} &= 5,691 + 0,304x_1 + 0,393x_3 + 1,827x_4.\end{aligned}$$

При  $\tau = 0,3$  коэффициенты несколько отличны от вычисленных при  $\tau = 0,2$  вследствие роста смещения оценок, однако модель не подвергается качественным изменениям и в целом правильно отражает закономерности влияния факторов на урожайность:

$$\begin{aligned}\hat{y} &= 4,548 + 0,324x_1 + 0,282x_3 + 4,935x_4, \\ \hat{y} &= 4,696 + 0,355x_1 + 0,392x_3 + 3,023x_4.\end{aligned}$$

Приведенные модели нуждаются в дополнительном исследовании их основных характеристик, более трудоемком, чем в классических моделях, но в условиях мультиколлинеарности метод ридж-регрессии может давать ценные результаты.

Более рациональным в условиях мультиколлинеарности может оказаться построение уравнения регрессии на главных компонентах, которые являются линейными функциями от всех исходных показателей и некоррелированы между собой.

### **Вопросы для контроля знаний**

1. Как связано поведение остатков уравнения регрессии с качеством МНК-оценок параметров регрессии?
2. Какие существуют методы диагностики наличия автокорреляции остатков?
3. Как устранить автокорреляцию в остатках?
4. В каком случае остатки называют гетероскедастичными? гомоскедастичными?
5. Какие существуют тесты проверки остатков регрессионной модели на гетероскедастичность? В чем их суть?
6. Как устранить гетероскедастичность в остатках?
7. К каким последствиям в уравнении регрессии приводит наличие мультиколлинеарности факторов?

8. Какие существуют методы диагностики мультиколлинеарности факторов?
9. Как устранить мультиколлинеарность факторов?

### Практическое занятие № 6

#### Выявление мультиколлинеарности и методика ее устранения

**Цель работы:** отработать практические навыки устранения мультиколлинеарности.

#### Методические рекомендации

По данным 20 сельскохозяйственных районов региона требуется построить регрессионную модель урожайности. Исходные данные приведены в таблице.

Для анализа будут использованы следующие показатели:

$y$  – урожайность зернобобовых культур (ц/га);

$x_1$  – балл плодородия почвы;

$x_2$  – число недождливых дней в напряженный период;

$x_3$  – число колесных тракторов (приведенной мощности) и зерноуборочных комбайнов на 100 га;

$x_4$  – число орудий поверхностной обработки почвы на 100 га;

$x_5$  – количество химических удобрений, расходуемых на гектар, ц.

*Таблица*

Исходные данные

№	$y_i$	$x_{i1}$	$x_{i2}$	$x_{i3}$	$x_{i4}$	$x_{i5}$
1	26,9	49	53	5,4	2,8	0,09
2	25,6	46	56	3,4	1,7	0,23
3	26,2	49	52	1,2	1,8	0,15
4	25,8	48	58	3,0	1,7	0,34
5	22,6	46	62	1,9	0,9	0,18
6	24,5	44	57	1,4	0,7	0,09
7	28,5	53	54	4,7	2,1	0,12
8	26,5	46	55	1,2	1,1	0,61

№	$y_i$	$x_{i1}$	$x_{i2}$	$x_{i3}$	$x_{i4}$	$x_{i5}$
9	27,8	51	62	1,7	0,9	0,48
10	24,9	49	60	3,1	1,6	0,57
11	23,2	44	57	3,1	2,5	0,26
12	29,5	52	52	1,2	0,7	0,35
13	27,0	50	61	1,2	0,6	0,47
14	25,5	47	64	3,9	1,6	0,41
15	27,1	49	66	2,0	1,0	0,31
16	22,6	46	54	1,7	0,8	0,09
17	29,2	50	56	4,7	3,3	0,48
18	21,8	44	55	2,9	1,5	0,36
19	25,1	49	54	1,8	1,0	0,62
20	33,5	51	66	5,9	3,9	0,64

1. Вводим данные в Excel (колонки А–G), заголовок вводится в первую строку.

2. Строим таблицу парных коэффициентов корреляции  $R$  (**Анализ данных/Корреляция**).

3. Анализ матрицы парных коэффициентов корреляции показывает, что результативный показатель наиболее тесно связан с фактором  $x_1$  – количеством удобрений, расходуемых на 1 га ( $r_{yx_1} = 0,794$ ).

В то же время связь между признаками-аргументами достаточно тесная: существует практически функциональная связь между  $x_3$  и  $x_4$  ( $r_{x_3, x_4} = 0,872$ ).

Учитывая тесную взаимосвязь этих факторов, в регрессионную модель урожайности следует включить только один из них.

Чтобы продемонстрировать отрицательное влияние мультиколлинеарности, рассмотрим регрессионную модель урожайности, включив в нее все исходные показатели:

$$\hat{y} = -16,701 + 0,728x_1 + 0,109x_2 - 0,526x_3 + 1,694x_4 + 0,985x_5;$$

$$t_{a_1} = 5,469; t_{a_2} = 1,273; t_{a_3} = -1,068; t_{a_4} = 2,171; t_{a_5} = 0,474; F_{\text{набл}} = 10,805.$$

Для проверки гипотезы о значимости отдельных коэффициентов регрессии сравним критическое значение  $t_{кр} = 2,1448$ , найденное по таблице  $t$ -распределения на уровне значимости  $\alpha = 0,05$  при числе степеней свободы  $\nu = 14$ , с расчетными значениями  $t$ -статистик (практическое занятие № 2, табл. 1). Из уравнения следует, что статистически значимым являются коэффициент регрессии при  $x_1$  и  $x_4$ .

Не поддается экономической интерпретации отрицательный знак коэффициента регрессии при  $x_3$ , из чего следует, что повышение насыщенности сельского хозяйства колесными тракторами и зерноуборочными комбайнами отрицательно сказывается на урожайности. Таким образом, полученное уравнение регрессии неприемлемо.

Для получения уравнения регрессии со всеми значимыми коэффициентами используем пошаговый алгоритм регрессионного анализа. Первоначально используем пошаговый алгоритм с исключением переменных.

Исключим из модели переменную  $x_3$ , мультиколлинеарную переменную  $x_4$ , имеющую экономически недопустимый знак и незначимый коэффициент регрессии.

Для оставшихся переменных построим уравнение регрессии.  $F_{набл} = 13,098$ , следовательно, полученное уравнение значимо, так как  $F_{набл} > F_{кр} = 2,90$ .  $F_{кр}$  определено по таблице  $F$ -распределения на уровне значимости  $\alpha = 0,05$  при числе степеней свободы числителя и знаменателя, соответственно,  $\nu_1 = 4$  и  $\nu_2 = 15$ . Однако в уравнении значимы только коэффициенты регрессии при  $x_1$  и  $x_4$ . Модули расчетных значений  $t$ -статистик для остальных коэффициентов меньше  $t_{кр} = 2,131$ , найденного по таблице  $t$ -распределения при  $\alpha = 0,05$  и  $\nu = 15$ .

Проведя пошаговое исключение переменных, получим значимое уравнение регрессии со значимыми и интерпретируемыми коэффициентами при регрессорах:

$$\hat{y} = -10,731 + 0,733x_1 + 1,001x_4;$$
$$t_{a_1} = 5,693; t_{a_4} = 2,656; F_{набл} = 24,063.$$

Однако данное уравнение является не самой лучшей моделью урожайности для приведенного примера.



В условиях мультиколлинеарности применяется также пошаговый алгоритм с включением переменных. На первом шаге в модель урожайности включим объясняющую переменную  $x_4$ , имеющую самый высокий коэффициент корреляции с зависимой переменной  $y$ , равный 0,794. На втором шаге включаем в уравнение наряду с  $x_1$  переменные  $x_4$  и  $x_3$ , получаем модель, которая по экономическим соображениям и статистическим характеристикам превосходит полученную ранее.

Включение в уравнение регрессии дополнительно любой из трех оставшихся переменных ухудшает его свойства.

Множественный коэффициент детерминации  $\hat{R}^2 = 0,739$  свидетельствует о том, что 73,9 % вариации урожайности объясняется вошедшими в модель факторами  $x_1$  и  $x_4$ . Оставшаяся часть вариации обусловлена действием неучтенных факторов. Средняя относительная ошибка аппроксимации  $\bar{\delta} = 10,5\%$  и величина остаточной дисперсии  $s^2 = 1,97$  характеризуют адекватность модели.

### **Задания для самостоятельной работы**

Продолжить работу с индивидуальным заданием практического занятия № 5.

Выполнить расчеты в Excel. Проанализировать полученные результаты. Оформить отчет.

## **Практическое занятие № 7**

### **Тестирование модели на наличие гетероскедастичности и автокорреляции остатков**

**Цель работы:** отработать практические тестирования эконометрической модели на наличие гетероскедастичности и автокорреляции остатков.

### **Методические рекомендации**

1. Проверить на наличие гетероскедастичности данные о зависимости настрига шерсти от живой массы овец (практическое занятие № 3). Для получения значений остатков  $e_i$  использовать

инструмент **Регрессия**, при построении модели ставим «птичку» в пункте **Остатки**.

2. Для визуального определения наличия или отсутствия гетероскедастичности построить график зависимости  $e_i$  от исходных данных  $x_i$  (рисунок).

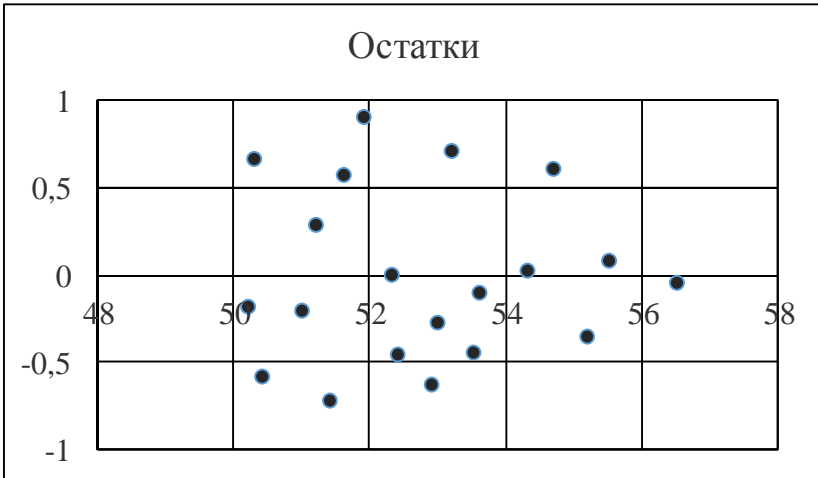


Рисунок. График остатков

3. Полученный результат позволяет предположить, что гетероскедастичность отсутствует, однако для достоверного анализа выполняем проверку с помощью теста ранговой корреляции Спирмена.

4. На листе Excel оформляем таблицу, в которую заносим значения  $x_i$  и  $e_i$ .

5. Поскольку  $e_i$  имеет как положительные, так и отрицательные значения, рассчитываем модуль  $e_i$  (функция ABS, категория Математические).

6. Ранжируем значения переменной  $x_i$ , начисляя ранг 1 наименьшему значению, в соответствии с правилами ранжирования.

7. Ранжируем значения модуля  $e_i$ .

8. Подсчитываем разности между рангами по каждой строке.

9. Возводим каждую разность  $d$  в квадрат.

10. Подсчитываем сумму квадратов (таблица).

Расчет ранговой корреляции

$x_i$	$e_i$	$ e_i $	$d_1$	$d_2$	$d = d_1 - d_2$	$d^2$
50,2	-0,174	0,174	12	20	-8	64
50,3	0,673	0,673	5	19	-14	196
50,4	-0,581	0,581	1	18	-17	289
51,0	-0,202	0,202	6	17	-11	121
51,2	0,292	0,292	2	16	-14	196
51,4	-0,715	0,715	20	15	5	25
51,6	0,578	0,578	17	14	3	9
51,9	0,917	0,917	9	13	-4	16
52,3	0,003	0,003	16	12	4	16
52,4	-0,450	0,450	3	11	-8	64
52,9	-0,618	0,618	7	10	-3	9
53,0	-0,271	0,271	11	9	2	4
53,2	0,722	0,722	14	8	6	36
53,5	-0,438	0,438	18	7	11	121
53,6	-0,092	0,092	10	6	4	16
54,3	0,034	0,034	4	5	-1	1
54,7	0,620	0,620	19	4	15	225
55,2	-0,347	0,347	15	3	12	144
55,5	0,092	0,092	8	2	6	36
56,5	-0,043	0,043	13	1	12	144
					Сумма	1732

11. Рассчитаем коэффициент ранговой корреляции по формуле

$$r_{x,e} = 1 - \frac{6 \sum_{i=1}^n d^2}{n(n^2 - 1)},$$

где  $n$  – число наблюдений.

Получено абсолютное значение коэффициента ранговой корреляции, равное 0,302.

12. Рассчитаем  $t$ -статистику по формуле

$$t = \frac{r_{x,e} \sqrt{n-2}}{\sqrt{1-r_{x,e}^2}},$$

где  $t = 1,345$ .

13. Определим  $t_{\text{крит}}$ , используя двухстороннее обратное распределение Стьюдента (функция **СТЮДЕНТ.ОБР.2Х**, категория **Статистические**) с вероятностью 0,025 и степенью свободы  $n - 2$ ,  $t_{\text{крит}} = 2,445$ .

14. Выполняется условие  $t < t_{\text{крит}}$ , следовательно, принимается гипотеза об отсутствии гетероскедастичности на уровне значимости 5 %.

### Задания для самостоятельной работы

Используем данные примера методических рекомендаций практического занятия № 5. Для вариантов 1, 3, 5, 7, 9, 11 строим однофакторную модель зависимости урожайности гречихи от количества химических средств для удобрения и обработки почвы, расходуемых на гектар; для вариантов 2, 4, 6, 8, 10, 12 строим однофакторную модель зависимости урожайности гречихи от балла плодородия почвы.

Анализируем полученную модель на наличие гетероскедастичности.

Выполнить расчеты в Excel. Проанализировать полученные результаты. Оформить отчет.

## Тема 6. НЕЛИНЕЙНЫЕ МОДЕЛИ РЕГРЕССИИ. ИСПОЛЬЗОВАНИЕ ФИКТИВНЫХ ПЕРЕМЕННЫХ В КОЛИЧЕСТВЕННОМ АНАЛИЗЕ

### *Вопросы*

- 6.1. Основные модели нелинейной регрессии.
- 6.2. Регрессионные модели с качественными переменными.
- 6.3. LOGIT- и PROBIT-модели.

### **6.1. Основные модели нелинейной регрессии**

Зависимости между экономическими показателями часто описываются нелинейными связями. Из экономической теории известны зависимости объема производства и факторов производства (функция Кобба–Дугласа, CES-функция); кривые Энгеля, описывающие зависимость объема продаж от денежных доходов; кривые Лаффера, устанавливающие связь между поступлениями в бюджет и ставкой налога, и пр. В частности, зависимости объема производства от факторов производства (труд и капитал), спроса на товар от денежных доходов описываются степенной функцией  $y = ax^b$ ; на основе экономического анализа несложно показать, что гиперболола  $y = a + \frac{b}{x}$  в случае, когда  $b > 0$ , хорошо описывает динамику нормативных показателей: с увеличением факторов-индикаторов ресурсосбережения нормы расхода асимптотически приближаются к своему предельно возможному уровню, в случае, когда  $b < 0$ , гиперболола хорошо описывает функции насыщения: после достижения некоторого значения очень медленное повышение результативного показателя с ростом значений фактора.

Большой класс нелинейных функций с помощью тождественных преобразований может быть приведен к линейному виду. В табл. 6.1 рассматривается класс нелинейных функций, допускающих сведение их к линейным с помощью тождественных преобразований переменных.

Таблица 6.1

Функции, наиболее часто используемые в экономических исследованиях

Название функции	Аналитический вид функции	Преобразования переменных, приводящие функцию к линейному виду
Линейно-логарифмическая	$y = a + b \ln x$	$x_1 = \ln x$ $\Downarrow$ $y = a + bx_1$
Степенная	$y = ax^b$	$x_1 = \ln x$ } $y_1 = \ln y$ } $\Rightarrow$ $\Rightarrow y_1 = \ln a + bx_1$
Показательная (лог-линейная)	$y = a \cdot b^x$	$y_1 = \ln y$ $\Downarrow$ $y = \ln a + x \cdot \ln b$
Параболическая	$y = a + b_1x + b_2x^2$	$x_1 = x$ } $x_2 = x^2$ } $\Rightarrow$ $\Rightarrow y = a + b_1x_1 + b_2x_2$
Гипербола (обратная)	$y = a + \frac{b}{x}$	$x_1 = \frac{1}{x}$ $\Downarrow$ $y = a + bx_1$

Для данного класса функций методы линейного регрессионного анализа применяются к преобразованным данным. В некоторых случаях экономические зависимости сложно привести к линейному виду, например: известная в экономической теории CES-функция. В этом случае для построения модели используются методы нелинейного регрессионного анализа.

В данной теме рассматривается методика построения нелинейных регрессий, которые с помощью тождественных преобразований приводятся к линейному виду.

*Степенная функция.* Пусть некоторая экономическая зависимость моделируется степенной функцией:

$$Y = A \cdot X^\beta, \quad (6.1)$$

где  $A$  и  $\beta$  – параметры модели (т. е. константы, подлежащие определению).

Стандартным и широко используемым подходом к анализу функций данного рода в эконометрике является логарифмирование по экспоненте (по основанию  $e = 2,718\ 28\dots$ ). Прологарифмировав обе части (6.1), имеем:

$$\ln Y = \ln A + \beta \ln X . \quad (6.2)$$

После замены  $\ln A = \beta_0$  выражение (6.2) примет вид:

$$\ln Y = \beta_0 + \beta \ln X .$$

С целью статистической оценки коэффициентов добавим в модель случайную погрешность  $\varepsilon$  и получим двойную логарифмическую модель:

$$\ln Y = \beta_0 + \beta \ln X + \varepsilon . \quad (6.3)$$

Не являясь линейным относительно  $X$  и  $Y$ , данное уравнение является линейным относительно  $\ln X$  и  $\ln Y$ , а также относительно параметров  $\beta_0$  и  $\beta$ . Вводя замены  $Y^* = \ln Y$  и  $X^* = \ln X$ , формулу (6.3) можно переписать в виде:

$$Y^* = \beta_0 + \beta X^* + \varepsilon .$$

Коэффициент  $\beta$  определяет эластичность переменной  $Y$  по переменной  $X$ , т. е. процентное изменение  $Y$  для данного процентного изменения  $X$ . Продифференцировав левую и правую части формулы (6.3) по  $X$ , получим:

$$\frac{1}{Y} \frac{dY}{dX} = \beta \frac{1}{X} \Rightarrow \frac{dY}{dX} \frac{X}{Y} = E(Y) .$$

Коэффициент  $\beta$  является константой, указывая на постоянную эластичность.

*Полулогарифмические функции.* Модели вида:

$$\ln Y = \beta_0 + \beta X + \varepsilon , \quad (6.4)$$

$$Y = \beta_0 + \beta \ln X + \varepsilon \quad (6.5)$$

описываются полулогарифмическими функциями; их обычно используют в тех случаях, когда необходимо определять темп роста или прироста каких-либо экономических показателей, например:

при анализе банковского вклада по первоначальному вкладу и процентной ставке, прироста объема выпуска от относительного (процентного) увеличения затрат ресурса, бюджетный дефицит от темпа роста ВВП, темп роста инфляции от объема денежной массы и т. д.

*Показательная функция.* Рассмотрим зависимость в виде показательной функции, хорошо известную в банковском и финансовом анализе:

$$Y_t = Y_0(1+r)^t, \quad (6.6)$$

где  $Y_t$  – значение величины  $Y$  в момент времени  $t$  (вклад в банке в момент времени  $t$ )

$Y_0$  – начальная величина переменной  $Y$  (первоначальный вклад в банке);

$r$  – сложный темп прироста величины  $Y$  (процентная ставка).

Модель (формула (6.6)) легко сводится к полулогарифмической модели (формула (6.4)).

Прологарифмировав (формула (6.6)), имеем:

$$\ln Y_t = \ln Y_0 + t \cdot \ln(1+r). \quad (6.7)$$

Введем обозначения:  $\ln Y_0 = \beta_0$ ,  $\ln(1+r) = \beta$ . Тогда формула (6.7) примет вид:

$$\ln Y_t = \beta_0 + \beta t + \varepsilon_t. \quad (6.8)$$

Полулогарифмическая модель (формула (6.4)) легко сводится к линейной модели заменой  $Y^* = \ln Y$ .

Коэффициент  $\beta$  в модели (формула (6.5)) имеет смысл темпа прироста переменной  $Y$  по переменной  $X$ , т. е. характеризует отношение относительного изменения  $Y$  к абсолютному изменению  $X$ . Действительно, продифференцировав (формула (6.5)) по  $X$ , имеем:

$$\frac{1}{Y} \cdot \frac{dY}{dX} = \beta \Rightarrow \beta = \frac{dY/Y}{dX} = \frac{\text{относительное изменение } Y}{\text{относительное изменение } X}.$$

Умножив полученное  $\beta$  на 100, получим процентное изменение переменной  $Y$  (темп прироста переменной  $Y$ ). Поэтому полулогарифмическая модель (формула (6.4)) обычно используется для



измерения темпа прироста экономических показателей. Заметим, что из соотношения (6.7) определяется темп прироста  $r$  показателя  $Y$ :

$$1 + r = e^\beta \Rightarrow r = e^\beta - 1. \quad (6.9)$$

Отметим, что коэффициент  $\beta$  (формула (6.18)) определяет мгновенный темп прироста,  $r$  (формула (6.19)) определяет обобщенный (сложный) темп прироста. Поэтому в общем случае они отличаются друг от друга.

Иногда показательную функцию удобно представить в виде, часто применяемом в эконометрическом анализе:

$$Y = \beta_0 e^{\beta x}.$$

В ситуации, когда анализируется изменение переменной  $Y$  с постоянным темпом прироста во времени, в качестве независимой переменной выступает фактор времени  $t$ :

$$Y = \beta_0 e^{\beta t}.$$

Данная функция путем логарифмирования ( $\ln e^{\beta t} = \beta t$ ) сводится к модели вида (формула (6.8)):  $\ln y = \ln \beta_0 + \beta t$ .

*Линейно-логарифмическая функция* имеет следующий вид:

$$Y = \beta_0 + \beta \ln X + \varepsilon. \quad (6.10)$$

Она сводится к линейной модели заменой  $X^* = \ln X$ . В данной модели коэффициент  $\beta$  определяет изменение переменной  $Y$  вследствие единичного относительного прироста  $X$  (например на 1 %), т. е. характеризует отношение абсолютного изменения  $Y$  к относительному изменению  $X$ . Действительно, продифференцировав (формула (6.10)), имеем:

$$\frac{dY}{dX} = \beta \frac{1}{X} \Rightarrow \beta = \frac{dY}{dX / X} = \frac{\text{абсолютное изменение } Y}{\text{относительное изменение } X} \Rightarrow$$

$$\Rightarrow dY = \beta \frac{dX}{X} \Rightarrow \Delta Y \approx \beta \frac{dX}{X}.$$

Умножив последнее соотношение на 100, получим абсолютный прирост  $Y$  при процентном изменении  $X$ . Таким образом, если  $\frac{\Delta X}{X}$  увеличилось на 1 % (0,01), то  $Y$  изменилось на  $0,01 \cdot \beta$ .

*Гипербола.* Модели вида:

$$Y = \beta_0 + \beta_1 \frac{1}{X} \ln X + \varepsilon,$$

описываются гиперболой, которая сводится к линейной функции заменой  $X^* = \frac{1}{X}$ . Данная модель применяется в тех случаях, когда неограниченное увеличение объясняющей переменной  $X$  асимптотически приближает зависимую переменную  $Y$  к некоторому пределу (в данном случае – к  $\beta_0$ ). В зависимости от знаков  $\beta_0$  и  $\beta_1$  – три вида гиперболы (рис. 6.1, а, б, в).

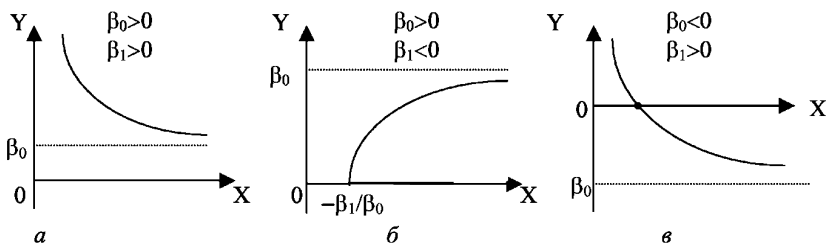


Рис. 6.1. Вид гиперболы в зависимости от ее параметров

*Полиномиальная функция.* Модели вида:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \dots + \beta_m X^m + \varepsilon,$$

описываются полиномиальной функцией. Например, кубическая функция:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon,$$

в микроэкономике моделирует зависимость общих издержек ( $Y$ ) от объема выпуска. Заметим, что степенная и параболическая функции являются частным случаем полиномиальной.

## 6.2. Регрессионные модели с качественными переменными

Зачастую в регрессионных моделях в качестве объясняющих переменных приходится использовать не только количественные (определяемые численно), но и качественные переменные. Например, спрос на некоторое благо может определяться ценой данного блага, ценой на заменители данного блага, ценой дополняющих благ, доходом потребителей и т. д. (эти показатели определяются количественно). Но спрос может также зависеть от вкусов потребителей, их ожиданий, национальных и религиозных особенностей и т. п. Эти показатели представить в численном виде не всегда возможно. Возникает проблема отражения в модели влияния таких переменных на исследуемую величину, что является достаточно сложной задачей. Обычно в моделях влияние качественного фактора выражается в виде фиктивной (искусственной) переменной, которая отражает два противоположных состояния качественного фактора. Например, «фактор действует—фактор не действует», «курс валюты фиксированный—курс валюты плавающий», «сезон летний—сезон зимний» и т. д. В этом случае фиктивная переменная может выражаться в двоичной форме:

$$D = \begin{cases} 0, & \text{фактор не действует;} \\ 1, & \text{фактор действует.} \end{cases}$$

Например,  $D = 0$ , если потребитель не имеет высшего образования,  $D = 1$ , если потребитель имеет высшее образование.

Переменная  $D$  называется *фиктивной* (искусственной, двоичной) *переменной* (индикатором).

Таким образом, кроме моделей, содержащих только количественные объясняющие переменные (обозначаемые  $X_i$ ), в регрессионном анализе рассматриваются также модели, содержащие лишь качественные переменные (обозначаемые  $D_i$ ), либо и те и другие одновременно.

Регрессионные модели, содержащие лишь качественные объясняющие переменные, называются *ANOVA-моделями* (моделями дисперсионного анализа).

При наличии фиктивных переменных исходная модель линейной регрессии записывается в виде:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m + \alpha_1 D_1 + \dots + \alpha_k D_k + \varepsilon,$$

где количество фиктивных переменных на единицу меньше количества уровней качественного признака. Это нужно для того, чтобы обеспечить независимость столбцов матрицы наблюдений. Если включить количество  $k + 1$  фиктивных переменных со значениями 0 или 1, равное количеству уровней признака, то первый столбец матрицы  $X$ , состоящий из единиц, вместе с  $k + 1$  столбцами переменных  $D$  будут линейно зависимыми.

*Пример 6.1.* Изучается зависимость заработной платы работника  $Y$  от возраста  $X$  с использованием фиктивной переменной по качественному фактору пола  $D$  по 20 работникам одного предприятия (табл. 6.2).

Таблица 6.2

Исходные данные

№ п/п	$Y$ – заработная плата работника, усл. ден. ед.	$X$ – возраст работника, лет	$D$ – пол, м/ж (1/0)
1	300	29	0
2	400	40	1
3	300	36	0
4	320	32	0
5	200	23	1
6	350	45	0
7	350	38	0
8	400	40	1
9	380	50	1
10	400	47	1
11	250	28	0
12	350	30	1
13	200	25	1
14	400	48	1
15	220	30	0
16	320	40	1
17	390	40	1

№ п/п	Y – заработная плата работника, усл. ден. ед.	X – возраст работника, лет	D – пол, м/ж (1/0)
18	360	38	1
19	260	29	0
20	250	25	1

*Решение.* Введем в модель фиктивную переменную  $D$ , которая принимает два значения: 1 – если пол мужской; 0 – если пол женский. Оценим параметры модели  $Y = \beta_0 + \beta_1 X_1 + \alpha_1 D_1 + \varepsilon$  МНК. Для вычислений воспользуемся Пакетом анализа в Excel (рис. 6.2).

	F	G	H	I	J
2	Вывод итогов				
3					
4	<i>Регрессионная статистика</i>				
5	Множественный R	0,861699165			
6	R-квадрат	0,742525451			
7	Нормированный R-квадрат	0,712234327			
8	Стандартная ошибка	37,20624062			
9	Наблюдения	20			
10					
11	<i>Дисперсионный анализ</i>				
12		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
13	Регрессия	2	67866,82621	33933,4131	24,51297168
14	Остаток	17	23533,17379	1384,304341	
15	Итого	19	91400		
16					
17		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
18	Y-пересечение	60,70849922	38,13432111	1,591964861	0,129815164
19	X	6,982516877	1,072470179	6,510686274	5,33458E-06
20	D	17,27462351	17,46232369	0,989251134	0,33641244

Рис. 6.2. Протокол регрессионного анализа

Уравнение множественной регрессии примет вид:

$$\hat{y} = 60,71 + 6,98x + 17,27d.$$

Коэффициент детерминации  $R^2$  равен 0,74, что говорит о высоком качестве модели. Уравнение регрессии значимо по  $F$ -критерию на 5%-м уровне, так как  $F_{\text{расч}} = 24,5 > F_{(0,05;2;17)} = 3,59$ .

Поскольку коэффициент при фиктивной переменной  $d$  не является значимым (его  $t$ -статистика, равная 0,989, меньше критического значения 2,11), уровень зарплаты не зависит от пола сотрудника, что позволяет сделать вывод, что заработная плата работников предприятия определяется преимущественно возрастом работников.

Отметим особенности моделей с фиктивными переменными. Фиктивные переменные позволяют разбить выборку данных на подгруппы, выделяемые уровнями качественных свойств, и проводить регрессию для каждой из подгрупп. Если имеется различие в отклике зависимой переменной на вариации качественных переменных в различных подгруппах, то это различие будет проявляться в уравнениях регрессии. Однако указанный метод надо применять осмотрительно. Во-первых, если в уравнениях присутствует постоянный член, то количество фиктивных переменных должно быть меньше количества уровней каждой качественной переменной. Во-вторых, коэффициент, соответствующий фиктивной переменной, должен всегда интерпретироваться связью со своей группой, что означает его обращение в ноль. Наконец, если модель имеет много качественных переменных с многими уровнями, введение фиктивных переменных приводит к большому количеству степеней свободы. Поэтому необходимо оценить сначала количество фиктивных переменных, чтобы затем оценивать число наблюдений для анализа.

В случае, когда качественные переменные являются эндогенными (объясняющими), при оценке коэффициентов используется метод МНК.

### 6.3. LOGIT- и PROBIT-модели

В эконометрических социальных исследованиях часто необходимо решать задачи, в которых зависимая переменная может принимать только целочисленное значение. В этом случае переменная  $Y$  называется фиктивной, а ее значения – уровнями. Различаются модели бинарного выбора, когда уровней два, и множественного выбора, когда их больше. При этом переменная  $X$  может быть как

количественной, так и качественной, и независимых переменных может быть несколько. Будем рассматривать модели бинарного выбора.

Принятие значения  $Y$ , равного 1, называется положительным решением. Логистическая регрессия выражает статистическую зависимость условной вероятности положительного решения от значений независимых переменных.

В силу того, что для переменной  $Y$  возможны только значения 1 и 0, условная вероятность значения 1 совпадает с математическим ожиданием переменной  $X$ . Поэтому значение функции регрессии в этом случае совпадает с названной условной вероятностью:  $P_i = M(Y=1/X_i) = b_0 + b_1 X_i$ . Эта модель называется линейной вероятностной моделью. Непосредственное применение этой регрессии для модели бинарного выбора вызывает нарушение классических предположений регрессии:

- случайный остаток не подчиняется нормальному закону распределения;
- наблюдается его гетероскедастичность;
- коэффициент детерминации  $R^2$  не является корректной мерой адекватности модели;
- оценки зависимой переменной, полученные по МНК, могут принимать значения, отличные от нуля и единицы.

В связи с этим в качестве зависимой переменной рассматривается непрерывная переменная, полученная из исходной на основе логистической функции распределения вероятностей:

$$F(Y) = \frac{e^Y}{1 + e^Y}.$$

Значения этой переменной равны логарифмам отношения шансов  $\ln \frac{\hat{P}_i}{1 - \hat{P}_i}$ , где  $\hat{P}_i$  – условная вероятность принятия положительного решения при  $i$ -м значении независимой переменной  $X$ .

Оценки  $\hat{P}_i$  равны относительным частотам появления положительных решений, которые вычисляются по таблице значений выборки по формуле

$$\hat{P}_i = \frac{n_i}{N_i},$$

где  $n_i$  – количество положительных решений при этом значении;  
 $N_i$  – количество появлений значения  $X_i$ .

Таким образом, для построения регрессии получаем два ряда переменных:  $X_i$  и  $\ln \frac{\hat{P}_i}{1 - \hat{P}_i}$ .

Построенная таким образом регрессия называется логистической, модель – LOGIT-моделью.

Оценка коэффициентов указанной регрессии проводится по методу максимального правдоподобия, при этом выполняются основные классические предположения. Исходя из оценки полученного уравнения регрессии  $\hat{Z}_i = b_0 + b_1 X_i$ , предсказанные значения положительного решения  $\text{pre} \hat{Y}_i$  находятся по формуле

$$\text{pre} \hat{Y}_i = \frac{e^{Z_i}}{1 + e^{Z_i}}.$$

Они принадлежат промежутку  $[0;1]$ .

Переходя к бинарным значениям 0 и 1 после выбора порога отсечения  $\mu$ , получаем прогноз:

$$\hat{Y}_i = \begin{cases} 1, & \text{pre} \hat{Y}_i \geq \mu, \\ 0, & \text{pre} \hat{Y}_i < \mu. \end{cases}$$

Часто в качестве порога отсечения выбирается  $\mu = 0,5$ . Однако его выбор определяется важностью положительного решения. При этом используется ROC-анализ.

*Пример 6.2.* При приеме на работу претенденты проходят испытательный срок, а затем выполняют тест. Исследуется зависимость успешного выполнения теста от продолжительности испытательного срока. Данные приведены в табл. 6.3:  $X_i$  ( $i = 1, 2, \dots, 10$ ) – продолжительность испытания (мес.),  $N_i$  – число всех претендентов со сроком  $x_i$  месяцев, число  $n_i$  – прошедшие тест после  $N_i$  месяцев испытания. Возникает стохастическая зависимость количественной



переменной  $X_i$ , когда распределение вероятностей бинарной переменной  $Y$ :  $Y = 1$  – тест сдан,  $Y = 0$  – тест не сдан. Значение функции регрессии совпадает с условной вероятностью  $P_i = E(Y = 1|X_i)$ . Оценкой этой вероятности  $P_i$  служит относительная частота (частота)  $\hat{P}_i = \frac{n_i}{N_i}$ . Согласно логистической регрессии проводится регрессия логарифма отношения шансов  $\hat{L}_i = \ln \frac{\hat{P}_i}{1 - \hat{P}_i}$ ,  $\hat{z}_i = b_0 + b_1 x_i$ .

Предсказанные значения  $Y$  находятся по формуле  $\text{pre } \hat{Y} = \frac{e^{\hat{z}_i}}{1 + e^{\hat{z}_i}}$ .

После выбора порога отсечения  $\mu$  получаем прогноз  $Y$ :

$$\hat{Y} = \begin{cases} 1, & \text{pre } \hat{Y} \geq \mu, \\ 0, & \text{pre } \hat{Y} < \mu. \end{cases}$$

Таблица 6.3

Данные для логистической регрессии

№ исследования	$X_i$	$N_i$	$n_i$
1	3	15	3
2	6	10	4
3	9	12	5
4	12	18	8
5	15	16	8
6	18	10	6
7	21	6	4
8	24	14	11
9	27	9	8
10	30	9	8

Провести логистическую регрессию: найти  $\hat{P}_i$ ,  $\hat{L}_i$ , провести и оценить парную регрессию  $\hat{z}_i = b_0 + b_1 x_i$ , найти предсказанные  $\hat{Y}_i$  для порога  $\mu = 0,5$ .

Найти прогноз  $\hat{Y}$  при  $X = 20$  мес.

Основа: стохастическая зависимость  $X \rightarrow Y$  при условии, что  $Y$  принимает только два значения: 1 или 0. Особенность ее в том, что  $E(Y|X) = \text{Prob}(Y = 1|X)$ . В этом случае функция регрессии имеет вид:

$$X_i \text{ а } P_i = \text{Prob}(Y = 1 | X_i),$$

где  $X_i$  – одна переменная или их набор.

Если строить линейную оценку функции регрессии по обычной схеме, то наблюдаем нарушения классических ограничений и, как следствие, плохое качество оценки.

Другой путь: преобразование зависимой переменной  $P$  так, чтобы ее поведение больше соответствовало свойствам функции распределения вероятностей. Практически используются три способа такой замены: логистическая регрессия, PROBIT-модель и замена Вейбулла.

PROBIT-модель для модели бинарного выбора строится по аналогичной схеме на основе функции стандартного нормального распределения.

Различия между LOGIT- и PROBIT-моделями в случае малых выборок практически отсутствуют, поскольку коэффициенты регрессий отличаются практически постоянным множителем.

### **Вопросы для контроля знаний**

1. Какие виды нелинейных зависимостей вы знаете? При описании каких экономических зависимостей они используются?
2. Как оценить параметры нелинейной регрессии?
3. Охарактеризуйте экономический смысл параметров логарифмической и полупологарифмической моделей.
4. В каких случаях используют регрессионную модель с качественными переменными?
5. В каких пределах могут изменяться качественные переменные?
6. С помощью каких ППП можно строить регрессионные модели?
7. Дайте определение понятий LOGIT- и PROBIT-моделей.

## Практическое занятие № 8

### Оценивание параметров LOGIT- и PROBIT-моделей методом максимального правдоподобия

**Цель работы:** отработать практические навыки построения LOGIT- и PROBIT-моделей и использования их для прогноза.

#### Методические рекомендации

1. На опытной станции при выведении элитного сорта растений проводится подогрев почвы с семенами растения. Продолжительность времени подогрева является объясняющей (экзогенной) переменной  $X$  (ч). Фиктивная качественная переменная  $Y$  принимает значение 1 в случае всхожести семян и значение 0 – в противном случае.

2. Оценкой условной вероятности положительного исхода при заданном значении  $X_i$  является величина  $\hat{P}_i = P(Y = 1 | X = X_i) = \frac{n_i}{N_i}$  – частотъ взошедших семян среди всех семян  $N_i$ , прошедших подогрев продолжительностью  $X_i$  ( $i = 1, 2, \dots, k$ ).

3. Для того чтобы сделать вид изменения величин  $\hat{P}_i$  похожим на изменение вероятностей, необходимо преобразовать их с помощью обратной функции нормального стандартного распределения вероятностей  $\Phi(T) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^T e^{-t^2/2} dt$ . Табличный процессор Excel содержит такую функцию под именем НОРМ.СТ.ОБР.

4. Для перехода к положительным значениям к полученным результатам  $\Phi^{-1}(\hat{P}_i)$  прибавляем число 5.

5. Проводим регрессию  $X_i$  на  $\Phi^{-1}(\hat{P}_i) + 5$ :  $Z_i = 4,207\ 809 + 0,382\ 788X_i$ . Предсказанные значения  $\text{pre}Z_i$  находятся в разделе

#### **Вывод остатка.**

6. Из предсказанных значений  $Z_i$  вычитаем число 5. Применением функции НОРМ.СТ.РАСП( $\text{pre}Z_i - 5$ ; ИСТИНА) из значений получим предсказанные условные вероятности принятия положительного решения  $\text{pre}P_i(Y = 1 | X = X_i)$ .

7. Для прогнозного значения при  $X = 2,7$  предсказанное значение  $\text{pre}P_i(Y = 1 | X = X_i)$   $\text{pre}Z_i - 5 = 0,24136$ , предсказанное (прогнозное) значение вероятности положительного решения равно  $0,595$  (табл. 1).

Таблица 1

Результаты вычислений PROBIT-модели

Последовательность расчета	Номер наблюдения									
	1	2	3	4	5	6	7	8	9	10
$X_i$	0,5	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0
$N_i$	20	25	30	10	13	40	50	35	45	45
$n_i$	5	10	13	5	7	24	34	26	38	40
$\hat{P}_i$	0,25	0,4	0,453	0,5	0,538	0,6	0,68	0,743	0,844	0,889
$\Phi^{-1}(\hat{P}_i)$	-0,675	-0,254	-0,169	0	0,095	0,254	0,468	0,653	1,011	1,221
$\Phi^{-1}(\hat{P}_i) + 5$	4,326	4,747	4,831	5	5,1	5,253	5,467	5,623	6,011	6,221
$\text{pre}Z_i$	4,399	4,591	4,782	4,973	5,165	5,356	5,548	5,739	5,93	6,122
$\text{pre}Z_i - 5$	-0,6018	-0,409	-0,218	-0,027	0,165	0,356	0,548	0,739	0,930	1,122
$\text{pre}\hat{P}_i$	0,274	0,341	0,414	0,489	0,565	0,639	0,708	0,770	0,824	0,870

### Задания для самостоятельной работы

**Задание 1.** В соответствии с номером варианта (табл. 2) и по аналогии с примером 6.2 построить LOGIT-модель, используя данные табл. 3.

Таблица 2

№ варианта	Включаемые номера	№ варианта	Включаемые номера
1, 5, 9	1–15	3, 7, 11	4–19
2, 6, 10	2–17	4, 8, 12	5–20

Таблица 3

Исходные данные для логистической регрессии

Номер наблюдения	Наблюдаемые значения		
	$X_i$	$N_i$	$n_i$
1	2	20	5
2	2	25	5
3	4	20	12
4	6	20	5
5	2	25	5
6	8	20	12
7	8	25	10
8	3	20	6
9	3	25	8
10	3	20	12
11	3	20	5
12	3	25	8
13	6	20	12
14	6	25	10
15	7	20	8
16	7	25	10
17	7	30	6
18	7	25	8
19	7	30	10
20	7	25	9

Выполнить расчеты в Excel. Проанализировать полученные результаты. Оформить отчет.

**Задание 2.** Построить PROBIT-модель по аналогии с методическими рекомендациями, по данным табл. 2 и в соответствии с номером варианта.

Выполнить расчеты в Excel и дать прогноз для среднего показателя. Проанализировать полученные результаты. Оформить отчет.

## **Тема 7. МОДЕЛИ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ НА ОСНОВЕ ВРЕМЕННЫХ РЯДОВ**

### *Вопросы*

- 7.1. Временной ряд и его компоненты. Стационарные временные ряды и автокорреляционная функция.
- 7.2. Моделирование тренда и сезонной компоненты.
- 7.3. ARIMA-модели временных рядов.
- 7.4. Изучение взаимосвязей по временным рядам.
- 7.5. Статическая модель Леонтьева.

Одним из важных разделов предмета «Количественные методы анализа в АПК» является раздел, посвященный временным рядам. По специфике своих значений показатели временного ряда, как правило, отличаются от пространственных данных в силу того, что данные временного ряда не являются одинаково распределенными и статистически независимыми. Применение традиционных методов корреляционно-регрессионного анализа к изучению причинно-следственных зависимостей показателей, представленных в форме временных рядов, может привести к ряду серьезных проблем, в том числе к ложным выводам о наличии связи. Поэтому построение регрессионной модели на временных рядах начинается с анализа временных рядов с последующим, при необходимости, их преобразованием.

### **7.1. Временной ряд и его компоненты. Стационарные временные ряды и автокорреляционная функция**

Показатели значений временного ряда образуют серию наблюдений, проведенных через равные промежутки времени: день, неделя, месяц, квартал, год и т. д.

Например, временными рядами будут: серия ежедневных наблюдений в течение некоторого периода за ценами товара при закрытии торгов на бирже, дневные объемы выпуска товара,

месячные показатели инфляции или индекса потребительских цен, ежеквартальные оценки валового национального продукта или средних зарплат, ежегодные данные об объеме, выручке и прибыли компании. Основным положением, на котором базируется использование временных рядов для прогнозирования, является то, что факторы, влияющие на отклик изучаемой системы, действовали некоторым образом в прошлом и настоящем, и ожидается, что они будут действовать схожим образом и в не слишком далеком будущем. Поэтому основной целью анализа временных рядов является оценка и вычленение этих влияющих факторов с целью прогноза дальнейшего поведения системы и выработки рациональных управленческих решений.

Временной ряд рассматривается как случайный процесс, т. е. последовательность случайных величин:

$$\{Y_1, Y_2, \dots, Y_n\} = \{Y_t\}, t = 1, \dots, n,$$

зависящая от показателя времени  $t$  (через равные промежутки). Конкретные показатели ряда, его уровней  $\{y_1, y_2, \dots, y_n\}$  рассматриваются как реализация случайного процесса.

Спецификация модели временного ряда обычно включает следующие составляющие.

*Тренд (T)* – плавно меняющаяся с течением времени составляющая, которая описывает влияние долговременных факторов, эффект которых сказывается постепенно (например, рост населения, рост потребления, изменения структуры потребления, экономическое развитие).

*Сезонная компонента (S)* описывает регулярно изменяющиеся наблюдения. Она определяет короткопериодические колебания, связанные с изменениями внутригодовой активности, и повторяющиеся через более или менее продолжительные моменты времени, недели, месяцы, кварталы, сезоны и т. д.

Циклическая составляющая описывает длительные периоды относительного подъема и спада и состоит из циклов, которые меняются по амплитуде и протяженности. На циклическую составляющую оказывают влияние такие факторы, как рост и истощение ресурсов, продолжительные периоды погодных условий, циклы солнечной активности, изменения в финансовой и налоговой политике и т. д. Факторы, влияющие на циклическую составляющую,

трудно идентифицируются формальными методами. В связи с такими особенностями циклические составляющие исключаем из рассмотрения.

*Случайная компонента* ( $\varepsilon$ ) вызывает отклонения от хода отклика, определяемого трендовой, циклической и сезонной составляющими. В терминах статистики  $\varepsilon$  считается случайным остатком, и на нее накладываются соответствующие ограничения.

Если временной ряд представляется в виде суммы компонент:

$$y_t = T_t + S_t + \varepsilon_t, \quad t = 1, \dots, n,$$

то модель ряда называется аддитивной.

Представление ряда в виде произведения компонент:

$$y_t = T_t \cdot S_t \cdot \varepsilon_t, \quad t = 1, \dots, n,$$

носит название мультипликативной модели.

Модель смешанного типа имеет следующий вид:

$$y_t = T_t \cdot S_t + \varepsilon_t, \quad t = 1, \dots, n.$$

Задачи анализа и прогнозирования временных рядов начинаются, как правило, с построения графика. На этом этапе можно исследовать компонентный состав ряда и провести выбор модели.

Стационарные временные ряды – это ряды, у которых вероятностные свойства не изменяются с течением времени. На интуитивном уровне можно дать следующее определение стационарного временного ряда: это такой ряд, который имеет постоянную среднюю, а значения ряда колеблются вокруг этой средней с некоторой постоянной дисперсией.

Хотя на практике временные ряды часто не являются стационарными, для их исследования используются стационарные временные ряды.

Временной ряд  $\{Y_t\}$ , ( $t = 1, \dots, n$ ) называется стационарным (в широком смысле), если его среднее значение и дисперсия постоянны (не зависят от времени  $t$ ), а его ковариации между значениями и  $Y_{t+\tau}$  с лагом  $\tau$  зависят только от величины лага  $\tau$  и не зависят от времени.

Примером стационарного временного ряда является белый шум – ряд случайных остатков классической линейной модели



регрессии, удовлетворяющий условиям Гаусса–Маркова. Для такого ряда среднее и ковариации равны нулю, а дисперсия постоянна.

Примером нестационарного временного ряда является классический пример случайного блуждания, модель которого представляется в виде  $y_t = y_{t-1} + \varepsilon_t$ , где  $\varepsilon_t$  – белый шум с нулевым средним и постоянной дисперсией.

Для стационарности временного ряда необходимо (но недостаточно) отсутствие в его структуре тренда и сезонной составляющей, а также условие гомоскедастичности случайного остатка. Стационарный временной ряд представляется в виде:

$$y_t = m + \varepsilon_t,$$

где  $m$  – ожидаемое значение (математическое ожидание) уровней ряда;

$\varepsilon_t$  – случайный остаток с нулевым средним, постоянной дисперсией  $\sigma^2$  и ковариацией, зависящей от лага.

Термин «автоковариация» для величины  $\gamma(\tau)$  применяется в связи с тем, что он характеризует статистическую зависимость между уровнями одного и того же ряда, отстоящими на  $\tau$  лагов. На практике для стационарных временных рядов вместо коэффициента автоковариации чаще используется безразмерная величина – коэффициент автокорреляции

$$\rho_t = \frac{\text{cov}\{y_t, y_{t+\tau}\}}{\sqrt{D\{y_t\}D\{y_{t+\tau}\}}} = \frac{\gamma(\tau)}{\gamma(0)}. \quad (7.1)$$

На основании значений функции автокорреляции и ее графика (коррелограммы) можно делать заключение о выборе вида модели ряда. Если колебания коррелограммы периодические и их амплитуда практически постоянна, то принимается аддитивная модель. Если же амплитуда колебаний коррелограммы возрастает (убывает), то принимается мультипликативная модель.

Большинство экономических временных рядов являются нестационарными, поскольку их вероятностные характеристики зависят от времени. Если математическое ожидание временного ряда изменяется во времени в соответствии с некоторым детерминированным или стохастическим трендом, то ряд нестационарен

по среднему значению. Для описания таких рядов используются два основных класса моделей:

- с детерминированным трендом, т. е. модели с трендом в виде детерминированной функции времени;
- модели авторегрессии интегрированного скользящего среднего (ARIMA-модели).

Разностно-стационарные процессы имеют стационарные ряды первых (или последующих) разностей. Тренд-стационарные процессы будут стационарными относительно линии тренда (детерминированного или стохастического).

## 7.2. Моделирование тренда и сезонной компоненты

*Моделирование тренда.* Одной из важнейших задач исследования показателей экономического временного ряда является выявление основной тенденции изучаемого процесса, выраженной неслучайной составляющей (тренда либо тренда с циклической и (или) сезонной компонентой). Для этого необходимо выбрать вид функции  $f(t)$  (линейная, полиномиальная, экспоненциальная, логистическая, Гомперца и т. д.). Выбор той или иной функции является наиболее важным этапом анализа временного ряда, так как неправильный выбор приводит к ошибкам при прогнозировании уровней ряда. Из нескольких вариантов предпочтение обычно отдается той функции, при которой меньше сумма квадратов отклонений фактических данных от расчетных на основе этих функций.

Для выявления основной тенденции чаще всего используется МНК. Значения временного ряда  $y_t$  рассматриваются как зависимая переменная, время  $t$  – как объясняющая:

$$y_t = f(t) + \varepsilon_t.$$

Согласно МНК в случае линейной зависимости параметры прямой  $\hat{y}_t = f(t) = b_0 + b_1t$  находим из системы нормальных уравнений, в которой в качестве  $x_i$  берем  $t$ .

Полиномиальный тренд имеет вид:

$$y_t = b_0 + b_1t + b_2t^2 + \dots + b_mt^m.$$

Характерной особенностью линейного тренда является то, что приросты ординат остаются постоянными. Для полинома второй степени приросты ординат описываются линейной зависимостью  $y_t - y_{t-1} = b_1 - b_2 + 2b_2t$ , приросты второго порядка постоянны. Для выбора порядка полинома используются последовательные разности значений ряда. Нужная степень полинома соответствует случаю, когда разности постоянны (точнее – их минимальный порядок).

Экспоненциальный тренд имеет вид  $y_t = ab^t$ . Характерной особенностью этой модели являются постоянные темпы роста и прироста, при этом приросты уровней ряда зависят от величины самой функции.

Логистическая модель тренда имеет вид:

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

Одним из критериев проверки наличия тренда служит знаковый критерий Кокса–Стьюарта. Для временного ряда  $y_t$ ,  $t = 1, \dots, n$  проверяется гипотеза об отсутствии тренда:

$$H_0 : E\{y_t\} = a = \text{const},$$

при альтернативной гипотезе

$$H_1 : |E\{y_{t+1}\} - E\{y_t\}| \neq 0.$$

Гипотеза проверяется по следующему алгоритму.

1. Временной ряд разбивается на три части:  $n = n_1 + n_2 + n_3$ , таких, что  $n_1 = n_3$ .

2. Вычисляются разности всех наблюдений с одинаковыми номерами из третьей и первой частей и фиксируются их знаки. Пусть  $n_1^+$  – число положительных разностей,  $n_1^-$  – число отрицательных разностей:

$$S = \max\{n_1^+, n_1^-\}.$$

Если  $n_1^+ > n_1^-$ , то принимается решение о том, что тренд возрастающий, если  $n_1^+ < n_1^-$  – убывающий.

3. При  $n > 30$  статистика  $S$  аппроксимируется нормальным распределением со средним  $\frac{n}{6}$  и дисперсией  $\frac{n}{12}$ , значит, статистика имеет стандартное нормальное распределение:

$$S^* = \frac{|S - \frac{n}{6}|}{\sqrt{\frac{n}{12}}}$$

При малых объемах выборки ( $n \leq 30$ ) с учетом поправки на непрерывность:

$$S^* = \frac{|S - \frac{n}{6}| - 0,5}{\sqrt{\frac{n}{12}}}$$

4. Гипотеза об отсутствии тренда принимается на уровне  $\alpha$ , если  $|S^*| < u_{1-\frac{\alpha}{2}}$ , где  $u_{1-\frac{\alpha}{2}}$  – квантиль уровня  $1 - \frac{\alpha}{2}$  стандартного нормального распределения.

Особенностью теста Кокса–Стюарта является тот факт, что его применение не требует предположений о распределении вероятностей уровней ряда. Такие тесты называются непараметрическими.

Для длинных временных рядов невозможно указать подходящую аппроксимирующую параметрическую кривую (функцию). Одним из методов выделения тренда в этом случае является непараметрический метод скользящего среднего. 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.

Скользящее среднее позволяет сгладить как случайные, так и периодические колебания, выявить имеющуюся тенденцию в развитии показателей, и поэтому служит важным инструментом при фильтрации компонент временного ряда. Сглаживание по простой скользящей средней представляется в виде следующей последовательности шагов.

1. Сначала определяется длина интервала сглаживания  $m$ , включающего  $m$  последовательных уровней ряда. Чем сильнее колебания показателей, тем большим должен быть интервал сглаживания.

2. Затем исходный ряд разбивается на участки по  $m$  уровней. Каждый последующий участок сдвигается по отношению к предыдущему на один лаг времени.

3. Срединное значение каждого участка заменяется средней арифметической (или взвешенной средней) уровней из данного участка.

Интервал  $m$  удобнее брать из нечетного числа наблюдений  $m = 2p + 1$ . Если период колебаний ряда является четным,  $m$  можно брать также и четным. Если представить уровни участка сглаживания для  $m = 2p + 1$  в виде  $y_{t-p}, \dots, y_{t-1}, y_t, y_{t+1}, \dots, y_{t+p}$ , то скользящая средняя определяется формулой

$$\% = \frac{1}{2p+1}(y_{t-p} + \dots + y_{t-1} + y_t + y_{t+1} + \dots + y_{t+p}).$$

При четном  $m = 2p$  скользящая средняя находится по формуле

$$\% = \frac{1}{2p} \left( \frac{1}{2} y_{t-p} + y_{t-p+1} + \dots + y_{t-1} + y_t + y_{t+1} + \dots + y_{t+p-1} + \frac{1}{2} y_{t+p} \right).$$

Указанные формулы применяются, когда основная тенденция ряда является линейной. Когда же тренд сглаживаемого ряда имеет колебания, целесообразно использовать взвешенную скользящую среднюю. На каждом из участков слагаемые берутся со своими весовыми коэффициентами. Например, если сглаживание производится с помощью полинома, то на основе метода наименьших квадратов можно показать, что для интервала длины  $m = 5$  весовые коэффициенты будут следующими:  $-\frac{3}{35}, \frac{12}{35}, \frac{17}{35}, \frac{12}{35}, -\frac{3}{35}$ .

Для интервала длины  $m = 7$ :  $-\frac{2}{21}, \frac{3}{21}, \frac{6}{21}, \frac{7}{21}, \frac{6}{21}, \frac{3}{21}, -\frac{2}{21}$ .

*Моделирование сезонной компоненты.* Приведем два критерия для обнаружения сезонной компоненты временного ряда. Первый из них – дисперсионный, второй основан на вычислении коэффициента автокорреляции уровней ряда. Для применения дисперсионного критерия выдвигается гипотеза о том, что во временном ряде, из которого отфильтрован тренд, отсутствуют сезонные колебания. В случае справедливости данной гипотезы

$F$ -статистика будет иметь  $F$ -распределение с  $(m - 1)$  и  $(n - m)$  степенями свободы:

$$F = k \cdot \frac{n - m}{m} \cdot \frac{\sum_{j=1}^m (\bar{s}_j - \bar{s})^2}{\sum_{j=1}^m \sum_{i=1}^k (s_{ij} - \bar{s}_j)^2}, \quad (7.2)$$

где  $k$  – число лет наблюдений;

$n = k \cdot m$  – количество всех наблюдений;

$m$  – период колебаний ( $m = 4$  для квартальных данных,  $m = 12$  для месячных данных);

$\bar{s}_j = \frac{1}{k} \sum_{i=1}^k s_{ij}$  – среднее для периода с номером  $j$ ;

$\bar{s} = \frac{1}{m} \sum_{j=1}^m \bar{s}_j = \frac{1}{n} \sum_{j=1}^m \sum_{i=1}^k s_{ij}$  – общее среднее,  $s_{ij}$  – значения ряда

после выделения из него тренда,  $i = 1, \dots, k, i = 1, \dots, m$ .

Проверяемая гипотеза отвергается (т. е. сезонные колебания присутствуют) на уровне значимости  $\alpha$ , если  $F$ -статистика (формула (7.2)) больше критического значения  $F(\alpha, m - 1, n - m)$ .

Для обнаружения сезонных колебаний с использованием коэффициента автокорреляции его значения рассчитываются по формуле (7.1).

Расчетные значения коэффициентов автокорреляции сравниваются с табличными значениями с заданным уровнем значимости, и в случае, когда расчетное значение больше табличного, соответствующий коэффициент признается значимым, что может свидетельствовать о наличии сезонных колебаний. Критические значения коэффициента корреляции  $r(1)$  в зависимости от количества наблюдений приведены в табл. 7.1. Заключение о наличии или отсутствии сезонных колебаний точнее можно сделать на основании коррелограммы. Если значения коэффициента корреляции  $r(\tau)$  изменяются периодически, то этот период и будет периодом сезонных колебаний.

Критические значения коэффициента корреляции

Число наблюдений $n$	$r(1)$
10	0,360
15	0,328
20	0,300
25	0,276
30	0,257

### **Выделение сезонной компоненты**

*Способ 1 (сглаживание).* Для расчета сезонных компонент воспользуемся методом скользящей средней. Просуммируем уровни ряда последовательно за каждые четыре квартала со сдвигом на один квартал и определим условные годовые объемы потребления электроэнергии (табл. 7.2, графа 3). Полученные суммы разделим на длину периода (в нашем случае – на 4) и найдем скользящие средние, которые уже не зависят от сезонности. Чтобы привести эти значения в соответствие с фактическими моментами времени, найдем средние значения из каждых двух соседних скользящих средних (табл. 7.2, графа 5). Оценку сезонной компоненты найдем, вычитая из фактического значения уровня ряда  $u_t$  центрированную скользящую среднюю (табл. 7.2, графа 6).

На следующем этапе подготовим вторую вспомогательную таблицу. Занесем в нее оценки сезонных компонент, распределив их по кварталам. За каждый квартал найдем среднюю оценку сезонной компоненты. Например, для 1-го квартала:  $\bar{S}_1 = (0,575 + 0,55 + 0,675) / 3 = 0,6$ .

Сезонные воздействия за период должны взаимопогашаться. В аддитивной модели это выражается в том, что сумма всех сезонных компонент за период должна быть равна нулю. Рассчитаем корректирующий коэффициент по формуле:  $k = \sum_{i=1}^n \bar{S}_i / n$ , где  $n$  – длина периода. Для нашего примера  $k = (0,6 - 1,958 - 1,275 + 2,708) / 4 = 0,01875$ .

Таблица 7.2

Расчет оценок сезонных компонент методом скользящей средней

№ квартала, $t$	Потребление электроэнергии (кВт·ч), $y_t$	Итого за 4 квартала	Скользящая средняя за 4 квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	6,0	24,4	6,10	6,25	-1,25
2	4,4				
3	5,0	25,6	6,40	6,45	2,55
4	9,0	26,0	6,50	6,63	0,58
5	7,2	27,0	6,75	6,88	-2,08
6	4,8	28,0	7,00	7,10	-1,10
7	6,0	28,8	7,20	7,30	2,70
8	10,0	29,6	7,40	7,45	0,55
9	8,0	30,0	7,50	7,63	-2,03
10	5,6	31,0	7,75	7,88	-1,48
11	6,4	32,0	8,00	8,13	2,88
12	11,0	33,0	8,25	8,33	0,68
13	9,0	33,6	8,40	8,33	-1,78
14	6,6	33,4	8,35		
15	7,0				
16	10,8				

Скорректированные значения сезонной компоненты рассчитываем как разность между средним значением сезонной компоненты и корректирующим коэффициентом:  $S_i = \bar{S}_i - k, i = \overline{1, n}$  (табл. 7.3).



Таблица 7.3

Поиск сезонных компонент для аддитивной модели

№ квартала	№ года				Средняя оценка сезонной компоненты для $i$ -го квартала, $\bar{S}_i$	Скорректированная сезонная компонента, $S_i$
	1	2	3	4		
1 квартал		0,575	0,55	0,675	0,6	0,6
2 квартал		-2,075	-2,025	-1,775	-1,958	-2,0
3 квартал	-1,25	-1,1	-1,475		-1,275	-1,3
4 квартал	2,55	2,7	2,875		2,708	2,7
Корректирующий коэффициент					0,01875	

Исключим сезонную компоненту из исходного ряда, т. е. считаем  $y - S$ . С этой целью заполним рабочую таблицу (табл. 7.4).

Таблица 7.4

Исключение из уровней ряда сезонности

Номер периода	Исходный ряд	Сезонная компонента	Преобразованный ряд
$t$	$y$	$S$	$y - S$
1	6,0	0,6	5,4
2	4,4	-2,0	6,4
3	5,0	-1,3	6,3
4	9,0	2,7	6,3
5	7,2	0,6	6,6
6	4,8	-2,0	6,8
7	6,0	-1,3	7,3
8	10,0	2,7	7,3
9	8,0	0,6	7,4
10	5,6	-2,0	7,6
11	6,4	-1,3	7,7
12	11,0	2,7	8,3
13	9,0	0,6	8,4
14	6,6	-2,0	8,6
15	7,0	-1,3	8,3
16	10,8	2,7	8,1

Далее в преобразованном ряду  $y - S$  выделим линейный тренд:

$$T = 5,7 + 0,2t, R^2 = 0,91, DW = 1,29.$$

$$t_{ст}(38,9)(12,3)$$

Визуальный анализ графика остатков позволяет предположить их гомоскедастичность.

В данном случае статистика  $DW$  лежит в области неопределенности (при уровне значимости 5 %  $d_U = 1,37$ ,  $d_L = 1,1$ ). Поэтому нельзя ни отклонить, ни принять нулевую гипотезу относительно автокорреляции остатков. Если условно принять допущение об отсутствии статистически значимой автокорреляции остатков, то с использованием последней модели найдем потребление электроэнергии в каждом квартале.

Зная значения сезонных компонент:

$$S_t = \begin{cases} 0,6; t = 1,5,9,13; \\ -2; t = 2,6,10,14; \\ -1,3; t = 3,7,11,15; \\ 2,7; t = 4,8,12,16 \end{cases}$$

и тренд  $T = 5,7 + 0,2t$ , можно прогнозировать потребление электроэнергии в последующих кварталах с использованием модели  $\hat{y}_t = 5,7 + 0,2t + S_t$ :

$$\hat{y}_{17} = 5,7 + 0,2 \cdot 17 + 0,6 = 9,7 \text{ млн кВт} \cdot \text{ч};$$

$$\hat{y}_{18} = 5,7 + 0,2 \cdot 18 - 2 = 7,3 \text{ млн кВт} \cdot \text{ч};$$

$$\hat{y}_{19} = 5,7 + 0,2 \cdot 19 - 1,3 = 8,2 \text{ млн кВт} \cdot \text{ч};$$

$$\hat{y}_{20} = 5,7 + 0,2 \cdot 20 + 2,7 = 12,4 \text{ млн кВт} \cdot \text{ч}.$$

И в целом по году:  $9,7 + 7,3 + 8,2 + 12,4 = 37,6$  млн кВт·ч.

*Способ 2 (с включением фиктивной сезонной компоненты).*

В качестве факторов в модель включают показатели: время  $t$  и три фиктивные сезонные компоненты.

### 7.3. ARIMA-модели временных рядов

Среди временных рядов особо выделяются стационарные временные ряды. К рядам с такими свойствами могут быть приведены нестационарные ряды после выделения тренда, фильтрации сезонной компоненты или взятия разностей. В частности, ряд случайных

остатков является стационарным при выполнении условий Гаусса–Маркова. Для моделирования стационарных рядов часто применяются модели авторегрессии и скользящего среднего. Авторегрессионная модель порядка  $p$  ( $AR(p)$ ) представляется в виде

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t.$$

Для порядка  $p = 1$   $AR(1)$  модель авторегрессии называется марковским процессом:

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t,$$

где  $|\alpha_1| < 1$ , случайные остатки образуют белый шум (с постоянной дисперсией  $\sigma_0^2$ ).

Для такого процесса математическое ожидание и дисперсия равны, соответственно:

$$My_t = 0; D(y_t) = \frac{\sigma_0^2}{1 - \alpha^2},$$

автокорреляционная функция –  $\rho(\tau) = \alpha^\tau$ .

Таким образом, коэффициент  $\alpha$  является коэффициентом автокорреляции первого порядка:  $\alpha = \rho(1)$ . В силу равенства  $\rho(\tau) = \alpha^\tau$  ( $|\alpha| < 1$ ) степень тесноты корреляционной связи между членами ряда и их смещенными значениями экспоненциально убывает. Для ряда  $AR(1)$  существует отличительный признак: значения его частной автокорреляционной функции равны нулю для лага  $\tau > 1$ . Указанные свойства автокорреляционной функции и частной автокорреляционной функции могут служить признаками процесса  $AR(1)$ . Выборочные коэффициенты автокорреляции частной автокорреляции могут несколько отличаться от теоретических.

Для авторегрессионной модели порядка  $p$  частная автокорреляционная функция обращается в нуль, начиная со значения аргумента  $\tau = p + 1$ .

На практике часто встречаются модели  $AR(2)$  с остатками в виде белого шума.

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t.$$

Они называются процессами Юла.

Модель скользящего среднего порядка  $q$  ( $MA(q)$ ) имеет вид:

$$y_t = \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \dots + \gamma_q \varepsilon_{t-q}.$$

Характерной особенностью модели  $MA(1)$  является условие обращения в нуль автокорреляционной функции ряда  $y_t$  для лагов  $\tau \geq 2$ .

На практике чаще используется объединенная модель  $ARMA(1,1)$ :

$$y_t = \alpha_1 y_{t-1} + \varepsilon_t + \gamma \varepsilon_{t-2}.$$

Стационарность этого процесса выполняется при условии  $|\alpha| < 1$ .

Условие  $|\gamma| < 1$  обеспечивает возможность выражения текущих значений  $y_t$  в виде сходящегося ряда своих прошлых значений  $y_{t-1}$ ,  $y_{t-2}$  и т. д.

Для процесса  $ARMA(1,1)$  значения автокорреляционной функции будут экспоненциально убывать от значений  $\rho(1)$ , причем, если  $\alpha$  имеет положительное значение – то убывать монотонно, если отрицательное – то знакопеременно.

Поведение частной автокорреляционной функции определяется начальным значением  $\rho_q(1)$ , после которого функция экспоненциально убывает. Если  $\gamma$  имеет положительное значение, то функция убывает монотонно, если отрицательное – то знакопеременно.

Анализ поведения автокорреляционной функции и частной автокорреляционной функции оказывают существенную помощь в выборе вида модели.

Нестационарным однородным называется временной ряд, для которого случайный остаток, полученный вычитанием из исходного ряда его неслучайной составляющей, представляет собой стационарный временной ряд (см. п. 7.2: стационарный относительно тренда ряд).

#### 7.4. Изучение взаимосвязей по временным рядам

При построении регрессии на временных рядах следует учитывать следующие особенности:

1. Наличие сезонной компоненты в уровнях ряда приводит к искажению истинных показателей силы и тесноты связи изучаемых временных рядов, поэтому на этапе анализа временных рядов

проводится устранение сезонной компоненты. Методы исключения сезонной компоненты описаны в п. 7.2.

2. Наличие тенденции в каждом временном ряде приводит к «ложной корреляции», что искажает статистические характеристики, описывающие причинно-следственную связь между показателями, поэтому на этапе анализа временных рядов проводится устранение тенденции.

При построении моделей регрессии по временным рядам с детерминированным трендом для устранения тенденции используются следующие методы.

*Метод отклонений от тренда* предполагает вычисление трендовых значений для каждого временного ряда модели, например  $\hat{y}_t$  и  $\hat{x}_t$ , и расчет отклонений от трендов:  $y_t - \hat{y}_t$  и  $x_t - \hat{x}_t$ . Для дальнейшего анализа используют отклонения от тренда.

*Метод последовательных разностей* заключается в следующем: если ряд содержит линейный тренд, тогда исходные данные заменяются первыми разностями:

$$\Delta_t = y_t - y_{t-1};$$

если параболический тренд – вторыми разностями:

$$\Delta_t^2 = \Delta_t - \Delta_{t-1}.$$

В случае экспоненциального и степенного тренда метод последовательных разностей применяется к логарифмам исходных данных.

*Включение в модель регрессии фактора времени* предполагает построение модели следующего вида:

$$y_t = a + b_1 x_t + b_2 t + \varepsilon_t.$$

Параметры  $a$  и  $b$  этой модели определяются обычным МНК. Модель с включением фактора времени используют в случае, если ряд  $x_t$  тенденции не содержит, ряд  $y_t$  содержит линейную тенденцию.

Для оценки адекватности регрессионной модели по временным рядам, в сравнении с традиционным корреляционно-регрессионным анализом, дополнительно требуется анализ остатков  $\varepsilon_t$  для подтверждения предположений относительно их некоррелированности

и постоянства дисперсии. С этой целью рассчитывается статистика Дарбина–Уотсона и оценивается гомоскедастичность остатков.

Регрессионные модели, содержащие не только текущие, но и лаговые значения факторных переменных, называются *моделями с распределенным лагом*. Модель с распределенным лагом в предположении, что максимальная величина лага конечна, имеет вид:

$$y_t = a + b_0x_t + b_1x_{t-1} + \dots + b_px_{t-p} + \varepsilon_t.$$

Модели, содержащие в качестве факторов лаговые значения зависимой переменной, называются моделями *авторегрессии*, например:

$$y_t = a + b_0x_t + b_1x_{t-1} + \varepsilon_t.$$

В этом случае автокорреляция остатков модели оценивают с помощью *h*-статистики Дарбина:

$$h = (1 - 0,5DW) \sqrt{\frac{n}{1 - n \cdot S^2(b)}},$$

где  $S^2(b)$  – квадрат стандартной ошибки коэффициента регрессии при лаговой зависимой переменной ( $b$ ).

При большой выборке гипотеза отсутствия автокорреляции остатков может быть отклонена при уровне значимости 5 %, если  $|h| > 1,96$ . При этом метод наименьших квадратов для оценки параметров данного уравнения корректно применять в случае, если  $|b| < 1$ .

*Пример 7.1.* По годовым данным (табл. 7.5) построить модель, описывающую взаимосвязь спроса и цены на продукцию. Выполнить прогноз объема производства при условии, что цена увеличится на 1 %.

Таблица 7.5

Годовые данные среднедневного спроса и темпа прироста цены

Год	Среднедневной спрос на продукцию (ед.)	Темп прироста цены на продукцию (%)
1	5	–2,0
2	7	–3,5
3	4	–1,5

Год	Среднедневной спрос на продукцию (ед.)	Темп прироста цены на продукцию (%)
4	3	-1,0
5	4	-1,5
6	7	-4,0
7	6	-2,5
8	6	-2,7
9	7	-3,5
10	5	-2,1
11	3	-1,0
12	4	-1,6
13	7	-4,0
14	6	-2,5
15	7	-4,1
16	5	-2,0
17	4	-1,5

*Решение.* В данной задаче в качестве зависимого показателя  $y$  выступает спрос, в качестве фактора  $x$  – темп прироста цены. Вначале исследуем структуру временных рядов. Поскольку данные годовые, то сезонность отсутствует. Для определения наличия тренда рассчитаем для каждого ряда коэффициенты автокорреляции первого порядка:  $r_1^y = 0,22$ ,  $r_1^x = 0,004$ . Эти характеристики свидетельствуют об отсутствии в рядах линейного тренда.

Далее построим линейную функцию для описания зависимости спроса от цены:

$$\hat{y} = 2,1 - 1,32x, R^2 = 0,94, DW = 2,3, A = 6,2 \%;$$

$$t_{\text{ст}}(9,1)(-15)F = 223,7.$$

График остатков визуально позволяет сделать предположение об их гомоскедастичности.

Все проверяемые статистические характеристики являются удовлетворительными, из чего заключаем, что уравнение может быть использовано для прогноза. Используя данное уравнение, можно рассчитать, что спрос снизится до  $2,1 - 1,32 \cdot 1 = 0,78$  (ед.).

Чтобы усилить в прогнозе влияние последних наблюдений, удобно ориентироваться на параметр регрессии, показывающий, что с ростом цены на 1 % спрос в среднем уменьшится на 1,32 ед., следовательно, в прогнозном периоде составит  $4,0 - 1,32 = 2,68$ .

## 7.5. Статическая модель Леонтьева

Рассмотрим наиболее простой вариант модели межотраслевого баланса (другое ее название – модель Леонтьева, или модель «затраты–выпуск»).

Различают отчетный и плановый межотраслевые балансы. Такие балансы могут составляться для страны, региона и предприятия.

*Отчетный межотраслевой баланс* отражает структуру производства и потребления продукции, произведенной в стране за отчетный год. *Плановый межотраслевой баланс* предназначен для планирования производства валового внутреннего продукта. В некоторых странах с рыночной экономикой такой план разрабатывается, но является индикативным, т. е. необязательным, а нацеливающим субъектов экономики на рациональные, с точки зрения общества, действия.

Эффективное функционирование экономики предполагает наличие баланса между отдельными отраслями. Каждая отрасль при этом выступает как производитель некоторой продукции и как потребитель продуктов, вырабатываемых другими отраслями.

Алгебраическая теория анализа «затраты–выпуск» сводится к системе линейных уравнений, в которых параметрами являются коэффициенты затрат на производство продукции.

Пусть весь производственный сектор народного хозяйства разбит на  $n$  чистых отраслей. Условное понятие «чистая отрасль» – некоторая часть народного хозяйства, относительно цельная (например, энергетика, машиностроение, сельское хозяйство и т. п.). Введем следующие обозначения:

$x_{ij}$  – количество продукции  $i$ -й отрасли, расходуемое в  $j$ -й отрасли;  
 $X_i$  – объем производства  $i$ -й отрасли за данный промежуток времени, или валовой выпуск  $i$ ;

$Y_i$  – объем потребления продукции  $i$ -й отрасли в непродуцированной сфере, объем конечного потребления;



$Z_j$  – условно чистая продукция  $j$ -й отрасли, включающая оплату труда, чистый доход и амортизацию.

Единицы измерения всех указанных величин могут быть или натуральными ( $m^3$ , т, шт. и т. п.) или стоимостными. В зависимости от этого различают межотраслевые балансы (МОБ): натуральный и стоимостной. Рассмотрим принципиальную схему стоимостного МОБ (табл. 7.6).

Таблица 7.6

Принципиальная схема МОБ в стоимостном выражении

Производящие отрасли	Потребляющие отрасли				Конечный продукт	Валовой продукт
	1	2	...	$n$		
1	$x_{11}$	$x_{12}$	...	$x_{1n}$	$Y_1$	$X_1$
2	$x_{21}$	$x_{22}$	...	$x_{2n}$	$Y_2$	$X_2$
...	...	...	...	...	...	...
$n$	$x_{n1}$	$x_{n2}$	...	$x_{nn}$	$Y_n$	$X_n$
Условно чистая продукция	$Z_1$	$Z_2$	...	$Z_n$	$\sum_{i=1}^n Y_i = \sum_{j=1}^n Z_j$	
Валовой продукт	$X_1$	$X_2$	...	$X_n$		$\sum_{i=1}^n X_i = \sum_{j=1}^n X_j$

1. Анализируя схему баланса по столбцам, можно сделать вывод, что итог материальных затрат любой потребляющей отрасли и ее условно чистой продукции равен валовой продукции этой отрасли. Данный вывод можно записать в виде соотношения:

$$X_j = \sum_{i=1}^n x_{ij} + Z_j, \quad j = \overline{1, n}. \quad (7.3)$$

Соотношение (7.3) охватывает систему из  $n$  уравнений, отражающих стоимостной состав продукции всех отраслей материальной сферы.

2. Анализируя схему МОБ по строкам для каждой производящей отрасли, можно видеть, что валовая продукция той или иной отрасли равна сумме материальных затрат потребляющих ее продукцию отраслей и конечной продукции данной отрасли:

$$X_j = \sum_{i=1}^n x_{ij} + Y_j, \quad i = \overline{1, n}. \quad (7.4)$$

Формула (7.4) описывает систему из  $n$  уравнений, которые называются уравнениями распределения продукции отраслей материального производства по направлениям использования.

Балансовый характер таблицы выражается в том, что

$$\sum_{i=1}^n X_i = \sum_{j=1}^n X_j,$$

$$\sum_{i=1}^n Y_i = \sum_{j=1}^n Z_j.$$

Основу экономико-математической модели МОБ составляет матрица коэффициентов прямых затрат  $A = (a_{ij})$ . Коэффициент прямых затрат  $a_{ij}$  показывает, какое количество продукции  $i$ -й отрасли необходимо (если учитывать только прямые затраты) для производства единицы продукции  $j$ -й отрасли:

$$a_{ij} = \frac{x_{ij}}{X_j}, \quad i, j = \overline{1, n}.$$

Для дальнейшего рассмотрения модели Леонтьева сделаем два важных предположения.

Первое состоит в том, что сложившуюся технологию производства считаем неизменной. Таким образом, матрица  $A = (a_{ij})$  постоянна.

Второе состоит в постулировании свойства линейности существующих технологий, т. е. для выпуска  $j$ -й отраслилю любого объема продукции  $X_j$  необходимо затратить продукцию отрасли  $i$  в количестве  $a_{ij}X_j$ , т. е. материальные издержки пропорциональны объему производимой продукции:

$$x_{ij} = a_{ij}X_j. \quad (7.5)$$

Подставляя формулу (7.5) в балансовое соотношение (формула (7.4)), получаем:

$$X_i = \sum_{j=1}^n a_{ij}X_j + Y_i,$$

или в матричной форме:

$$X = AX + Y. \quad (7.6)$$

С помощью этой модели можно выполнять три вида плановых расчетов:

- задав в модели величины валовой продукции каждой отрасли ( $X_j$ ), можно определить объемы конечной продукции каждой отрасли ( $Y_j$ ):

$$Y = (E - A)X; \quad (7.7)$$

- задав величины конечной продукции всех отраслей ( $Y_j$ ), можно определить величины валовой продукции каждой отрасли ( $X_j$ ):

$$X = (E - A)^{-1}Y; \quad (7.8)$$

- задав величины валовой продукции для ряда отраслей и объемы конечной продукции для всех остальных отраслей, можно найти величины конечной продукции первых отраслей и объемы валовой продукции вторых.

В формулах (7.7) и (7.8)  $E$  – единичная матрица  $n$ -го порядка,  $(E - A)^{-1}$  – матрица, обратная к матрице  $(E - A)$ . Если определитель матрицы  $(E - A)$  не равен нулю (т. е. эта матрица невырожденная), то обратная к ней матрица существует. Обозначим эту обратную матрицу через  $B = (E - A)^{-1}$ , тогда систему уравнений (7.8) в матричной форме можно записать в виде:

$$X = B \cdot Y.$$

Элементы матрицы  $B$  называются коэффициентами полных затрат. Они показывают, сколько всего нужно произвести продукции

$i$ -й отрасли для выпуска в сферу конечного использования единицы продукции  $j$ -й отрасли.

Плановые расчеты по модели Леонтьева можно проводить, если выполняется условие продуктивности.

Будем называть неотрицательную матрицу  $A$  продуктивной, если существует такой неотрицательный вектор  $X \geq 0$ , что

$$X > AX . \quad (7.9)$$

Очевидно, что условие (формула (7.9)) означает существование положительного вектора конечной продукции  $Y > 0$  для модели межотраслевого баланса (формула (7.6)).

Для того чтобы матрица коэффициентов прямых затрат  $A$  была продуктивной, необходимо и достаточно выполнение одного из условий:

1) матрица  $(E - A)$  неотрицательно обратима, т. е. существует обратная матрица  $(E - A)^{-1} \geq 0$ ;

2) матричный ряд  $E + A + A^2 + A^3 + \dots = \sum_{k=0}^{\infty} A^k$  сходится, причем его сумма равна обратной матрице  $(E - A)^{-1}$ ;

3) наибольшее по модулю собственное значение  $\lambda$  матрицы  $A$ , т. е. решение характеристического уравнения  $|\lambda E - A| = 0$  строго меньше единицы;

4) все главные миноры матрицы  $(E - A)$ , т. е. определители матриц, образованные элементами первых строк и первых столбцов этой матрицы, порядка от 1 до  $n$  положительны.

Более простым, но только достаточным условием продуктивности матрицы  $A$  является ограничение на величину ее нормы, т. е. на величину наибольшей из сумм элементов матрицы  $A$  в каждом столбце: если норма матрицы  $A$  строго меньше единицы, то эта матрица продуктивна. Поскольку данное условие является только достаточным, матрица  $A$  может оказаться продуктивной и в случае, когда ее норма больше единицы.

*Пример 7.2.* Даны коэффициенты прямых затрат  $a_{ij}$  и конечный продукт  $Y_j$  для трехотраслевой экономической системы:

$$A = \begin{pmatrix} 0,3 & 0,1 & 0,4 \\ 0,2 & 0,5 & 0 \\ 0,3 & 0,1 & 0,2 \end{pmatrix}, Y = \begin{pmatrix} 200 \\ 100 \\ 300 \end{pmatrix}.$$

1. Определить коэффициенты полных затрат, вектор валового выпуска, межотраслевые поставки продукции.
2. Проверить продуктивность матрицы  $A$ .
3. Заполнить схему межотраслевого баланса.

*Решение.* Воспользуемся функциями Excel. В табл. 7.7 приведены результаты решения задачи.

Таблица 7.7

Результаты решения

Номер строки	A	B	C	D	E	F	G
1							
2		0,3	0,1	0,4			
3	<b>A</b>	0,2	0,5	0			
4		0,3	0,1	0,2			
5							
6		0,7	-0,1	-0,4			
7	$E - A$	-0,2	0,5	0			
8		-0,3	-0,1	0,8			
9	1.						
10		2,04	0,61	1,02			200
11	<b>B</b>	0,82	2,24	0,41		<b>Y</b>	100
12		0,87	0,51	1,68			300
13							
14	2.						
15		775					
16	<b>X</b>	511					
17		729					
18							
19	3.						

Номер строки	A	B	C	D	E	F	G
20		232,5	51,1	291,6			
21	$(X_{ij})$	155	255,5	0			
22		2325	51,1	145,8			

В ячейки B6:D8 (см. табл. 7.7) запишем элементы матрицы  $E-A$ . Массив  $E-A$  задан как диапазон ячеек. Выделим диапазон B10:D12 для размещения обратной матрицы  $B = (E - A)^{-1}$  и введем формулу для вычислений МОБР(B6:D8). Затем следует нажать на клавиши Ctrl+Shift+Enter. Все элементы матрицы коэффициентов полных затрат B неотрицательны, следовательно, матрица A продуктивна.

В ячейки G10:G12 запишем элементы вектора конечного продукта Y. Выделим диапазон B15:B17 для размещения вектора валового выпуска X, вычисляемого по формуле  $X = (E - A)^{-1}Y$ , введем формулу для вычислений МУМНОЖ(B10:D12, G10:G12). Затем следует нажать на клавиши Ctrl+Shift+Enter.

Межотраслевые поставки  $x_{ij}$  вычисляем по формуле  $x_{ij} = a_{ij}X_j$ .

Заполняем схему МОБ (табл. 7.8).

Таблица 7.8

Схема МОБ

Производящие отрасли	Потребляющие отрасли			Конечный продукт	Валовой продукт
	1	2	3		
1	232,5	51,1	291,6	200	775
2	155	255,5	0	100	511
3	232,5	51,1	145,8	300	729
Условно чистая продукция	155	153	291,6	<b>600</b>	–
Валовой продукт	775	511	729	–	<b>2015</b>

## Вопросы для контроля знаний

1. Каковы основные компоненты временного ряда?
2. В чем отличие между аддитивной и мультипликативной моделями временного ряда?
3. Какой ряд называют стационарным?
4. Что такое «автокорреляционная функция временного ряда»?
5. Что такое «частная автокорреляционная функция временного ряда»?
6. Опишите алгоритм построения аддитивной модели с включением сезонной компоненты.
7. В чем различия методик построения регрессионных моделей на пространственных данных и временных рядах?
8. Опишите способы устранения тренда в моделях регрессии на временных рядах.
9. Сформулируйте принципиальную схему МОБ в стоимостном выражении.
10. Что такое «понятие продуктивной матрицы»?

## Практическое занятие № 9

### Моделирование компонент временных рядов для построения трендовых и сезонных моделей

**Цель работы:** отработать практические навыки выделения тренда и сезонной компоненты, выполнить проверку тренд-стационарности.

### Методические рекомендации 1

1. Исходные данные для построения трендовой модели – урожайность яровых за 44 года, вводим в Excel (табл. 1). Но для удобства последующего отображения на графике год заменим порядковым номером.

2. Строим график с выделением линии тренда – проведенный визуальный анализ показывает наличие линейного тренда. Однако можно предположить нарушение гомоскедастичности остатков.

3. Для оценки качества модели проводим регрессионный анализ с выводом остатков (выходной интервал – А66). Качество модели достаточно хорошее. Коэффициент детерминации  $R$ -квадрат равен

0,699, *t*-статистики значительно превосходят критическое значение – коэффициенты уравнения значимы.

Таблица 1

Исходные данные урожайности яровых зерновых

Год	Яровые зерновые культуры, ц/га	Год	Яровые зерновые культуры, ц/га
1972	16,4	1994	45,7
1973	33,8	1995	41,3
1974	32,8	1996	48,6
1975	20,4	1997	41,4
1976	25,3	1998	40,9
1977	28,1	1999	35,0
1978	28,1	2000	57,8
1979	37,3	2001	53,2
1980	22,4	2002	51,1
1981	32,1	2003	48,2
1982	35,7	2004	60,3
1983	37,2	2005	58,4
1984	46,1	2006	53,4
1985	39,7	2007	52,6
1986	43,6	2008	62,5
1987	48,3	2009	43,1
1988	32,8	2010	51,9
1989	41,4	2011	55,8
1990	43,7	2012	58,2
1991	46,2	2013	48,0
1992	45,6	2014	72,9
1993	54,0	2015	83,9

4. Для проверки независимости остатков (условие Гаусса–Маркова) проведем оценку их корреляции по статистике Дарбина–Уотсона. В ячейку E90 введем формулу для ее вычисления: СУММКВРАЗН(C91:C133; C90:C132)/СУММКВ(C90:C133).

Полученное значение DW-статистики равно 1,612. По таблице Дарбина–Уотсона (прилож. 3) (при 44 наблюдениях принимаем



$n = 45, k = 1$ ) определяем интервал  $(d_U; 4-d_U)$ , равный  $(1,544; 2,456)$ , принятия гипотезы  $H_0$  об отсутствии автокорреляции остатков. Следовательно, условия Гаусса–Маркова выполняются.

5. Построим прогноз урожайности на 2018 год (порядковый номер года – 45). По уравнению регрессии имеем:

$$y_{\text{пр}} = 24,780 + 0,874 \cdot 45 = 65,839 .$$

6. Для нахождения ошибки прогноза находим  $\bar{x} = 22,5$  (функция СРЗНАЧ), выборочную дисперсию  $s_x^2 = 165$  (функция ДИСП.В). Стандартную ошибку регрессии берем из протокола регрессии:  $s = 7,443$ .

7. Рассчитываем ошибку прогноза:

$$s_{\hat{y}} = s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{\text{пр}} - \bar{x})^2}{n \cdot s_x^2}} = 7,443 \cdot \sqrt{1 + \frac{1}{44} + \frac{(47 - 22,5)^2}{44 \cdot 165}} = 7,825.$$

8. По формуле «=СТЮД.ОБР.2X(0,05;42)» или по таблице Стьюдента рассчитываем значение  $t_{\text{кр}} = 2,018$ .

9. Тогда предельная ошибка прогноза  $\Delta_{\hat{y}} = t_{\text{кр}} \cdot s_{\hat{y}} = 2,018 \cdot 7,825 = 15,791$ .

Доверительный интервал прогноза представляет собой отклонение точечного прогноза на величину предельной ошибки прогноза:  $65,839 \pm 15,791$ .

10. С вероятностью 95 % урожайность в 2018 году будет находиться в интервале от 50,05 до 81,63 ц/га.

## Методические рекомендации 2

1. Исходные данные: сбор меда с пчелосемьи  $y_i$  за 15 лет за сезон  $D$  (0 – весна, 1 – лето) вводим в Excel (табл. 2).

Таблица 2

Исходные данные сбора меда

Номер наблюдения	Сбор меда с пчелосемьи $y_i$ , кг	Сезон $D$ , (0, 1)
1	7,040	0
2	7,976	1
3	6,791	0

Номер наблюдения	Сбор меда с пчелосемьи $y_t$ , кг	Сезон $D_t$ , (0, 1)
4	7,976	1
5	7,011	0
6	8,009	1
7	6,919	0
8	7,963	1
9	7,063	0
10	7,894	1
11	6,968	0
12	8,096	1
13	7,289	0
14	8,071	1
15	6,980	0
16	7,871	1
17	7,257	0
18	8,132	1
19	6,823	0
20	8,131	1
21	6,927	0
22	8,007	1
23	6,951	0
24	8,046	1
25	7,051	0
26	7,837	1
27	6,966	0
28	7,943	1
29	6,952	0
30	8,006	1

2. В качестве объясняющей переменной, принимающей значения от 1 до 30, вводим  $t$  – номер сезона.

3. Для оценки качества модели проводим регрессионный анализ с выводом остатков. Качество модели достаточно хорошее. Коэффициент детерминации  $R$ -квадрат равен 0,953. Уравнение в целом

значимо, статистика Фишера, равная 272,855, значима. Однако  $t$ -статистика номера сезона  $t$ , равная 0,123, значительно ниже критического значения. Поэтому коэффициент при номере сезона  $t$  незначим, что указывает на отсутствие трендовой составляющей. В то же время коэффициент регрессии при сезоне  $D$ , равный 0,995, значим, что указывает на наличие сезонной компоненты. Визуальный анализ графика остатков по  $D$  указывает на их гомоскедастичность (рис. 1).

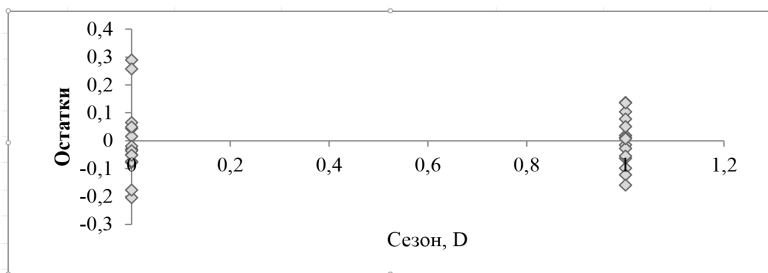


Рис. 1. График остатков для сезонной компоненты

4. Для подтверждения наличия сезонной компоненты и ее периодичности строим матрицу коэффициентов корреляции. Для этого несколько раз (в примере – 5) сдвигаем столбец значений  $y_t$  и строим корреляционную матрицу по заполненным столбцам данных сверху и снизу.

5. Строим график АКФ по первому столбцу корреляционной матрицы для столбцов 2–5 (рис. 2).

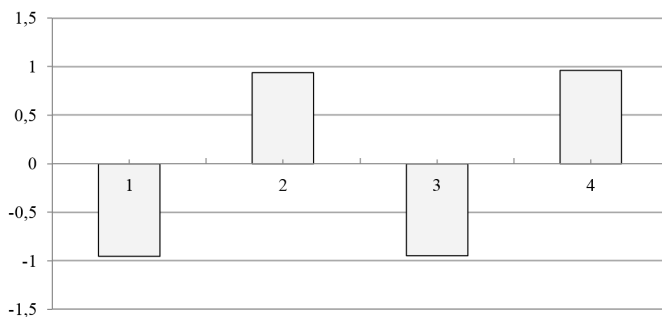


Рис. 2. График АКФ

6. Построенный график с периодом два подтверждает наличие сезонных колебаний.

7. Для построения прогноза медосбора на два сезона вперед используем уравнение регрессии:  $y_{\text{пр}} = 6,995 + 0,995D$ ;  $y_{\text{пр}}(31) = 6,995$ ;  $y_{\text{пр}}(32) = 7,989$ .

8. Для нахождения ошибки прогноза  $\bar{x} = 0,5$  находим выборочную дисперсию  $s_x^2 = 165$  (функция ДИСП.В). Стандартную ошибку регрессии берем из протокола регрессии:  $s = 0,117$ .

9. По формуле «=СТЫЮД.ОБР.2X(0,05;28)» или по таблице Стьюдента рассчитываем значение  $t_{\text{кр}} = 2,05$ .

10. Рассчитываем ошибку прогноза

$$s_{\hat{y}} = s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{\text{пр}} - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 0,117 \cdot \sqrt{1 + \frac{1}{30} + \frac{0,5^2}{30 \cdot 0,5^2}} =$$

$$= 0,117 \cdot \sqrt{\frac{16}{15}} = 0,121.$$

10. Предельная ошибка прогноза  $\Delta \hat{y} = t_{\text{кр}} \cdot s_{\hat{y}} = 2,05 \cdot 0,121 = 0,248$ .

21. Доверительные интервалы прогноза для сезонов  $t = 31$  ( $D = 0$ ) и  $t = 32$  ( $D = 1$ ) равны [6,747;7,243] и [7,741;8,237] соответственно.

### Задания для самостоятельной работы

**Задание 1.** В соответствии с номером варианта (табл. 3), по аналогии с приведенным примером, построить трендовую модель по данным табл. 4.

Таблица 3

№ варианта	Годы	№ варианта	Годы
1, 5, 9	1975–2016	3, 7, 11	1979–2017
2, 6, 10	1978–2018	4, 8, 12	1980–2018

## Исходные данные

Год	Зерновые и зернобобовые культуры, ц/га	Год	Зерновые и зернобобовые культуры, ц/га
1975	16,6	1997	45,3
1976	33,6	1998	40,2
1977	34,1	1999	43,7
1978	24,7	2000	39,3
1979	28,2	2001	39,2
1980	28,1	2002	31,2
1981	27,3	2003	47,1
1982	34,5	2004	48,1
1983	23,8	2005	51,9
1984	31,0	2006	45,1
1985	32,9	2007	56,3
1986	36,5	2008	57,5
1987	41,0	2009	47,0
1988	34,6	2010	50,7
1989	40,3	2011	57,4
1990	47,3	2012	44,2
1991	32,5	2013	46,0
1992	42,2	2014	48,3
1993	42,6	2015	58,4
1994	43,1	2016	48,1
1995	47,0	2017	63,8
1996	53,7	2018	69,7

Выполнить расчеты в Excel. Проанализировать полученные результаты. Оформить отчет.

**Задание 2.** Провести анализ ряда на сезонность, его сглаживание. Построить регрессионную модель с фиктивной переменной. Использовать данные табл. 2 в соответствии с номером варианта.

№ варианта	Включаемые наблюдения	№ варианта	Включаемые наблюдения
1, 5, 9	1–20	3, 7, 11	7–27
2, 6, 10	3–23	4, 8, 12	11–30

## Практическое занятие № 10

### Анализ развития отраслей материального производства с использованием МОБ

**Цель работы:** отработать практические навыки составления схемы межотраслевого баланса.

#### Методические рекомендации

Для пяти отраслей (промышленность, сельское хозяйство, прочие отрасли материального производства, торговля, прочие отрасли сферы услуг) за отчетный период известны межотраслевые потоки  $x_{ij}$  и вектор объемов конечного использования  $\bar{y}_{\text{отч}}$  (таблица).

*Таблица*

Исходные данные

Отрасли	Межотраслевые потоки					$\bar{y}_{\text{отч}}$	$\bar{y}_{\text{пз}}$
	1	2	3	4	5		
1	53	28	91	49	45	189	153
2	35	57	99	78	83	74	95
3	31	72	41	19	65	144	132
4	40	61	44	85	50	133	162
5	63	45	23	32	65	158	94
$k = 4$				$p = 60 \%$			

Предполагая, что в плановом периоде технология производства останется неизменной, требуется:

1. Составить схему планового межотраслевого баланса, если задан вектор конечного использования  $\bar{y}_{\text{пз}}$ .

2. Определить, как изменятся цены в отраслях, если известно, что номинальный объем валовой добавленной стоимости 4-й отрасли увеличится на 60 % по сравнению с отчетным периодом, а в других отраслях останется неизменным.

Подготовительная работа.

1. Сформируем в Excel рабочую таблицу «Отчетный МОБ» (рис. 1) и занесем в нее данные межотраслевых потоков и конечного использования (КИ) по отраслям.

	A	B	C	D	E	F	G	H	I
1									
2	Отрасли- производители	Отрасли-потребители					ПП	КИ	ВВ
3		1	2	3	4	5			
4	1	53	28	91	49	45	266	189	455
5	2	35	57	99	78	83	352	74	426
6	3	31	72	41	19	65	228	144	372
7	4	40	61	44	85	50	280	133	413
8	5	63	45	23	32	65	228	158	386
9	ПЗ	222	263	298	263	308	1354	698	2052
10	ВДС	233	163	74	150	78	698		
11	ВВ	455	426	372	413	386	2052		

Рис. 1. Рабочая таблица «Отчетный МОБ»

2. Элементы столбца ПП (промежуточное потребление) рассчитаем по формуле  $\sum_{j=1}^n x_{ij}$ . Для этого курсор помещаем в ячейку G4

столбца ПП и вызываем функцию СУММ из меню **Вставка–Функция–Математические**. В качестве аргумента берем элементы межотраслевых потоков первой строки. Затем копируем эту формулу в остальные ячейки столбца ПП.

3. Рассчитываем элементы столбца ВВ (валовой выпуск) по формуле  $ВВ=ПП+КИ$ . В ячейку I4 вводим формулу  $I4=G4+H4$ . Затем копируем введенную формулу в ячейки I5:I8, аналогично столбцу ПП.

4. Строка ПЗ (промежуточные затраты) формируется суммированием межотраслевых потоков по столбцам:  $В9=СУММ(В4:В8)$  и т. д.

5. Элементы строки ВЗ (валовые затраты) должны соответствовать элементам столбца ВВ (валовой выпуск). Используем функцию ТРАНСП. Выделяем ячейки B11:F11 и вызываем функцию ТРАНСП (**Вставка–Функция–Ссылки и массивы**). В качестве аргумента введем массив I4:I8. Далее нажимаем сочетание клавиш Ctrl+Shift+OK. Если значение появилось только в одной ячейке, то необходимо заново выделить массив B11:F11, нажать на кнопку F2 и снова нажать на сочетание клавиш Ctrl+Shift+OK. Причем первые две нужно нажать и удерживать.

6. В последнюю очередь рассчитываем строку ВДС (валовая добавленная стоимость):  $В10=В11-В9$  и т. д.


*Шаг 1.* Составить схему планового межотраслевого баланса, если известен вектор конечного использования  $\bar{y}_{\text{пз}}$ .

Построим матрицу прямых затрат А (рис. 2). В ячейку В13 введем формулу В13=В4/В\$11. Знак доллара ставим, чтобы при копировании до ячейки F17 не изменялся адрес строки 11. Главную диагональ массива, отведенного под единичную матрицу, заполним единицами. Матрицы складываются и вычитаются поэлементно, поэтому в ячейку В19 матрицы  $(E - A)$  записываем соответствующую формулу и копируем ее до ячейки F23.

	0,116484	0,065728	0,244624	0,118644	0,11658		1									
	0,076923	0,133803	0,266129	0,188862	0,215026			1								
A	0,068132	0,169014	0,110215	0,046005	0,168394	E			1							
	0,087912	0,143192	0,11828	0,205811	0,129534					1						
	0,138462	0,105634	0,061828	0,077482	0,168394						1					
	0,883516	-0,06573	-0,24462	-0,11864	-0,11658		1,298556	0,307864	0,523293	0,338524	0,420336					
	-0,07692	0,866197	-0,26613	-0,18886	-0,21503		0,32399	1,461458	0,635574	0,494122	0,628969					
E-A	-0,06813	-0,16901	0,889785	-0,046	-0,16839	B	0,232784	0,379321	1,365195	0,247545	0,445713					
	-0,08791	-0,14319	-0,11828	0,794189	-0,12953		0,285972	0,403442	0,42625	1,467306	0,459271					
	-0,13846	-0,10563	-0,06183	-0,07748	0,831606		0,301314	0,30269	0,309074	0,274245	1,4283					

Рис. 2. Матрицы прямых и полных затрат

Матрицу В рассчитываем по формуле  $B = (E - A)^{-1}$ . Для ее расчета используем встроенную функцию **МОБР**. Выделяем массив Н19:L23, вызываем функцию **МОБР**, вводим ячейки матрицы  $(E - A)$ , используя сочетание клавиш Ctrl+Shift+ОК.

*Важно! Проверьте все введенные формулы на правильность. Если курсор стоит в строке формулы и вы не можете из нее выйти, потому что нельзя изменять часть массива, то нажмите на кнопку  слева от строки формул.*

Составим схему межотраслевого баланса на плановый период. В ячейки Н28:Н32 запишем известные значения конечного спроса, прогнозируемого на плановый период. Валовой выпуск найдем, используя прогнозную модель  $\bar{x} = B \cdot \bar{y}$ . Выделяем ячейки I28:I32, вызываем математическую функцию **МУМНОЖ**. Вводим адреса двух массивов: массив 1 – матрица В, массив 2 – вектор значений конечного использования.



Строки В3, П3, DLC формируем аналогично отчетному периоду. Межотраслевые потоки рассчитываем, используя элементы матрицы  $A$  по формуле  $x_{ij} = a_{ij} X_j$ . Проверим правильность расчетов, используя несколько вариантов расчета ВВП.

*Ответьте на вопросы:*

1. На сколько процентов изменились ВВП и валовой выпуск в целом по экономике в плановом периоде по отношению к отчетному?

2. Каковы значение элемента  $a_{ki}$  (см. условие) матрицы прямых затрат и его экономический смысл?

3. Каковы значение элемента  $b_{ii}$  (см. условие) матрицы полных затрат и его экономический смысл?

4. Чему равны материальные затраты  $k$ -й отрасли в отчетном периоде? материалоемкость  $l$ -й отрасли?

*Шаг 2.* Определить, как изменятся цены в отраслях, если известно, что номинальный объем валовой добавленной стоимости  $k$ -й отрасли увеличится на  $p$  % по сравнению с отчетным периодом, а в других отраслях останется неизменным:  $k = 4, p = 60$  %.

1. Самостоятельно составьте модель для определения индексов цен в отраслях:

$p_1 \dots p_5$  – цены за единицу продукции в отраслях;

$455p_1$  – стоимость продукции I отрасли;

$455p_1 = 53p_1 + 35p_2 + \dots + 65p_5 + 233$ .

В таблице отчетного межотраслевого баланса (см. рис. 1) имеются данные о валовых выпусках отраслей (ВВ) и валовой добавленной стоимости (ВДС) отраслей и межотраслевых потоках на производственное потребление. Согласно заданию НДС 4-й отрасли вместо значения 150 примет значение  $240 = 150(1 + 60 / 100)$ . Обозначая цену единицы продукции  $i$ -й отрасли как  $p_i$  ( $i = \overline{1,5}$ ), получим систему уравнений:

$$\begin{cases} 455p_1 = 53p_1 + 35p_2 + 31p_3 + 40p_4 + 65p_5 + 233, \\ 426p_2 = 28p_1 + 57p_2 + 72p_3 + 61p_4 + 45p_5 + 163, \\ 372p_3 = 91p_1 + 99p_2 + 41p_3 + 44p_4 + 25p_5 + 74, \\ 372p_4 = 49p_1 + 78p_2 + 19p_3 + 85p_4 + 32p_5 + 240, \\ 386p_5 = 45p_1 + 83p_2 + 65p_3 + 50p_4 + 65p_5 + 78. \end{cases}$$

Перенесем слагаемые с неизвестными  $p_i$  в левые части равенств, обозначим векторы-столбцы  $(\mathbf{P})^T = (p_1, p_2, p_3, p_4, p_5)$  и  $(\mathbf{b})^T = (233, 163, 74, 240, 78)$ . После приведения подобных запишем систему в матричном виде:  $\mathbf{M}(\mathbf{p}) = (\mathbf{b})$ . С помощью функций Excel МОБР и МУМНОЖ получим решение системы  $\bar{\mathbf{p}} = \mathbf{M}^{-1}(\mathbf{b})$ .

Индексом цены  $p_i$  называется отношение  $\frac{\bar{p}_i - p_i}{p_i} \cdot 100\%$ . Запишите систему равенств для расчета индексов цен.

*Ответьте на вопрос: чему равны индексы цен?*

Оформите отчет о проделанной работе.

### **Задания для самостоятельной работы**

Выполнить расчеты в Excel. Проанализировать полученные результаты. Оформить отчет.

## Тема 8. СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ

### Вопросы

- 8.1. Понятие и примеры систем одновременных уравнений.
- 8.2. Зависимость эндогенных переменных от экзогенных.
- 8.3. Счетное правило и условие ранга для идентифицируемости модели.
- 8.4. Методы наименьших квадратов, максимального правдоподобия и инструментальных переменных для оценивания коэффициентов структурной модели.

### 8.1. Понятие и примеры систем одновременных уравнений

При построении моделей экономических систем эконометрические соотношения включают целый набор уравнений и тождеств, что вызвано сложностью моделей динамики систем. Предварительные понятия и два примера систем уравнений (модель спроса-предложения и модель потребления Самуэльсона–Хикса) приведены в теме 1.

Наибольшее распространение в эконометрических исследованиях получили системы одновременных уравнений, в которых одни и те же зависимые переменные могут входить и в левую, и в правую часть уравнений системы:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \dots + b_{1m}y_m + a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \dots + b_{2m}y_m + a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + \dots + b_{3m}y_m + a_{31}x_1 + a_{32}x_2 + \dots + a_{3n}x_n + \varepsilon_3, \\ \dots \\ y_m = b_{m1}y_1 + b_{m2}y_2 + \dots + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n + \varepsilon_m. \end{cases} \quad (8.1)$$

Вид связей между переменными и вид уравнений строятся на основе законов экономической теории. В качестве объясняющих переменных могут выступать не только экзогенные переменные, но и лаговые значения эндогенных переменных. В зависимости от вида матрицы  $(b_{ij})$  различают следующие виды систем эконометрических уравнений:

- система независимых уравнений характеризуется тем, что матрица  $B = (b_{ij})$  является диагональной, и каждая эндогенная переменная выражается через экзогенные переменные;
- система рекурсивных уравнений характеризуется тем, что каждая эндогенная переменная является объясняющей переменной в следующих за ней уравнениях:

$$\begin{cases} y_1 = a_{01} + a_{11}x_1 + a_{12}x_2 \dots + \varepsilon_1, \\ y_2 = a_{02} + a_{21}y_1 + a_{21}x_1 + a_{22}x_2 \dots + \varepsilon_2, \\ y_3 = a_{03} + a_{31}y_1 + a_{32}y_2 + a_{31}x_1 + a_{32}x_2 \dots + \varepsilon_3, \\ \dots \end{cases}$$

- система взаимозависимых уравнений, в отличие от предыдущих, содержит обратные связи между эндогенными переменными, т. е. одни и те же переменные рассматриваются как зависимые в одних уравнениях и как объясняющие – в других.

Система взаимозависимых уравнений получила название системы совместных, одновременных уравнений. Тем самым подчеркивается, что в системе одни и те же переменные одновременно рассматриваются как зависимые в одних уравнениях и как независимые – в других. В эконометрике эта система уравнений называется также структурной формой модели. В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим. С этой целью используются специальные приемы оценивания. Необходимость их использования в системе рекурсивных и одновременных уравнений можно аргументировать коррелированностью случайных остатков с переменными в правых частях уравнений модели.

## 8.2. Зависимость эндогенных переменных от экзогенных

Разделение переменных на эндогенные и экзогенные зависит от теоретической концепции принятой модели. Экономические переменные могут выступать в одних моделях как эндогенные, в других – как экзогенные переменные. Переменные не экономического характера (например, климатические условия, социальное положение, пол, возрастная категория) входят в систему только как



Коэффициенты приведенной формы модели представляют собой, как правило, нелинейные функции коэффициентов структурной формы модели. Рассмотрим это положение на примере простейшей структурной модели, выразив коэффициенты приведенной формы модели через коэффициенты структурной модели.

Для структурной модели вида:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2, \end{cases} \quad (8.3)$$

приведенная форма модели имеет вид:

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

Из первого уравнения системы (8.3) можно выразить  $y_2$  следующим образом (ради упрощения опускаем случайные остатки):

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

Подставив значение  $y_2$  во второе уравнение (формула (8.3)), имеем:

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

откуда

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

Поступая аналогично со вторым уравнением (формула (8.3)), получим

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

т. е. система уравнений (8.3) принимает вид:

$$\begin{cases} y_1 = \frac{a_{11}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}b_{12}}{1 - b_{12}b_{21}}x_2, \\ y_2 = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}}{1 - b_{12}b_{21}}x_2. \end{cases}$$

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

$$\delta_{11} = \frac{a_{11}}{1 - b_{12}b_{21}}, \quad \delta_{12}y_2 = \frac{a_{22}b_{12}}{1 - b_{12}b_{21}};$$

$$\delta_{21} = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}, \quad \delta_{22} = \frac{a_{22}b_{21}}{1 - b_{12}b_{21}}.$$

Приведенная форма модели хотя и позволяет получить значения эндогенных переменных через значения экзогенных переменных, но она уступает структурной форме модели, так как, с одной стороны, в ней отсутствуют оценки взаимосвязи между эндогенными переменными, с другой стороны, экономический смысл коэффициентов с трудом поддается интерпретации.

### 8.3. Счетное правило и условие ранга для идентифицируемости модели

При переходе от приведенной формы модели к структурной возникает проблема идентификации, т. е. возможности выражения (единственным образом) структурных коэффициентов модели через приведенные.

Структурная модель (формула (8.1)) в полном виде содержит  $m(m + n - 1)$  параметров, а приведенная форма модели в полном виде содержит  $m \cdot n$  параметров, т. е. в полном виде структурная модель содержит большее число параметров, чем приведенная форма модели. Соответственно,  $m(m + n - 1)$  параметров структурной модели не могут быть однозначно определены на основе  $m \cdot n$  параметров приведенной формы модели. При оценке системы одновременных уравнений подлежит оценке также и ковариационная матрица случайных остатков, поэтому для идентификации модели вводится априорная информация о виде матрицы  $B = (b_{ij})$ , введением известных тождеств между переменными, линейных или нелинейных соотношений между коэффициентами, исключением некоторых переменных из уравнений, требованием обнуления соответствующих элементов ковариационной матрицы остатков.

По выполнению свойства идентифицируемости структурные модели можно подразделить на три вида:

- идентифицируемые,
- неидентифицируемые,
- сверхидентифицируемые.

Модель идентифицируема, если все ее структурные коэффициенты однозначно выражаются через коэффициенты приведенной формы модели. В этом случае оценки приведенных коэффициентов позволяют однозначно получить оценки структурных коэффициентов.

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

Модель сверхидентифицируема, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. Сверхидентифицируемая модель, в отличие от неидентифицируемой модели, практически решаема, но требует для этого специальных методов оценок параметров.

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

Выполнение условия идентифицируемости модели проверяется для каждого уравнения системы. Чтобы уравнение было идентифицируемо, необходимо, чтобы число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без единицы.

Если обозначить число эндогенных переменных в  $i$ -м структурном уравнении системы через  $H$ , а число экзогенных (предопределенных) переменных, которые содержатся в системе, но не входят в данное уравнение, – через  $D$ , то необходимое условие идентифи-



цируемости модели может быть записано в виде счетного правила (условия порядка) (табл. 8.1).

Таблица 8.1

Условия идентифицируемости

<i>Если <math>D + 1 = H</math></i>	уравнение идентифицируемо
<i>Если <math>D + 1 &lt; H</math></i>	уравнение неидентифицируемо
<i>Если <math>D + 1 &gt; H</math></i>	уравнение сверхидентифицируемо

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

Рассмотренное счетное правило отражает необходимое, но недостаточное условие идентификации. Более точно условия идентификации определяются, если накладывать ограничения на коэффициенты матриц параметров структурной модели. Проиллюстрируем счетное правило идентификации на следующем примере системы одновременных уравнений:

$$\begin{cases} y_1 = a_1 + b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2 + \xi_1, \\ y_2 = a_2 + b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \xi_2, \\ y_3 = a_3 + b_{31}y_1 + b_{32}y_1 + a_{33}x_3 + a_{34}x_4 + \xi_3. \end{cases} \quad (8.4)$$

В системе уравнений (8.4) три эндогенные переменные ( $y_1$ ,  $y_2$  и  $y_3$ ) и четыре экзогенные переменные ( $x_1$ ,  $x_2$ ,  $x_3$  и  $x_4$ ). Используем счетное правило для проверки на идентификацию каждого уравнения системы (8.4). В первом уравнении  $H = 3$ , так как в нем присутствуют 3 эндогенные переменные,  $D = 2$ , так как отсутствуют переменные  $x_3$  и  $x_4$ . Тогда имеем равенство  $D + 1 = H$  ( $2 + 1 = 3$ ), означающее, что первое уравнение идентифицируемо. К аналогичным выводам придем, рассматривая второе и третье уравнение системы (8.4):

- во втором уравнении  $H = 2$  ( $y_1$  и  $y_2$ ) и  $D = 1$  ( $x_4$ ), т. е.  $D + 1 = H$ ;
- в третьем уравнении  $H = 3$  ( $y_1$ ,  $y_2$ ,  $y_3$ ) и  $D = 2$  ( $x_1$  и  $x_2$ ), т. е.  $D + 1 = H$ .

Таким образом, система уравнений (8.4) по счетному правилу является точно идентифицируемой, так как каждое уравнение в ней идентифицируемо.

Предположим, что в модели (формула (8.4))  $a_{21} = a_{33} = 0$ . Тогда система имеет вид:

$$\begin{cases} y_1 = a_1 + b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2 + \xi_1, \\ y_2 = a_2 + b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + \xi_2, \\ y_3 = a_3 + b_{31}y_1 + b_{32}y_1 + a_{34}x_4 + \xi_3. \end{cases} \quad (8.5)$$

Так как система уравнений (8.5) также содержит три эндогенные и четыре экзогенные переменные, то первое уравнение системы точно идентифицируемо (оно не изменилось). Второе уравнение системы (8.5) теперь имеет  $H = 2$  и  $D = 2$  (отсутствуют  $x_1$  и  $x_4$ ), и так как  $D + 1 > H$  ( $2 + 1 > 2$ ), то данное уравнение оказывается сверхидентифицируемым. Третье уравнение системы (8.5) также является сверхидентифицируемым:  $H = 3$ , ( $y_1$ ,  $y_2$  и  $y_3$ ) и  $D = 3$  (отсутствуют  $x_1$ ,  $x_2$  и  $x_3$ ), т. е.  $D + 1 > H$ .

Если предположим, что последнее уравнение системы (8.5) примет вид:

$$y_3 = a_3 + b_{31}y_1 + b_{32}y_1 + a_{31}x_1 + a_{32}x_2 + a_{34}x_4 + \xi_3,$$

то оно становится неидентифицируемым, так как в нем  $H = 3$  и  $D = 1$  (отсутствует только  $x_3$ ), т. е.  $D + 1 < H$ . Соответственно, его параметры не имеют статистического решения. Рассмотренное счетное правило отражает необходимое, но недостаточное условие идентификации.

Счетное правило, являющееся необходимым условием идентифицируемости, легко проверяется по структурной форме системы одновременных уравнений. Достаточное условие идентифицируемости, или условие ранга, проверяется после нахождения приведенной формы системы уравнений: для идентифицируемости уравнения, входящего в систему, необходимо и достаточно, чтобы ранг матрицы коэффициентов системы по отсутствующим в данном уравнении переменным (эндогенным и экзогенным) был на единицу меньше числа эндогенных переменных в системе и определитель этой матрицы не был равен нулю.

Целесообразность проверки условия идентификации модели через определитель матрицы коэффициентов, отсутствующих в данном уравнении, но присутствующих в других, объясняется тем, что возможна ситуация, когда для каждого уравнения системы

выполнено счетное правило, а определитель матрицы названных коэффициентов равен нулю. В этом случае соблюдается лишь необходимое, но недостаточное условие идентификации.

В эконометрических моделях часто наряду с уравнениями, параметры которых должны быть статистически оценены, используются балансовые тождества переменных, коэффициенты при которых равны плюс или минус единице. В этом случае само тождество не требует проверки на идентификацию, так как коэффициенты при его переменных известны. Вместе с тем при проверке на идентификацию собственно структурных уравнений системы тождества участвуют.

Рассмотрим упрощенный вариант модели Клейна:

$$\begin{cases} CN_t = a_0 + a_1 y_t + a_2 T + u_1, & (\text{функция потребления}), \\ I = b_0 + b_1 y_t + b_2 K_{t-1} + u_2, & (\text{уравнение инвестиций}), \\ y_t = CN_t + I_t, & (\text{тождество}), \end{cases} \quad (8.6)$$

где  $CN_t$  – конечное потребление в период времени  $t$ ;

$y_t$  – национальный доход в период времени  $t$ ;

$T$  – фактор времени, годы;

$I_t$  – инвестиции в период времени  $t$ ;

$K_{t-1}$  – капитал в период времени  $t$ .

В данной модели три эндогенные переменные ( $CN_t$ ,  $I_t$ ,  $y_t$ ) и две предопределенные переменные ( $T$  и  $K_{t-1}$ ). Первое и второе уравнения (8.6) модели Клейна точно идентифицированы по счетному правилу и по достаточному условию идентификации (ранг матрицы равен 2) (табл. 8.2).

Таблица 8.2

Матрица коэффициентов при отсутствующих переменных

В первом уравнении	$I_t$	$K_{t-1}$	Во втором уравнении	$CN_t$	$T$
Уравнение II	1	$-b_2$	Уравнение I	1	$-a_2$
Тождество	$-1$	0	Тождество	$-1$	0
detA $\neq$ 0			detA $\neq$ 0		

Исследуем на идентификацию классическую модель спроса и предложения:

$$\begin{cases} Q_t^d = a_0 + a_1 P_t + U_t, \\ Q_t^S = b_0 + b_1 P_t + V_t, \\ Q_t^S = Q_t^d, \end{cases} \quad (8.7)$$

где  $Q_t^d$  – спрашиваемое количество благ (объем спроса) в момент времени  $t$ ;

$Q_t^S$  – предлагаемое количество благ (объем предложения) в момент времени  $t$ ;

$P_t$  – цена в момент времени  $t$ .

В системе уравнений (8.7) три эндогенные переменные –  $Q_t^d$ ,  $Q_t^S$  и  $P_t$ . Если  $Q_t^d$  и  $Q_t^S$  расположены в левой части системы и являются эндогенными, исходя из структуры системы, то  $P_t$  является эндогенной по экономическому содержанию (цена зависит от спроса и предложения).

Приравняв первое и второе уравнения системы (8.7), можно показать, что  $P_t$  является зависимой (эндогенной) переменной:

$$a_0 + a_1 P_t + U_t = b_0 + b_1 P_t + V_t.$$

$$\text{Отсюда } P_t = \frac{b_0 - a_0}{a_1 - b_1} + \frac{V_t - U_t}{a_1 - b_1}.$$

Модель спроса и предложения (уравнения (8.7)) не содержит ни одной экзогенной переменной и является неидентифицируемой: не выполняется порядковое условие идентификации – в первом и втором уравнениях системы  $D + 1 = H$  так как  $H = 2$ ,  $D = 0$ .

Чтобы модель имела статистическое решение, в нее вводятся экзогенные переменные.

Предположим, что в правую часть уравнения спроса включен совокупный располагаемый доход  $y_t$ , тогда система примет вид:

$$\begin{cases} Q_t^d = a_0 + a_1 P_t + a_2 y_t + U_t, \\ Q_t^S = b_0 + b_1 P_t + V_t, \\ Q_t^d = Q_t^S. \end{cases} \quad (8.8)$$

В системе уравнений (8.7) уравнение спроса неидентифицируемо: не соблюдено порядковое условие идентификации ( $H = 2$ ;  $D = 0$  и  $D + 1 = H$ ); уравнение предложения идентифицируемо, исходя из необходимого и достаточного условия идентификации:  $H = 2$ ;  $D = 1$  и  $D + 1 = H$ , и ранг матрицы равен 2 при трех эндогенных переменных в системе.

Чтобы уравнение спроса в системе уравнений (8.8) было идентифицируемо, дополним уравнение предложения, включив в него экзогенную переменную  $W_t$ . Так, при изучении спроса на сельскохозяйственную продукцию в качестве экзогенной переменной может выступать размер осадков. Экзогенной переменной может быть и спрос на взаимозаменяемый товар. В этом случае система уравнений примет вид:

$$\begin{cases} Q_t^d = a_0 + a_1 P_t + a_2 y_t + U_t, \\ Q_t^s = b_0 + b_1 P_t + b_2 W_t + V_t, \\ Q_t^d = Q_t^s, \end{cases} \quad (8.9)$$

где  $W_t$  – погодные условия.

Модель (уравнения (8.9)) становится точно идентифицируемой как по достаточному, так и по необходимому условию идентификации:  $D + 1 = H$ , и ранг матрицы равен 2 при трех эндогенных переменных в системе.

Рассмотрим пример макроэкономической модели следующего вида:

$$\begin{cases} C_t = a_1 + b_{11} Y_t + b_{12} C_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21} r_t + b_{22} I_{t-1} + \varepsilon_2, \\ r_t = a_3 + b_{31} Y_t + b_{32} M_t + \varepsilon_3, \\ Y_t = C_t + I_t + G_t, \end{cases} \quad (8.10)$$

где  $C_t$  – расходы на потребление в период  $t$ ;

$I_t$  – инвестиции в период  $t$ ;

$r_t$  – процентная ставка в период  $t$ ;

$M_t$  – денежная масса в период  $t$ ;

$Y_t$  – совокупный доход в период  $t$ ;

$G_t$  – государственные расходы в период  $t$ .

Первое уравнение – функция потребления, второе – функция инвестиций, третье – функция денежного рынка, четвертое уравнение – тождество дохода.

Проверим каждое уравнение системы одновременных уравнений на идентификацию.

Модель системы уравнений (8.10) включает в себя четыре эндогенные переменные ( $C_t, I_t, Y_t, r_t$ ) и четыре предопределенные переменные (две экзогенные переменные –  $M_t$  и  $C_t$  и две лаговые переменные –  $C_{t-1}$  и  $I_{t-1}$ ).

Проверим необходимое условие идентификации для каждого из уравнений модели.

I уравнение  $C_t = a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_1$  содержит две эндогенные переменные  $C_t$  и  $Y_t$  и одну предопределенную переменную  $C_{t-1}$ . Таким образом,  $H = 2, D = 4 - 1 = 3$ , т. е. выполняется условие  $D > H$ . Уравнение сверхидентифицируемо.

II уравнение  $I_t = a_1 + b_{21}r_t + b_{22}I_{t-1} + \varepsilon_2$  включает две эндогенные переменные  $I_t$  и  $r_t$  и одну экзогенную переменную  $I_{t-1}$ . Выполняется условие  $D + 1 = 3 + 1 > H = 2$ . Уравнение сверхидентифицируемо.

III уравнение  $r_t = a_3 + b_{31}Y_t + b_{32}M_t + \varepsilon_3$  включает две эндогенные переменные  $Y_t$  и  $r_t$  и одну экзогенную переменную  $M_t$ . Выполняется условие  $D + 1 = 3 + 1 > H = 2$ . Уравнение сверхидентифицируемо.

IV уравнение  $Y_t = C_t + I_t + G_t$  представляет собой тождество, параметры которого известны. Необходимости в идентификации нет.

Проверим для каждого уравнения достаточное условие идентификации, для чего составим матрицу коэффициентов при переменных модели (табл. 8.3).

Таблица 8.3

Матрица коэффициентов при переменных модели системы одновременных уравнений

	$C_t$	$I_t$	$r_t$	$Y_t$	$C_{t-1}$	$I_{t-1}$	$M_t$	$G_t$
I уравнение	-1	0	0	$b_{11}$	$b_{12}$	0	0	0
II уравнение	0	-1	$b_{21}$	0	0	$b_{22}$	0	0
III уравнение	0	0	-1	$b_{31}$	0	0	$b_{31}$	0
Тождество	1	1	0	-1	0	0	0	1

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

*I уравнение.* Матрица коэффициентов при переменных, не входящих в уравнение (табл. 8.4).

Таблица 8.4

Матрица коэффициентов при переменных,  
не входящих в I уравнение

	$I_t$	$r_t$	$I_{t-1}$	$M_t$	$G_t$
II уравнение	-1	$b_{21}$	$b_{22}$	0	0
III уравнение	0	-1	0	$b_{32}$	0
Тождество	1	0	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

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

Достаточное условие идентификации для данного уравнения выполняется.

*II уравнение.* Матрица коэффициентов при переменных, не входящих в уравнение (табл. 8.5).

Таблица 8.5

Матрица коэффициентов при переменных,  
не входящих во II уравнение

	$C_t$	$Y_t$	$C_{t-1}$	$M_t$	$G_t$
I уравнение	-1	$b_{11}$	$b_{12}$	0	0
III уравнение	0	$b_{31}$	0	$b_{32}$	0
Тождество	1	-1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы  $3 \times 3$  не равен нулю:

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

Достаточное условие идентификации для данного уравнения выполняется.

*III уравнение.* Матрица коэффициентов при переменных, не входящих в уравнение (табл. 8.6).

Таблица 8.6

Матрица коэффициентов при переменных,  
не входящих в III уравнение

	$C_t$	$I_t$	$C_{t-1}$	$I_{t-1}$	$G_t$
I уравнение	-1	0	$b_{12}$	0	0
II уравнение	0	-1	0	$b_{22}$	0
Тождество	1	1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы  $3 \times 3$  не равен нулю:

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

Достаточное условие идентификации для данного уравнения выполняется.

Таким образом, все уравнения модели системы одновременных уравнений свержидентифицируемы. Приведенная форма модели в общем виде будет выглядеть следующим образом:

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



#### 8.4. Методы наименьших квадратов, максимального правдоподобия и метод инструментальных переменных для оценивания коэффициентов структурной модели

В зависимости от вида системы одновременных уравнений и вида связей между ее переменными для оценок структурных коэффициентов модели требуются различные способы построения оценок. Наибольшее распространение получили следующие методы оценивания коэффициентов структурной модели:

- 1) косвенный метод наименьших квадратов;
- 2) двухшаговый метод наименьших квадратов;
- 3) трехшаговый метод наименьших квадратов;
- 4) метод максимального правдоподобия с полной информацией;
- 5) метод максимального правдоподобия при ограниченной информации;
- 6) метод инструментальных переменных.

Рассмотрим сущность каждого из этих методов.

Косвенный МНК применяется в случае точно идентифицируемой структурной модели. Процедура применения косвенного МНК следующая:

1. Структурная модель преобразуется в приведенную форму модели.
2. Для каждого уравнения приведенной формы модели обычным МНК оцениваются приведенные коэффициенты  $\delta_{ij}$ .
3. Коэффициенты приведенной формы модели преобразуются в параметры структурной модели.

*Пример 8.1.* Исследовать зависимость спроса и предложения некоторого товара от его цены ( $P$ ), дохода на душу населения ( $y$ ) и инвестиций в производство ( $I$ ). Модель спроса и предложения имеет вид:

$$\begin{cases} Q_t^d = a_0 + a_1 P_t + a_2 y_t + \xi_t, \\ Q_t^S = b_0 + b_1 P_t + b_2 I_t + \xi_t, \\ Q_t^d = Q_t^S, \end{cases}$$

где  $Q_t^d$  – спрос в момент времени  $t$ ;

$Q_t^S$  – предложение в момент времени  $t$ .

Примем, что  $Q_t^d = Q_t^S = Q_t$ .

В данной модели  $Q_t$  и  $P_t$  – эндогенные переменные,  $y_t$  и  $I_t$  – экзогенные переменные. Каждое уравнение системы точно идентифицируемо. Применим косвенный МНК, используя исходные данные (табл. 8.7).

Таблица 8.7

Исходные данные

$Q_t$	20	33	28	41	40	36	42	38	51
$P_t$	3	3	5	4	5	6	6	7	7
$y_t$	34	43	51	49	55	62	70	68	78
$I_t$	5	6	6	7	7	6	8	8	12

Приведенная форма модели следующая:

$$\begin{cases} Q_t = A_1 + \delta_1 y_t + \delta_2 I_t + U_t, \\ P_t = A_2 + \delta_3 y_t + \delta_4 I_t + V_t, \end{cases}$$

где  $U_t$  и  $V_t$  – случайные ошибки.

Оценку параметров каждого уравнения приведенной модели произведем обычным МНК:

$$\begin{cases} Q_t = 6,022 + 0,234 y_t + 2,394 I_t + U_t, \\ P_t = -0,692 + 0,127 y_t - 0,189 I_t + V_t. \end{cases}$$

Найдем уравнение спроса  $Q_t^d = a_0 + a_1 P_t + a_2 y_t + \xi_t$ , используя коэффициенты приведенной модели. Для этой цели из второго уравнения приведенной модели выразим  $I_t$  и подставим его в первое уравнение:

$$I_t = \frac{-P_t - 0,692 + 0,127 y_t}{0,189};$$

$$Q_t^d = 6,022 + 0,234 y_t + 2,394 \cdot \frac{-P_t - 0,692 + 0,127 y_t}{0,189} =$$

$$= -2,743 - 12,667 P_t + 1,842 y_t.$$

Аналогично найдем второе уравнение структурной модели, характеризующее функцию предложения:

$$Q_t^S = b_0 + b_1 P_t + b_2 I_t + \xi_t.$$

Для этого из первого уравнения приведенной структурной модели исключим  $y_t$ , выразив его через второе уравнение и подставив в первое:

$$y_t = \frac{P_t + 0,692 + 0,189I_t}{0,127};$$

$$Q_t^S = 6,022 + 0,234 \cdot \frac{P_t + 0,692 + 0,189I_t}{0,127} + 2,394I_t = 7,297 + 1,843P_t + 2,742I_t.$$

Таким образом, модель спроса и предложения имеет вид:

$$\begin{cases} Q_t^d = -2,743 - 12,667P_t + 1,842y_t + \xi_t, \\ Q_t^S = 7,297 + 1,843P_t + 2,742I_t + \xi_t, \\ Q_t^d = Q_t^S. \end{cases}$$

Если к исходным данным применить обычный МНК, то результаты окажутся следующими:

$$Q_t^d = 1,424 - 7,238P_t + 1,273y_t + \xi_1,$$

$$Q_t^S = 8,824 + 0,663P_t + 3,371I_t + \xi_2.$$

Очевидно, что оценки, полученные косвенным МНК и обычным МНК, различаются. Эти различия могут быть как более, так и менее существенными в зависимости от характера исследуемой информации. В рассматриваемом примере знаки при переменных согласуются с теоретической концепцией функции спроса и предложения. Так, в модели спроса рост цены способствует снижению спроса, а рост доходов – увеличению спроса. В модели предложения рост инвестиций и цен ведет к росту объема выпуска продукции. Эти выводы характерны как при использовании косвенного МНК, так и обычного МНК. Вместе с тем необходимо учитывать, что оценки, полученные применением обычного МНК к каждому уравнению структурной модели, несостоятельны.

Если система сверхидентифицируема, то косвенный МНК не используется, так как он не дает однозначных оценок для параметров структурной модели. В этом случае могут использоваться разные методы оценивания, среди которых наиболее распространенным и простым является двухшаговый МНК.

Основная идея двухшагового МНК – получение, на основе приведенной формы модели для сверхидентифицируемого уравнения, теоретических значений эндогенных переменных, содержащихся в правой части уравнения.

Далее, подставив их вместо фактических значений, можно применить обычный МНК к структурной форме сверхидентифицируемого уравнения. На первом шаге – при определении приведенной формы модели и нахождении на ее основе оценок теоретических значений эндогенной переменной  $\hat{y}_t = \delta_{i1}x_1 + \delta_{i2}x_2 + \dots + \delta_{in}x_n$  и на втором шаге – применительно к структурному сверхидентифицируемому уравнению при определении структурных коэффициентов модели по данным теоретических (расчетных) значений эндогенных переменных.

Сверхидентифицируемая структурная модель может быть двух типов:

- 1) все уравнения системы сверхидентифицируемы;
- 2) система содержит, наряду со сверхидентифицируемыми, точно идентифицируемые уравнения.

Если все уравнения системы сверхидентифицируемые, то для оценки структурных коэффициентов каждого уравнения используется двухшаговый МНК. Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

Для примера макроэкономической модели, рассмотренного выше, необходимо применить именно двухшаговый метод наименьших квадратов.

Если из модели исключить тождество дохода, число эндогенных переменных модели снизится на единицу – переменная  $Y_t$  станет экзогенной. Число predetermined переменных модели не изменится, так как из модели будет исключена эндогенная переменная  $G_t$ , но ее место займет переменная  $Y_t$ . В правых частях функции потребления и функции денежного рынка будут находиться только

предопределенные переменные. Функция инвестиций постулирует зависимость эндогенной переменной  $I_t$  от эндогенной переменной  $r_t$  (которая зависит только от предопределенных переменных) и предопределенной переменной  $I_{t-1}$ . Таким образом, будет получена рекурсивная система. Ее параметры можно оценивать обычным МНК, и нет необходимости исследования уравнения на идентификацию.

Косвенный и двухшаговый МНК рассматриваются как традиционные методы оценки коэффициентов структурной модели и легко реализуются.

*Метод максимального правдоподобия* рассматривается как наиболее общий метод оценивания, результаты которого при нормальном распределении признаков совпадают с МНК. Однако при большом числе уравнений системы этот метод приводит к сложным вычислительным процедурам. Поэтому в качестве модификации используется метод максимального правдоподобия при ограниченной информации (метод наименьшего дисперсионного отношения), разработанный Т. Андерсоном и Н. Рубиным в 1949 г.

В отличие от метода максимального правдоподобия в данном методе сняты ограничения на параметры, связанные с функционированием системы в целом. Это делает решение более простым, но трудоемкость вычислений остается высокой. Несмотря на его значительную популярность, к середине 60-х гг. XX в. он был практически вытеснен двухшаговым МНК в связи с гораздо большей простотой последнего.

Дальнейшим развитием двухшагового МНК является трехшаговый МНК, в 1962 г. предложенный А. Зельнером и Г. Тейлом. Этот метод оценивания пригоден для всех видов уравнений структурной модели. Однако при некоторых ограничениях на параметры более эффективным оказывается двухшаговый МНК.

*Метод инструментальных переменных* применяется для точно идентифицируемых систем одновременных уравнений. Предопределенные переменные структурных уравнений коррелируют со случайными остатками, поэтому классические МНК-оценки параметров не обладают нужными свойствами, например, не являются состоятельными. Для получения улучшенных оценок вводятся новые переменные. Для  $j$ -го структурного уравнения рассматривается

вектор наблюдений  $j$ -й эндогенной переменной и вектор значений соответствующей компоненты случайных остатков. По причине нарушения условий некоррелированности регрессоров и случайных остатков в этом уравнении классическая МНК-оценка параметров является несостоятельной. Вектором инструментальных переменных называется случайный вектор с невырожденной ковариационной матрицей, который некоррелирован с вектором случайных остатков и имеет невырожденную ковариационную матрицу с эндогенными и predetermined переменными, входящими в  $j$ -е структурное уравнение. На основе такого вектора строится состоятельная МИП-оценка вектора параметров  $j$ -го структурного уравнения. Точность такой оценки тем выше, чем сильнее вектор инструментальных переменных коррелирует с эндогенными и predetermined переменными, входящими в  $j$ -е структурное уравнение.

### **Вопросы для контроля знаний**

1. В чем заключается проблема идентифицируемости систем одновременных уравнений (СОУ)?
2. Опишите необходимые и достаточные условия идентифицируемости СОУ.
3. Как связаны между собой структурная и приведенная формы СОУ?
4. Какие методы применяются для оценки параметров СОУ?
5. В каком случае применяется и в чем заключается косвенный метод оценки параметров СОУ?
6. В чем заключается идея двухшагового МНК?
7. В чем суть метода инструментальных переменных?

### **Практическое занятие № 11 Применение счетного правила, условия ранга и оценка структурных коэффициентов**

**Цель работы:** отработать практические навыки применения счетного правила, условия ранга и оценки структурных коэффициентов.

## Методические рекомендации 1

Требуется проверить на идентификацию следующую структурную модель:

$$\begin{cases} y_1 = b_{13}y_3 + a_{11}x_1 + a_{13}x_3, \\ y_2 = b_{21}y_1 + b_{23}y_3 + a_{22}x_2, \\ y_3 = b_{32}y_2 + a_{31}x_1 + a_{33}x_3. \end{cases}$$

Найдем структурные коэффициенты модели, исходя из приведенной формы модели уравнений:

$$\begin{cases} y_1 = 2x_1 + 4x_2 + 10x_3, \\ y_2 = 3x_1 - 6x_2 + 2x_3, \\ y_3 = -5x_1 + 8x_2 + 5x_3. \end{cases}$$

*Решение.* 1. Модель имеет 3 эндогенные переменные ( $y_1$ ,  $y_2$  и  $y_3$ ) и 3 экзогенные переменные ( $x_1$ ,  $x_2$ ,  $x_3$ ). Проверим каждое уравнение системы на необходимое (Н) и достаточное (Д) условия идентификации.

*Первое уравнение*

(Н) В первом уравнении – 2 эндогенные переменные ( $y_1$ ,  $y_3$ ) и 1 отсутствующая экзогенная переменная ( $x_2$ ). Выполняется необходимое равенство:  $2 = 1 + 1$ , следовательно, уравнение точно идентифицируемо.

(Д) В первом уравнении отсутствуют  $y_2$  и  $x_2$ . Построим матрицу из коэффициентов при них в других уравнениях системы (табл. 1).

Таблица 1

Идентификация первого уравнения

Уравнение	Отсутствующие переменные	
	$y_2$	$x_2$
Второе	-1	$a_{22}$
Третье	$b_{32}$	0

$$DetA = -1 \cdot 0 - b_{32} \cdot a_{22} \neq 0.$$

Определитель матрицы не равен 0, ранг матрицы равен 2; следовательно, выполняется достаточное условие идентификации, и первое уравнение точно идентифицируемо.

### *Второе уравнение*

(Н) Во втором уравнении – 3 эндогенные переменные ( $y_1, y_2$  и  $y_3$ ) и 2 отсутствующие экзогенные ( $x_1, x_3$ ). Выполняется необходимое равенство:  $3 = 2 + 1$ , следовательно, уравнение точно идентифицируемо.

(Д) Во втором уравнении отсутствуют  $x_1$  и  $x_3$ . Построим матрицу из коэффициентов при них в других уравнениях системы (табл. 2).

Таблица 2

Идентификация второго уравнения

Уравнение	Отсутствующие переменные	
	$x_1$	$x_3$
Первое	$a_{11}$	$a_{13}$
Третье	$a_{31}$	$a_{33}$

$$DetA = a_{11} \cdot a_{33} - a_{31} \cdot a_{13} \neq 0.$$

Определитель матрицы не равен 0, ранг матрицы равен 2; следовательно, выполняется достаточное условие идентификации, и второе уравнение точно идентифицируемо.

### *Третье уравнение*

(Н) В третьем уравнении – 2 эндогенные переменные ( $y_2, y_3$ ) и 1 отсутствующая экзогенная ( $x_2$ ). Выполняется необходимое равенство:  $2 = 1 + 1$ , следовательно, уравнение точно идентифицируемо.

(Д) В третьем уравнении отсутствуют  $y_1$  и  $x_2$ . Построим матрицу из коэффициентов при них в других уравнениях системы (табл. 3).

Таблица 3

Идентификация третьего уравнения

Уравнение	Отсутствующие переменные	
	$y_1$	$x_2$
Первое	-1	0
Второе	$b_{21}$	$a_{22}$

$$DetA = -1 \cdot a_{22} - b_{21} \cdot 0 \neq 0.$$



Определитель матрицы не равен 0, ранг матрицы равен 2; следовательно, выполняется достаточное условие идентификации, и третье уравнение точно идентифицируемо.

Таким образом, исследуемая система точно идентифицируема и может быть решена косвенного МНК.

2. Вычислим структурные коэффициенты модели:

1) из третьего уравнения приведенной формы выразим  $x_2$  (так как его нет в первом уравнении структурной формы):

$$x_2 = \frac{y_3 + 5x_1 - 5x_3}{8}.$$

Данное выражение содержит переменные  $y_3$ ,  $x_1$  и  $x_3$ , которые нужны для первого уравнения. Подставим полученное выражение  $x_2$  в первое уравнение:

$$y_1 = 2x_1 + 4 \cdot \frac{y_3 + 5x_1 - 5x_3}{8} + 10x_3 \Rightarrow$$

$$\Rightarrow y_1 = 0,5y_3 + 4,5x_1 + 7,5x_3;$$

2) во втором уравнении СФМ нет переменных  $x_1$  и  $x_3$ . Структурные параметры второго уравнения СФМ можно определить в два этапа.

*Первый этап:* выразим  $x_1$  в данном случае из первого уравнения:

$$x_1 = \frac{y_1 - 4x_2 - 10x_3}{2} = 0,5y_1 - 2x_2 - 5x_3.$$

Подстановка данного выражения во второе уравнение не решила бы задачи до конца, так как в выражении присутствует  $x_3$ .

Выразим  $x_3$  из третьего уравнения:

$$x_3 = \frac{y_3 + 5x_1 - 8x_2}{5}.$$

Подставим его в выражение  $x_1$ :

$$x_1 = 0,5y_1 - 2x_2 - 5 \cdot \left( \frac{y_3 + 5x_1 - 8x_2}{5} \right) = 0,5y_1 - y_3 + 6x_2 - 5x_1;$$

$$x_1 = \frac{0,5y_1 - y_3 + 6x_2 - 5x_1}{6}.$$

*Второй этап:* чтобы выразить  $x_3$  через искомые  $y_1$ ,  $y_2$  и  $x_2$ , заменим в выражении  $x_3$  значение  $x_1$  на полученное из первого уравнения ПФМ:

$$x_3 = \frac{y_3 + 5(0,5y_1 - 2x_2 - 5x_3) - 8x_2}{5} = 0,2y_3 + 0,5y_1 - 3,6x_2 - 5x_3.$$

Следовательно:

$$x_3 = 0,033y_3 + 0,083y_1 - 0,6x_2.$$

Подставим полученные  $x_1$  и  $x_3$  во второе уравнение ПФМ:

$$y_2 = 3 \cdot \frac{0,5y_1 - y_3 + 6x_2}{6} - 6x_2 + 2 \cdot (0,033y_3 + 0,083y_1 - 0,6x_2) \Rightarrow$$

$$\Rightarrow y_2 = 0,416y_1 - 0,434y_3 - 4,2x_2 \text{ – второе уравнение СФМ.}$$

Это уравнение можно получить из ПФМ иным путем.

Суммируя все уравнения, получим:

$$y_1 = 2x_1 + 4x_2 + 10x_3,$$

$$y_2 = 3x_1 - 6x_2 + 2x_3,$$

$$y_3 = -5x_1 + 8x_2 + 5x_3,$$

$$y_1 + y_2 + y_3 = 6x_2 + 17x_3.$$

Далее из первого и второго уравнений ПФМ исключим  $x_1$ , умножив первое уравнение на 3, а второе – на минус 2 и просуммировав их:

$$3y_1 = 6x_1 + 12x_2 + 30x_3,$$

$$-2y_2 = -6x_1 + 12x_2 - 4x_3.$$

$$3y_1 - 2y_2 = 24x_2 + 26x_3.$$

Аналогичным путем из полученных уравнений исключаем  $x_3$ :

$$\begin{cases} y_1 + y_2 + y_3 = 6x_2 + 17x_3, & (-26), \\ 3y_1 - 2y_2 = 24x_2 + 26x_3, & (17), \end{cases}$$

$$-26y_1 - 26y_2 - 26y_3 = -156x_2 - 44x_3,$$

$$51y_1 - 34y_2 = 408x_2 + 442x_3,$$

$$25y_1 - 60y_2 - 26y_3 = 252x_2 \Rightarrow 60y_2 = 25y_1 - 26y_3 - 252x_2 \Rightarrow$$

$$\Rightarrow y_2 = 0,416y_1 - 0,433y_3 - 4,2x_2;$$

3) из второго уравнения ПФМ выразим  $x_2$ , так как его нет в третьем уравнении СФМ:

$$x_2 = \frac{-y_2 + 3x_1 + 2x_3}{6} = -0,167y_2 + 0,5x_1 + 0,333x_3.$$

Подставим полученное выражение в третье уравнение ПФМ:

$$y_3 = -5x_1 + 8 \cdot (-0,167y_2 + 0,5x_1 + 0,033x_3) + 5x_3 \Rightarrow$$

$$\Rightarrow y_3 = -1,336y_2 - x_1 + 7,664x_3.$$

Таким образом, СФМ примет вид:

$$\begin{cases} y_1 = 0,5y_3 + 4,5x_1 + 7,5x_3, \\ y_2 = 0,416y_1 - 0,434y_3 - 4,2x_2, \\ y_3 = -1,336y_2 - x_1 + 7,664x_3. \end{cases}$$

## Методические рекомендации 2

Рассматривается следующая модель:

$$\begin{cases} C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + U_1, & (\text{функция потребления}); \\ I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + U_2, & (\text{функция инвестиций}); \\ r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + U_3, & (\text{функция денежного роста}); \\ \dot{Y}_t = \dot{C}_t + \dot{I}_t + \dot{G}_t, & (\text{тождество дохода}), \end{cases}$$

где  $C_t$  – расходы на потребление в период  $t$ ;

$C_{t-1}$  – расходы на потребление в период  $t-1$ ;

$I_t$  – инвестиции в период  $t$ ;

$I_{t-1}$  – инвестиции в период  $t-1$ ;

$r_t$  – процентная ставка в период  $t$ ;

$M_t$  – денежная масса в период  $t$ ;  
 $Y_t$  – совокупный доход в период  $t$ ;  
 $G_t$  – государственные расходы в период  $t$ ;  
 $U_1, U_2, U_3$  – случайные ошибки.

Требуется:

1. Предложить способ оценки параметров модели по имеющимся временным рядам данных по всем переменным.
2. Как изменится ответ на вопрос п. 1, если из модели исключить тождество дохода?

*Решение.* Модель представляет собой систему одновременных уравнений. Проверим каждое ее уравнение на идентификацию.

Модель включает четыре эндогенные переменные ( $C_t, I_t, Y_t, r_t$ ) и четыре predetermined переменные (две экзогенные переменные –  $M_t$  и  $C_t$  и две лаговые переменные –  $C_{t-1}$  и  $I_{t-1}$ ).

Проверим необходимое условие идентификации для каждого из уравнений модели.

I уравнение  $C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + \varepsilon_1$  содержит две эндогенные переменные  $C_t$  и  $Y_t$  и одну predetermined переменную  $C_{t-1}$ . Таким образом,  $H = 2$ ,  $D = 4 - 1 = 3$ , т. е. выполняется условие  $D + 1 > H$ . Уравнение сверхидентифицируемо.

II уравнение  $I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + \varepsilon_2$  включает две эндогенные переменные  $I_t$  и  $r_t$  и одну экзогенную переменную  $I_{t-1}$ . Выполняется условие  $D + 1 = 3 + 1 > H = 2$ . Уравнение сверхидентифицируемо.

III уравнение  $r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + \varepsilon_3$  включает две эндогенные переменные  $Y_t$  и  $r_t$  и одну экзогенную переменную  $M_t$ . Выполняется условие  $D + 1 = 3 + 1 > H = 2$ . Уравнение сверхидентифицируемо.

IV уравнение  $Y_t = C_t + I_t + G_t$  представляет собой тождество, параметры которого известны. Необходимости в идентификации нет.

Проверим для каждого уравнения достаточное условие идентификации, для чего составим матрицу коэффициентов при переменных модели (табл. 4–7).

Таблица 4

Матрица коэффициентов при переменных

	$C_t$	$I_t$	$r_t$	$Y_t$	$C_{t-1}$	$I_{t-1}$	$M_t$	$G_t$
I уравнение	-1	0	0	$b_{11}$	$b_{12}$	0	0	0
II уравнение	0	-1	$b_{21}$	0	0	$b_{22}$	0	0
III уравнение	0	0	-1	$b_{31}$	0	0	$b_{32}$	0
Тождество	1	1	0	-1	0	0	0	1

В соответствии с достаточным условием идентификации ранг матрицы коэффициентов при переменных, не входящих в исследуемое уравнение, должен быть равен числу эндогенных переменных модели без одного, т. е.  $4 - 1 = 3$ .

I уравнение. Матрица коэффициентов при переменных, не входящих в уравнение.

Таблица 5

Матрица коэффициентов без I уравнения

	$I_t$	$r_t$	$I_{t-1}$	$M_t$	$G_t$
II уравнение	-1	$b_{21}$	$b_{22}$	0	0
III уравнение	0	-1	0	$b_{32}$	0
Тождество	1	0	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы  $3 \times 3$  не равен нулю:

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

II уравнение. Матрица коэффициентов при переменных, не входящих в уравнение.

Таблица 6

Матрица коэффициентов без II уравнения

	$C_t$	$Y_t$	$C_{t-1}$	$M_t$	$G_t$
I уравнение	-1	$b_{11}$	$b_{12}$	0	0
III уравнение	0	$b_{31}$	0	$b_{32}$	0
Тождество	1	-1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы  $3 \times 3$  не равен нулю:

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

Достаточное условие идентификации для данного уравнения выполняется.

III уравнение. Матрица коэффициентов при переменных, не входящих в уравнение.

Таблица 7

Матрица коэффициентов без III уравнения

	$C_t$	$I_t$	$C_{t-1}$	$I_{t-1}$	$G_t$
I уравнение	-1	0	$b_{12}$	0	0
II уравнение	0	-1	0	$b_{22}$	0
Тождество	1	1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы  $3 \times 3$  не равен нулю:

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

Достаточное условие идентификации для данного уравнения выполняется.

Таким образом, все уравнения модели сверхидентифицируемы. Для оценки параметров каждого из уравнений будем применять двухшаговый МНК.

Шаг 1. Запишем приведенную форму модели в общем виде:

$$\begin{cases} C_t = A_1 + A_2C_{t-1} + A_3I_{t-1} + A_4M_t + A_5G_t + V_1, \\ I_t = B_1 + B_2C_{t-1} + B_3I_{t-1} + B_4M_t + B_5G_t + V_2, \\ r_t = D_1 + D_2C_{t-1} + D_3I_{t-1} + D_4M_t + D_5G_t + V_3, \\ \dot{Y}_t = E_1 + E_2C_{t-1} + E_3I_{t-1} + E_4M_t + E_5G_t + V_4, \end{cases}$$

где  $V_1, V_2, V_3$  – случайные ошибки.

Определим параметры каждого из приведенных выше уравнений в отдельности обычным методом МНК. Затем найдем расчетные значения эндогенных переменных  $\hat{Y}_t, \hat{r}_t$ , используемых в правой части структурной модели, подставляя в каждое уравнение приведенной формы соответствующее значение предопределенных переменных.

*Шаг 2.* В исходных структурных уравнениях заменим эндогенные переменные, выступающие в качестве факторных признаков, их расчетными значениями:

$$C_t = a_1 + b_{11}\hat{Y}_t + b_{12}C_{t-1} + U_1^*, \text{ где } U_1^* = U_1 + b_{11}V_1;$$

$$I_t = a_2 + b_{21}\hat{r}_t + b_{22}I_{t-1} + U_2^*, \text{ где } U_2^* = U_2 + b_{21}V_2;$$

$$r_t = a_3 + b_{31}\hat{Y}_t + b_{32}M_t + U_3^*, \text{ где } U_3^* = U_3 + b_{31}V_3.$$

Применяя к каждому из полученных уравнений в отдельности обычный МНК, определим структурные параметры  $a_1, b_{11}, b_{12}, a_2, b_{21}, b_{22}, a_3, b_{31}$  и  $b_{32}$ .

Если из модели исключить тождество дохода, число предопределенных переменных модели уменьшится на 1 (будет исключена переменная  $G_t$ ). Число эндогенных переменных модели также снизится на единицу – переменная  $Y_t$  станет экзогенной. В правых частях функции потребления и функции денежного рынка будут находиться только предопределенные переменные. Функция инвестиций постулирует зависимость эндогенной переменной  $I_t$  от эндогенной переменной  $r_t$  (которая зависит только от предопределенных переменных) и предопределенной переменной  $I_{t-1}$ . Таким образом, получена рекурсивная система. Ее параметры можно оценивать обычным МНК, и нет необходимости исследования системы уравнения на идентификацию.

### **Задания для самостоятельной работы**

#### ***Задача 1***

$$R_t = a_1 + b_{11}M_t + b_{12}Y_t + e_1,$$

$$Y_t = a_2 + b_{21}R_t + b_{22}I_t + e_2,$$

где  $R$  – процентная ставка;  
 $M$  – денежная масса;  
 $Y$  – ВВП;  
 $I$  – внутренние инвестиции;  
 $t$  – текущий период.

### **Задача 2**

Одна из версий модифицированной модели Кейнса имеет вид:

$$C_t = a_1 + b_{11}Y_t + b_{12}Y_{t-1} + e_1,$$

$$I_t = a_2 + b_{21}Y_t + b_{22}I_{t-1} + e_2,$$

$$Y_t = C_t + I_t + G_t,$$

где  $C$  – расходы на потребление;  
 $Y$  – доход;  
 $I$  – инвестиции;  
 $G$  – государственные расходы;  
 $t$  – текущий период;  
 $t-1$  – предыдущий период.

### **Задача 3**

Макроэкономическая модель (упрощенная версия модели Клейна):

$$C_t = a_1 + b_{12}Y_t + b_{13}T_t + e_1,$$

$$I_t = a_2 + b_{21}Y_t + b_{24}K_{t-1} + e_2,$$

$$Y_t = C_t + I_t,$$

где  $C$  – потребление;  
 $Y$  – доход;  
 $T$  – налоги;  
 $I$  – инвестиции;  
 $K$  – запас;  
 $t$  – текущий период;  
 $t-1$  – предыдущий период.



#### **Задача 4**

Модель денежного и товарного рынков:

$$R_t = a_1 + b_{12}Y_t + b_{14}M_t + e_1 \quad (\text{функция денежного рынка});$$

$$Y_t = a_2 + b_{21}R_t + b_{23}I_t + b_{25}G_t + e_2 \quad (\text{функция товарного рынка});$$

$$I_t = a_3 + b_{31}R_t + e_3 \quad (\text{функция инвестиций}),$$

где  $R$  – процентная ставка;

$Y$  – реальный ВВП;

$M$  – денежная масса;

$I$  – внутренние инвестиции;

$G$  – реальные государственные расходы.

#### **Задача 5**

Модель спроса и предложения на деньги:

$$R_t = a_1 + b_{11}M_t + b_{12}Y_t + e_1,$$

$$Y_t = a_2 + b_{21}R_t + \varepsilon_2,$$

где  $R_t$  – процентные ставки в период  $t$ ;

$M_t$  – денежная масса в период  $t$ ;

$Y_t$  – ВВП в период  $t$ .

#### **Задача 6**

Модель денежного рынка:

$$R_t = a_1 + b_{11}M_t + b_{12}Y_t + e_1,$$

$$Y_t = a_2 + b_{21}R_t + b_{22}I_t + e_2,$$

$$I_t = a_3 + b_{33}R_t + e_3,$$

где  $R_t$  – процентные ставки;

$M_t$  – денежная масса;

$Y_t$  – ВВП;

$I_t$  – внутренние инвестиции.

## Тема 9. ПРИКЛАДНЫЕ МОДЕЛИ В ЭКОНОМИКЕ

### *Вопросы*

- 9.1. Понятие и свойства производственных функций.
- 9.2. Модели спроса.
- 9.3. Модели рынка межбанковских кредитов.

Построение прикладных эконометрических моделей реализуется по единой схеме: выбор теоретической модели как инструмент обоснования спецификации эконометрической модели, информационное обеспечение, обоснование эконометрического инструментария, интерпретация результатов моделирования. В данной теме рассматриваются вопросы построения прикладной эконометрической модели в комплексе с привлечением аппарата экономической теории (модели микро- и макроэкономики), благодаря чему представляется возможным на основе эконометрической модели делать теоретически обоснованные аналитические выводы.

### **9.1. Понятие и свойства производственных функций**

*Производственная функция* (ПФ) – экономико-математическое уравнение, связывающее переменные величины затрат (ресурсов)  $(x_1, \dots, x_m)$  с величинами продукции (выпуска)  $y = f(x_1, \dots, x_m)$ . ПФ применяются для анализа влияния различных сочетаний факторов на объем выпуска в определенный момент времени (статический вариант), для анализа и прогнозирования соотношения объемов факторов и объема выпуска в разные моменты времени (динамический вариант) на различных уровнях экономики – от фирмы (предприятия) до народного хозяйства в целом (агрегированная функция, в которой выпуском служит показатель совокупного общественного продукта или национального дохода). На уровне предприятия функция описывает объем выпуска продукции  $Y$ , которую они в состоянии произвести при каждом сочетании используемых факторов производства (например, численность занятых  $L$  и объем основных фондов  $K$ ).

Характеристики ПФ:

$$y = \frac{Y}{L}$$

– средняя производительность труда;

$$l = \frac{L}{Y}$$

– средние затраты труда (трудоемкость);

$$\frac{Y}{K}$$

– средняя капиталоотдача (фондоотдача);

$$\frac{K}{Y}$$

– средняя капиталоемкость (фондоемкость) продукции;

$$k = \frac{K}{L}$$

– средняя капиталовооруженность (фондовооруженность) труда;

$$w_k = \frac{\partial Y}{\partial L}$$

– предельная производительность труда, выражающая величину дополнительного выпуска продукта при использовании дополнительной единицы труда;

$$w_k = \frac{\partial Y}{\partial K}$$

– предельная капиталоотдача (фондоотдача), выражающая величину дополнительного выпуска продукта при использовании дополнительной единицы капитала;

$$El_{Y/K} = \frac{\partial Y / Y}{\partial K / K}$$

– коэффициент эластичности выпуска по капиталу (фондам), характеризующий темп прироста объема производства при увеличении объема капитала на 1 % и неизменной численности занятых;

$$El_{Y/L} = \frac{\partial Y / Y}{\partial L / L}$$

– коэффициент эластичности выпуска по труду, характеризующий темп прироста объема производства при увеличении труда на 1 % и неизменном объеме капитала;

$$\Psi_{K/L} = \frac{dK}{dL} = -\frac{w_L}{w_K}$$

– предельная норма замещения труда основными фондами (капиталом), показывающая количество единиц капитала, замещающего единицу труда при неизменном объеме выпуска продукции;

$\Psi_{L/K} = \frac{dL}{dK} = -\frac{w_K}{w_L}$  – предельная норма замещения капитала трудом, показывающая количество единиц труда, замещающего единицу капитала при неизменном объеме выпуска продукции;

$\varepsilon = \frac{d(K/L)/(K/L)}{d\Psi_{K/L}}$  – эластичность замещения труда основными фондами (капиталом), показывающая, на сколько процентов должна измениться капиталовооруженность труда при изменении отношения предельных производительностей труда и капитала на 1 %, чтобы выпуск не изменился.

Среди всего набора статических ПФ в прикладных эконометрических исследованиях наибольшее использование нашла степенная производственная функция (типа Кобба–Дугласа):

$$Y = AK^\alpha L^\beta,$$

где  $A$ ,  $\alpha$ ,  $\beta$  – положительные числа.

Рассмотрим основные свойства ПФ Кобба–Дугласа, которые полезны в прикладном эконометрическом анализе.

*Свойство 1* (свойство эластичности выпуска продукции). Эластичность выпуска продукции по капиталу и труду равна, соответственно,  $\alpha$  и  $\beta$ , причем коэффициенты эластичности постоянны.

Следовательно, увеличение затрат капитала на 1 % приведет к росту выпуска продукции на  $\alpha$  %, увеличение затрат труда на 1 % приведет к росту выпуска на  $\beta$  %. Можно предположить, что обе величины в эффективной экономике находятся между 0 и 1. Они должны быть положительными, что следует из свойства 1 ПФ. В то же время они будут меньше 1, так как логично предположить более медленный рост производства в сравнении с затратами производственных факторов, при условии, что другие факторы постоянны.

*Свойство 2* (свойство эффекта от масштаба). Если  $\alpha$  и  $\beta$  в сумме превышают 1, то функция имеет возрастающий эффект от масштаба производства: если  $K$  и  $L$  увеличиваются в  $r$  раз, то объем производства растет больше, чем в  $r$  раз. Если сумма коэффициентов

равна 1, то речь идет о постоянном эффекте от масштаба производства. Если сумма коэффициентов меньше 1, то имеет место убывающий эффект от масштаба производства.

Таким образом, сумма коэффициентов  $\alpha$  и  $\beta$  определяет тип экономического роста:  $(\alpha + \beta) > 1$  соответствует интенсивному экономическому росту, причем при  $\alpha > \beta$  имеет место трудоинтенсивный экономический рост, при  $\alpha < \beta$  – фондоинтенсивный экономический рост;  $(\alpha + \beta) < 1$  означает, что выпуск продукции растет медленнее, чем рост факторов, т. е. экономический рост отсутствует; при  $(\alpha + \beta) = 1$  имеет место экстенсивный экономический рост.

*Свойство 3.* В ПФ типа Кобба–Дугласа темп прироста выпуска продукции может быть разложен на три составляющие:

$$\frac{dY}{Y} = j + \alpha \frac{dK}{K} + \beta \frac{dL}{L},$$

где  $j$  – вклад НТП;

$\alpha \frac{dK}{K}$  – вклад от прироста капитала;

$\beta \frac{dL}{L}$  – вклад увеличения численности занятых.

*Свойство 4.* Эластичность замещения факторов производства как основная характеристика специфики технологии в функции Кобба–Дугласа равна 1,  $\varepsilon = 1$ .

*Свойство 5.* Логарифмирование ПФ Кобба–Дугласа приводит ее к линейному виду относительно параметров  $\alpha$ ,  $\beta$ , что позволяет проводить их оценку с использованием методов линейного регрессионного анализа.

Приведенные свойства ПФ являются полезными при оценке параметров регрессии стандартными ППП и при проведении содержательного экономического анализа.

*Пример 9.1.* По данным предприятия (таблица) проведена оценка ПФ в виде:

$$\ln Y = 0,72 \ln K + 0,66 \ln L, \quad R_{\text{норм}}^2 = 0,85, \quad F = 1033, \quad DW \approx 2.$$

$t - \text{ст.}(4,9) \quad (44,6)$

Исходные данные для примера 9.1

Годы	Объем выпуска $Y$ , тыс. ед.	Основной капитал $K$ в сопоставимых ценах (млн усл. ден. ед.)	Численность ППП $L$ , чел.
1	94,7	0,2	5636
2	112,4	0,3	5520
3	224,5	1,16	4502
4	302,0	1,16	4515
...	...	...	...
26	486,9	1,159	5454
27	283,0	2,4	4109
28	392,2	2,4	3897
29	493,9	2,5	3958
30	696,4	1,6	4086

По данной функции охарактеризовать степень влияния факторов производства на объем производства.

*Решение.* По набору статистических характеристик модель является удовлетворительной. Экономическая интерпретация параметров модели следующая:  $\alpha = 0,72$  (эластичность выпуска по основным фондам), т. е. при увеличении фондов, на 1 % выпуск увеличится на 0,72 %;  $\beta = 0,66$  (эластичность выпуска по труду), т. е. при увеличении численности работников на 1 % выпуск увеличится на 0,66 %.

Так как  $\alpha > \beta$ , в исследуемом периоде наблюдался трудосберегающий (интенсивный) рост в экономике. Обратная ситуация характеризовалась бы как фондосберегающий (экстенсивный) рост. И так как  $\alpha + \beta > 1$ , имеет место растущая экономика.

Полученная регрессионная модель позволяет вычислить основные характеристики ПФ для существующего способа производства, который, например, характеризуется данными последнего года:

- объем основных фондов  $K = 1,6$  (млн усл. ден. ед.);
- численность ППП  $L = 4086$  (чел.);
- предельная производительность труда

$$w_l = \frac{\partial Y}{\partial L} = 0,66 K^{0,72} L^{-0,34} = 0,66 \cdot (1,6)^{0,72} \cdot (4086)^{-0,34} =$$

$$= 0,055 \text{ (тыс. ед./чел.)},$$

или увеличение на одного человека всей численности занятых способствует росту производства на 0,055 тыс. ед.;

– предельная капиталоотдача (фондоотдача)

$$w_K = \frac{\partial Y}{\partial K} = 0,72K^{-0,28}L^{0,66} = 0,72 \cdot (1,6)^{-0,28} \cdot (4086)^{0,66} =$$

$$= 152,7(\text{тыс. ед./млн усл. ден. ед.}),$$

или увеличение основных фондов на 1 млн усл. ден. ед. способствует росту производства на 152,7 тыс. ед.;

– предельная норма замещения капитала трудом

$$\Psi_{L/K} = \frac{dL}{dK} = -\frac{w_K}{w_L} = -\frac{152,7}{0,055} = 2776,3 \text{ (усл. ден. ед./чел.)},$$

или, с учетом влияния на объем производства, увеличение основных фондов на 1 млн усл. ден. ед. равносильно дополнительному привлечению трудовых ресурсов в количестве 2776,3 чел.

В исследуемом периоде объем производства увеличился в  $t_Y = \frac{Y_{30}}{Y_1} = \frac{696,4}{94,7} = 7,354$  раза, и средний за период объем производства

составил  $\bar{Y} = 342,9$  тыс. ед.; объем основных фондов увеличился в  $t_K = \frac{K_{30}}{K_1} = \frac{1,6}{0,2} = 8$  раз, и средний объем основных фондов в стои-

мостном выражении составил  $\bar{K} = 1,4$  млн усл. ден. ед.; численность занятых возросла в  $t_L = \frac{L_{30}}{L_1} = \frac{4086}{5636} = 0,72$  раза, т. е. снизилась на 28 %, и средняя численность занятых в периоде составила  $\bar{L} = 4631$  чел.

За исследуемый период средняя производительность труда составила  $y = \frac{\bar{Y}}{\bar{L}} = 0,074$  тыс. ед./чел.; средние затраты труда (трудоемкость)

составили  $l = \frac{\bar{L}}{\bar{Y}} = \frac{4630,8}{342,9} = 13,5$  чел./тыс. ед.; средняя капиталоот-

дача (фондоотдача) –  $\frac{\bar{Y}}{\bar{K}} = \frac{342,9}{1,4} = 239,6$  тыс. ед./млн усл. ден. ед. ;

средняя капиталоемкость (фондоёмкость) продукции  $\frac{\bar{K}}{Y} = \frac{1,4}{342,9} = 0,00417$  млн усл. ден. ед./тыс. ед.; средняя капиталовооруженность (фондовооруженность) труда –  $k = \frac{\bar{K}}{L} = \frac{1,4}{4630,8} = 0,00031$  млн усл. ден. ед./чел. = 310 усл. ден. ед./чел.

Рассчитаем относительную эластичность по фондам и труду:

$$\alpha_1 = \frac{\alpha}{\alpha + \beta} = \frac{0,72}{1,38} = 0,52, \quad \beta_1 = \frac{\beta}{\alpha + \beta} = \frac{0,66}{1,38} = 0,48,$$

затем представим темпы роста производства следующим образом:

$$\begin{aligned} t_Y = \frac{Y_{30}}{Y_1} &= \left( \frac{Y_{30}}{Y_1} \right)^{\alpha_1} \left( \frac{Y_{30}}{Y_1} \right)^{\beta_1} = \left( \frac{\frac{Y_{30}}{Y_1} \cdot \frac{K_{30}}{K_1}}{\frac{K_{30}}{K_1}} \right)^{\alpha_1} \cdot \left( \frac{\frac{Y_{30}}{Y_1} \cdot \frac{L_{30}}{L_1}}{\frac{L_{30}}{L_1}} \right)^{\beta_1} = \\ &= (t_K^{\alpha_1} \cdot t_L^{\beta_1}) \cdot \left[ \left( \frac{t_Y^1}{t_K^1} \right)^{\alpha_1} \left( \frac{t_Y^1}{t_L^1} \right)^{\beta_1} \right] = \\ &= (8^{0,52} \cdot 0,72^{0,48}) \cdot \left[ \left( \frac{7,354}{8} \right)^{0,52} \left( \frac{7,354}{0,72} \right)^{0,48} \right] = (2,52) \cdot (2,92). \end{aligned}$$

Смножитель в первых скобках представляет собой темп роста производства за счет факторов масштаба производства, множитель во вторых скобках представляет собой темп роста производства за счет факторов эффективности производства. На основании расчетов можно заключить, что рост производства в исследуемом периоде произошел за счет масштаба производства в 2,52 раза, за счет эффективности производства – в 2,92 раза.

Простейшей модификацией ПФ Кобба–Дугласа является построение модели с экзогенных НТП вида  $Y = A(t)K^\alpha L^\beta$ . Методика оценки параметров модели аналогична классической ПФ с той лишь разницей, что появляется еще один фактор – время.



*Пример 9.2.* В результате исследования взаимосвязи факторов роста, проведенного за период 1950–1970 гг. на основе динамической функции Кобба–Дугласа, наилучшей спецификацией для модели была признана модель вида:

$$y = \frac{39,2}{t + 20} + 0,315k + 1,8735l,$$

где  $y$ ,  $k$ ,  $l$  – показатели годовых темпов прироста, соответственно, конечного продукта, основных фондов и численности занятых.

Оценить влияние НТП в данной зависимости.

*Решение.* В данном случае имеет место ПФ с экзогенным НТП.

Множитель  $A(t) = \frac{39,2}{t + 20}$  описывает монотонное падение темпов технического прогресса по гиперболическому закону.

## 9.2. Модели спроса

Теоретической базой построения моделей спроса выступает функция полезности. В соответствии с последней уровень полезности товаров и услуг можно выразить целевой функцией потребления  $f(Y)$ , где  $Y = (y_1, y_2, \dots, y_n)$  – вектор переменных, включающий разнообразные виды товаров и услуг. В пространстве потребительских благ каждому уравнению  $f(Y) = C$  соответствует определенная поверхность равноценных, или безразличных, благ, которая называется поверхностью безразличия.

Экономическая теория, послужившая методологической базой разработки ПФ в соответствии с классической схемой основных факторов производства, пока не в состоянии обеспечить четкие рекомендации по спецификации функциональных зависимостей для уравнения спроса. Это означает, что исследователь-практик свободен в выборе спецификаций для уравнений. В то же время из теории полезности следует, что в качестве основных факторов спроса выступают цены на товары и доходы потребителей.

*Модели, описывающие зависимость спроса от дохода.* Модель вида:

$$y = ax^b,$$

где  $y$  – представляет собой спрос на товар,  $x$  – доход, в экономической теории описывает общую форму кривых Энгеля. Данная зависимость является степенной, и в такой форме связи параметр  $b$  представляет собой коэффициент эластичности. Таким образом, если имеется кривая Энгеля вида:  $y = 0,01x^{0,3}$ , то это означает, что эластичность спроса по доходу равна 0,3, или что увеличение дохода на 1 % вызывает увеличение спроса на 0,3 %.

Выделяют три группы товаров в зависимости от коэффициента эластичности спроса на них от дохода: товары с малой эластичностью ( $0 < b < 1$ ), товары со средней эластичностью ( $b$  близко к 1), товары с высокой эластичностью ( $b > 1$ ). Формы кривых Энгеля для различных групп товаров представлены на рис. 9.1, *а*, *б*, *в*.

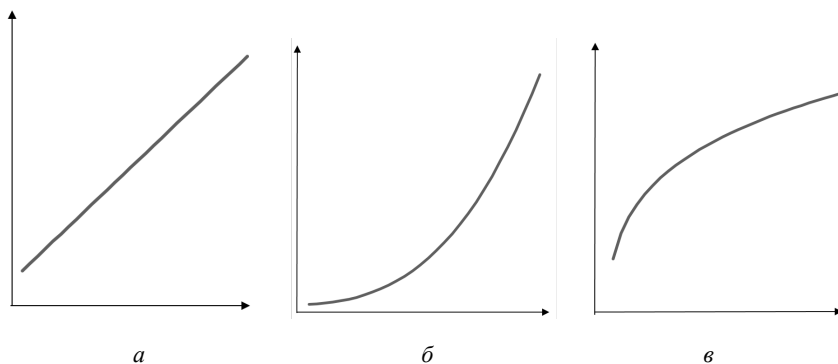


Рис. 9.1. Кривые Энгеля для различных групп товаров

К малоценным товарам, т. е. товарам с отрицательной эластичностью спроса от дохода, относятся такие, как хлеб, низкосортные товары. По результатам обследований коэффициенты эластичности для основных продуктов питания находятся в интервале от 0,4 до 0,8; по одежде, обуви – в интервале от 1,1 до 1,3. По мере увеличения дохода структура спроса перемещается с товаров первой и второй групп к товарам третьей и последующих групп, при этом потребление товаров первой группы по абсолютным размерам сокращается.

Так же, как и в ПФ Кобба–Дугласа, логарифмирование кривой Энгеля приводит ее к линейному виду относительно параметра  $b$ ,

что позволяет проводить оценку функции с использованием методов линейного регрессионного анализа:

$$\ln y = \ln(a) + b \ln x.$$

Аналогичный кривым Энгеля принцип разграничения групп товаров по типам функций спроса от дохода использовал шведский экономист Торнквист, который предложил специальные виды функций спроса для трех групп товаров: первой, второй необходимости и предметов роскоши. Для товаров первой необходимости функция Торнквиста имеет следующий вид:

$$y = \frac{a_1 x}{x + c_1}. \quad (9.1)$$

Она отражает тот факт, что рост спроса на первоочередные товары с ростом дохода постепенно замедляется и имеет некоторый предел  $a_1$ . Функция Торнквиста второго порядка описывает спрос на товары второй необходимости и также имеет предел, но более высокого порядка  $a_2$ , причем спрос на эту группу товаров появляется лишь после того, как доход достигнет некоторой величины  $b_2$ :

$$y = \frac{a_2(x - b_2)}{x + c_2}, \quad x \geq b_2. \quad (9.2)$$

Функция Торнквиста третьего порядка не имеет предела. Спрос на эту группу товара возникает после того, как доход достигает величины  $b_2$  и далее быстро возрастает:

$$y = \frac{a_3 x(x - b_3)}{x + c_3}, \quad x \geq b_3. \quad (9.3)$$

Графически вид функций Торнквиста представлен на рис. 9.2.

Чтобы применить методы линейного регрессионного анализа к функции Торнквиста, требуется предварительно провести их линеаризацию путем тождественных преобразований. Для функции первого порядка это возможно на основе процедуры обращения функции (9.1):

$$\frac{1}{y} = \frac{x + c_1}{ax}.$$

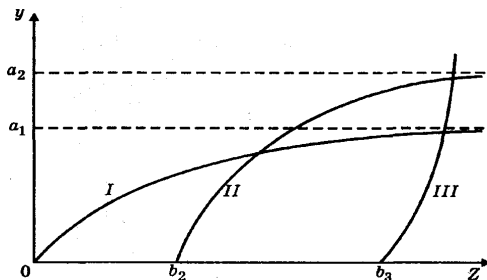


Рис. 9.2. Графики функций Торнквиста для различных групп товаров

Соответствующие преобразования переменных  $y_1 = \frac{1}{y}, x_1 = \frac{1}{x}$  позволяют получить линейную функцию относительно преобразованных переменных:

$$y_1 = \frac{1}{a_1} + \frac{c_1}{a_1} \cdot x_1,$$

параметры которой оцениваются на основе методов линейного регрессионного анализа.

Для функций второго и третьего порядка задача значительно усложняется, поскольку для преобразования требуется предварительная оценка уровней дохода, после которого возникает спрос на товар –  $b_2, b_3$ . Линеаризация (формулы (9.2), (9.3)) осуществляется путем умножения правой и левой частей уравнений, соответственно, на выражения:

$$\frac{1}{y} \cdot \frac{x + c_2}{a_2} \text{ и } \frac{1}{y} \cdot \frac{x + c_3}{a_3}.$$

После преобразования переменных:

$$y_2 = \frac{x - b_2}{y}, y_3 = \frac{x - b_3}{y}, x_3 = \frac{1}{x},$$

приходим, соответственно, к линейным функциям относительно преобразованных переменных:

$$y_2 = \frac{1}{a_2} \cdot x + \frac{c_2}{a_2}, y_3 = \frac{1}{a_3} + \frac{c_3}{a_3} \cdot x_3.$$

В отличие от функций Энгеля, где эластичность спроса по доходу является величиной постоянной, для функций Торнквиста эластичность спроса по доходу не является постоянной и зависит от величины дохода, что более соответствует реальной ситуации. Несложно показать, что коэффициенты эластичности спроса по доходу для функций Торнквиста первого и второго порядка, соответственно, имеют вид:

$$El^1 = \frac{c_1}{x + c_1}, El^2 = x \left( \frac{1}{x - b_2} - \frac{1}{x + c_2} \right).$$

*Пример. 9.3.* По данным объема продаж (шт.) и доходов на одного члена семьи (усл. ден. ед.) построена модель в первых разностях:

$$\ln \Delta y_1 = 0,74 + 0,73 \ln \Delta x_1 \quad R^2 = 0,89 \quad DW = 2,26 \quad t - \text{ст.} (4,1) (5,1).$$

Оценить статистическую значимость дохода как фактора спроса. Спрогнозировать динамику спроса, если в следующем году предполагается рост среднедушевого дохода на 3 усл. ден. ед.

*Решение.* В модели все статистические характеристики являются удовлетворительными, и ее можно использовать для анализа и прогноза. Все характеристики модели указывают на статистическую значимость дохода как фактора спроса. Эластичность прироста спроса по приросту доходу составляет 0,73. Следовательно, в прогнозном периоде при увеличении дохода на 3 ден. ед. можно ожидать увеличение спроса на 4 шт.:  $(3 \cdot (3 - 2) / 2) \cdot 0,73 + 3$ .

*Модели, описывающие зависимость спроса от цены.* Закон спроса гласит, что количество товара, на которое предъявляется спрос, возрастает при снижении цены и сокращается при ее повышении. При этом нет строго одинакового соотношения между снижением цены и повышением спроса.

Количественной характеристикой влияния цены на спрос выступает коэффициент эластичности. При этом различают эластичный и неэластичный спрос. Эластичный спрос характеризуется коэффициентом эластичности по абсолютному значению больше 1 (товары роскоши). Неэластичный спрос имеет коэффициент эластичности спроса по цене по абсолютному значению меньше 1 (товары первой необходимости).

Постоянной эластичностью обладает функция спроса, представленная в виде степенной функции с показателем степени  $b$ :

$$y = ap^b. \quad (9.4)$$

Если имеется кривая спроса вида:  $y = 0,01p^{-0,5}$ , это означает, что эластичность спроса по цене равна  $(-0,5)$ , или что увеличение цены на 1 % вызывает уменьшение спроса на 0,5 %.

Как и в ПФ Кобба–Дугласа, логарифмирование модели (формула (9.4)) приводит ее к линейному виду относительно параметра  $b$ , что позволяет проводить оценку функции с использованием методов линейного регрессионного анализа:

$$\ln y = \ln(a) + b \ln(p).$$

Если эластичность спроса на товар – переменная величина, то в качестве иллюстрации функции спроса по цене рассмотрим линейную функцию спроса:

$$y = a - bp.$$

Любую функцию спроса можно аппроксимировать в виде кусочно-линейной функции.

Найдем коэффициент эластичности спроса по цене для случая, когда спрос и цена находятся в линейной зависимости:

$$E_{y/p} = \frac{\partial y}{y} / \frac{\partial p}{p} = \frac{\partial y}{\partial p} \cdot \frac{p}{y} = -b \cdot \frac{p}{y}.$$

Из последнего выражения следует, что величина коэффициента эластичности  $E_{y/p}$  изменяется от точки к точке линии спроса  $y = a - bp$

в силу наличия множителя  $\frac{p}{y}$ .

При  $p \rightarrow 0$ ,  $E_{y/p} \rightarrow 0$ , т. е. при приближении цены к нулю спрос на товар остается практически неизменным. Более того, в некоторых случаях (группа товаров Гиффина, Веблена) при росте цен на определенные товары наблюдается рост спроса вместо ожидаемого уменьшения.

При  $p \rightarrow 0$ ,  $E_{y/p} \rightarrow -\infty$ , т. е. если товар является очень дорогим, то даже небольшое изменение цены может серьезно отразиться на объеме продаж.

При единичной эластичности спроса по цене ( $E_{y/p} = 1$ ) в случае линейной функции спроса от цены  $y = a - bp$  параметры цены и объема продаж имеют вид:  $p = \frac{a}{2b}$ ,  $y = \frac{a}{2}$ .

Несложно показать, что в этой точке достигается и максимальная выручка от реализации.

Пусть имеет место линейная зависимость объема продаж товара и его цены:  $y = a - bp$ . Тогда максимальная выручка от реализации достигается при уровне цены  $p = \frac{a}{2b}$ .

Определим аналитически выручку от реализации:  $f(p) = p y = p(a - bp)$ . Функция непрерывна, дифференцируема и выпукла. Необходимым условием максимума функции является равенство нулю ее производной:  $f'(p) = p \cdot y = (a - 2bp) = 0$ . Отсюда  $p = \frac{a}{2b}$ .

Чем ниже цена, тем меньше от ее изменения зависит количество проданного товара. И наоборот, чем выше цена, тем сильнее влияет ее изменение на объем проданного товара. Повышением цены на товар первой необходимости можно перевести его в товар эластичного спроса. Так, при значительном повышении цены на хлеб его будут покупать уже не целиком, а частями.

При изучении спроса по цене следует иметь в виду, что на спрос влияют также цены на сопряженный товар. Приведет ли изменение цены на сопряженный товар к повышению или снижению спроса на рассматриваемый продукт, зависит от того, является ли этот родственный товар заменителем продукта или сопутствует ему. Когда два продукта взаимозаменяемы, между ценой на один из них и спросом на другой существует прямая связь. Когда два товара являются взаимодополняющими, между ценой на один из них и спросом на другой существует обратная связь. Многие пары товаров являются независимыми, самостоятельными товарами, изменение цены на один очень мало влияет или вообще не влияет на спрос на другой, например: снижение тарифов на пассажирские авиаперевозки сокращает спрос на поездки автобусным транспортом; снижение цены на компьютеры повышает спрос на принтеры. Количественно влияние цены на спрос сопряженного товара можно оценить с использованием перекрестной эластичности спроса по цене.

*Перекрестный коэффициент эластичности* – это коэффициент, который показывает, на сколько процентов изменится спрос на товар  $A(Y_A)$  при изменении на 1 % цены на товар  $B(p_B)$  при условии, что другие цены и доходы покупателей остаются неизменными, и рассчитывается в виде:

$$E_{AB}^Y = \frac{\Delta y_A / y_A}{\Delta p_B / p_B}.$$

В зависимости от значений перекрестного коэффициента эластичности различают следующие связи между товарами  $A$  и  $B$ :  
 $E_{AB}^Y > 0$  – рост цены на продукт  $B$  приводит к увеличению спроса на продукт  $A$ , что свидетельствует о взаимозаменяемости товара;  
 $E_{AB}^Y < 0$  – рост цены на продукт  $B$  приводит к снижению спроса на продукт  $A$ , что свидетельствует о том, что товары являются сопутствующими (например, ноутбуки и комплектующие к ним);  
 $E_{AB}^Y = 0$  – товары независимы друг от друга.

*Пример 9.4.* Изучалась зависимость спроса на свинину  $Y_A$  от ее цены  $p_A$  и цены на говядину  $p_B$ . В результате было построено уравнение регрессии в линейно-логарифмической форме связи:  $\ln y_A = 0,127 - 0,814 \ln p_A + 0,82 \ln p_B + e_t$ . Требуется дать экономическую интерпретацию полученному решению.

*Решение.* В данном случае эластичность спроса на свинину по ее цене составляет  $(-0,814)$ , т. е. при увеличении цены на свинину на 1 % спрос на нее снижается на 0,814 %. Перекрестный коэффициент эластичности составил 0,82, следовательно, при росте цен на говядину на 1 % спрос на свинину увеличивается на 0,82 %.

Сравнение коэффициентов эластичности показывает примерно одинаковое влияние на спрос на свинину со стороны цены товара и его заменителя – говядины. В этом случае воздействие на спрос возможно не только на основе изменения цены товара, но и его заменителя.

*Модели, описывающие зависимость спроса от набора факторов.* В экономической литературе известны и другие факторы, оказывающие влияние на спрос: потребительские вкусы; число покупателей; ожидание; расходы на рекламу.



При построении эконометрических моделей спроса особое внимание необходимо обращать на формирование информационной базы, которая не всегда совпадает с фактическими объемами продаж. Например, при приобретении или потере клиента необходимо внести соответствующие корректировки в ряд потребления, описывающие динамику его потребления за последний год. Или, известно, что промоакции вызывают эффект бумеранга, выражающийся в падении продаж после проведения: надо полагать, что тот, кто собирался купить товар, сделал это в период кампании. По результатам мониторинга промоакций необходимо соответствующим образом уменьшить показатели потребления в период промоакции и увеличить после ее окончания.

Итоговая схема построения прикладной эконометрической модели содержит следующие этапы.

Этап 1. Выбор теоретической базы модели.

Этап 2. Модификация теоретической модели как обоснование формы связи эконометрической модели.

Этап 3. Формирование информационной базы модели.

Этап 4. Анализ временных рядов.

Этап 5. Математическое обеспечение модели и оценка адекватности модели.

Этап 6. Экономическая интерпретация модельных расчетов.

### **Вопросы для контроля знаний**

1. Каковы свойства ПФ?
2. Каковы основные положения теории полезности, необходимые для построения моделей спроса?
3. Как определить и описать кривые Энгеля?
4. Как принято оценивать параметры функции Торнквиста?
5. Перечислите виды функций эластичности.
6. Каков потенциально возможный набор факторов, формирующих спрос на товар?
7. Как построить эконометрическую модель спроса в зависимости от набора факторов?
8. Опишите общую схему построения прикладных эконометрических моделей. Охарактеризуйте сущность каждого этапа.

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Айвазян, С. А. Методы эконометрики : учебник / С. А. Айвазян. – М. : Магистр: ИНФРА-М, 2015. – 509 с.
2. Белько, И. В. Теория вероятностей, математическая статистика, математическое программирование : учебное пособие / И. В. Белько, И. М. Морозова, Е. А. Криштапович. – Минск : Новое знание; М. : ИНФРА-М, 2016. – 299 с.
3. Кремер, Н. Ш. Эконометрика : учебник для студентов вузов / Н. Ш. Кремер, Б. А. Путко ; под. ред. Н. Ш. Кремера. – 3 изд., перераб. и доп. – М. : ЮНИТИ-ДАНА, 2013. – 328 с.
4. Новиков, А. И. Эконометрика : учебное пособие / А. И. Новиков. – 3-е изд., перераб. и доп. – М. : ИНФРА-М, 2016. – 272 с.
5. Орлова, И. В. Экономико-математические методы и модели: компьютерное моделирование : учебное пособие / И. В. Орлова, В. А. Половников. – М. : Вузовский учебник, 2008. – 365 с.
6. Подашевский, И. Я. Эконометрика : учебно-методическое пособие / И. Я. Подашевский. – Минск : БИП-С Плюс, 2010. – 84 с.
7. Хацкевич, Г. А. Эконометрика : учебно-методический комплекс / Г. А. Хацкевич, А. Б. Гедранович. – 2-е изд., испр. и доп. – Минск : Изд-во МИУ, 2007. – 272 с.
8. Эконометрика и экономико-математические методы и модели : учебное пособие / Г. О. Читая [и др.]; под ред. Г. О. Читая, С. Ф. Миксюк. – Минск : БГЭУ, 2018. – 511 с.
9. Эконометрика : учебник для магистров / И. И. Елисеева [и др.]; под ред. И. И. Елисеевой. – М. : Юрайт, 2014. – 449 с.
10. Gujarati, Damodar N. Basic econometrics / Damodar N. Gujarati. – 3rd ed. – N.Y. : International Editions, 1995. – 838 p.

## ПРИЛОЖЕНИЕ 1

Значения  $t$ -критерия Стьюдента при уровне значимости 0,05

v	$\alpha$	v	$\alpha$
	0,05		0,05
1	12,706	18	2,1009
2	4,3027	19	2,093
3	3,1825	20	2,086
4	2,7764	21	2,0796
5	2,5706	22	2,0739
6	2,4469	23	2,0687
7	2,3646	24	2,0639
8	2,306	25	2,0595
9	2,2622	26	2,0555
10	2,2281	27	2,0518
11	2,201	28	2,0484
12	2,1788	29	2,0452
13	2,1604	30	2,0423
14	2,1448	40	2,0211
15	2,1315	60	2,0003
16	2,1199	120	1,9799
17	2,1098	$\infty$	1,96

## ПРИЛОЖЕНИЕ 2

Шкала Чеддока

Количественная мера тесноты связи $R$	Качественная характеристика силы связи
0,1–0,3	Слабая
0,3–0,5	Умеренная
0,5–0,7	Заметная
0,7–0,9	Высокая
0,9–0,99	Весьма высокая

### ПРИЛОЖЕНИЕ 3

Статистика Дарбина–Уотсона  $d_L$  и  $d_U$ ,  
уровень значимости 5 %

$n^*$	$k^* = 1$		$k^* = 2$		$k^* = 3$		$k^* = 4$		$k^* = 5$	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77

Окончание приложения 3

$n^*$	$k^* = 1$		$k^* = 2$		$k^* = 3$		$k^* = 4$		$k^* = 5$	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

\*  $n$  – число наблюдений,  $k$  – число объясняющих переменных (без учета постоянного члена).

## ПРИЛОЖЕНИЕ 4

### МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА И ПРОДОВОЛЬСТВИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования  
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ  
АГРАРНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра информационных технологий  
и моделирования экономических систем

### ЛАБОРАТОРНАЯ РАБОТА

по дисциплине «Количественные методы анализа в АПК»

Выполнил  
магистрант группы \_\_\_\_\_ И. И. Иванов  
(дата) (подпись)

Проверил \_\_\_\_\_ П. П. Петров  
(дата) (подпись)

Минск  
20\_\_

Учебное издание

**Белько** Иван Васильевич,  
**Сапун** Оксана Леонидовна,  
**Подашевская** Елена Игоревна и др.

## КОЛИЧЕСТВЕННЫЕ МЕТОДЫ АНАЛИЗА В АПК

Учебное пособие

Ответственный за выпуск *О. Л. Сапун*  
Редактор *Т. В. Каркоцкая*  
Компьютерная верстка *Д. О. Бабаковой*  
Дизайн обложки *Д. О. Бабаковой*

Подписано в печать 20.02.2020. Формат 60×84<sup>1</sup>/<sub>16</sub>.  
Бумага офсетная. Ризография.  
Усл. печ. л. 13,48. Уч.-изд. л. 10,55. Тираж 98 экз. Заказ 56.

Издатель и полиграфическое исполнение:  
Учреждение образования  
«Белорусский государственный аграрный технический университет».  
Свидетельство о государственной регистрации издателя, изготовителя,  
распространителя печатных изданий  
№ 1/359 от 09.06.2014.  
№ 2/151 от 11.06.2014.  
Пр-т Независимости, 99–2, 220023, Минск.