Учреждение образования
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

В. М. Марченко, Н. П. Можей, Е. А. Шинкевич

ЭКОНОМЕТРИКА И ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ

Допущено
Министерством образования Республики Беларусь в качестве учебного пособия для студентов учреждений высшего образования по экономическим специальностям

В 2-х частях

Часть 1. Эконометрика

Минск 2011

УДК 519.2:330.46(075.8)
ББК 22.172
М30

Рецензенты:
доктор экономических наук, профессор, проректор
по научной работе и инновациям Гродненского государственного
университета им. Янки Купалы *Г. А. Хацкевич*;
кафедра высшей математики Белорусского государственного
экономического университета (доктор физико-математических
наук, профессор, заведующий кафедрой *М. П. Дымков*)

Марченко, В. М.
М30 Эконометрика и экономико-математические методы и модели. В 2 ч. Ч. 1. Эконометрика : учеб. пособие для студентов учреждений высшего образования по экономическим специальностям / В. М. Марченко, Н. П. Можей, Е. А. Шинкевич. – Минск : БГТУ, 2011. – 157 с.
ISBN 978-985-530-123-4.

Предлагаемое учебное пособие – первая часть учебно-методического комплекса по курсу «Эконометрика и экономико-математические методы и модели (ЭиЭММ)», написанного в соответствии с уровневой методологией преподавания математических дисциплин.

Содержит программу курса ЭиЭММ, конспект лекций, практикум, лабораторный практикум, теоретический и практический минимум с примерами решения типовых задач, задания для самоконтроля, образцы контрольных работ и тестов по эконометрике.

Предназначено для студентов экономических специальностей. Будет полезно всем, кто интересуется ЭиЭММ.

УДК 519.2:330.46(075.8)
ББК 22.172

ISBN 978-985-530-123-4 (Ч. 1)
ISBN 978-985-530-122-7

ПРЕДИСЛОВИЕ

На современном этапе развития образовательных технологий одним из важнейших факторов повышения качества подготовки специалистов в высших учебных заведениях является рационализация учебного процесса посредством оптимальных учебных планов, программ нового поколения, новых форм и методов преподавания.

Предлагаемый учебно-методический комплекс написан в соответствии с уровневой технологией преподавания математических дисциплин.

Целью уровневой технологии организации учебного процесса является *создание условий для включения каждого обучаемого в деятельность, соответствующую зоне его ближайшего развития*, а также условий для самостоятельного (и/или под контролем преподавателя) усвоения программного материала в том размере и с той глубиной, которые позволяют индивидуальные особенности обучаемого.

Отметим некоторые принципиальные моменты уровневой технологии организации учебного процесса по математике в БГТУ. Материал классифицируется по трем уровням: **А**, **Б**, **С**. Первый уровень **А** (базовый) – обязательное поле знаний по предмету, программа-минимум – уровень знаний, необходимый для успешного продолжения обучения. Второй уровень **Б** (или *) содержит задания, расширяющие представление студента об изучаемых темах, устанавливает связи между понятиями и методами различных разделов, дает их строгое математическое обоснование, а также примеры применения при решении прикладных задач. Материал **А + Б** (профильный) уровней **А** и **Б** охватывает всю стандартную программу – программу-максимум. Уровень **С** (или ** – необязательный) содержит материал повышенной трудности. Материал **А + Б + С** трех уровней – углубленная программа – открывает путь исследованиям в области ЭиЭММ и их приложений.

Структурно комплекс организован следующим образом: программа курса, конспект лекций, практикум, лабораторный практикум, теоретический и практический минимум с примерами решения типовых задач, задания для самоконтроля, образцы контрольных работ и тестов.

ВВЕДЕНИЕ

Постоянно усложняющиеся экономические процессы привели к необходимости создания и совершенствования методов их изучения и анализа. При этом широкое распространение получили моделирование и количественный анализ.

Изучение экономических процессов (взаимосвязей) в эконометрике осуществляется через математические (эконометрические) модели. В этом заключается ее родство с математической экономикой. Но если математическая экономика строит и анализирует эти модели без использования реальных числовых значений, то эконометрика концентрируется, главным образом, на изучении моделей на базе эмпирических данных.

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

Развитие компьютерных систем и создание специальных прикладных программ, совершенствование методов анализа сделали эконометрику мощнейшим инструментом экономических исследований.

1. УЧЕБНАЯ ПРОГРАММА

1.1. ТЕМАТИЧЕСКИЙ ПЛАН КУРСА

Курс «Эконометрика и экономико-математические методы и модели» базируется на программах курсов высшей математики, статистики, экономической теории. Весь материал классифицирован по трем уровням глубины – базовый (без звездочек), дополнительный * (одна звездочка) и углубленный ** (две звездочки). Материал, отмеченный **, является необязательным.

На изучение курса эконометрики и экономико-математических методов и моделей учебным планом специальностей 1-25 01 07 «Экономика и управление на предприятии», 1-25 01 08 «Бухгалтерский учет, анализ и аудит», 1-26 02 03 «Маркетинг» предусмотрено 26 часов лекций, 17 часов практических и 8 часов лабораторных занятий.

В табл. 1 приводится примерный тематический план курса «Эконометрика и экономико-математические методы и модели» с распределением изучаемого материала по лекциям, практическим и лабораторным занятиям.

Таблица 1

Примерный тематический план курса ЭиЭММ

Тема	Количество аудиторных часов		
	Лекции	Практические занятия	Лабораторные работы
1	2	3	4
Введение	**1**	**–**	**–**
Раздел 1. Эконометрика	**11**	**7**	**4**
1.1. Основные понятия эконометрики	1	–	–
1.2. Элементы корреляционно-регрессионного анализа	4	3	2
1.3. Эконометрический анализ при нарушении классических предположений. Временные ряды	4	2	2
1.4. Системы одновременных уравнений	2	2	–
Раздел 2. Экономико-математические методы и модели	**14**	**10**	**4**
2.1. Модели оптимального планирования	2	–	2

Окончание табл. 1

1	2	3	4
2.2. Модели межотраслевого баланса	3	2	2 (БУ)
2.3. Методы и модели массового обслуживания	2	2	–
2.4. Элементы теории игр	3	2	2 (МК)
2.5. Модели управления запасами	2	2	–
2.6. Сетевое планирование и управление. Инвестиционные модели	2	2	2 (ЭУП)
Итого	26	17	8

Примечание. Лабораторные работы, выполняемые студентами разных специальностей, могут отличаться.

Примерная тематика практических занятий

1. Элементы корреляционно-регрессионного анализа.
2. Эконометрический анализ при нарушении классических предположений. Временные ряды.
3. Системы одновременных уравнений.
4. Модели межотраслевого баланса.
5. Методы и модели систем массового обслуживания.
6. Элементы теории игр.
7. Модели управления запасами.
8. Сетевое планирование и управление.

Примерная тематика лабораторных занятий

1. Построение и анализ уравнения парной линейной регрессии.
2. Построение и анализ уравнения множественной регрессии.
3. Выбор вида зависимости, построение уравнения регрессии, его анализ.
4. Модели оптимального планирования.
5. Модели межотраслевого баланса. Построение и анализ межотраслевого баланса.
6. Системы массового обслуживания. Расчет основных параметров.
7. Теория игр. Нахождение оптимальных стратегий.
8. Сетевое планирование и управление. Расчет временных параметров сетевого графика. Построение графика Ганта.

1.2. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Введение

Основные понятия, предмет и область применения эконометрики и экономико-математического моделирования. *Теоретические основы экономико-математического моделирования.**

Раздел 1. Эконометрика

Основные понятия эконометрики

Предмет и методы эконометрики. Понятие эконометрической модели, классификация моделей. Основные этапы построения эконометрической модели.

Элементы корреляционно-регрессионного анализа

Виды функциональной и корреляционной зависимости. *Задачи построения качественного уравнения регрессии.** Линейная регрессия. Метод наименьших квадратов. *Основные положения регрессионного анализа.** Оценки параметров регрессионной модели и их свойства. *Интервальная оценка функции регрессии.** Коэффициент детерминации. Модель множественной регрессии. *Спецификация эконометрической модели.* Методы выбора экзогенных переменных.** Методы выбора вида зависимости, нелинейная регрессия. *Модели с качественными переменными.***

Эконометрический анализ при нарушении классических предположений. Временные ряды

Эконометрический анализ при нарушении классических предположений. Мультиколлинеарность, ее обнаружение и устранение. Гетероскедастичность и автокорреляция остатков модели: обнаружение, устранение и анализ последствий. *Модели и методы анализа стационарных* и нестационарных** временных рядов. Автокорреляционная функция.**

Системы одновременных уравнений

Системы одновременных уравнений. *Построение и анализ многомерных эконометрических моделей.**

Раздел 2. Экономико-математические методы и модели

Модели оптимального планирования

Экономико-математические методы и модели оптимального планирования в промышленности. Экономико-математические методы и модели оптимального планирования в АПК. Экономико-математические методы и модели финансов и кредита. Экономико-математические методы и модели в коммерческой деятельности. Экономико-математические методы и модели в управлении социально-культурной сферой.

Модели межотраслевого баланса

Модели межотраслевого баланса (МОБ), основные понятия, методы построения МОБ *и их использование в анализе.**

Методы и модели массового обслуживания

Методы и модели массового обслуживания, основные понятия и классификация систем массового обслуживания (СМО), графическое представление СМО, расчет основных характеристик. Уравнения Колмогорова. Финальные вероятности состояний СМО. СМО с отказами. *СМО с неограниченной очередью.**

Элементы теории игр

Моделирование конфликтных ситуаций с помощью теории игр, основные понятия и классификация. Матричные игры с нулевой суммой. Решение матричных игр в чистых стратегиях. Игры с «природой». Критерии Байеса, Лапласа, Вальда, Сэвиджа, Гурвица. *Решение матричных игр в смешанных стратегиях.**

Модели управления запасами

Задачи и модели управления запасами и сбытом готовой продукции. Модель Уилсона. *Модель производственных поставок.**

Сетевое планирование и управление. Инвестиционные модели

Математические методы сетевого планирования и управления (СПУ). Основные понятия СПУ. *Правила построения сетевых графиков.** Расчет основных параметров сетевого графика. Построение календарного графика. *Инвестиционные модели.***

2. КОНСПЕКТ ЛЕКЦИЙ

Лекция 1

2.1. ОСНОВНЫЕ ПОНЯТИЯ ЭКОНОМЕТРИКИ

1А1 (*Эконометрика*). *Эконометрика* – наука, объединяющая совокупность математико-статистических методов, которые позволяют дать количественное выражение взаимосвязей экономических явлений и процессов.

Эконометрика позволяет найти количественное подтверждение (либо неподтверждение) того или иного экономического закона или гипотезы. Одним из важнейших направлений эконометрики является построение прогнозов по различным экономическим показателям.

Эконометрика как научная дисциплина возникла и получила развитие на основе экономической теории, математической экономики, экономической и математической статистики. Эконометрика делает упор на количественные, а не на качественные аспекты явлений.

1А2 (*Цель эконометрики*). Эконометрика занимается разработкой способов моделирования и количественного анализа реальных экономических объектов.

1А+Б3 (*Задачи эконометрики*). Задачами эконометрики являются:

3.1) *спецификация модели* – построение эконометрических моделей для эмпирического анализа;

3.2) *параметризация модели* – оценка параметров модели;

3.3) *верификация модели* – проверка качества параметров модели и самой модели в целом;

3.4) *прогнозирование модели* – составление прогноза и рекомендаций для конкретных экономических явлений по результатам моделирования.

1А4 (*Эконометрическая модель*). *Эконометрическая модель* – совокупность математических соотношений между входными (объясняющими, независимыми, экзогенными) и выходными (объясняемыми, зависимыми, эндогенными) переменными изучаемого экономического явления или процесса, основанная на реальных статистических данных. *Эконометрическое моделирование* –

исследование экономических явлений и процессов посредством их эконометрических моделей.

Эконометрическая модель учитывает реальные условия существования объекта и не противоречит общим законам экономики.

1А+Б5 (*Этапы эконометрического моделирования*). Эконометрическое моделирование представляет собой комплексное решение целого ряда задач, поэтому весь процесс можно разделить на следующие этапы:

5.1) *постановочный*: формулировка цели исследования (анализ, прогноз, имитация развития, управленческое решение и т. д.), определение экономических переменных модели;

5.2) *априорный*: анализ изучаемого экономического явления, формирование и формализация информации, известной до начала моделирования;

5.3) *параметризации*: определение вида экономической модели, выражение в математической форме взаимосвязи между ее переменными, формулирование исходных предпосылок и ограничений модели;

5.4) *информационный*: сбор необходимой статистической информации;

5.5) *идентификации модели*: оценка параметров модели, проведение ее статистического анализа;

5.6) *верификации модели*: проверка истинности модели, определение степени соответствия построенной модели реальному экономическому явлению.

Обычно для построения работоспособной модели и сравнения ее с другими возможными моделями необходимо учитывать следующие свойства: модель должна быть максимально простой; для любого набора статистических данных определяемые коэффициенты должны вычисляться однозначно; уравнение тем лучше, чем большую часть разброса зависимой переменной оно может объяснить; уравнение должно удовлетворять известным теоретическим предпосылкам; модель может быть признана качественной, если полученные на ее основе прогнозы подтверждаются реальностью.

1А6 (*Классы эконометрических моделей*). Все эконометрические модели условно делят на три класса:

6.1) *регрессионные модели с одним уравнением.* Результативный признак представлен в виде функции от факторных признаков $Y = f(X) + \varepsilon = f(X_1, X_2, \ldots, X_k) + \varepsilon$, где Y – *наблюдаемое значение* объясняемой (зависимой, *эндогенной*) переменной (результата);

$f(X)$ – *объясненная часть*, которая зависит от значений объясняющих (независимых, *экзогенных*) переменных (факторов); ε – *случайная составляющая* (ошибка, возмущение).

Объясняемая переменная Y – случайная величина (СВ) при заданных значениях объясняющих переменных X_i, $i = \overline{1, k}$. Объясняющие переменные в модели могут иметь случайные или определенные значения.

Например, модель зависимости цены от объема поставки, модель зависимости спроса от цены на отдельный товар, модель зависимости спроса от реальных доходов потребителей, модель зависимости объема производства от производственных факторов;

6.2) *системы одновременных уравнений.* Они состоят из тождеств и регрессионных уравнений, в которые наряду с факторными признаками включены результативные признаки из других уравнений системы. Одни и те же переменные одновременно рассматриваются как зависимые переменные в одних уравнениях и независимые – в других. Например, модель спроса и предложения (см. 1А+Б7), кейнсианская модель формирования доходов;

6.3) *модели временных рядов.* Результативный признак является функцией переменной времени или переменных, относящихся к другим моментам времени.

1А+Б7 (*Пример*). Модель спроса и предложения представляет собой систему одновременных уравнений:

$$\begin{cases} Q^D = \alpha_{01} + \alpha_{11}P + \alpha_{12}I + \varepsilon_1, \\ Q^S = \alpha_{02} + \alpha_{21}P + \varepsilon_2, \\ Q^D = Q^S, \end{cases}$$

где Q^D – спрос на товар; Q^S – предложение товара; P – цена товара; I – доход населения.

Цена товара P, спрос на товар Q^D и предложение товара Q^S являются зависимыми (объясняемыми, эндогенными) переменными и определяются из уравнений модели, а величина дохода I является независимой (объясняющей, экзогенной) переменной.

1А8 (*Типы данных*). В эконометрическом моделировании рассматриваются следующие типы данных:

8.1) *пространственные данные* – набор сведений по разным объектам, взятым за один и тот же период времени (объем производства предприятий региона, численность сотрудников институтов города);

8.2) *временные данные* – набор сведений, характеризующий один и тот же объект за разные периоды времени (индекс потребительских цен, численность занятых за последние годы).

2.2. ЭЛЕМЕНТЫ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА

2.2.1. Основные понятия корреляционного анализа

1А9 (*Определение*). *Корреляционный анализ* – раздел математической статистики, изучающий тесноту связи между признаками (между двумя признаками при парной связи и между результативным и множеством факторных признаков при многофакторной связи). *Регрессионный анализ* – раздел математической статистики, изучающий форму связи между признаками.

1А10 (*Типы зависимостей*). Различают следующие типы зависимостей между явлениями и их признаками:

10.1) *функциональная зависимость* – связь, при которой каждому значению независимой переменной X соответствует точно определенное значение зависимой переменной Y (зависимость выработки продукции на одного рабочего от объема выпущенной продукции и численности рабочих);

10.2) *статистическая зависимость* – связь, при которой каждому значению независимой переменной X соответствует множество значений зависимой переменной Y и изменение которой происходит в условиях неопределенности, имеющей, как правило, случайный характер (зависимость всхожести семян некоторых культур от количества микроэлементов при их обработке, зависимость производительности труда на предприятии от его энерговооруженности и т. д.);

10.3) *корреляционная зависимость* – частный случай статистической зависимости – связь, при которой каждому значению независимой переменной X соответствует определенное математическое ожидание (среднее значение) зависимой переменной Y.

1А11 (*Определение*). *Условным математическим ожиданием* (*условной средней*) $M_x(Y) = M(Y|X = x) = \overline{Y}_x$ называется математическое ожидание СВ Y, вычисленное в предположении, что СВ X приняла значение x.

1А12 (*Виды корреляционной зависимости*). Корреляционная зависимость бывает трех видов:

12.1) *парная* – связь между двумя признаками (результативным Y и факторным X или двумя факторными);

12.2) *частная* – зависимость между результативным и одним факторным признаком или между двумя факторными признаками при фиксированных значениях других факторных признаков;

12.3) *множественная* – зависимость между результативным признаком и двумя или более факторными признаками, включенными в исследование.

Теснота связи количественно выражается величиной коэффициента корреляции.

1А13 (*Виды связей в зависимости от направления действия*). В зависимости от направления действия выделяют два вида функциональной и корреляционной связей:

13.1) *прямая*: с увеличением (уменьшением) значений факторного признака происходит увеличение (уменьшение) результативного признака;

13.2) *обратная*: с увеличением (уменьшением) значений факторного признака происходит уменьшение (увеличение) результативного признака.

1А14 (*Виды связей в зависимости от количества признаков*). В зависимости от количества признаков, включенных в модель, связи подразделяются на два вида:

14.1) *однофакторные* – связи между одним признаком-фактором и результативным признаком (при абстрагировании от влияния других факторов);

14.2) *многофакторные* – связи между несколькими факторными признаками и результативным признаком (факторы действуют комплексно, т. е. одновременно и во взаимосвязи).

2.2.2. Понятие о регрессионной модели

1А15 (*Определение*). *Теоретическим уравнением* (или просто *уравнением*) *регрессии* Y на X называется уравнение $\overline{Y}_x = f(x)$. Функция $f(x)$ называется *теоретической регрессией* (или просто *регрессией*) Y на X, а ее график – *линией регрессии* СВ Y на СВ X. При этом X

является независимой (объясняющей) переменной, Y – зависимой (объясняемой) переменной. При рассмотрении зависимости двух СВ говорят о *парной* регрессии.

Зависимость нескольких переменных, выражаемую функцией

$$M(Y \mid x_1, x_2, ..., x_k) = f(x_1, x_2, ..., x_k),$$

где $M(Y \mid x_1, x_2, ..., x_k)$ – условное математическое ожидание (математическое ожидание СВ Y при условии, что СВ X_1, X_2, ..., X_k приняли значения x_1, x_2, ..., x_k соответственно), называют *множественной регрессией*.

Поскольку реальные значения зависимой переменной не всегда совпадают с ее условными математическими ожиданиями и могут быть различными при одном и том же значении объясняющей переменной (наборе объясняющих переменных), фактическая зависимость должна учитывать ошибку (погрешность) ε, которая также является СВ. Таким образом, связи между зависимой и объясняющей(ими) переменными можно описать соотношениями:

$$Y = M(Y \mid x) + \varepsilon;$$

$$Y = M(Y \mid x_1, x_2, ..., x_k) + \varepsilon.$$

1Б16 (*Наличие случайного фактора*). Наиболее существенные причины присутствия в регрессионных моделях случайного фактора (ошибки, отклонения):

16.1) *невключение в модель всех объясняющих переменных.* Любая регрессионная (в частности, эконометрическая) модель является упрощением реальной ситуации, которая всегда представляет собой сложнейшее переплетение различных факторов. Многие из этих факторов в модели не учитываются, что порождает отклонение реальных значений зависимой переменной от ее модельных значений;

16.2) *неправильный выбор функциональной формы модели.* Из-за слабой изученности исследуемого процесса либо из-за его переменчивости может быть неверно подобрана функция, его моделирующая. Например, *производственная функция (Y) одного фактора (X)* может моделироваться функцией $Y = a + bX$, хотя должна была использоваться другая модель: $Y = aX^b$ $(0 < b < 1)$, учитывающая *закон убывающей эффективности*. Кроме того, неверным может быть подбор объясняющих переменных;

16.3) *агрегирование переменных.* Во многих моделях рассматриваются зависимости между факторами, которые сами представляют сложную комбинацию других, более простых переменных. Например, при рассмотрении в качестве зависимой переменной совокупного спроса проводится анализ зависимости, в которой объясняемая переменная является сложной композицией индивидуальных спросов, оказывающих на нее определенное влияние помимо факторов, учитываемых в модели. Это также может оказаться причиной отклонения реальных значений от модельных;

16.4) *ошибки измерений.* Какой бы качественной ни была модель, ошибки измерений переменных могут привести к несоответствию модельных значений эмпирическим данным, что также отразится на величине случайной переменной;

16.5) *ограниченность статистических данных*. Как правило, модели, описываемые непрерывными функциями, строятся по данным, имеющим дискретную структуру. Данное несоответствие также находит свое выражение в случайном отклонении;

16.6) *непредсказуемость человеческого фактора.* Даже при правильном выборе формы модели и подборе объясняющих переменных невозможно спрогнозировать поведение каждого индивидуума.

2.2.3. Задачи и этапы корреляционно-регрессионного анализа

1А+Б17 (*Основные задачи корреляционно-регрессионного анализа*). Перед корреляционно-регрессионным анализом стоят следующие основные задачи:

17.1) установление формы корреляционной связи, т. е. установление вида функции регрессии (линейная, квадратичная, показательная и т. д.);

17.2) оценка тесноты корреляционной связи Y и X (степени рассеяния значений СВ Y около $\bar{Y}_x$). Большое рассеяние означает слабую зависимость Y от X либо вообще ее отсутствие. Малое рассеяние указывает на существование достаточно сильной зависимости Y от X;

17.3) оценка неизвестных параметров регрессионной модели, проверка гипотез об их значимости и адекватности модели рассматриваемому экономическому объекту.

1А+Б18 (*Этапы корреляционно-регрессионного анализа*). Корреляционно-регрессионный анализ проводится поэтапно в определенной логической последовательности:

18.1) предварительный анализ явлений и выяснение причин возникновения взаимосвязей между признаками, характеризующими эти явления, разделение признаков на факторные и результативные, выбор наиболее существенных признаков;

18.2) предварительная оценка формы уравнения регрессии и определение уравнения регрессии, расчет теоретически ожидаемых значений результативного признака, оценка тесноты связи между признаками, включенными в регрессионную модель;

18.3) общая оценка качества модели, отсев несущественных (или включение дополнительных) факторов, построение исправленной модели;

18.4) статистическая оценка достоверности параметров уравнения регрессии, построение доверительных границ для теоретически ожидаемых по уравнению регрессии значений функции и получение практических выводов из проведенного анализа.

Отметим, что в корреляционном анализе изучается в основном сила (теснота) корреляционной зависимости, а в регрессионном анализе – ее форма.

2.2.4. Проблемы спецификации

1А19 (*Определение*). Выбор формулы связи переменных называется *спецификацией* уравнения регрессии. В случае парной регрессии выбор формулы обычно осуществляется по графическому изображению реальных статистических данных.

1А+Б20 (*Этапы спецификации модели*). Стандартная схема анализа зависимостей включает следующие этапы:

20.1) *подбор начальной модели.* Осуществляется на основе экономической теории, предыдущих знаний об объекте исследования, опыта исследователя и его интуиции;

20.2) *оценка параметров модели* на основе имеющихся статистических данных;

20.3) *осуществление тестов проверки качества модели* (обычно используются t-статистики для коэффициентов регрессии, F-статистика для коэффициента детерминации, статистика Дарбина – Уотсона для анализа отклонений и ряд других тестов). При наличии хотя бы одного неудовлетворительного ответа по какому-либо тесту *модель совершенствуется* с целью устранения выявленного недостатка. При положительных ответах по всем проведенным тестам модель считается *качественной*. Она используется для анализа и прогноза объясняемой переменной.

1Б21 (*Типы ошибок спецификации*). Правильная спецификация уравнения регрессии означает, что оно в целом верно отражает соотношение между экономическими показателями, участвующими в модели. Неправильный выбор функциональной формы или набора объясняющих переменных называется *ошибкой спецификации*. Выделяют следующие основные типы ошибок спецификации:

21.1) *отбрасывание значимой переменной.* Оценки, полученные по такому уравнению, могут быть смещенными и несостоятельными даже при большом числе испытаний. Возможные интервальные оценки и результаты проверки соответствующих гипотез будут ненадежными;

21.2) *добавление незначимой переменной.* В некоторых случаях в уравнения регрессии включают слишком много объясняющих переменных, причем не всегда обоснованно. Оценки коэффициентов, как правило, остаются несмещенными и состоятельными, однако их точность уменьшается, т. е. оценки становятся неэффективными, что отражается на их устойчивости. Увеличение дисперсии оценок может привести к ошибочным результатам проверки гипотез относительно значений коэффициентов регрессии, расширению интервальных оценок;

21.3) *выбор неправильной функциональной формы.* Обычно такая ошибка приводит либо к получению смещенных оценок, либо к ухудшению статистических свойств оценок коэффициентов регрессии и других показателей качества уравнения. Прогнозные качества модели в этом случае очень низки.

1С22 (*Пути совершенствования моделей*). Одно из главных направлений эконометрического анализа – постоянное совершенствование моделей. Какого-то общего подхода, заранее определяющего возможные пути совершенствования, не существует. В силу постоянно изменяющихся условий протекания экономических процессов не может быть и постоянно качественных моделей. Распространены два пути совершенствования моделей:

22.1) *начинать с самой простой модели и постоянно усложнять ее.* Иногда приводит к обыкновенной подгонке модели под эмпирические данные;

22.2) *начинать с максимально сложной модели и упрощать ее на основе проводимых исследований.* Поиск возможных направлений совершенствования модели зачастую сводится к полному перебору, что делает проводимый анализ неэффективным. На этапах упрощения модели возможно также отбрасывание объясняющих переменных, которые были бы весьма полезны в упрощенной модели.

Построение модели является индивидуальным в каждой конкретной ситуации и опирается на знания экономической теории и статистического анализа. При всех недостатках моделей принятие решений на их основе приводит в целом к более точным результатам, чем при принятии решений лишь на основе интуиции и экономической теории.

Лекция 2

2.2.5. Линейная парная регрессия

1А23 (*Теоретическое уравнение регрессии*). *Теоретическим уравнением* (или просто *уравнением*) *регрессии* Y на X называется уравнение $\overline{Y}_x = f(x)$. Функция $f(x)$ называется *теоретической регрессией* (или просто *регрессией*) Y на X. При этом X является независимой (объясняющей) переменной, Y – зависимой (объясняемой) переменной.

1А24 (*Эмпирическое уравнение линейной регрессии*). По выборке ограниченного объема можно искать регрессионную зависимость в определенном виде, например в виде линейной зависимости (эмпирическое линейное уравнение регрессии):

$$\hat{y}_i = b_0 + b_1 x_i + e_i, \qquad (*)$$

где $\hat{y}_i$ – *оценка* условного математического ожидания $M(Y \mid X = x_i)$; b_0 и b_1 – *оценки* неизвестных параметров, называемые *эмпирическими коэффициентами линейной регрессии*; отклонение e_i – оценка теоретического случайного отклонения ε_i.

Модель линейной регрессии (линейное уравнение) является наиболее распространенным (и простым) видом зависимости между экономическими переменными. Кроме того, построенное линейное уравнение может служить начальным этапом эконометрического анализа.

1А25 (*Задачи линейного регрессионного анализа*). Выделяют следующие задачи линейного регрессионного анализа:

25.1) по имеющимся статистическим данным $(x_i;\, y_i)$, $i = \overline{1,\, n}$, получить наилучшие оценки неизвестных параметров;

25.2) проверить статистические гипотезы о параметрах модели;

25.3) проверить, достаточно ли хорошо модель согласуется со статистическими данными (адекватность модели данным наблюдений).

2.2.6. Метод наименьших квадратов

1А26 (*Метод наименьших квадратов*). Различные выборки из одной и той же генеральной совокупности обычно приводят к определению отличающихся друг от друга оценок. Требуется по конкретной выборке $(x_i;\, y_i)$, $i=\overline{1,\, n}$, найти оценки b_0 и b_1 неизвестных параметров уравнения (*) так, чтобы соответствующая линия регрессии (прямая) являлась бы наилучшей в некотором смысле среди всех других прямых. Другими словами, построенная прямая должна быть «ближайшей» к совокупности точек наблюдений. Мерами качества найденных оценок могут служить определенные функции отклонений (невязок) $e_i = y_i - \hat{y}_i$, $i=\overline{1,\, n}$ (рис. 1).

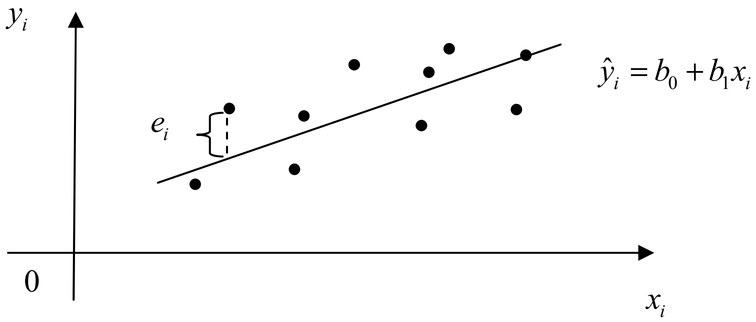


Рис. 1. Линейная регрессия

Самым распространенным методом нахождения коэффициентов (оценок) b_0 и b_1 уравнения эмпирической линейной регрессии является метод наименьших квадратов (МНК). Согласно МНК, эти коэффициенты выбираются таким образом, чтобы минимизировать функцию (сумму квадратов отклонений)

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

Необходимым условием минимума данной функции является равенство нулю ее частных производных по параметрам b_0 и b_1, откуда после преобразований получается система уравнений для определения параметров линейной регрессии.

1А27 (*Система уравнений для определения параметров линейной регрессии*). Коэффициенты b_0 и b_1, обеспечивающие минимум функции $s(b_0; b_1)$, удовлетворяют системе линейных алгебраических уравнений вида

$$\begin{cases} nb_0 + b_1 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} y_i, \\ b_0 \sum\limits_{i=1}^{n} x_i + b_1 \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i y_i. \end{cases}$$

Коэффициент b_1 называется *выборочным коэффициентом регрессии* Y на X. Он показывает, на сколько единиц в среднем изменяется переменная Y при увеличении переменной X на одну единицу.

1А+Б28 (*Коэффициент эластичности*). Коэффициент b_1 нельзя непосредственно использовать для оценки влияния факторного признака x на результативный признак y из-за различия единиц измерения исследуемых показателей. Для этих целей применяется коэффициент эластичности. Для эмпирической линейной регрессии коэффициент эластичности

$$Э_{yx} = b_1 \frac{\overline{x}}{\overline{y}},$$

где $\overline{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$, $\overline{y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$ – средние значения независимой и зависимой переменных.

Напомним, что в общем случае коэффициент эластичности определяется по формуле:

$$Э_{yx} = y' \frac{x}{y}.$$

Коэффициент эластичности показывает, на сколько процентов изменяется результативный признак y при изменении факторного признака x на 1%.

1С29 (*Замечание*). Среди других методов определения оценок коэффициентов регрессии отметим метод наименьших модулей (МНМ), метод моментов (ММ) и метод максимального правдоподобия (ММП).

2.2.7. Основные положения регрессионного анализа. Теорема Гаусса – Маркова. Оценки параметров регрессионной модели и их свойства

МНК обеспечивает оптимальные свойства оценок лишь при выполнении следующих классических предположений.

1А+Б30 (*Предпосылки МНК*). Для получения по МНК наилучших в классе линейных несмещенных оценок результатов необходимо, чтобы выполнялся ряд предпосылок относительно случайного отклонения. Предпосылки МНК (условия Гаусса – Маркова) для парной линейной регрессии:

30.1) *математическое ожидание случайного отклонения* ε_i *равно нулю для всех наблюдений:* $M(\varepsilon_i) = 0$. Данное условие означает, что случайное отклонение в среднем не оказывает влияния на зависимую переменную;

30.2) *дисперсия случайных отклонений* ε_i *постоянна для любых* i *и* j*:* $D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2$. Выполнимость данной предпосылки называется *гомоскедастичностью* (постоянством дисперсий отклонений). Так как $D(\varepsilon_i) = M(\varepsilon_i - M(\varepsilon_i))^2 = M(\varepsilon_i^2)$, то данное условие можно записать в виде: $M(\varepsilon_i^2) = \sigma^2$;

30.3) *отсутствие корреляции между случайными отклонениями* ε_i *и* ε_j *для* $i \neq j$. Если данное условие выполняется, то говорят об отсутствии *автокорреляции*;

30.4) *случайное отклонение должно быть независимо от объясняющих переменных.* Обычно это условие выполняется автоматически, если объясняющая переменная x_i не является случайной в данной модели. Без этого требования нельзя решить задачу построения доверительных интервалов для оценки и проверки гипотез о численных значениях регрессионных коэффициентов;

30.5) *случайное отклонение* ε_i *есть нормально распределенная случайная величина.*

1А+С31 (*Теорема Гаусса – Маркова*). Если предпосылки МНК выполнены, то оценки b_0 и b_1, полученные по этому методу, являются:

31.1) несмещенными, т. е. математические ожидания отклонений оценок равны нулю, что говорит об отсутствии систематической ошибки в определении положения линии регрессии;

31.2) состоятельными (т. е. дисперсии оценок параметров при возрастании числа n наблюдений стремятся к нулю), если $\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n} \to \text{const}$ при $n \to \infty$. Другими словами, при увеличении объема выборки надежность оценок увеличивается (см. с. 95 пособия [16]);

31.3) эффективными, т. е. они имеют наименьшую дисперсию по сравнению с любыми другими оценками данных параметров, линейными относительно величин y_i.

В англоязычной литературе такие оценки называются BLUE (*Best Linear Unbiased Estimators*) – *наилучшие линейные несмещенные оценки.*

1Б+С32 (*Замечания*). Замечания:

32.1) если дисперсии отклонений ε_i непостоянны и/или значения ε_i и ε_j связаны друг с другом, то свойства несмещенности и состоятельности сохраняются, а свойство эффективности – нет;

32.2) для получения уравнения регрессии достаточно выполнения предпосылок 30.1–30.4. Требование выполнения предпосылки 30.5 необходимо для оценки точности уравнения регрессии и его параметров;

32.3) эмпирическое уравнение регрессии определяется на основе конечного числа статистических данных. Поэтому коэффициенты эмпирического уравнения регрессии являются СВ, изменяющимися от выборки к выборке. Для сравнения эмпирических коэффициентов b_0 и b_1 регрессии с теоретически ожидаемыми значениями этих коэффициентов используется схема статистической проверки гипотез.

1А+Б33 (*Критерий Стьюдента для проверки гипотез о статистической значимости коэффициентов регрессии*). Для проверки гипотез о статистической значимости коэффициента регрессии, т. е. гипотез

$$H_0\text{: } b_1 = 0;\ H_1\text{: } b_1 \neq 0,$$

используется t-статистика:

$$t = \frac{b_1}{S_{b_1}},$$

которая, при выполнении исходных предпосылок модели, имеет распределение Стьюдента с числом степеней свободы $\gamma = n - k - 1$, где n – число наблюдений; k – число независимых переменных в уравнении регрессии (если речь идет о парной линейной регрессии, то $k = 1$); S_{b_1} – стандартная ошибка коэффициента регрессии, т. е.

$$S_{b_1} = \frac{S_e}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}; \quad S_e = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-k-1}},$$

где S_e – стандартная ошибка регрессии.

Гипотеза H_0 отклоняется, если $\left|t_{\text{расч}}\right| = \left|\frac{b_1}{S_{b_1}}\right| \geq t_{\text{табл}} = t_{\alpha;n-k-1}$, где α – требуемый уровень значимости.

Если гипотеза H_0 принимается, то есть основания считать, что величина Y не зависит от X. В этом случае говорят, что коэффициент b_1 статистически незначим. При отклонении гипотезы H_0 коэффициент b_1 считается статистически значимым, что указывает на наличие линейной зависимости между Y и X.

Для парной регрессии более важным является анализ статистической значимости коэффициента b_1, т. к. именно он позволяет оценить влияние объясняющей переменной X на зависимую переменную Y.

Воздействие неучтенных факторов и ошибок наблюдений определяется с помощью дисперсии случайных отклонений $D(\varepsilon_i)$. Несмещенной оценкой этой дисперсии является выборочная остаточная дисперсия.

1А34 (*Выборочная остаточная дисперсия*). Выборочная остаточная дисперсия определяется по формуле:

$$S_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1}.$$

1А35 (*Стандартная ошибка регрессии*). Корень квадратный из выборочной остаточной дисперсии, т. е. $S_e = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-k-1}}$, называется *стандартной ошибкой регрессии* (стандартной ошибкой оценки).

2.2.8. Интервальная оценка функции регрессии

1А36 (*Точечный прогноз*). Прогнозируемое значение переменной y вычисляется по формуле:

$$y^*_{\text{прогн}} = b_0 + b_1 x_{\text{прогн}}.$$

Данный прогноз называется *точечным*. Вероятность реализации точечного прогноза практически равна нулю, поэтому рассчитывается доверительный интервал прогнозируемого значения.

1А+Б37 (*Доверительный интервал для функции регрессии*). *Доверительный интервал для функции регрессии* (т. е. для условного математического ожидания $M_x(Y)$), который с заданной надежностью (доверительной вероятностью) $\gamma = 1 - \alpha$ накрывает неизвестное значение математического ожидания, определяется по формуле:

$$\hat{y} - t_{\alpha;\gamma} \cdot S_{\hat{y}} \le M_x(Y) \le \hat{y} + t_{\alpha;\gamma} \cdot S_{\hat{y}},$$

где $S^2_{\hat{y}} = \dfrac{\sum\limits_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}\left(\dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2}\right) = S_e^2\left(\dfrac{1}{n} + \dfrac{(x-\overline{x})^2}{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2}\right)$; $t_{\alpha;\gamma}$ определяется по таблице распределения Стьюдента; $\gamma = k = n - 2$ – число степеней свободы.

Доверительные интервалы прогноза зависят от стандартной ошибки регрессии, удаления x от своего среднего значения $\overline{x}$, количества наблюдений n и уровня значимости прогноза α.

Величина доверительного интервала зависит от значения объясняющей переменной x: при $x = \overline{x}$ она минимальна, а по мере удаления $x_{\text{прогн}}$ от $\overline{x}$ величина доверительного интервала увеличивается. Использование уравнения регрессии вне пределов обследованного диапазона значений объясняющей переменной (даже если оно оправдано смыслом задачи) может привести к значительным погрешностям.

2.2.9. Модель множественной регрессии

1А38 (*Уравнение множественной линейной регрессии*). *Уравнение множественной эмпирической линейной регрессии* имеет вид:

$$y_i = b_0 + b_1 x_{i1} + \ldots + b_k x_{ik} + \varepsilon_i,$$

где y_i – i-е наблюдение зависимой переменной; x_{i1}, x_{i2}, ..., x_{ik} – i-е наблюдения независимых переменных x_1, x_2, ..., x_k; n – количество наблюдений (объем выборки); k – количество независимых переменных в уравнении. Случайные отклонения ε_i, $i = \overline{1,\ n}$, удовлетворяют приведенным выше предпосылкам 30.1–30.3.

1А+Б39 (*Оценка параметров регрессии*). Оценка параметров b_0, b_1, b_2, ..., b_k обычно осуществляется по методу наименьших квадратов:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_{i1} + \ldots + b_k x_{ik}))^2 \to \min.$$

Метод наименьших квадратов предусматривает нахождение параметров b_0, b_1, b_2, ..., b_k из условия минимума суммы квадратов отклонений. Используя необходимое условие экстремума, получим систему уравнений для определения неизвестных коэффициентов b_0, b_1, b_2, ..., b_k:

$$\begin{cases} nb_0 + b_1\sum\limits_{i=1}^{n} x_{i1} + \ldots + b_k \sum\limits_{i=1}^{n} x_{ik} = \sum\limits_{i=1}^{n} y_i, \\ b_0\sum\limits_{i=1}^{n} x_{i1} + b_1\sum\limits_{i=1}^{n} x_{i1}^2 + \ldots + b_k \sum\limits_{i=1}^{n} x_{i1}x_{ik} = \sum\limits_{i=1}^{n} x_{i1}y_i, \\ \ldots \\ b_0\sum\limits_{i=1}^{n} x_{ik} + b_1\sum\limits_{i=1}^{n} x_{i1}x_{ik} + \ldots + b_k \sum\limits_{i=1}^{n} x_{ik}^2 = \sum\limits_{i=1}^{n} x_{ik}y_i. \end{cases}$$

1А+Б40 (*Оценка параметров регрессии в матричной форме*). Оценку параметров модели можно провести в матричной форме. Уравнение линейной множественной регрессии *в матричной форме* имеет вид:

$$Y = XB + \varepsilon,$$

где $Y = (y_1, y_2, ..., y_n)'$ – вектор значений зависимой переменной размерности $n \times 1$; $X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix}$ – матрица значений независимых переменных $x_1, x_2, ..., x_k$; $B = (b_0, b_1, ..., b_k)'$ – подлежащий оценке вектор неизвестных параметров; $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ – вектор случайных отклонений. Знаком «′» обозначена операция транспонирования матрицы.

Тогда формула для вычисления параметров регрессионного уравнения по методу наименьших квадратов имеет вид:

$$B = (X'X)^{-1} \cdot X'Y,$$

где X' – транспонированная матрица X; $(X'X)^{-1}$ – обратная матрица (матрица $X'X$ является невырожденной, если матрица X имеет максимальный ранг).

$$X'X = \begin{bmatrix} n & \sum x_{i1} & \dots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \dots & \sum x_{i1}x_{ik} \\ \dots & \dots & \dots & \dots \\ \sum x_{ik} & \sum x_{ik}x_{i1} & \dots & \sum x_{ik}^2 \end{bmatrix}; \quad X'Y = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \dots \\ \sum x_{ik}y_i \end{bmatrix}.$$

1A41 (*Матричное уравнение для двухфакторной модели*). В частном случае *для двухфакторной модели* получаем *матричное уравнение*

$$(X'X)B = X'Y,$$

где $X'X = \begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i2}x_{i1} & \sum x_{i2}^2 \end{bmatrix}$; $X'Y = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{bmatrix}$; $B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$.

Коэффициенты $b_1, b_2, ..., b_k$ показывают количественное воздействие каждого фактора на результативный показатель при неизменности значений других факторов.

Проверка значимости коэффициентов регрессии осуществляется по критерию Стьюдента

$$t_{\text{расч}.b_i} = \frac{b_i}{S_{b_i}},$$

где $S_{b_i}^2 = c_{ii}S_e^2$; S_e^2 – остаточная дисперсия (выборочная); c_{ii} – соответствующий $(i + 1)$-й элемент диагонали матрицы $(X'X)^{-1}$.

Коэффициент b_i при уровне значимости α признается значимым, если

$$\left|t_{\text{расч.}b_i}\right| > t_{\text{табл}} = t_{\alpha;n-k-1}.$$

1А42 (*Коэффициенты эластичности для множественной линейной регрессии*). Коэффициенты эластичности рассчитываются по формуле:

$$Э_i = b_i \frac{\overline{x}_i}{\overline{y}}.$$

Коэффициенты эластичности показывают, на сколько процентов в среднем изменяется функция с изменением аргумента на 1%.

2.2.10. Коэффициент корреляции. Коэффициент детерминации

1А+Б43 (*Коэффициент парной корреляции*). *Коэффициент парной корреляции* используется в качестве меры, характеризующей степень линейной связи двух переменных. Он представляет собой ковариацию двух наборов данных, деленную на произведение их стандартных отклонений:

$$r_{yx_j} = \frac{n\sum_{i=1}^{n} x_{ij} y_i - \sum_{i=1}^{n} x_{ij} \cdot \sum_{m=1}^{n} y_m}{\sqrt{n\sum_{i=1}^{n} x_{ij}^2 - \left(\sum_{i=1}^{n} x_{ij}\right)^2} \cdot \sqrt{n\sum_{m=1}^{n} y_m^2 - \left(\sum_{m=1}^{n} y_m\right)^2}};$$

$$r_{x_j x_l} = \frac{n\sum_{i=1}^{n} x_{ij} x_{il} - \sum_{i=1}^{n} x_{ij} \cdot \sum_{m=1}^{n} x_{ml}}{\sqrt{n\sum_{i=1}^{n} x_{ij}^2 - \left(\sum_{i=1}^{n} x_{ij}\right)^2} \cdot \sqrt{n\sum_{m=1}^{n} x_{ml}^2 - \left(\sum_{m=1}^{n} x_{ml}\right)^2}}.$$

Коэффициент корреляции принимает значения от –1 до +1. Если $r > 0$, то полагаем, что корреляционная связь между переменными является прямой, если $r < 0$ – обратной.

Если $r = \pm 1$, корреляционная связь представляется линейной функциональной зависимостью. При $r = 0$ линейная корреляционная связь отсутствует.

1А44 (*Характеристики тесноты связи*). Качественные характеристики связи приведены в табл. 2.

Таблица 2

Качественные характеристики связи

Значение r	Характер связи
От 0 до $\lvert\pm 0{,}3\rvert$	Практически отсутствует
От $\lvert\pm 0{,}3\rvert$ до $\lvert\pm 0{,}5\rvert$	Слабая
От $\lvert\pm 0{,}5\rvert$ до $\lvert\pm 0{,}7\rvert$	Умеренная
От $\lvert\pm 0{,}7\rvert$ до $\lvert\pm 1\rvert$	Сильная (тесная)

Множественная корреляция возникает от взаимодействия нескольких факторов с результативным показателем.

Значительный интерес представляют коэффициенты корреляции, характеризующие взаимосвязь факторов между собой. В корреляционную модель следует подбирать независимые между собой факторы. Если коэффициент корреляции двух факторов выше 0,8, то один из этих факторов рекомендуется исключить из модели.

В случае парной линейной регрессии коэффициент корреляции также можно вычислить по формуле:

$$r_{xy} = b_1 \frac{\sigma_x}{\sigma_y},$$

где σ_x, σ_y – средние квадратические отклонения случайных величин x и y:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}; \quad \sigma_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n}}.$$

1А45 (*Матрица коэффициентов парной корреляции*). Матрица коэффициентов парной корреляции (корреляционная матрица) имеет вид:

$$K = \begin{bmatrix} 1 & r_{yx_1} & \dots & r_{yx_k} \\ r_{x_1y} & 1 & \dots & r_{x_1x_k} \\ \dots & \dots & \dots & \dots \\ r_{x_ky} & r_{x_kx_1} & \dots & 1 \end{bmatrix}.$$

По данным этой матрицы можно примерно оценить, какие факторы существенно влияют на переменную y, а какие – несущественно, а также выявить взаимосвязь между факторами.

1А46 (*Коэффициент множественной корреляции*). Для линейной множественной регрессии *коэффициент множественной корреляции* определяется по формуле:

$$R = \sqrt{1 - \frac{\det K}{K_{11}}},$$

где $\det K$ – определитель корреляционной матрицы; K_{11} – алгебраическое дополнение элемента первой строки и первого столбца матрицы K. Коэффициент множественной корреляции принимает значения от 0 до 1. Чем ближе его значение к единице, тем в большей степени учтены факторы, влияющие на зависимую переменную, тем более точной является построенная на основе отобранных факторов модель.

1А47 (*Индекс корреляции*). *Индекс корреляции* (*коэффициент множественной корреляции*) вычисляется по формуле:

$$R = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2}}.$$

Чем выше значение R, тем вероятнее близость расчетных значений результативного признака к фактическим. Данный показатель используется при любой форме связи переменных.

1А48 (*Коэффициент детерминации*). Долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака характеризует *коэффициент детерминации* $D = R^2$, получаемый возведением в квадрат коэффициента множественной корреляции.

О полноте связи можно судить по величине коэффициентов множественной корреляции и детерминации. Например, если R = 0,92, а D = 0,85, то это значит, что вариация результативного признака на 85% зависит от изменения исследуемых факторов, а на долю других факторов приходится 15% вариации результативного показателя. Следовательно, в корреляционную модель удалось включить наиболее существенные факторы.

1А+Б49 (*Критерий Стьюдента для проверки гипотез о статистической значимости коэффициентов корреляции*). Для проверки гипотез о статистической значимости коэффициента корреляции, т. е. гипотез

$$H_0: r = 0;\ H_1: r \neq 0,$$

при заданном уровне значимости α и объеме выборки n используется t-статистика:

$$t_{\text{расч}} = \frac{|r|}{\sqrt{1-r^2}}\sqrt{n-1-1}.$$

По таблице критических точек распределения Стьюдента по заданному уровню значимости α и числу степеней свободы $\gamma = n-2$ находят $t_{\text{кр}} = t_{\text{табл}} = t_{\alpha;n-2}$. Если $t_{\text{расч}} \leq t_{\text{кр}}$, нет оснований отвергнуть гипотезу H_0. Если $t_{\text{расч}} > t_{\text{кр}}$, то гипотезу H_0 о равенстве коэффициента корреляции нулю отвергают. Другими словами, r значительно отличается от нуля, т. е. СВ X и СВ Y коррелированны.

1А+Б50 (*Критерий Фишера для проверки гипотез об адекватности модели*). Значимость построенной модели проверяется следующим образом. Выдвигаем гипотезу H_0: модель незначима. Конкурирующая гипотеза H_1: модель значима. Гипотеза проверяется по критерию Фишера. Фактическая величина

$$F_{\text{расч}} = \frac{R^2}{1-R^2}\frac{(n-k-1)}{k}$$

сопоставляется с табличной и делается заключение о надежности связи. Здесь k – количество независимых переменных в уравнении связи. Если $F_{\text{расч}} \geq F_{\text{табл}} = F_{\alpha;\gamma_1;\gamma_2}$ со степенями свободы $\gamma_1 = k$; $\gamma_2 = n-k-1$ при заданном уровне значимости α, то линейную модель можно считать адекватной, гипотеза о случайной природе зависимости между оцениваемыми характеристиками (нулевая гипотеза) отклоняется и признается статистическая значимость и надежность модели.

Также фактическое значение критерия Фишера можно рассчитать по формуле:

$$F_{\text{расч}} = \frac{\sum\limits_{i=1}^{n}(\hat{y}_i - \overline{y})^2 \Big/ k}{\sum\limits_{i=1}^{n}(y_i - \hat{y}_i)^2 \Big/ (n-k-1)},$$

где y_i – фактические индивидуальные значения результативного показателя; $\overline{y}$ – среднее значение результативного показателя; $\hat{y}_i$ – индивидуальные значения результативного показателя, рассчитанные по уравнению регрессии; n – количество наблюдений (объем выборки); k – количество независимых переменных в уравнении связи.

1А+Б51 (*Определение меры точности модели*). Определение меры точности модели производится с помощью расчета *точностных характеристик*:

51.1) *средней абсолютной ошибки*

$$\bar{\varepsilon}_{\text{абс}} = \frac{1}{n}\sum_{i=1}^{n}|e_i|,$$

где $e_i = y_i - \hat{y}_i$.

Ошибка показывает, насколько в среднем отклоняются фактические значения от модели;

51.2) *средней относительной ошибки аппроксимации*

$$E_{\text{отн}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_i}{y_i}\right| \cdot 100\%.$$

Допустимый предел значений $E_{\text{отн}}$ составляет не более 8–15%.

51.3) *дисперсии ряда остатков.*

Если какая-либо из описанных характеристик не является удовлетворительной, то есть основания сомневаться в качестве данной модели (неправильно выбрана функциональная форма уравнения; не учтена важная объясняющая переменная; имеется объясняющая переменная, не оказывающая значимого влияния на зависимую переменную).

1А+Б52 (*Исследование остаточного члена модели*). Графическое представление поведения остаточного члена e:

$$e_i = y_i - \hat{y}_i,\ i = \overline{1,\ n},$$

позволяет проанализировать наличие автокорреляции и гетероскедастичности (непостоянства дисперсий отклонений). С помощью графического представления отклонений может быть обнаружена неправильная спецификация уравнения. Для этого строится график зависимости величин отклонений e от номера наблюдения. Если зависимость, получившаяся на графике, имеет неслучайный характер, то исследуемое уравнение регрессии неверно специфицировано.

1Б+С53 (*Обнаружение и корректировка ошибок спецификации*). Ошибки спецификации допускаются из-за поверхностных знаний об исследуемых экономических процессах или из-за недостаточно глубоко проработанной теории, или из-за погрешностей сбора и обработки статистических данных при построении эмпирического уравнения регрессии. Сложность процедуры определяется типом ошибки и знаниями об исследуемом объекте.

Если в уравнении регрессии имеется несущественная переменная, то она обнаружит себя по низкой t-статистике. Следует построить другое уравнение регрессии без незначимых переменных. Затем необходимо сравнить коэффициенты детерминации для первоначального и дополнительного уравнений регрессии.

Существует ряд тестов обнаружения ошибок спецификации, среди которых можно выделить:

1. Тест Рамсея RESET (*Regression specification error test*).
2. Тест (критерий) максимального правдоподобия (*The Likelihood ratio test*).
3. Тест Валда (*The Wald test*).
4. Тест множителя Лагранжа (*The Lagrange multiplier test*).
5. Тест Хаусмана (*The Hausman test*).
6. Box-Cox преобразование (*Box-Cox transformation*).

Суть указанных тестов состоит либо в осуществлении преобразований случайных отклонений, либо в масштабировании зависимой переменной, с тем чтобы можно было сравнить начальное и преобразованное уравнения регрессии на основе известного критерия.

Лекция 3

2.2.11. Нелинейная эмпирическая регрессия

1А+Б54 (*Нелинейные модели, допускающие сведение их к линейным*). Многие экономические зависимости не являются линейными по своей сути, и поэтому их моделирование линейными уравнениями регрессии не дает положительного результата. Если между экономическими явлениями существуют нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных функций.

Нелинейность может проявляться как относительно переменных, так и относительно входящих в функцию коэффициентов.

Для оценки параметров нелинейных моделей используют два подхода. Первый подход основан на линеаризации модели и заключается в том, что с помощью подходящих преобразований исходных переменных исследуемую зависимость представляют в виде линейного соотношения между преобразованными переменными.

Второй подход обычно применяют в случаях, когда подобрать соответствующее линеаризующее преобразование не удается. Тогда используют методы нелинейной оптимизации на основе исходных переменных.

Оценка параметров регрессии, нелинейной по переменным, включенным в анализ, но линейной по оцениваемым параметрам, проводится с помощью МНК путем решения системы линейных алгебраических уравнений. Рассмотрим следующие модели:

54.1) *степенные модели* вида $y = bx^a$, где a, b – параметры модели. Эта функция может отражать зависимость спроса y от цены x (в данном случае $a < 0$) или от дохода x (в данном случае $a > 1$). Эта функция может отражать также зависимость объема выпуска y от использования ресурса x (производственная функция), в которой $0 < a < 1$ (рис. 2), а также ряд других зависимостей. Для упрощения выкладок случайное отклонение ε введем в соотношение позднее.

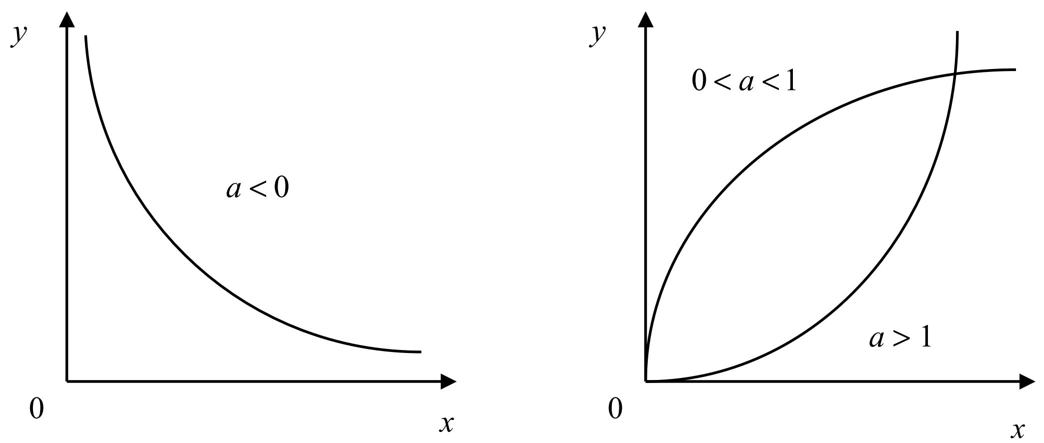


Рис. 2. Зависимость $y = bx^a$, $b > 0$

Прологарифмируем выражение $y = bx^a$: $\ln y = \ln b + a \ln x$. Выполним замену: $Y = \ln y$; $X = \ln x$; $b_0 = \ln b$; $b_1 = a$. Тогда получим: $Y = b_0 + b_1 X$.

С целью статистической оценки коэффициентов добавим в модель случайную погрешность ε, получим уравнение:

$$Y = b_0 + b_1 X + \varepsilon.$$

Полученная модель является линейной моделью, подробно рассмотренной ранее (п. 2.2.5–2.2.8). Если все необходимые предпосылки классической линейной регрессионной модели выполнены, то по МНК можно определить наилучшие линейные несмещенные оценки коэффициентов.

Коэффициент b_1 определяет эластичность переменной y по переменной x и является константой. Поэтому эта модель также называется *моделью постоянной эластичности*.

Данная модель легко обобщается на большее число переменных. Например, $\ln y = b_0 + b_1 \ln x_1 + b_2 \ln x_2 + \varepsilon$. Здесь коэффициенты b_1 и b_2 являются эластичностями переменной y по переменным x_1 и x_2 соответственно;

54.2) *показательную модель* $y = be^{ax}$, $b > 0$ (рис. 3). Наиболее важным ее приложением является ситуация, когда анализируется изменение переменной у с постоянным темпом прироста во времени.

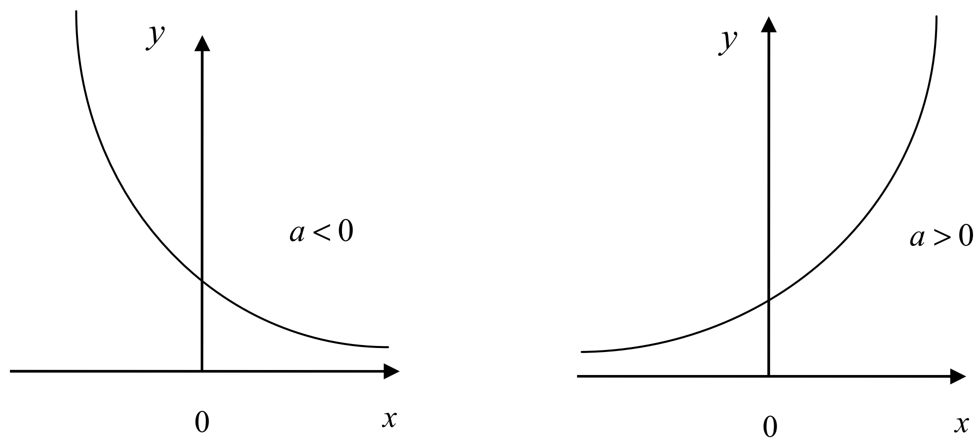


Рис. 3. Зависимость $y = be^{ax}$, $b > 0$

Данная модель путем логарифмирования сводится к логлинейной модели.

Прологарифмируем выражение $y = be^{ax}$: $\ln y = \ln b + ax$. Выполним замену: $Y = \ln y$; $X = x$; $b_0 = \ln b$; $b_1 = a$. Получим линейную модель $Y = b_0 + b_1 X$;

54.3) *логарифмические модели* вида $y = a \ln x + b$ (рис. 4). Они сводятся к линейной модели заменой $Y = y$; $X = \ln x$. В данных моделях коэффициент a определяет изменение переменной y вследствие единичного относительного прироста x (например, на 1%), т. е. характеризует отношение абсолютного изменения y к относительному изменению x.

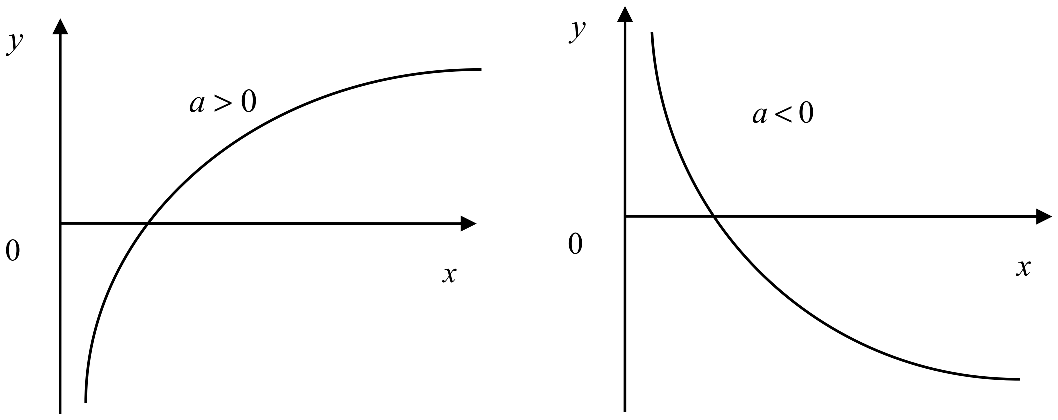


Рис. 4. Зависимость $y = a \ln x + b$

Модель используется обычно в тех случаях, когда необходимо исследовать влияние процентного изменения независимой переменной на абсолютное изменение зависимой переменной;

54.4) *обратную модель.* Модель вида $y = \dfrac{1}{ax+b}$ (рис. 5) приводится к линейной модели $Y = aX + b$ заменой $Y = \dfrac{1}{y}$; $X = x$, а модель вида $y = \dfrac{x}{ax+b}$ (рис. 6) – заменой $Y = \dfrac{x}{y}$; $X = x$.

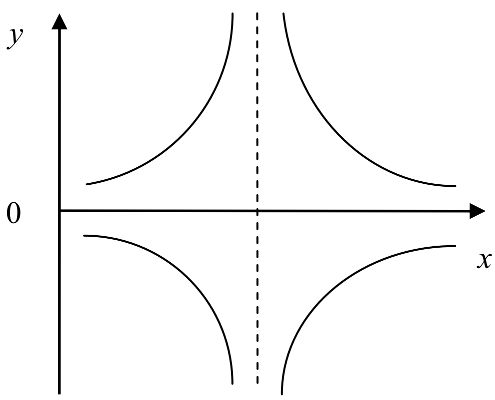


Рис. 5. Зависимость $y = \dfrac{1}{ax+b}$

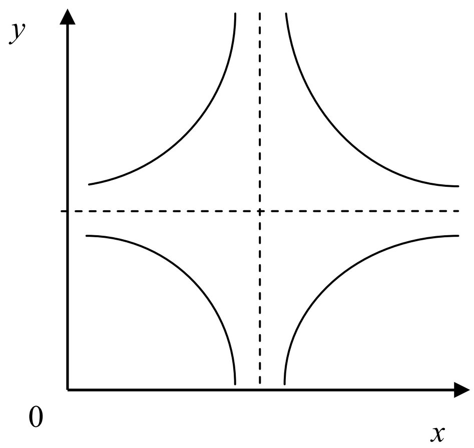


Рис. 6. Зависимость $y = \dfrac{x}{ax+b}$

1Б+С55 (*Преобразование случайного отклонения*). Для получения качественных оценок существенную роль играет выполнимость определенных предпосылок МНК для случайных отклонений. Наиболее важные из них требуют, чтобы отклонения ε являлись нормально распределенными случайными величинами с нулевым математическим ожиданием и постоянной дисперсией, а также не коррелировали друг с другом. При невыполнимости указанных предпосылок проводимые для оценок тесты окажутся ненадежными.

1А+Б56 (*Пример*). Степенная функция вида $y = b + b_1x + b_2x^2 + \ldots + b_mx^m$ часто отражает ту или иную экономическую зависимость. Например, квадратичная функция

$$y = b + b_1x + b_2x^2$$

может отражать зависимость между объемом выпуска и средними (либо предельными) издержками или между расходами на рекламу и прибылью и т. д.

Как и ранее рассмотренные модели, эта модель является линейной относительно коэффициентов регрессии. Следовательно, ее можно свести к линейной регрессионной модели. Заменяя x на X_1, x^2 на X_2, ..., x^m на X_m, получаем модель множественной линейной регрессии

$$y = b + b_1X_1 + b_2X_2 + \ldots + b_mX_m.$$

1А+Б57 (*Пример*). Ряд экономических показателей моделируется через функции, являющиеся композицией перечисленных функций, что позволяет также свести их к линейным. Например, широко известна производственная функция Кобба – Дугласа с учетом научно-технического прогресса:

$$y = AK^{\alpha}L^{\beta}e^{\gamma t}.$$

Прологарифмировав данную функцию, получим следующее соотношение:

$$\ln y = \ln A + \alpha \ln K + \beta \ln L + \gamma t \ln e,$$

которое сводится к линейному.

Модели вида $\ln y = b_0 + b_1x + \varepsilon$ (полулогарифмические) обычно используют в тех случаях, когда необходимо определить темп роста или прироста каких-либо экономических показателей. Например, при анализе банковского вклада по первоначальному вкладу и процентной ставке, при исследовании зависимости прироста объема выпуска от относительного (процентного) увеличения затрат ресурса, бюджетного дефицита от темпа роста ВНП, темпа роста инфляции от объема денежной массы и т. д.

Полулогарифмическая модель легко сводится к линейной модели заменой $Y = \ln y$.

Коэффициент b_1 в модели $\ln y = b_0 + b_1 x + \varepsilon$ имеет смысл темпа прироста переменной y по переменной x, т. е. характеризует отношение относительного изменения y к абсолютному изменению x. Действительно, продифференцировав по x, имеем:

$$\frac{1}{y}\frac{dy}{dx} = b_1,$$

следовательно,

$$b_1 = \frac{dy/y}{dx} = \frac{\text{Относительное изменение } y}{\text{Абсолютное изменение } x}.$$

Умножив b_1 на 100, получим процентное изменение переменной y (темп прироста переменной y).

1А+Б58 (*Пример*). Модель вида $y = a + b\frac{1}{x}$ также называется обратной моделью. Данная модель обычно применяется в тех случаях, когда неограниченное увеличение переменной x асимптотически приближает зависимую переменную y к некоторому пределу.

Если $a > 0$; $b > 0$, то функция может отражать зависимость между объемом выпуска (x) и средними фиксированными издержками (y).

Если $a > 0$; $b < 0$, то функция может описывать зависимость между доходом x и спросом y (например, на товары первой необходимости либо товары относительной роскоши); это так называемые функции Торнквиста (в этом случае x – минимально необходимый уровень дохода).

Графиком функции $y = a + b\frac{1}{x}$ при $a < 0$; $b > 0$ является кривая Филлипса, отражающая зависимость между уровнем безработицы (x) в процентах и процентным изменением заработной платы (y). При этом точка пересечения кривой с осью OX определяет естественный уровень безработицы.

2.3. ЭКОНОМЕТРИЧЕСКИЙ АНАЛИЗ ПРИ НАРУШЕНИИ КЛАССИЧЕСКИХ ПРЕДПОЛОЖЕНИЙ. ВРЕМЕННЫЕ РЯДЫ

2.3.1. Основные проблемы при нарушении классических предположений

1А+Б59 (*Основные предпосылки регрессионного анализа*). Условия Гаусса – Маркова (сравните с 1А+Б30):

59.1) математическое ожидание ε_i равно нулю: $M(\varepsilon_i) = 0$;

59.2) дисперсии ε_i постоянны для всех i: $D(\varepsilon_i) = \sigma^2$;

59.3) отклонения ε_i и ε_j не коррелированны при $i \neq j$: $M(\varepsilon_i \varepsilon_j) = 0$;

59.4) зависимая переменная y_i есть величина случайная, а объясняющая переменная x_i – величина неслучайная;

59.5) отклонение ε_i – это нормально распределенная СВ.

Для получения уравнения линейной регрессии достаточно предпосылок 1–4. Выполнение требования 5 необходимо для оценки точности уравнения регрессии и его параметров. В этом случае оценки параметров, полученные по МНК, будут обладать свойствами несмещенности и эффективности.

1А+Б60 (*Проблемы при нарушении классических предположений*). При моделировании реальных экономических процессов возникают ситуации, в которых условия классической модели регрессии оказываются нарушенными, а при их нарушении МНК может давать оценки с плохими статистическими свойствами:

60.1) если имеется линейная связь экзогенных переменных, например $x_2 = b_0 + b_1 x_1$, то МНК-оценки не будут существовать, т. к. матрица $X'X$ становится вырожденной. Такая ситуация в эконометрике носит название проблемы *мультиколлинеарности*;

60.2) если нарушается гипотеза о взаимной независимости случайных отклонений: $M(\varepsilon_i \varepsilon_j) \neq 0$, то возникает проблема *автокорреляции*, в рамках которой МНК-оценки не обладают несмещенностью;

60.3) одной из ключевых предпосылок МНК является *условие постоянства дисперсий случайных отклонений*. Выполнимость данной предпосылки называется *гомоскедастичностью*. Невыполнимость данной предпосылки называется *гетероскедастич-*

ностью (непостоянством дисперсий отклонений), при которой существуют $i \neq j$, такие что $\sigma_i^2 \neq \sigma_j^2$, где $D(\varepsilon_i) = \sigma_i^2$, $i = \overline{1, n}$.

Для исследования этих проблем необходимо проанализировать источники возникновения и последствия наличия проблем, провести диагностику и применить определенные методы для устранения проблем.

2.3.2. Мультиколлинеарность

1А61 (*Мультиколлинеарность*). Если в модель включаются два или более тесно взаимосвязанных фактора, то наряду с уравнением изучаемой регрессии появляется и другая зависимость. *Мультиколлинеарность* – тесная зависимость между факторными признаками, включенными в модель. Она искажает величину коэффициентов регрессии и затрудняет их экономическую интерпретацию. Мультиколлинеарность возникает лишь в случае множественной регрессии.

1Б62 (*Последствия мультиколлинеарности*). Обычно выделяются следующие *последствия мультиколлинеарности*:

62.1) большие дисперсии оценок. Это затрудняет нахождение истинных значений определяемых величин и расширяет интервальные оценки, ухудшая их точность;

62.2) уменьшаются *t*-статистики коэффициентов, что может привести к неоправданному выводу о существенности влияния соответствующей объясняющей переменной на зависимую переменную;

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

62.4) затрудняется определение вклада каждой из объясняющих переменных в объясняемую уравнением регрессии дисперсию зависимой переменной;

62.5) возможно получение неверного знака у коэффициента регрессии.

1А+Б63 (*Этапы решения проблемы мультиколлинеарности*). В решении проблемы мультиколлинеарности можно выделить несколько этапов:

63.1) установление наличия мультиколлинеарности;

63.2) определение причин возникновения мультиколлинеарности;

63.3) разработка мер по устранению мультиколлинеарности.

1Б64 (*Частный коэффициент корреляции*). *Частный коэффициент корреляции* характеризует тесноту связи между двумя переменными при фиксированных значениях остальных факторов. Если парный коэффициент корреляции между двумя наблюдаемыми величинами оказался больше по модулю частного коэффициента корреляции между этими же факторами, то фиксированные факторы усиливают взаимосвязь между изучаемыми величинами. Более низкое значение парного коэффициента корреляции в сравнении с соответствующим частным коэффициентом корреляции свидетельствует об ослаблении связи между изучаемыми величинами действиями фиксированных величин.

Частный коэффициент корреляции, например $r_{yx_1(x_2)}$, характеризует степень линейной зависимости между двумя величинами y и x_1 при исключенном влиянии третьей величины x_2, включенной в модель. Он определяется по формуле:

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}}.$$

Можно рассчитать и взаимосвязь факторных признаков при устранении влияния результативного признака.

Частный коэффициент корреляции изменяется в пределах от –1 до +1.

Частные коэффициенты корреляции как статистические величины подвергаются в анализе оценке на достоверность. С этой целью используется t-критерий Стьюдента.

Если частные коэффициенты корреляции возвести в квадрат, то получатся *частные коэффициенты детерминации.* Частный коэффициент детерминации показывает долю вариации признака под действием одного фактора при неизменном значении другого фактора.

1А+Б65 (*Причины возникновения мультиколлинеарности*). Причины возникновения мультиколлинеарности между признаками:

65.1) факторные признаки, характеризующие одну и ту же сторону явления или процесса (например, показатели объема произведенной продукции и среднегодовой стоимости основных фондов одновременно включать в модель не рекомендуется, т. к. оба признака характеризуют размер предприятия);

65.2) факторные признаки, суммарное значение которых представляет собой постоянную величину (например, коэффициент годности и коэффициент износа основных фондов);

65.3) факторные признаки, являющиеся элементами друг друга (например, затраты на производство продукции и себестоимость единицы продукции);

65.4) факторные признаки, по экономическому смыслу дублирующие друг друга (например, прибыль и рентабельность продукции).

1А+Б66 (*Способы определения мультиколлинеарности*). Способы определения наличия мультиколлинеарности:

66.1) *анализ матрицы коэффициентов парной корреляции*. Факторы x_i и x_j могут быть признаны коллинеарными, если $\left|r_{x_i x_j}\right| > 0{,}8.$ Данный признак будет надежным лишь в случае двух объясняющих переменных. При большем их количестве целесообразным является использование частных коэффициентов корреляции;

66.2) *исследование матрицы $X'X$*. Если определитель матрицы $X'X$ близок к нулю, то это свидетельствует о возможности наличия мультиколлинеарности;

66.3) *выявление статистически незначимых коэффициентов регрессии (т. е. имеющих низкие t-статистики) при достаточно высоком коэффициенте детерминации* R^2;

66.4) *выявление высоких частных коэффициентов корреляции.*

1А+Б67 (*Методы устранения мультиколлинеарности*). Единого метода устранения мультиколлинеарности не существует, т. к. причины и последствия мультиколлинеарности неоднозначны и во многом зависят от выборочных данных. Выделяют следующие методы устранения или уменьшения мультиколлинеарности:

67.1) *сравнение значений линейных коэффициентов корреляции.*

При отборе факторов предпочтение отдается тому, который более тесно, чем другие факторы, связан с результативным признаком, причем желательно, чтобы связь данного факторного признака с результатом y была выше, чем его связь с другими факторными признаками;

67.2) *метод включения факторов (метод пошаговой регрессии).*

В модель включаются факторы по одному в определенной последовательности. На первом шаге в модель вводится тот фактор, который имеет наибольший коэффициент корреляции с зависимой переменной. На последующих шагах в модель включается фактор, который имеет наибольший коэффициент корреляции с остатками модели. После включения каждого фактора в модель рассчитывают ее характеристики и проверяют модель на достоверность;

67.3) *метод исключения факторов.*

В модель включаются все факторы. Затем после построения уравнения регрессии из модели исключают фактор, коэффициент при котором незначим и имеет наименьшее значение t-критерия. Получают новое уравнение регрессии и снова проводят оценку значимости всех оставшихся коэффициентов регрессии. Процесс исключения факторов продолжается до тех пор, пока модель не станет удовлетворять определенным условиям и все коэффициенты регрессии не будут значимы;

67.4) *получение дополнительных данных или новой выборки.*

Поскольку мультиколлинеарность напрямую зависит от выборки, то, возможно, при другой выборке мультиколлинеарности не будет либо она не будет столь серьезной. Увеличение количества данных сокращает дисперсии коэффициентов регрессии и тем самым увеличивает их статистическую значимость;

67.5) *изменение спецификации модели.*

В ряде случаев проблема мультиколлинеарности может быть решена путем изменения спецификации модели: либо изменяется форма модели, либо добавляются объясняющие переменные, не учтенные в первоначальной модели, но существенно влияющие на зависимую переменную;

67.6) *использование предварительной информации о некоторых параметрах.*

Иногда при построении модели множественной регрессии можно воспользоваться предварительной информацией, в частности известными значениями некоторых коэффициентов регрессии. Вполне возможно, что значения коэффициентов, рассчитанные для каких-либо предварительных (обычно более простых) моделей или для аналогичной модели по ранее полученной выборке, могут быть использованы для разрабатываемой в данный момент модели.

Лекция 4

2.3.3. Автокорреляция

1А68 (*Суть и причины автокорреляции*). Важной предпосылкой построения качественной регрессионной модели по МНК является независимость значений случайных отклонений ε_i от значений отклонений во всех других наблюдениях. Отсутствие зависимости гарантирует отсутствие коррелированности между любыми отклонениями, и в частности между соседними отклонениями. *Автокорреляция* (*последовательная корреляция*) определяется как корреляция между наблюдаемыми показателями, упорядоченными во времени (временные ряды) или в пространстве (перекрестные данные).

Автокорреляция остатков (отклонений) обычно встречается в регрессионном анализе при использовании данных временных рядов, поэтому в дальнейшем вместо символа i номера наблюдения будем использовать символ t, отражающий момент наблюдения. Объем выборки при этом можно обозначать символом T вместо символа n. В экономических задачах так называемая *положительная автокорреляция* встречается значительно чаще, чем *отрицательная автокорреляция*. Отрицательная автокорреляция означает, что за положительным отклонением следует отрицательное и наоборот.

1А+Б69 (*Пример*). Исследуется спрос Y на прохладительные напитки в зависимости от дохода X по ежемесячным данным. Трендовая зависимость, отражающая увеличение спроса с ростом дохода, может быть представлена линейной функцией, изображенной на рис. 7. Однако фактические точки наблюдений обычно будут превышать трендовую линию в летние периоды и будут ниже ее зимой. Аналогичная картина может иметь место в макроэкономическом анализе с учетом циклов деловой активности.

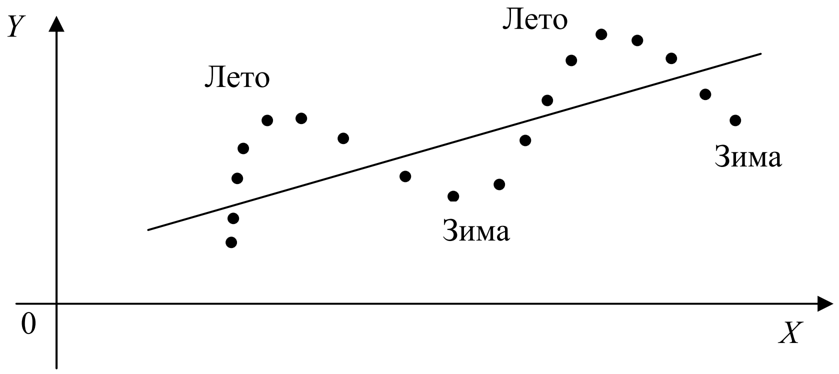


Рис. 7. Зависимость спроса
на прохладительные напитки от дохода

1А+Б70 (*Причины появления автокорреляции*). Среди основных причин, вызывающих появление автокорреляции, можно выделить:

70.1) *ошибки спецификации.* Неучет в модели какой-либо важной объясняющей переменной или неправильный выбор формы зависимости обычно приводит к системным отклонениям точек наблюдений от линии регрессии, что может обусловить автокорреляцию;

70.2) *инерция*. Многие экономические показатели (например, инфляция, безработица, ВНП и т. п.) обладают определенной цикличностью, связанной с волнообразностью деловой активности. Действительно, экономический подъем приводит к росту занятости, сокращению инфляции, увеличению ВНП и т. д. Этот рост продолжается до тех пор, пока изменение конъюнктуры рынка и ряда экономических характеристик не приведет к замедлению роста, затем остановке и движению вспять рассматриваемых показателей. В любом случае эта трансформация происходит не мгновенно, а обладает определенной инертностью;

70.3) *эффект паутины*. Во многих производственных и других сферах экономические показатели реагируют на изменение экономических условий с запаздыванием (временным лагом). Например, предложение сельскохозяйственной продукции реагирует на изменение цены с запаздыванием (равным периоду созревания урожая). Большая цена сельскохозяйственной продукции в прошедшем году, скорее всего, вызовет ее перепроизводство в текущем году, а следовательно, цена на нее снизится и т. д.;

70.4) *сглаживание данных*. Зачастую данные по некоторому продолжительному временному периоду получают усреднением данных по составляющим его подынтервалам. Это может привести к определенному сглаживанию колебаний, которые имелись внутри рассматриваемого периода, что в свою очередь может послужить причиной автокорреляции.

1А+Б71 (*Пример*). Анализируется зависимость предельных издержек *MC* от объема выпуска *Q*. Если для ее описания вместо реальной квадратичной модели выбрать линейную модель, то совершается ошибка спецификации. Ее можно рассматривать как неправильный выбор формы модели или как отбрасывание значимой переменной при линеаризации указанных моделей. Последствия данной ошибки выразятся в системном отклонении точек наблюдений от прямой регрессии и существенном преобладании последовательных отклонений одинакового знака над соседними отклонениями противоположных знаков. Налицо типичная картина, характерная для положительной автокорреляции (рис. 8).

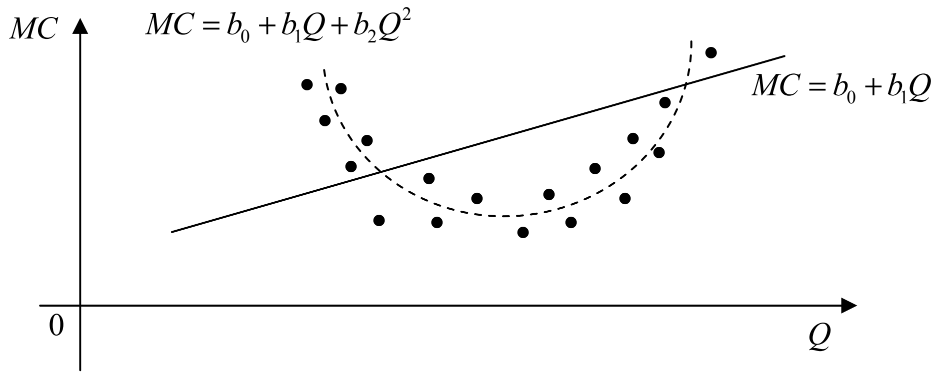


Рис. 8. Зависимость предельных издержек от объема выпуска

1Б72 (*Последствия автокорреляции*). Среди последствий автокорреляции при применении МНК обычно выделяются следующие:

72.1) оценки параметров, оставаясь линейными и несмещенными, перестают быть эффективными. Следовательно, они перестают обладать свойствами наилучших линейных несмещенных оценок;

72.2) дисперсии оценок являются смещенными. Часто дисперсии, вычисляемые по стандартным формулам, являются заниженными, что влечет за собой увеличение *t*-статистик. Это может привести к признанию статистически значимыми объясняющих переменных, которые в действительности таковыми могут и не являться;

72.3) оценка дисперсии регрессии является смещенной оценкой истинного значения и во многих случаях занижает его;

72.4) в силу вышесказанного выводы по *t*- и *F*-статистикам, определяющим значимость коэффициентов регрессии и коэффициента детерминации, возможно, будут неверными. Вследствие этого ухудшаются прогнозные качества модели.

1А+С73 (*Обнаружение автокорреляции*). В силу неизвестности значений параметров уравнения регрессии неизвестны также и истинные значения отклонений ε_t, $t = \overline{1, n}$. Поэтому выводы об их независимости делаются на основе оценок e_t, $t = \overline{1, n}$, полученных из эмпирического уравнения регрессии. Рассмотрим возможные методы определения автокорреляции:

73.1) *графический метод.* По оси абсцисс откладываются либо время (момент) получения статистических данных, либо порядковый номер наблюдения, а по оси ординат – отклонения (или оценки отклонений). По графику предполагают, имеются ли определенные связи между отклонениями, т. е. автокорреляция. Отсутствие зависимости, скорее всего, свидетельствует об отсутствии автокорреляции. Данный график можно также дополнить графиком зависимости e_t от e_{t-1}.

В современных компьютерных прикладных программах для решения задач по эконометрике аналитическое выражение регрессии дополняется графическим представлением результатов. На график реальных колебаний зависимой переменной накладывается график колебаний переменной по уравнению регрессии. Сопоставив эти два графика, можно выдвинуть гипотезу о наличии автокорреляции остатков. Если эти графики пересекаются редко, то можно предположить наличие положительной автокорреляции остатков;

73.2) *тест Дарбина – Уотсона.* Критерий оценки основан на решающей функции

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

где $e_i = y_i - \hat{y}_i$. Тест позволяет делать выводы о наличии либо отсутствии автокорреляции (подробнее о применении статистики Дарбина – Уотсона см. в учебнике [4]).

1Б+С74 (*Устранение автокорреляции*). Основной причиной наличия случайного члена в модели являются недостаточные знания о причинах и взаимосвязях, определяющих значение зависимой переменной. Автокорреляция чаще всего вызывается неправильной спецификацией модели, поэтому необходимо, прежде всего, скорректировать саму модель. Возможно, автокорреляция вызвана отсутствием в модели некоторой важной объясняющей переменной. Следует попытаться определить данный фактор и учесть его в уравнении регрессии. Также можно попробовать изменить формулу зависимости.

В качестве способов уменьшения (устранения) автокорреляции во временных рядах применяют следующие методы:

74.1) метод включения дополнительного фактора, например времени;

74.2) метод последовательных разностей;

74.3) метод авторегрессионных преобразований.

2.3.4. Гетероскедастичность

1А+Б75 (*Причины возникновения проблемы гетероскедастичности*). При рассмотрении выборочных данных мы имеем дело с конкретными реализациями зависимой переменной y_i и соответственно с определенными случайными отклонениями ε_i, $i = \overline{1,\ n}$. До осуществления выборки эти показатели могли принимать произвольные значения на основе некоторых вероятностных распределений. Одним из требований к этим распределениям является равенство дисперсий. Несмотря на то что при каждом конкретном наблюдении случайное отклонение может быть любым, не должно быть априорной причины, вызывающей большую ошибку (отклонение) при одних наблюдениях и меньшую – при других. На практике зачастую есть основания считать, что вероятностные распределения случайных отклонений ε_i при различных наблюдениях будут различными.

Проблема гетероскедастичности характерна для перекрестных данных и довольно редко встречается при рассмотрении временных рядов. В перекрестных данных учитываются экономические субъекты (потребители, домохозяйства, фирмы, отрасли, страны), имеющие различные доходы, размеры, потребности и т. д. В этом случае возможны проблемы, связанные с эффектом масштаба. Во временных рядах обычно рассматриваются одни и те же показатели в различные моменты времени (ВНП, чистый экспорт, темпы инфляции в определенном регионе за определенный период времени). Однако при увеличении (уменьшении) рассматриваемых показателей с течением времени также может возникнуть проблема гетероскедастичности.

1А+Б76 (*Пример*). Рассмотрим зависимость потребления от дохода. С ростом дохода растет среднее значение потребления. Субъекты с большим доходом в среднем потребляют больше, чем субъекты с меньшим доходом, причем люди с большим доходом имеют больший простор для его распределения, т. е. разброс в их потреблении более существен. Разброс значений потребления вызывает разброс точек наблюдения относительно линии регрессии, дисперсия потребления не остается постоянной, а увеличивается с ростом дохода, имеет место гетероскедастичность остатков.

1Б77 (*Последствия гетероскедастичности*). Последствия гетероскедастичности:

77.1) оценки коэффициентов по-прежнему останутся несмещенными и линейными;

77.2) оценки не будут эффективными (т. е. они не будут иметь наименьшую дисперсию по сравнению с другими оценками данного параметра). Они не будут даже асимптотически эффективными. Увеличение дисперсии оценок снижает вероятность получения максимально точных оценок;

77.3) дисперсии оценок будут рассчитываться со смещением;

77.4) вследствие вышесказанного все выводы, получаемые на основе соответствующих t- и F-статистик, а также интервальные оценки будут ненадежными. Следовательно, статистические выводы, получаемые при стандартных проверках качества оценок, могут быть ошибочными и приводить к неверным заключениям по построенной модели. Вполне вероятно, что стандартные ошибки коэффициентов будут занижены, а значит, t-статистики будут завышены. Это может привести к признанию статистически значимыми коэффициентов, которые на самом деле таковыми не являются.

1А+Б78 (*Обнаружение гетероскедастичности*). В ряде случаев, зная характер данных, можно предвидеть появление проблемы гетероскедастичности и попытаться устранить этот недостаток еще на этапе спецификации. Однако значительно чаще эту проблему приходится решать после построения уравнения регрессии.

Обнаружение гетероскедастичности в каждом конкретном случае является довольно сложной задачей, т. к. для знания дисперсий отклонений необходимо знать распределение СВ Y, соответствующее выбранному значению x_i СВ X. На практике часто для каждого конкретного значения x_i определяется единственное значение y_i, что не позволяет оценить дисперсию СВ Y для данного значения x_i.

Не существует однозначного метода определения гетероскедастичности. Однако для проверки разработано много тестов и критериев. Наиболее популярные и наглядные: графический анализ отклонений, тест ранговой корреляции Спирмена, тест Парка, тест Глейзера, тест Голдфелда – Квандта.

1А+Б79 (*Графический анализ остатков*). Использование графического представления отклонений позволяет определиться с наличием гетероскедастичности. В этом случае по оси абсцисс откладываются значения объясняющей переменной X (либо линейной комбинации объясняющих переменных), а по оси ординат – либо отклонения, либо их квадраты.

Если все отклонения находятся внутри полосы постоянной ширины, параллельной оси абсцисс, то это говорит о независимости дисперсий от значений переменной X и их постоянстве, т. е. в этом случае выполняются условия *гомоскедастичности*.

Если наблюдаются некоторые систематические изменения в соотношениях между значениями переменной X и квадратами отклонений (линейная, квадратичная, гиперболическая и другие зависимости), то такие ситуации отражают большую вероятность наличия *гетероскедастичности* для рассматриваемых статистических данных.

Графический анализ отклонений является удобным и достаточно надежным в случае парной регрессии. При множественной регрессии графический анализ возможен для каждой из объясняющих переменных X_j, $j = \overline{1, m}$, отдельно. Чаще же вместо объясняющих переменных по оси абсцисс откладывают значения $\hat{y}_i$, получаемые из эмпирического уравнения регрессии. Поскольку по уравнению множественной линейной регрессии $\hat{y}_i$ является линейной комбинацией x_{ij}, $j = \overline{1, m}$, то график может указать на наличие гетероскедастичности. Такой анализ наиболее целесообразен при большом количестве объясняющих переменных.

1А+Б80 (*Методы смягчения проблемы гетероскедастичности*). Гетероскедастичность приводит к неэффективности оценок, несмотря на их несмещенность. Это может обусловить необоснованные выводы по качеству модели. Поэтому при установлении гетероскедастичности возникает необходимость преобразования модели. Вид преобразования зависит от того, известны или нет дисперсии отклонений.

Метод взвешенных наименьших квадратов (МВНК) применяется при известных для каждого наблюдения значениях σ_i^2. В этом случае можно устранить гетероскедастичность, разделив каждое наблюдаемое значение на соответствующее ему значение среднего квадратического отклонения σ_i; тогда при оценке коэффициентов регрессии наблюдения с меньшими дисперсиями отклонений будут более значимыми, чем наблюдения с большими дисперсиями. Далее по МНК для преобразованных данных строится уравнение регрессии (без свободного члена) с гарантированным качеством оценок.

Во многих случаях дисперсии отклонений зависят от не включенных в уравнение регрессии объясняющих переменных, играющих существенную роль в исследуемой зависимости.

В ряде случаев для устранения гетероскедастичности необходимо изменить спецификацию модели.

На практике имеет смысл применить несколько методов определения гетероскедастичности и способов ее корректировки.

Лекция 5

2.3.5. Временные ряды

1А81 (*Временной ряд*). Для характеристики и анализа различных социально-экономических явлений за определенный период применяют показатели и методы, характеризующие эти процессы во времени (динамике). Под *временным рядом* в экономике понимается последовательность наблюдений некоторого признака (случайной величины) Y в последовательные моменты времени. Отдельные наблюдения, которые называются *уровнями* ряда, будем обозначать y_t, $t = \overline{1, n}$ (где n – число уровней). Последовательно расположенные во времени числовые показатели характеризуют уровень состояния и изменения явления или процесса.

1А+Б82 (*Классификация временных рядов*). Классификация временных рядов:

82.1) *в зависимости от показателя времени* временные ряды бывают *моментные* (на определенную дату) и *интервальные* (за определенный период);

82.2) *по форме представления* уровни во временном ряду могут быть представлены *абсолютными*, *средними* и *относительными* величинами;

82.3) *по расстоянию между уровнями* временные ряды подразделяются на ряды с *равноотстоящими* и *неравноотстоящими* уровнями по времени. В равноотстоящих рядах даты регистрации периода следуют друг за другом с равными интервалами, в неравноотстоящих рядах равные интервалы не соблюдаются;

82.4) *по содержанию* временные ряды подразделяют на состоящие из *частных* и *агрегированных* показателей. Частные показатели характеризуют явления изолированно, односторонне (например, динамика показателей среднесуточного объема потребленной воды). Агрегированные показатели являются производными от частных показателей и характеризуют изучаемое явление комплексно (например, динамика показателей экономической конъюнктуры).

1Б83 (*Несопоставимость уровней временного ряда*). При построении временных рядов необходимо соблюдать определенные правила. Показатели уровня временного ряда должны подчиняться единому закону развития. При нарушении этих правил уровни ряда становятся несопоставимыми.

Несопоставимость уровней временного ряда:

83.1) *по кругу охватываемых объектов* (возникает в результате неодинаковой полноты охвата объектов);

83.2) *по методологии расчета* (расчет показателей должен проводиться по единой методологии расчета);

83.3) *по единицам измерения* (возникает, если показатель может быть представлен в различных единицах измерения. Например, производительность труда измеряется в трудовых и стоимостных единицах);

83.4) *по достоверности* (возникает вследствие неодинаковой репрезентативности выборки по различным периодам);

83.5) *по территории* (возникает в результате изменения границ регионов и т. д.);

83.6) *по стоимостным показателям* (возникает вследствие изменения цен);

83.7) *по времени регистрации* (возникает из-за сезонных явлений. Например, в разные времена года потребление электроэнергии различно, и, соответственно, сравнение возможно только с учетом определенной даты).

1А84 (*Составляющие временного ряда*). Каждый уровень временного ряда формируется из трендовой, циклической, сезонной и случайной компонент. В случае относительно коротких временных рядов (например, 3–5 лет) циклическая компонента отдельно не выделяется, а происходит объединение циклической компоненты с трендом.

В общем виде модель экономического временного ряда может быть представлена следующим образом:

$$y_t = u_t + v_t + \varepsilon_t,$$

где u_t – *тренд*, плавно меняющаяся компонента, описывающая чистое влияние долговременных факторов, т. е. длительную тенденцию изменения признака (например, рост населения, экономическое развитие, изменение структуры потребления и т. п.); v_t – *сезонная компонента*, отражающая повторяемость экономических процессов в течение не очень длительного периода: года, иногда месяца, недели и т. д.

(например, объем продаж товаров или перевозок пассажиров в различные времена года); ε_t – *случайная компонента*, отражающая влияние не поддающихся учету и регистрации случайных факторов.

Следует обратить внимание на то, что в отличие от компоненты ε_t составляющие u_t, v_t являются закономерными, неслучайными.

1А+Б85 (*Модели временного ряда*). Модели, в которых временной ряд представлен как сумма перечисленных компонент, называются аддитивными; как произведение – мультипликативными моделями временного ряда.

Аддитивная модель имеет вид: $y_t = u_t + v_t + \varepsilon_t$.

Мультипликативная модель имеет вид: $y_t = u_t \cdot v_t \cdot \varepsilon_t$. Такую модель применяют в случае, если происходят существенные сезонные изменения.

1Б86 (*Этапы анализа временных рядов*). При исследовании экономических временных рядов основной задачей является выявление и статистическая оценка *основной тенденции* развития изучаемого процесса и отклонений от нее. Основные этапы анализа временных рядов:

86.1) графическое представление и описание поведения временного ряда;

86.2) выделение и удаление закономерных (неслучайных) составляющих временного ряда (тренда, сезонных и циклических составляющих);

86.3) сглаживание и фильтрация (удаление низко- или высокочастотных составляющих временного ряда);

86.4) исследование случайной составляющей временного ряда, построение и проверка адекватности математической модели для ее описания;

86.5) прогнозирование развития изучаемого процесса на основе имеющегося временного ряда;

86.6) исследование взаимосвязи между различными временными рядами.

Среди наиболее распространенных методов анализа временных рядов выделим корреляционный анализ, модели авторегрессии и скользящей средней.

1А+Б87 (*Стационарные временные ряды и их характеристики*). Важную роль в анализе временных рядов играют стационарные временные ряды, вероятностные свойства которых не изменяются во времени. Стационарные временные ряды применяются, в частности, при описании случайных составляющих анализируемых рядов. Временной ряд y_t называется *стационарным*, если совместное распределение вероятностей p наблюдений $y_1, y_2, \ldots, y_p$ такое же, как и распределение вероятностей p наблюдений $y_{1+\tau}, y_{2+\tau}, \ldots, y_{p+\tau}$ при любых p и τ. Иначе говоря, свойства стационарных рядов y_t не зависят от момента t, т. е. закон распределения и его числовые характеристики не зависят от t. Поэтому математическое ожидание и среднее квадратическое отклонение могут быть оценены по наблюдениям y_t, $t = \overline{1,\ n}$, по формулам:

$$\overline{y} = \frac{\sum_{t=1}^{n} y_t}{n}; \quad s^2 = \frac{\sum_{t=1}^{n} (y_t - \overline{y})^2}{n}.$$

1А+Б88 (*Коэффициент автокорреляции*). Степень тесноты связи между компонентами временного ряда y_t (сдвинутых относительно друг друга на τ единиц, или, как говорят, с *лагом* τ) может быть определена с помощью коэффициента корреляции

$$r_{y_t y_{t+\tau}} = r(\tau,\ t) = \frac{M((y_t - M(y_t))(y_{t+\tau} - M(y_{t+\tau})))}{\sigma_y(t)\sigma_y(t+\tau)}.$$

Так как коэффициент $r(\tau,\ t)$ измеряет корреляцию между членами одного и того же ряда, его называют *коэффициентом автокорреляции.* Для стационарного временного ряда коэффициент автокорреляции не зависит от t.

1А+Б89 (*Автокорреляционная функция*). Зависимость $r(\tau,\ t)$ – *автокорреляционная функция.* Для стационарного временного ряда y_t, $t = \overline{1,\ n}$, автокорреляционная функция $r_{y_t y_{t+\tau}}$ зависит только от лага τ, причем $r(-\tau) = r(+\tau)$, т. е. при изучении $r(\tau)$ можно ограничиться рассмотрением только положительных значений τ.

1А+Б90 (*Выборочная автокорреляционная функция*). Статистической оценкой $r(\tau)$ является выборочный коэффициент автокорреляции $r_B(\tau)$, определяемый по формуле коэффициента корреляции:

$$r_B(\tau) = \frac{(n-\tau)\sum_{t=1}^{n-\tau} y_t y_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{(n-\tau)\sum_{t=1}^{n-\tau} y_t^2 - \left(\sum_{t=1}^{n-\tau} y_t\right)^2}\sqrt{(n-\tau)\sum_{t=1}^{n-\tau} y_{t+\tau}^2 - \left(\sum_{t=1}^{n-\tau} y_{t+\tau}\right)^2}}.$$

Функцию $r_B(\tau)$ называют *выборочной автокорреляционной функцией*, а ее график – *коррелограммой*, иными словами, коррелограмма – это ломаная линия, соединяющая точки с координатами (τ; $r_B(\tau)$).

При расчете $r_B(\tau)$ следует помнить, что с увеличением τ число $n - \tau$ пар наблюдений y_t, $y_{t+\tau}$ уменьшается, поэтому лаг τ должен быть таким, чтобы число $n - \tau$ было достаточным для определения выборочного коэффициента автокорреляции $r_B(\tau)$. Обычно ориентируются на соотношение $\tau \leq n / 4$.

Для стационарного временного ряда с увеличением лага τ взаимосвязь членов временного ряда ослабевает, и автокорреляционная функция $r(\tau)$ должна убывать (по абсолютной величине). В то же время для ее выборочного (эмпирического) аналога $r_B(\tau)$, особенно при небольшом числе $n - \tau$ пар наблюдений, свойство монотонного убывания (по абсолютной величине) при возрастании τ может нарушаться.

1А+Б91 (*Основная компонента временного ряда*). Одной из важнейших задач исследования экономического временного ряда является выявление основной тенденции изучаемого процесса, выраженной неслучайной составляющей $f(t)$ (тренда либо тренда с циклической и/или сезонной компонентой). Основной компонентой временного ряда, как уже говорилось, является тренд. *Тренд* – это устойчивая тенденция во временном ряду, более или менее свободная от случайных колебаний.

1А+Б+С92 (*Основные типы трендов*). Тенденции изменения показателей сложных общественных явлений только приближенно можно выразить тем или иным уравнением, линией тренда. На практике чаще всего используют следующие основные типы трендов временных рядов: прямолинейный, параболический, экспоненциальный, гиперболический и др.

Для определения параметров функции тренда $f(t)$ чаще всего используется метод наименьших квадратов. Значения временного ряда y_t рассматриваются как зависимая переменная, а время t – как объясняющая:

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

где ε_t – возмущения, удовлетворяющие основным предпосылкам регрессионного анализа, т. е. представляющие собой независимые и одинаково распределенные случайные величины, распределение которых предполагается нормальным.

Напомним, что, согласно методу наименьших квадратов, параметры зависимости $f(t) = b_0 + b_1 t$ находятся из системы нормальных уравнений

$$\begin{cases} nb_0 + b_1 \sum_{t=1}^{n} t = \sum_{t=1}^{n} y_t, \\ b_0 \sum_{t=1}^{n} t + b_1 \sum_{t=1}^{n} t^2 = \sum_{t=1}^{n} y_t t. \end{cases}$$

Линейный тип тренда подходит для отображения тенденции примерно равномерного изменения уровней: равных в среднем величин абсолютного прироста или абсолютного сокращения уровней за равные промежутки времени.

Экспоненциальный и степенной тренды характерны для процессов, развивающихся в среде, не создающей никаких ограничений для роста уровней. На практике такие явления встречаются только в ограниченном промежутке времени, поскольку любая среда рано или поздно создает ограничения.

Методом выравнивания (сглаживания) временного ряда, т. е. выделения неслучайной составляющей, является *метод скользящих средних*. Он основан на переходе от начальных значений членов ряда к их средним значениям на интервале времени, длина которого определена заранее.

1Б93 (*Прогнозирование на основе моделей временных рядов*). Одна из важнейших задач анализа временного ряда состоит в прогнозировании на его основе развития изучаемого процесса. При этом исходят из того, что тенденция развития, установленная в прошлом, может быть распространена (экстраполирована) на будущий период.

Задача ставится следующим образом: имеется временной ряд y_t, $t = \overline{1, n}$, требуется дать прогноз уровня этого ряда на момент $n + \tau$.

Ранее были рассмотрены точечный и интервальный прогнозы значений зависимой переменной Y, т. е. определение точечных и интервальных оценок Y, полученных для парной и множественной регрессий для значений объясняющих переменных X, расположенных вне пределов обследованного диапазона значений X.

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

Далее будем полагать, что возмущения ε_t, $t=\overline{1,\ n}$, удовлетворяют предпосылкам регрессионного анализа, т. е. условиям нормальной классической регрессионной модели.

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

1А+Б94 (*Понятие об авторегрессионных моделях*). Для данного временного ряда далеко не всегда удается подобрать адекватную модель, для которой ряд возмущений ε_t будет удовлетворять основным предпосылкам регрессионного анализа. До сих пор мы рассматривали модели, в которых в качестве регрессора выступала переменная t – «время». В эконометрике достаточно широкое распространение получили и другие регрессионные модели, в которых регрессорами выступают лаговые переменные. *Лаговая переменная* – это переменная, влияние которой в эконометрической модели характеризуется некоторым запаздыванием.

Авторегрессионная модель p-го порядка (или модель AR (p)) имеет вид:

$$y_t=\beta_0+\beta_1 y_{t-1}+\beta_2 y_{t-2}+\ldots+\beta_p y_{t-p}+\varepsilon_t,\ t=\overline{1,\ n},$$

где β_0, β_1, …, β_p – некоторые константы.

Данная модель описывает изучаемый процесс в момент t в зависимости от его значений в предыдущие моменты $t-1$, $t-2$, ..., $t-p$.

Если исследуемый процесс y_t в момент t определяется его значениями только в предшествующий период $t-1$, то рассматривают *авторегрессионную модель 1-го порядка* (или модель AR (1) – марковский случайный процесс):

$$y_t=\beta_0+\beta_1 y_{t-1}+\varepsilon_t,\ t=\overline{1,\ n}.$$

Лекция 6

2.4. СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ

1А95 (*Системы уравнений в эконометрике*). Для изучения комплексных экономических явлений средствами эконометрики, как правило, применяют не отдельные уравнения регрессии, а системы уравнений.

1А+Б96 (*Виды систем эконометрических уравнений*). Различают следующие виды систем эконометрических уравнений:

96.1) *система независимых уравнений.* Каждый результативный признак (объясняемая переменная) y_j, $j = \overline{1, n}$, является функцией одной и той же совокупности факторов (объясняющих переменных) x_i, $i = \overline{1, m}$. Набор факторов в каждом уравнении системы может варьироваться в зависимости от изучаемого явления;

96.2) *система рекурсивных уравнений.* Результативный признак y_j, $j = \overline{1, n}$, одного уравнения системы в каждом последующем уравнении является фактором наряду с одной и той же совокупностью факторов x_i, $i = \overline{1, m}$;

96.3) *система одновременных уравнений.* Результативный признак y_j, $j = \overline{1, n}$, одного уравнения системы входит во все другие уравнения системы в качестве фактора наряду с одной и той же совокупностью факторов x_i, $i = \overline{1, m}$. Такие системы эффективны в эконометрических исследованиях и наиболее широко применяются в макроэкономике.

Параметры системы независимых или рекурсивных уравнений определяют с помощью МНК. Для исследования системы одновременных уравнений требуются другие, отличные от МНК методы. Их применение обуславливается тем, что результативный признак одного уравнения системы в другом уравнении этой системы используется в качестве фактора и коррелирует с соответствующей ошибкой.

1А97 (*Формы моделей систем эконометрических уравнений*). Система одновременных уравнений может быть представлена следующим образом:

97.1) в виде структурной формы модели;

97.2) в виде приведенной формы модели.

1А98 (*Виды переменных*). Основными составляющими обеих форм записи систем одновременных уравнений являются эндогенные и экзогенные переменные.

Эндогенные переменные (y) определяются внутри модели и являются зависимыми переменными. *Экзогенные переменные* (x) определяются вне системы и являются независимыми переменными. Предполагается, что экзогенные переменные не коррелируют с ошибкой в соответствующем уравнении. Под *предопределенными переменными* системы одновременных уравнений понимают экзогенные и лаговые (за предыдущие моменты времени) эндогенные переменные этой системы.

1А99 (*Структурная форма модели*). Структурная форма модели имеет вид:

$$\begin{cases} y_1 = c_{10} + b_{12}y_2 + b_{13}y_3 + \ldots + b_{1n}y_n + a_{11}x_1 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = c_{20} + b_{21}y_1 + b_{23}y_3 + \ldots + b_{2n}y_n + a_{21}x_1 + \ldots + a_{2m}x_m + \varepsilon_2, \\ \ldots \\ y_n = c_{n0} + b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{n,n-1}y_{n-1} + a_{n1}x_1 + \ldots + a_{nm}x_m + \varepsilon_n, \end{cases}$$

где c_{i0}, $i = \overline{1,\ n}$, – свободный член уравнения модели; b_{ij}, $i = \overline{1,\ n}$; $j = \overline{1,\ m}$, – коэффициент при эндогенной переменной модели; a_{ij}, $i = \overline{1,\ n}$; $j = \overline{1,\ m}$, – коэффициент при экзогенной переменной; ε_i, $i = \overline{1,\ n}$, – случайная составляющая (ошибка) i-го уравнения структурной формы модели.

1А100 (*Классы структурных уравнений модели*). Наряду с регрессионными уравнениями в модели могут быть записаны и тождества. Таким образом, структурные уравнения модели подразделяются на два класса:

100.1) *поведенческие уравнения.* Описывают взаимодействие между экзогенными и эндогенными переменными;

100.2) *тождества.* Устанавливают соотношения между эндогенными переменными, не содержат случайных составляющих и структурных коэффициентов модели.

1А+Б101 (*Пример*). Модели спроса и предложения – классические примеры систем одновременных уравнений. Выбор переменных, оценка параметров уравнений системы отображают степень взаимного влияния признаков модели. Рассмотрим модель 1:

$$\begin{cases} y_{1t} = c_{10} + b_{13}y_{3t} + \varepsilon_1, & (1) \\ y_{2t} = c_{20} + b_{23}y_{3t} + \varepsilon_2, & (2) \\ y_{1t} = y_{2t}, \end{cases}$$

где y_{1t} – спрос на товар в момент времени t; y_{2t} – предложение количества товара в момент t; y_{3t} – цена, по которой заключаются сделки в момент t.

Все переменные системы (y_{1t}, y_{2t}, y_{3t}) эндогенные в силу экономического содержания модели. Величины спрос-предложение и цена определяются одновременно. Так как $y_{1t} = y_{2t}$, то $c_{10} + b_{13}y_{3t} + \varepsilon_1 = c_{20} + b_{23}y_{3t} + \varepsilon_2$; $(b_{13} - b_{23})y_{3t} = c_{20} - c_{10} + \varepsilon_2 - \varepsilon_1$; $y_{3t} = \dfrac{c_{20} - c_{10} + \varepsilon_2 - \varepsilon_1}{b_{13} - b_{23}}$. Последнее равенство подтверждает зависимый характер цены.

Для того чтобы модель 1 была эконометрически значимой, необходимо ввести предопределенные (экзогенные и лаговые эндогенные) переменные. Например, в уравнение спроса (1) ввести экзогенную переменную x_{1t} – доход на душу населения в момент времени t, а в уравнение предложения (2) – лаговую эндогенную переменную $y_{3,t-1}$ – цену товара в момент $(t - 1)$. В результате получим модель 2 спроса и предложения кейнсианского типа.

1А+Б102 (*Пример*). Рассмотрим модель 2 (предложение и спрос кейнсианского типа):

$$\begin{cases} y_{1t} = c_{10} + b_{13}y_{3t} + a_{11}x_{1t} + \varepsilon_1, \\ y_{2t} = c_{20} + b_{23}y_{3t} + a_{23}y_{3,t-1} + \varepsilon_2, \\ y_{1t} = y_{2t}, \end{cases}$$

где y_{1t} – спрос на товар в момент времени t; y_{2t} – предложение товара в момент t; y_{3t} – цена товара в момент t; $y_{3,t-1}$ – цена товара в момент $(t - 1)$; x_{1t} – доход на душу населения в момент t; t – текущий период; $(t - 1)$ – предыдущий период.

В модели три эндогенные переменные (y_{1t}, y_{2t}, y_{3t}) и две предопределенные переменные (x_{1t} и $y_{3,t-1}$).

1А103 (*Приведенная форма модели*). Структурная форма модели может быть преобразована в приведенную форму:

$$\begin{cases} y_1 = \alpha_{10} + \alpha_{11}x_1 + \ldots + \alpha_{1m}x_m + \eta_1, \\ y_2 = \alpha_{20} + \alpha_{21}x_1 + \ldots + \alpha_{2m}x_m + \eta_2, \\ \ldots \\ y_n = \alpha_{n0} + \alpha_{n1}x_1 + \ldots + \alpha_{nm}x_m + \eta_n, \end{cases}$$

где α_{i0}, $i = \overline{1,\ n}$, – свободный член уравнения модели; α_{ij}, $i = \overline{1,\ n}$; $j = \overline{1,\ m}$, – коэффициент при предопределенной переменной, являющийся функцией коэффициентов структурной формы модели; η_i, $i = \overline{1,\ n}$, – случайная составляющая (ошибка) i-го уравнения приведенной формы модели.

1А+Б104 (*Причины построения приведенной формы модели*). Причины, по которым наряду со структурной формой модели строят ее приведенную форму:

104.1) оценки параметров структурной формы модели, найденные с помощью МНК, являются смещенными и несостоятельными (нарушаются предпосылки МНК) в силу того, что эндогенные переменные, как правило, коррелируют со случайным отклонением;

104.2) независимость уравнений в приведенной форме модели позволяет определять состоятельные оценки ее параметров с помощью МНК;

104.3) параметры (коэффициенты) приведенной формы модели связаны с параметрами ее структурной формы.

1А+Б105 (*Идентификация модели*). *Идентификация* – это установление соответствия между приведенной и структурной формами модели. Единственность соответствия между приведенной и структурной формами модели составляет задачу идентификации.

1А+Б106 (*Классы структурных моделей*). С точки зрения задачи идентификации выделяют следующие классы структурных моделей:

106.1) *идентифицируемая.* Все структурные коэффициенты однозначно определяются через приведенные коэффициенты;

106.2) *неидентифицируемая.* Структурные коэффициенты невозможно найти по приведенным коэффициентам;

106.3) *сверхидентифицируемая.* Структурные коэффициенты, выраженные через приведенные коэффициенты, имеют два и более числовых значения.

1А+Б107 (*Установление идентифицируемости модели*). Модель идентифицируема тогда и только тогда, когда идентифицируемо каждое ее уравнение. Идентификация не применяется для тождеств модели.

1А+Б108 (*Необходимое условие идентифицируемости уравнений системы*). Если уравнение модели идентифицируемо, то количество эндогенных переменных (n) этого уравнения на единицу больше количества предопределенных переменных (p) системы, не входящих в данное уравнение: $n = p + 1$. Если $n < p + 1$, то уравнение сверхидентифицируемо; если $n > p + 1$, то уравнение неидентифицируемо.

1А+Б109 (*Достаточное условие идентифицируемости уравнений системы*). Если определитель (Δ^*) матрицы коэффициентов (A) при переменных системы, не входящих в данное уравнение, не равен нулю $(\Delta^* \neq 0)$ и количество эндогенных переменных системы без единицы равно рангу этой матрицы $(\text{rank}\, A = n - 1)$, то уравнение модели идентифицируемо.

Проверка структурной модели на идентифицируемость позволяет установить степень возможности оценки коэффициентов структурных уравнений по коэффициентам приведенных уравнений.

1А+Б110 (*Методы решения систем одновременных уравнений*). Для получения качественных оценок параметров системы одновременных уравнений пользуются специальными методами. В настоящее время классическими для решения систем одновременных уравнений являются:

110.1) *косвенный метод наименьших квадратов*. Основан на получении оценок параметров структурной формы модели по оценкам параметров приведенной формы. Оценки являются состоятельными и несмещенными в силу применения к каждому уравнению приведенной формы МНК.

Алгоритм косвенного МНК:

1. Структурная форма модели преобразуется в приведенную форму.
2. С помощью МНК оцениваются параметры приведенной формы.
3. Приведенная форма преобразуется в структурную форму.

Область применения косвенного МНК ограничивается идентифицируемыми системами одновременных уравнений;

110.2) *двухшаговый метод наименьших квадратов*. Применяется как для идентифицируемых, так и для сверхидентифицируемых систем одновременных уравнений.

Алгоритм двухшагового МНК:

1. Структурная форма модели преобразуется в приведенную форму.
2. С помощью МНК оцениваются параметры приведенной формы.
3. В правой части сверхидентифицируемого уравнения структурной модели выбираются эндогенные переменные и рассчитываются их теоретические значения по соответствующим приведенным уравнениям.
4. С помощью МНК на основе фактических значений предопределенных и теоретических значений эндогенных переменных оцениваются параметры сверхидентифицируемого уравнения структурной модели.

3. ПРАКТИКУМ

3.1. ЗАНЯТИЕ 1. ЭЛЕМЕНТЫ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА

Задачи для аудиторной работы

3А1. Имеются следующие данные (табл. 3) о зависимости между среднедушевым денежным доходом населения и оборотом розничной торговли по районам (за месяц).

Таблица 3

Зависимость между среднедушевым денежным доходом населения и оборотом розничной торговли

x_i, шт.	10	19	15	23	11	22	18	20	22	20
y_i, ден. ед.	22	30	27	34	21	34	31	32	36	33

Необходимо:

1) построить корреляционное поле;

2) выбрать вид зависимости (линейная, степенная, показательная, логарифмическая);

3) оценить тесноту и направление связи между переменными с помощью коэффициента корреляции;

4) оценить значимость полученного коэффициента корреляции по критерию Стьюдента (уровень значимости $\alpha = 0{,}05$);

5) вычислить коэффициент детерминации;

6) найти уравнение регрессии Y по X;

7) вычислить коэффициент эластичности;

8) проверить адекватность полученной модели по критерию Фишера при $\alpha = 0{,}05$;

9) сделать выводы.

3А2. Даны следующие функции:

1. $y = a + bx^3 + \varepsilon.$
2. $y = a + b\ln x + \varepsilon.$
3. $\ln y = a + b\ln x + \varepsilon.$
4. $y = a + bx^c + \varepsilon.$
5. $y^a = b + cx^2 + \varepsilon.$
6. $y = 1 + a(1 - x^b) + \varepsilon.$
7. $y = a + b\dfrac{x}{10} + \varepsilon.$

Определите, какие из них:

1) линейны по переменным;

2) линейны по параметрам;

3) нелинейны ни по переменным, ни по параметрам.

3А3. Для двух видов продукции *A* и *B* зависимость расходов y (тыс. ден. ед.) предприятия от объема x (шт.) производства характеризуется данными, представленными в табл. 4.

Таблица 4

Зависимость расходов предприятия от объема производства

Уравнение регрессии	R	Число наблюдений
$y_A = 160 + 0{,}7x$	0,85	30
$y_B = 50x^{0,5}$	0,74	25

Необходимо:

1) пояснить смысл величин 0,7 и 0,5 в уравнениях регрессии;

2) сравнить эластичность расходов на производство продукции *A* и *B* при выпуске объемом 500 единиц;

3) определить, каким должен быть выпуск продукции *A*, чтобы эластичность расходов на ее производство совпадала с эластичностью расходов на производство продукции *B*;

4) оценить значимость каждого уравнения регрессии с помощью F-критерия Фишера (уровень значимости $\alpha = 0{,}05$).

Задачи для самостоятельной работы

3А+Б4. Известна зависимость производительности труда Y от уровня механизации работ X (табл. 5).

Таблица 5

Зависимость производительности труда от уровня механизации работ

x_i, %	32	30	36	40	41	47	56	54	60	55	61	67	69	76
y_i, тыс. ден. ед.	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Необходимо:

1) найти уравнение регрессии Y по X;

2) найти коэффициент детерминации D и пояснить его смысл;

3) проверить значимость уравнения регрессии на 5%-ном уровне по F-критерию;

4) оценить среднюю производительность труда на предприятиях и построить для нее 95%-ный доверительный интервал.

3А+Б5. Исследуя спрос на телевизоры марки N по данным, собранным для 19 торговых точек, аналитический отдел компании АВС выявил следующую зависимость:

$$\ln y = 10{,}5 - 0{,}8 \ln x + \varepsilon;$$
$$t = \quad (2{,}5) \ \ (-4),$$

где y – объем продаж телевизоров марки N в отдельной торговой точке; x – средняя цена телевизора в данной торговой точке. В скобках приведены фактические значения t-критерия Стьюдента для параметров уравнения регрессии (уровень значимости $\alpha = 0{,}05$).

До проведения этого исследования администрация компании предполагала, что эластичность спроса по цене для телевизоров марки N составляет –0,9. Подтвердилось ли предположение администрации результатами исследования?

3Б6. Зависимость объема продаж y (тыс. ден. ед.) от расходов на рекламу x (тыс. ден. ед.) для 12 предприятий концерна характеризуется следующим образом:

– уравнение регрессии: $y = 12{,}6 + 0{,}6x$;

– среднее квадратическое отклонение x: $\sigma_x = 4{,}7$;

– среднее квадратическое отклонение y: $\sigma_y = 3{,}4$.

Необходимо:

1) определить коэффициент корреляции;

2) оценить значимость коэффициента корреляции с помощью t-критерия Стьюдента ($\alpha = 0{,}05$).

3Б7. По данным 30 нефтяных компаний получено следующее уравнение регрессии между оценкой Y (ден. ед.) и фактической стоимостью X (ден. ед.) этих компаний: $y_x = 0{,}8750x + 295$.

Найти: 95%-ный доверительный интервал для оценки стоимости предприятий, фактическая стоимость которых составила 1300 ден. ед., если среднее значение, которое может принимать фактическая стоимость, равно 1100 ден. ед., среднее квадратическое отклонение переменной X равно 270 ден. ед., а стандартная ошибка регрессии составляет 0,189.

3Б8. Зависимость материалоемкости продукции от объема выпуска для 10 однородных предприятий представлена в табл. 6.

Таблица 6

Зависимость материалоемкости продукции от объема выпуска

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

Необходимо:

1) записать систему нормальных уравнений, предполагая, что представленные данные могут описываться зависимостями вида: а) $y = a + \dfrac{b}{x}$; б) $y = ax^b$;

2) оценить тесноту связи с помощью индекса корреляции;

3) охарактеризовать эластичность изменения материалоемкости продукции;

4) сделать вывод о значимости уравнения регрессии.

3.2. ЗАНЯТИЕ 2. ЭЛЕМЕНТЫ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА. ЭКОНОМЕТРИЧЕСКИЙ АНАЛИЗ ПРИ НАРУШЕНИИ КЛАССИЧЕСКИХ ПРЕДПОЛОЖЕНИЙ

Задачи для аудиторной работы

ЗА9. Зависимость спроса на продукцию от ее цены для 10 однородных предприятий представлена в табл. 7.

Таблица 7

Зависимость спроса на продукцию от ее цены

Показатель	Предприятие									
	1	2	3	4	5	6	7	8	9	10
Количество продукции, шт.	90	60	50	40	37	36	35	60	70	35
Цена, ден. ед.	10	20	30	40	50	60	70	15	12	25

Необходимо:

1) построить корреляционное поле;

2) выбрать вид зависимости (степенная, показательная, логарифмическая и др.);

3) линеаризовать зависимость;

4) построить корреляционное поле для преобразованных данных.

3А+Б10. Моделирование прибыли фирмы по уравнению $y = ab^x$ привело к результатам, представленным в табл. 8.

Таблица 8

Результаты моделирования прибыли

№ п/п	Прибыль фирмы y, тыс. ден. ед.		№ п/п	Прибыль фирмы y, тыс. ден. ед.	
	фактическая y_i	расчетная $\hat{y}_i$		фактическая y_i	расчетная $\hat{y}_i$
1	10	11	5	18	20
2	12	11	6	11	11
3	15	17	7	13	14
4	17	15	8	19	16

Требуется оценить качество модели:

1) определить ошибку аппроксимации;

2) найти показатель тесноты связи прибыли с исследуемым в модели фактором;

3) рассчитать F-критерий Фишера (уровень значимости $\alpha = 0,05$);

4) сделать выводы.

3А11. Для 20 однотипных предприятий изучалась зависимость рентабельности (y) от производительности труда (x_1) и фондоотдачи (x_2). Были получены следующие варианты уравнений регрессии:

$$y = 25 + 15x_1,\ R^2 = 0,75;$$

$$y = 42 + 28x_2,\ R^2 = 0,53;$$

$$y = 32 + 10x_1 + 8x_2,\ R^2 = 0,8,\ F = 34.$$

Средняя ошибка: (2,5) (3)

$$y = 23 + 15x_1 + 2x_2 + 0,7x_2^2,\ R^2 = 0,82,\ F = 24,3.$$

Средняя ошибка: (5) (12) (0,4).

Проанализируйте связь результата с каждым из факторов. Выберите лучшее уравнение регрессии и обоснуйте выбор.

Задачи для самостоятельной работы

3А+Б12. Зависимость среднемесячной производительности труда от возраста рабочих характеризуется моделью $y = a + bx + cx^2$. Ее использование привело к результатам, представленным в табл. 9.

Таблица 9

Результаты моделирования производительности труда

№ п/п	Производительность труда рабочих y, тыс. ден. ед.		№ п/п	Производительность труда рабочих y, тыс. ден. ед.	
	фактическая y_i	расчетная $\hat{y}_i$		фактическая y_i	расчетная $\hat{y}_i$
1	12	10	6	11	12
2	8	10	7	12	13
3	13	13	8	9	10
4	15	14	9	11	10
5	16	15	10	9	9

Требуется оценить качество модели, определив ошибку аппроксимации, индекс корреляции и F-критерий Фишера (уровень значимости $\alpha = 0{,}05$).

3А+Б13. Дана зависимость уровня рентабельности y (%) от производительности труда x_1 (тыс. ден. ед.), продолжительности оборота оборотных средств предприятия x_2 (дней) и материалоотдачи x_3 (тыс. ден. ед.) (табл. 10).

Таблица 10

Зависимость уровня рентабельности от производительности труда, продолжительности оборота оборотных средств и материалоотдачи

y, %	x_1, тыс. ден. ед.	x_2, дней	x_3, тыс. ден. ед.
21	7,9	18	66
23	10,0	19	88
20	7,9	17	55
21	8,5	18	59
23	9,0	21	71
21	7,8	17	62
22	7,7	20	33
20	8,0	17	59
22	8,3	18	60
21	7,0	17	39

Необходимо:

1) найти множественный коэффициент детерминации и пояснить его смысл;

2) найти уравнение множественной линейной регрессии y по x_1, x_2, x_3;

3) оценить значимость этого уравнения и его коэффициентов на уровне $\alpha = 0{,}05$;

4) сравнить раздельное влияние на зависимую переменную каждой из объясняющих переменных, используя коэффициенты эластичности.

3Б14. Для 40 точек оценена следующая модель производственной функции:

$$y = 0{,}6 + 0{,}4l + 0{,}32k,\ R^2 = 0{,}75;$$

$$t = (2{,}6)\ (0{,}75)\ (2{,}81),$$

где y, l, k – темпы прироста объема выпуска, затрат труда и капитала соответственно.

Какие из следующих выводов верны:

1) нужно ввести новую объясняющую переменную, т. к. доля объясненной дисперсии слишком мала;

2) нужно исключить фактор k, т. к. он оказался статистически незначимым;

3) нужно исключить фактор l, т. к. он оказался статистически незначимым;

4) модель имеет удовлетворительные статистики, поэтому нет смысла ее совершенствовать?

Ответ обоснуйте.

3Б15. В результате 26 наблюдений получена следующая модель производственной функции:

$$y = 0{,}46l + 0{,}32k + 0{,}3,\ \ R^2 = 0{,}17;$$

$$t = (3{,}81)\quad (2{,}87),$$

где y, l, k – темпы прироста объема выпуска, затрат труда и капитала соответственно.

Какие из следующих выводов верны:

1) нужно ввести новую объясняющую переменную, т. к. доля объясненной дисперсии слишком мала;

2) нужно исключить фактор k, т. к. он оказался статистически незначимым;

3) нужно исключить фактор l, т. к. он оказался статистически незначимым;

4) модель имеет удовлетворительные статистики, поэтому нет смысла ее совершенствовать?

Ответ обоснуйте.

3Б16. По выборке объемом $n = 50$ для x_1, x_2, x_3 построена следующая корреляционная матрица:

$$K = \begin{bmatrix} 1 & 0{,}45 & -0{,}35 \\ 0{,}45 & 1 & 0{,}52 \\ -0{,}35 & 0{,}52 & 1 \end{bmatrix}.$$

Найдите и оцените статистическую значимость частных коэффициентов корреляции $r_{12(3)}$, $r_{23(1)}$, $r_{13(2)}$.

3Б+С17. Имеются следующие данные об урожайности озимой пшеницы y (ц/га) за 10 лет (табл. 11).

Таблица 11

Урожайность озимой пшеницы

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

Необходимо:

1) найти среднее значение, среднее квадратическое отклонение и коэффициенты автокорреляции (для лагов $\tau = 1$; 2) временного ряда;

2) найти уравнение тренда временного ряда y_t, полагая, что он линейный, и проверить его значимость на уровне $\alpha = 0{,}05$.

3) провести сглаживание временного ряда y_t методом скользящих средних, используя простую среднюю арифметическую с интервалом сглаживания: а) $m = 3$; б) $m = 5$.

3Б18. В табл. 12 представлены данные, отражающие динамику роста доходов на душу населения y_t (ден. ед.) за восьмилетний период.

Таблица 12

Динамика роста доходов на душу населения

t, лет	1	2	3	4	5	6	7	8
y_t, ден. ед.	1133	1222	1354	1389	1342	1377	1491	1684

Полагая, что тренд линейный и условия классической модели выполнены, необходимо:

1) найти уравнение тренда и оценить его значимость на уровне $\alpha = 0{,}05$;

2) дать точечный и с надежностью 0,95 интервальный прогнозы среднего и индивидуального значений доходов на 9-й год.

3.3. ЗАНЯТИЕ 3. СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ

Задачи для аудиторной работы

3А19. Рассматривается система уравнений вида

$$\begin{cases} y_1 = b_1 y_2 + a_1 x + \varepsilon_1, \\ y_2 = b_2 y_1 + \varepsilon_2. \end{cases}$$

Какие из переменных данной модели являются экзогенными, эндогенными, предопределенными? Проверить, является ли данная система идентифицируемой. Изменится ли ответ, если в число регрессоров второго уравнения включить: а) константу; б) переменную x?

3А20. Проверить, идентифицируемы ли уравнения:

$$\begin{cases} y_{1t} = b_{13} y_{3t} + a_{13} y_{3,t-1} + c_{10} + \varepsilon_1, \\ y_{2t} = b_{23} y_{3t} + a_{21} x_{1t} + c_{20} + \varepsilon_2, \\ y_{1t} = y_{2t}, \end{cases}$$

где y_{1t} – спрос на товар в момент времени t; y_{2t} – предложение товара в момент t; y_{3t} – цена товара в момент t; $y_{3,t-1}$ – цена товара в момент $(t-1)$; x_{1t} – доход в момент t; t – текущий период; $(t-1)$ – предыдущий период.

3А+Б21. Определите метод оценки параметров модели. Запишите приведенную форму модели. Оцените параметры идентифицируемой структурной модели:

$$\begin{cases} y_1 = c_{10} + b_{12} y_2 + a_{11} x_1 + \varepsilon_1, \\ y_2 = c_{20} + b_{21} y_1 + a_{22} x_2 + \varepsilon_2, \end{cases}$$

где y_1 и y_2 – эндогенные переменные системы; x_1 и x_2 – экзогенные переменные этой системы.

Исходные данные приведены в табл. 13.

Таблица 13

Исходные данные

y_1	y_2	x_1	x_2
0	20	7	2
–1	22	8	2
5	20	6	4
11	18	4	6
9	12	2	4
12	26	7	8
17	21	4	9
12	21	5	7

Задачи для самостоятельной работы

ЗА22. Проверить, идентифицируема ли система:

$$\begin{cases} y_{1t} = b_{13}y_{3t} + a_{13}y_{3,t-1} + a_{11}x_{1t} + c_{10} + \varepsilon_1, \\ y_{2t} = b_{23}y_{3t} + a_{23}y_{3,t-1} + c_{20} + \varepsilon_2, \\ y_{1t} = y_{2t}, \end{cases}$$

где y_{1t} – спрос на товар в момент времени t; y_{2t} – предложение товара в момент t; y_{3t} – цена товара в момент t; $y_{3,t-1}$ – цена товара в момент $(t-1)$; x_{1t} – доход в момент t; t – текущий период; $(t-1)$ – предыдущий период.

ЗА23. Рассматривается модель предложения денег и спроса на деньги:

$$\begin{cases} y_{1t} = c_{10} + b_{12}y_{2t} + a_{11}x_{1t} + \varepsilon_1, \\ y_{2t} = c_{20} + b_{21}y_{1t} + a_{21}x_{1t} + \varepsilon_2, \end{cases}$$

где y_{1t} – процентные ставки в период t; y_{2t} – ВВП в период t; x_{1t} – денежная масса в период t.

Какие из переменных данной модели являются экзогенными, эндогенными, предопределенными? Применив необходимое и достаточное условия идентифицируемости, определите, идентифицируемо ли каждое из уравнений модели.

ЗА+Б24. Оценить параметры идентифицируемой структурной модели:

$$\begin{cases} y_1 = c_{10} + b_{12}y_2 + a_{12}x_2 + \varepsilon_1, \\ y_2 = c_{20} + b_{21}y_1 + a_{21}x_1 + \varepsilon_2. \end{cases}$$

где y_1 и y_2 – эндогенные переменные системы; x_1 и x_2 – экзогенные переменные этой системы.

Исходные данные приведены в табл. 14.

Таблица 14

Исходные данные

y_1	y_2	x_1	x_2
–8	12	3	7
2	12	5	3
8	14	7	1
0	8	3	3
–3	22	8	7
–1	6	2	3
9	16	8	1
4	16	7	3

3А+Б25. Эконометрическая модель содержит три уравнения, три эндогенные переменные (y) и три экзогенные переменные (x). Ниже представлена матрица коэффициентов при переменных в структурной форме этой модели (табл. 15).

Таблица 15

Матрица коэффициентов

Уравнение	Коэффициенты при переменных					
	y_1	y_2	y_3	x_1	x_2	x_3
1	–1	0	b_{13}	a_{11}	0	a_{13}
2	b_{21}	–1	b_{23}	0	a_{22}	0
3	0	b_{32}	–1	a_{31}	0	a_{33}

Запишите структурную форму модели. Решите проблему идентифицируемости для данной модели.

Исходя из приведенной формы модели:

$$\begin{cases} y_1 = -x_1 + 5x_2 + 2x_3, \\ y_2 = 3x_1 - x_2 + 6x_3, \\ y_3 = -2x_1 + 4x_2 - x_3, \end{cases}$$

рассчитайте, если это возможно, структурную форму.

3Б26. Эконометрическая модель содержит четыре уравнения, четыре эндогенные переменные (y) и три экзогенные переменные (x). Ниже представлена матрица коэффициентов при переменных в структурной форме этой модели (табл. 16).

Таблица 16

Матрица коэффициентов

Уравнение	Коэффициенты при переменных						
	y_1	y_2	y_3	y_4	x_1	x_2	x_3
1	–1	0	b_{13}	b_{14}	a_{11}	0	0
2	0	–1	b_{23}	0	a_{21}	0	0
3	0	b_{32}	–1	0	a_{31}	0	a_{33}
4	b_{41}	b_{42}	b_{43}	–1	0	a_{42}	a_{43}

Применив необходимое и достаточное условия идентифицируемости, определите, идентифицируемо ли каждое уравнение модели. Определите метод оценки параметров модели.

3Б27. Ниже приводятся результаты расчета параметров некоторой эконометрической модели. Структурная форма модели:

$$\begin{cases} y_1 = -4 + ???\,y_2 - 9{,}4x_2 + \varepsilon_1, \\ y_2 = 12{,}83 - 2{,}67y_1 + ???\,x_1 + \varepsilon_2, \\ y_3 = 1{,}36 - 1{,}76y_1 + 0{,}828y_2 + \varepsilon_3, \end{cases}$$

приведенная форма модели:

$$\begin{cases} y_1 = 2 + 4x_1 - 3x_2 + v_1, \\ y_2 = 7{,}5 + 5x_1 + 8x_2 + v_2, \\ y_3 = 4 + ???\,x_1 + ???\,x_2 + v_3. \end{cases}$$

Какими методами получены параметры структурной и приведенной форм модели? Обоснуйте возможность применения косвенного МНК для расчета структурных параметров модели. Восстановите пропущенные характеристики.

4. ЛАБОРАТОРНЫЙ ПРАКТИКУМ

4.1. ЛАБОРАТОРНАЯ РАБОТА «НЕЛИНЕЙНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ»

Тема: элементы корреляционно-регрессионного анализа.

Цели:

– изучить основные понятия корреляционно-регрессионного анализа;

– освоить методы выбора вида зависимости и оценки качества полученного уравнения;

– научиться использовать MS Excel при построении и анализе регрессионных моделей.

Контрольные вопросы

1. Из каких соображений выбирается вид зависимости?
2. Какой вид должна иметь точечная диаграмма преобразованных исходных данных?
3. Каким методом ищется линейная зависимость по преобразованным исходным данным?
4. Как оценить тесноту связи между экзогенной и эндогенной переменными? Как оценить значимость полученного коэффициента?
5. Каков экономический смысл параметров эмпирической зависимости? Как оценить значимость полученных параметров?
6. Как находятся параметры искомой эмпирической зависимости?
7. Что характеризует показатель R^2?
8. Каков экономический смысл коэффициента эластичности? Как он находится для нелинейной регрессии?
9. Что дает анализ точечной диаграммы и графика эмпирической зависимости?
10. Как проверяется качество построенной модели?

Ход работы

Теоретические сведения. Пусть заданы результаты наблюдений (табл. 17), такие что точки $M(x_i; y_i)$ не располагаются на прямой линии. Пусть далее геометрическим или каким-либо другим образом определен вид нелинейной зависимости y от x. Найдем, если это возможно, взаимно-однозначное преобразование $X = \varphi(x;\ y)$, $Y = \psi(x;\ y)$, при котором эта нелинейная зависимость переходит в линейную $Y = AX + B$.

Таблица 17

Результаты наблюдений

x	x_1	x_2	…	x_n
y	y_1	y_2	…	y_n

По формулам $X_i = \varphi(x_i;\ y_i)$, $Y_i = \psi(x_i;\ y_i)$ и исходным данным (табл. 17) найдем соответствующие значения новых переменных X и Y (табл. 18).

Таблица 18

Преобразованные исходные данные

X	X_1	X_2	…	X_n
Y	Y_1	Y_2	…	Y_n

Здесь точки $M(X_i;\ Y_i)$, $i = \overline{1, n}$, будут располагаться вблизи некоторой прямой. По исходным данным (табл. 18) с помощью МНК найдем коэффициенты A и B эмпирической зависимости $Y = AX + B$. Затем, переходя в этой формуле к исходным переменным x и y, найдем искомую зависимость y от x.

Использованные выше преобразования называются *методом выравнивания*. Этот метод используется не только для определения неизвестных параметров, но и для проверки правильности выбора вида функциональной зависимости. Действительно, *если вид зависимости y от x определен правильно, то точки $M(X_i;\ Y_i)$, $i = \overline{1, n}$, будут располагаться вблизи некоторой прямой*; если это не так, то вид зависимости y от x определен неправильно.

В табл. 19 указаны преобразования, с помощью которых можно «выровнять» некоторые классы функций, наиболее часто встречающиеся на практике.

Таблица 19

Классы функций

$y = f(x)$	Преобразование	Определяемые параметры	Ограничения
$y = bx^a$	$Y = \ln y;\ X = \ln x$	$a = A;\ b = e^B$	$x > 0;\ y > 0$
$y = be^{ax}$	$Y = \ln y;\ X = x$	$a = A;\ b = e^B$	$y > 0$
$y = \dfrac{1}{ax+b}$	$Y = \dfrac{1}{y};\ X = x$	$a = A;\ b = B$	$y \neq 0$
$y = \dfrac{x}{ax+b}$	$Y = \dfrac{x}{y};\ X = x$	$a = A;\ b = B$	$y \neq 0$
$y = a\ln x + b$	$Y = y;\ X = \ln x$	$a = A;\ b = B$	$x > 0$

На рис. 9−13 приведены графики некоторых функциональных зависимостей.

Функциональные зависимости:

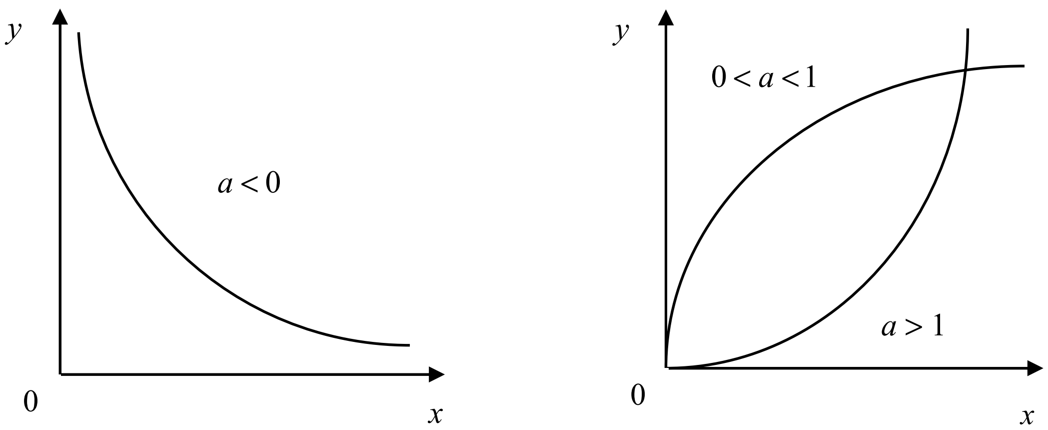


Рис. 9. Зависимость $y = bx^a$, $b > 0$

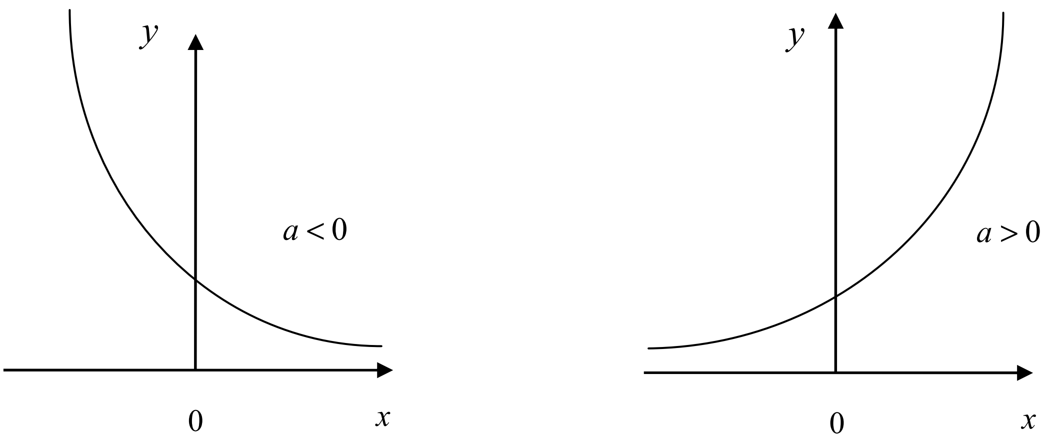


Рис. 10. Зависимость $y = be^{ax}$, $b > 0$

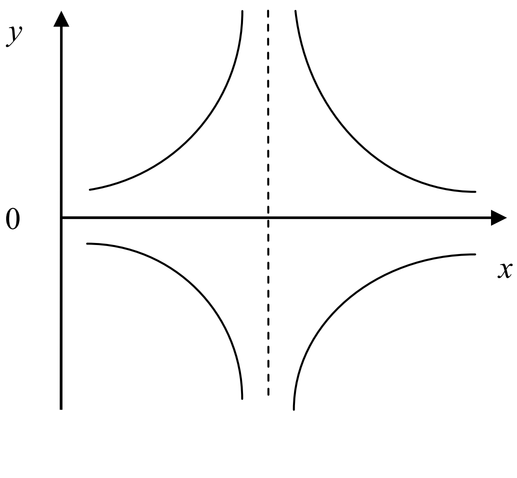


Рис. 11. Зависимость $y = \dfrac{1}{ax+b}$

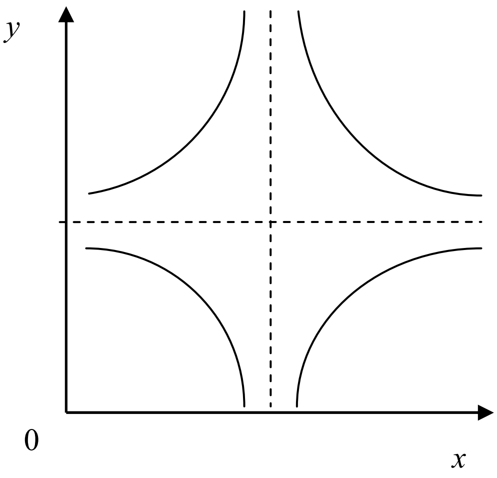


Рис. 12. Зависимость $y = \dfrac{x}{ax+b}$

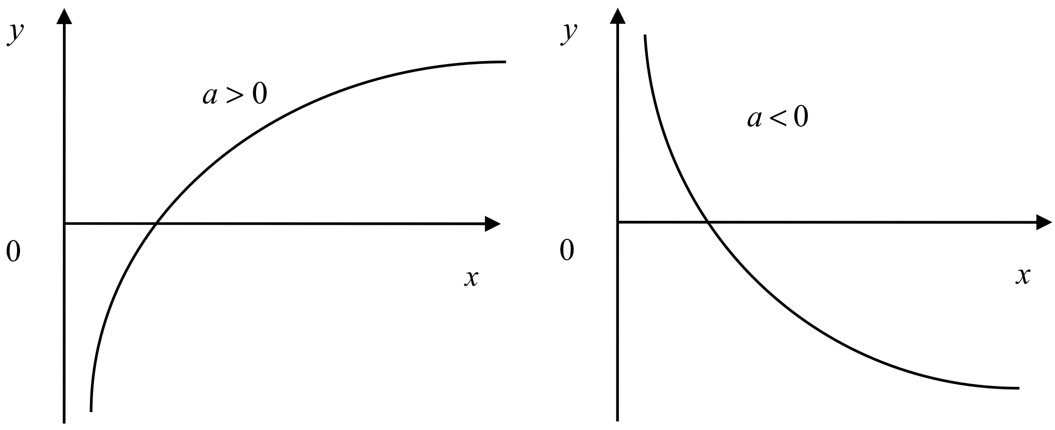


Рис. 13. Зависимость $y = a\ln x + b$

Параметры A и B линейной зависимости обычно ищут МНК. Затем находят параметры a и b нелинейной зависимости и записывают первоначальную зависимость между x и y.

Проверка качества построенного уравнения осуществляется с помощью *коэффициента множественной корреляции* (*индекса корреляции*)

$$R = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2}}.$$

Коэффициент R принимает значения от 0 до 1. Чем выше значение R, тем ближе расчетные значения результативного признака к фактическим. Коэффициент R используется при любой форме связи переменных.

Затем вычисляется средняя относительная ошибка аппроксимации

$$E_{\text{отн}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\varepsilon_i}{y_i}\right| \cdot 100\%.$$

Допустимый предел значений $E_{\text{отн}}$ составляет не более 8–15%.

Формулировка задачи. Для 10 районов за год известны доля расходов на покупку продовольственных товаров K (%) в общих расходах и средненедельная зарплата t (ден. ед.) одного работающего. Получены следующие экспериментальные данные зависимости между K и t, записанные в табл. 20.

Таблица 20

Зависимость расходов от зарплаты

t, ден. ед.	5,0	5,6	6,0	6,4	6,8	7,2	7,6	8,0	8,4	8,8
K, %	10,4	14,4	17,1	22,5	25,9	33,1	40,4	50,0	59,2	74,1

Необходимо:

1) построить корреляционное поле, выбрать вид зависимости (степенная, показательная, логарифмическая и др.), линеаризовать зависимость, построить корреляционное поле для преобразованных данных;

2) определить параметры линейной зависимости;

3) найти коэффициент корреляции и сделать вывод о тесноте связи между переменными. Проверить значимость коэффициента корреляции. Проверить адекватность полученной модели по критерию Фишера и определить среднюю относительную ошибку аппроксимации. Уровень значимости $\alpha = 0{,}1$;

4) получить зависимость для исходных данных. Для полученной модели определить коэффициент эластичности;

5) сделать проверку при помощи *Мастера диаграмм*.

Решение. Решим рассматриваемую задачу в MS Excel с помощью *Мастера диаграмм*. На основе диапазона данных (t_i; K_i) построим точечную диаграмму (рис. 14).

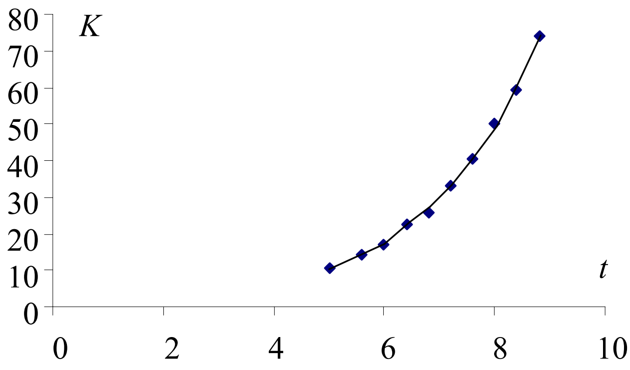


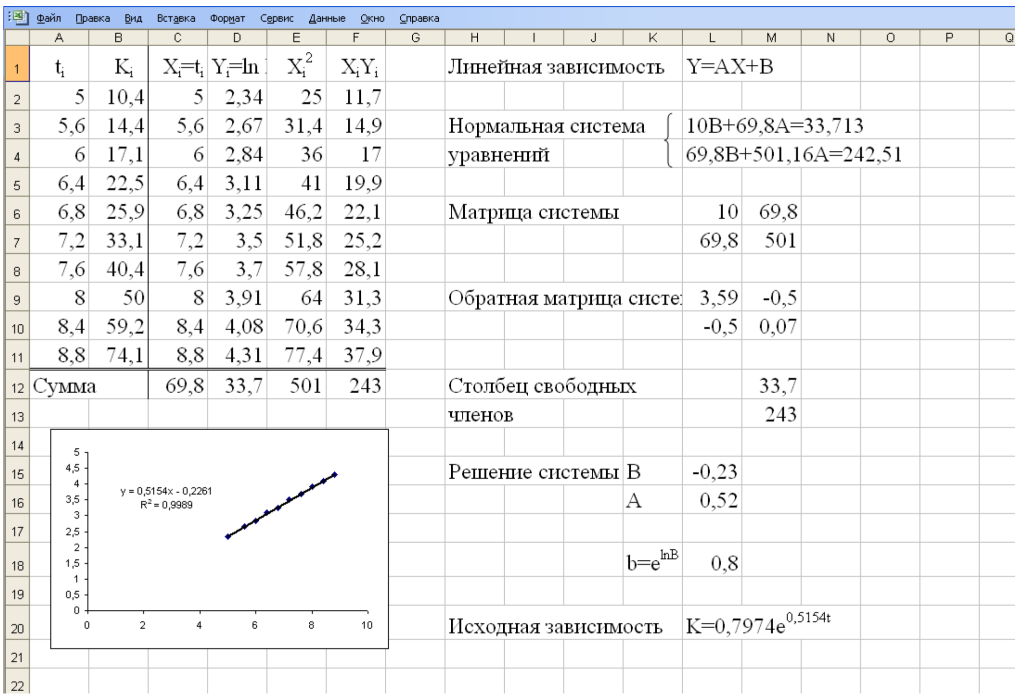
Рис. 14. Диаграмма исходных данных

Сравнивая полученный график с графиками на рис. 9−13, видим, что для данного случая могут подойти зависимости $y = be^{ax}$ или $y = \dfrac{1}{ax+b}$ (рис. 10, 11).

Выбираем первую зависимость $Y = be^{ax}$.

Делаем преобразование данных по формулам (см. табл. 19): $Y_i = \ln K_i$; $X_i = t_i$. Расчеты представлены на рабочем листе (рис. 15) в столбцах C и D.

По преобразованным данным (X; Y) снова строим диаграмму (см. рис. 15).



	A	B	C	D	E	F
1	t_i	K_i	X_i=t_i	Y_i=ln	X_i^2	X_iY_i
2	5	10,4	5	2,34	25	11,7
3	5,6	14,4	5,6	2,67	31,4	14,9
4	6	17,1	6	2,84	36	17
5	6,4	22,5	6,4	3,11	41	19,9
6	6,8	25,9	6,8	3,25	46,2	22,1
7	7,2	33,1	7,2	3,5	51,8	25,2
8	7,6	40,4	7,6	3,7	57,8	28,1
9	8	50	8	3,91	64	31,3
10	8,4	59,2	8,4	4,08	70,6	34,3
11	8,8	74,1	8,8	4,31	77,4	37,9
12	Сумма		69,8	33,7	501	243

	H	K	L	M
1	Линейная зависимость		Y=AX+B	
3	Нормальная система		10B+69,8A=33,713	
4	уравнений		69,8B+501,16A=242,51	
6	Матрица системы		10	69,8
7			69,8	501
9	Обратная матрица систе		3,59	-0,5
10			-0,5	0,07
12	Столбец свободных			33,7
13	членов			243
15	Решение системы	B	-0,23	
16		A	0,52	
18		$b=e^{\ln B}$	0,8	
20	Исходная зависимость		$K=0{,}7974e^{0{,}5154t}$	

Рис. 15. Рабочий лист 1

Из рис. 15 (см. диаграмму) видно, что точки $N_i(X_i;\ Y_i)$, $i=\overline{1,n}$, расположены вдоль прямой линии $Y = AX + B$. Это значит, что вид зависимости выбран правильно и преобразования переменных выполнены верно.

Параметры уравнения $Y = AX + B$ найдем по методу наименьших квадратов (МНК). Для этого нужно записать систему нормальных уравнений:

$$\begin{cases} nB + A\sum x_i = \sum y_i, \\ B\sum x_i + A\sum x_i^2 = \sum x_i y_i. \end{cases}$$

В этой системе $n = 10$ – число пар значений $(X_i;\ Y_i)$. Вычисление остальных значений системы представлено на рис. 15 в столбцах C–F. Получаем систему:

$$\begin{cases} 10B + 69{,}8A = 33{,}7, \\ 69{,}8B + 501A = 243. \end{cases}$$

Решим ее в MS Excel матричным методом. Запишем матрицу системы $\begin{bmatrix} 10 & 69{,}8 \\ 69{,}8 & 501 \end{bmatrix}$ и столбец свободных членов $\begin{bmatrix} 33{,}7 \\ 243 \end{bmatrix}$. Система в матричном виде: $\begin{bmatrix} 10 & 69{,}8 \\ 69{,}8 & 501 \end{bmatrix} \cdot \begin{bmatrix} B \\ A \end{bmatrix} = \begin{bmatrix} 33{,}7 \\ 243 \end{bmatrix}$.

Решение этой системы: $\begin{bmatrix} B \\ A \end{bmatrix} = \begin{bmatrix} 10 & 69{,}8 \\ 69{,}8 & 501 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 33{,}7 \\ 243 \end{bmatrix}$.

Обратную матрицу $\begin{bmatrix} 10 & 69{,}8 \\ 69{,}8 & 501 \end{bmatrix}^{-1}$ найдем в MS Excel, используя функцию МОБР.

В пустом месте экрана выделим область размером 2×2 (размер области такой, какой должна быть обратная матрица). Далее вызовем функцию МОБР и выделим ячейки, в которых расположена матрица $\begin{bmatrix} 10 & 69{,}8 \\ 69{,}8 & 501 \end{bmatrix}$ (это ячейки L6:M7 на рис. 15), нажмем на три клавиши: Ctrl+Shift+Enter. Получим обратную матрицу.

Далее используем функцию МУМНОЖ для того, чтобы умножить обратную матрицу на вектор-столбец свободных членов. Определим, какую размерность будет иметь результат.

Матрица размерности 2×2 умножается на матрицу размерности 2×1. Результат будет иметь размерность 2×1. Поэтому в пустом месте экрана выделим область размером 2×1 и вызовем функцию МУМНОЖ: ***fx*** → Математические → МУМНОЖ → ОК. В открывшемся окне зададим адреса перемножаемых матриц (L9:M10 и M12:M13) и нажмем Ctrl+Shift+Enter для выполнения действия. В результате получим решение системы:

$$\begin{bmatrix} B \\ A \end{bmatrix} = \begin{bmatrix} -0,23 \\ 0,52 \end{bmatrix}.$$

Таким образом, найдены коэффициенты линейной зависимости: $Y = 0{,}52X - 0{,}23$.

Коэффициент корреляции r для преобразованных данных X и Y можно вычислить с помощью функции КОРРЕЛ (r = = 0,9994). Связь между признаками прямая ($r > 0$) и достаточно сильная.

Проверим значимость коэффициента корреляции по критерию Стьюдента. При уровне значимости $\alpha = 0{,}1$ и с учетом того, что в примере количество степеней свободы равно: $\gamma = n - k - 1 =$ $= 10 - 1 - 1 = 8$, получим табличное значение критерия $t_{\text{табл}} = 1{,}86$ (функция СТЬЮДРАСПОБР). Теперь вычислим фактическое (расчетное) значение: $t_{\text{расч}} = \left|\dfrac{r\sqrt{n-1-1}}{\sqrt{1-r^2}}\right| = 84{,}32$. Поскольку фактическое значение $t_{\text{расч}}$ выше табличного, то связь между результативным и факторным показателями является надежной, а величина коэффициента корреляции – значимой.

Коэффициент детерминации равен: $D = R^2 = 0{,}9989$. Это значит, что изменение K на 99,89% зависит от изменения исследуемых факторов (t), а на долю других факторов приходится 0,11% изменения результативного показателя.

Для того чтобы убедиться в надежности уравнения связи и правомерности его использования для практических целей, необходимо дать статистическую оценку надежности показателей связи.

Критерий Фишера. Найдем фактическое значение критерия по формуле: $F_{\text{расч}} = (R^2 / k) / ((1 - R^2) / (n - k - 1)) = 7109{,}8$. Найдем табличное значение критерия: при уровне значимости $\alpha = 0{,}1$ и с учетом того, что в примере количество степеней свободы равно:

$\gamma_1 = k = 1$ и $\gamma_2 = n - k - 1 = 10 - 1 - 1 = 8$, получим: $F_{\text{табл}} = 3{,}458$ (функция FРАСПОБР). Так как $F_{\text{расч}} \geq F_{\text{табл}}$, то построенную модель можно считать адекватной.

Найдем среднюю относительную ошибку аппроксимации по формуле: $E_{\text{отн}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \cdot 100\% = 0{,}54\%$. Заметим, что $\hat{y}_i$ вычисляется по уравнению регрессии. Средняя относительная ошибка мала, что также свидетельствует об адекватности модели.

Найдем теперь коэффициенты нелинейной зависимости a, b по формулам (см. табл. 19): $a = A = 0{,}52$; $b = e^B = e^{-0{,}23} = 0{,}797$. Исходная зависимость имеет вид: $K = 0{,}797e^{0{,}52t}$.

Коэффициент эластичности находим по формуле: $Э = K'\frac{t}{K}$. Для этого вида зависимости Э = 0,5154t. Согласно полученным данным, K возрастает на 0,5154t (%) при увеличении t на 1%.

Сделаем проверку полученной формулы. На диаграмме, построенной по исходным данным (t_i; K_i), выделим любое значение и нажатием правой кнопки мыши вызовем меню, в котором выберем команду *Добавить линию тренда*. Будет открыто диалоговое окно *Линия тренда*, содержащее вкладку *Тип*, где задается вид тренда (уравнения): линейный, логарифмический, полиномиальный (от 2-й до 6-й степени включительно), степенной, экспоненциальный.

Замечание. В случае зависимостей $y = \frac{1}{ax+b}$ и $y = \frac{x}{ax+b}$ (табл. 19) следует с помощью *Мастера диаграмм* построить точечную диаграмму зависимости Y от X и выбрать линейную линию тренда (во вкладке *Тип* диалогового окна *Линия тренда* построение зависимостей $y = \frac{1}{ax+b}$ и $y = \frac{x}{ax+b}$ не предусмотрено).

Для того чтобы получить аналитическое выражение выбранного уравнения, необходимо во вкладке *Параметры* активизировать флажок *Показывать уравнение на диаграмме*.

Если поставить флажок *Поместить на диаграмму величину достоверности аппроксимации* R^2, то в области построения будет выделено значение показателя R^2, по которому можно судить, насколько хорошо выбранное уравнение аппроксимирует эмпирические данные. Чем ближе значение R^2 к единице, тем более адекватным исследуемому явлению или процессу является уравнение.

Выбирая предлагаемые виды зависимости, убеждаемся, что экспоненциальная зависимость $K = be^{at}$ наиболее хорошо аппроксимирует эмпирические данные (рис. 16).

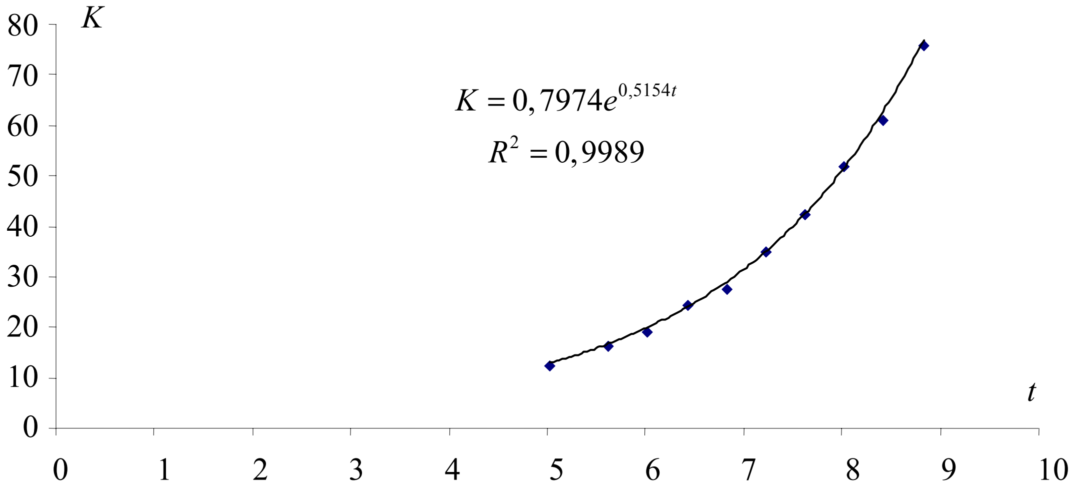


Рис. 16. Результат решения задачи
при помощи MS Excel

Индивидуальные задания

Исходные данные представлены в табл. 21.

Необходимо:

1) построить корреляционное поле, выбрать вид зависимости (степенная, показательная, логарифмическая и др.), линеаризовать зависимость, построить корреляционное поле для преобразованных данных;

2) определить параметры линейной зависимости;

3) найти коэффициент корреляции и сделать вывод о тесноте связи между переменными. Проверить значимость коэффициента корреляции. Проверить адекватность полученной модели по критерию Фишера и определить среднюю относительную ошибку аппроксимации. Уровень значимости $\alpha = 0,1$;

4) получить зависимость для исходных данных. Для полученной модели определить коэффициент эластичности;

5) сделать проверку при помощи *Мастера диаграмм*.

Таблица 21

Исходные данные

Вариант 1		Вариант 2		Вариант 3		Вариант 4		Вариант 5	
x	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*
−15	0,3	0,2	0,01	5	–6,4	7,8	–11	–2	19,5
–11	0,7	0,3	0,1	7	–3,6	8	–8,5	–1,6	15,2
−5	1,1	0,4	0,2	10	–0,3	8,5	–4,9	–1,1	12,5
–1	1,9	0,5	0,5	12	1,4	9	–4,2	–0,8	10,3
5	3,2	0,6	0,8	15	3,2	10	0,9	–0,1	7,5
9	5,1	0,7	1,5	20	6,1	11	1,9	0,4	5,6
13	7,2	0,8	2,2	25	8,1	13	2,9	1	4,4
18	12,2	0,9	3,5	30	9,5	16	5	1,9	2,6
21	16,1	1	5,1	35	11,1	20	8	2,8	1,8
25	24,5	1,1	7	40	12	25	10	3,6	0,9
Вариант 6		Вариант 7		Вариант 8		Вариант 9		Вариант 10	
x	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*
0,2	4,5	0,1	–1,5	0,2	0,4	15	5,1	0,001	0,7
1,3	8,2	0,2	–0,3	1,1	0,48	10	4,5	0,01	1,5
3	9,8	0,3	0,6	1,8	0,7	6	3,4	0,05	2,17
6	11,9	0,4	1,1	2,6	1	3,2	2,5	0,1	2,6
11	13,1	0,5	1,5	3	1,6	2,2	1,8	0,19	3
16	14,3	0,6	1,8	3,4	2,4	1,9	1	0,25	3,2
21	15	1	2,2	3,6	3,8	0,9	0,5	0,5	3,5
26	15,7	1,5	2,6	3,8	7,3	0,2	–0,5	1	3,7
31	16,8	2	2,9	3,85	10,1	0,03	–6	2	3,9
36	17,4	3	3,1	3,9	14,2	0,001	–10	3	4
Вариант 11		Вариант 12		Вариант 13		Вариант 14		Вариант 15	
x	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*
0,05	10	15,5	11,2	–7	1,2	0,7	0,01	0,4	15,9
0,071	5,5	15,7	7,8	–5,2	1,5	1,3	0,3	1,1	11,3
0,1	3,7	16	5,6	–2,3	2,1	2,1	0,8	2	7,6
0,15	2,6	16,5	3,5	–1,2	2,8	3,3	1,5	3,5	4,4
0,2	2,3	17,2	2,8	0,5	4,6	4,3	3,2	5	3,1
0,3	1,4	18	1,8	0,8	5,4	4,8	5,4	7	2,2
0,4	0,9	19	1,4	1,2	6,8	5,8	9	9	1,3
0,5	0,3	20	1,1	1,8	9,5	6	11,2	12	0,7
1	–0,2	22	0,65	2,2	12,1	6,1	15,9	15	0,3
2	–0,8	24	0,5	2,5	15,4	6,3	22,8	18	0,02

4.2. ЛАБОРАТОРНАЯ РАБОТА «МНОЖЕСТВЕННАЯ КОРРЕЛЯЦИЯ И РЕГРЕССИЯ. ЭКОНОМЕТРИЧЕСКИЙ АНАЛИЗ ПРИ НАРУШЕНИИ КЛАССИЧЕСКИХ ПРЕДПОЛОЖЕНИЙ»

Тема: элементы корреляционно-регрессионного анализа. Эконометрический анализ при нарушении классических предположений.

Цели:

– освоить методы отбора экзогенных переменных и оценки качества полученного уравнения;

– научиться использовать надстройку «Анализ данных» пакета MS Excel при построении и анализе моделей множественной регрессии;

– изучить предпосылки метода наименьших квадратов (условия Гаусса – Маркова), научиться обнаруживать и устранять нарушение этих предпосылок.

Контрольные вопросы

1. Какой метод используют, чтобы найти линейную зависимость по исходным данным?
2. Каков экономический смысл параметров эмпирической зависимости? Как оценить значимость полученных параметров?
3. Как оценить тесноту связи между экзогенными переменными?
4. Как определить наличие либо отсутствие мультиколлинеарности?
5. Назовите и опишите методы устранения мультиколлинеарности.
6. Как оценить тесноту связи между экзогенной и эндогенной переменными? Как оценить значимость полученных коэффициентов?
7. Что характеризует коэффициент детерминации?
8. Каков экономический смысл коэффициента эластичности? Как он находится для множественной регрессии?
9. Как проверяется качество построенной модели?
10. Как определить наличие либо отсутствие гетероскедастичности, автокорреляции?

Ход работы

Формулировка задачи. Зависимость уровня рентабельности y (%) от производительности труда x_1 (тыс. ден. ед.), продолжительности оборота оборотных средств предприятия x_2 (дней) и материалоотдачи x_3 (тыс. ден. ед.) приведена в табл. 22.

Таблица 22

Зависимость уровня рентабельности от производительности труда, продолжительности оборота оборотных средств предприятия и материалоотдачи

y, %	x_1, тыс. ден. ед.	x_2, дней	x_3, тыс. ден. ед.
21	7,9	18	66
23	10	19	88
20	7,9	17	55
21	8,5	18	59
23	9	21	71
21	7,8	17	62
22	7,7	20	33
20	8	17	59
22	8,3	18	60

1. Считая, что между результативным и факторными признаками имеет место линейная связь, найти линейное уравнение связи (регрессии), проверить значимость его коэффициентов.

2. Найти парные коэффициенты корреляции и составить корреляционную матрицу. По полученным данным сделать вывод о тесноте связи между рассматриваемыми переменными. Проверить значимость коэффициентов корреляции и проанализировать полученные данные. Сделать вывод о наличии либо отсутствии мультиколлинеарности и при необходимости устранить мультиколлинеарность.

3. Найти линейное уравнение регрессии для преобразованной модели. Для полученной линейной модели определить коэффициенты эластичности. Сделать выводы.

4. Проверить адекватность полученной модели по критерию Фишера и определить среднюю относительную ошибку аппроксимации. Уровень значимости $\alpha = 0,1$.

Решение. Уравнение множественной линейной регрессии имеет следующий вид:

$$y = b_0 + b_1x_1 + \ldots + b_kx_k.$$

Оценка параметров b_0, b_1, b_2, …, b_k обычно осуществляется по методу наименьших квадратов:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_{i1} + \ldots + b_k x_{ik}))^2 \to \min.$$

Для получения уравнения регрессии используем команды меню Сервис → Анализ данных → Регрессия. Укажем исходные эндогенные и экзогенные переменные, а также заданный уровень значимости. Получим следующий результат (табл. 23).

Таблица 23

Вывод итогов

Регрессионная статистика			
Множественный R		0,916743892	
R-квадрат		**0,840419363**	
Нормированный R-квадрат		0,744670981	
Стандартная ошибка		0,571186262	
Наблюдения		9	
Дисперсионный анализ			
	df	*SS*	*MS*
Регрессия	3	8,590953491	2,863651164
Остаток	5	1,631268731	0,326253746
Итого	8	10,22222222	
	Коэффициенты	*Стандартная ошибка*	*t-статистика*
Y-пересечение	**6,812993527**	2,996852471	2,273383022
Переменная $X1$	**0,464232385**	0,630385569	**0,736426098**
Переменная $X2$	**0,556517393**	0,195782166	**2,842533636**
Переменная $X3$	**0,009030445**	0,029112725	**0,31018894**

Запишем уравнение регрессии:

$$\hat{y} = 6{,}81299 + 0{,}46423x_1 + 0{,}55652x_2 + 0{,}00903x_3.$$

Найдем коэффициенты корреляции с помощью функции КОРРЕЛ (или по формулам) и составим корреляционную матрицу. Коэффициенты $r_{yx_1} = 0{,}70235$; $r_{yx_2} = 0{,}83406$; $r_{yx_3} = -0{,}39061$ показывают связь между результативным признаком y и факторами x_1, x_2, x_3 соответственно; коэффициенты $r_{x_1x_2} = 0{,}4359$; $r_{x_1x_3} = 0{,}8072$; $r_{x_2x_3} = 0{,}04064$ показывают связь между факторными признаками.

Коэффициент парной корреляции является безразмерной величиной и не зависит от выбора единиц обеих переменных. Значение коэффициента корреляции лежит в интервале от –1 (в случае строгой линейной отрицательной связи) до +1 (в случае строгой линейной положительной связи). Соответственно, положительное значение коэффициента корреляции свидетельствует о прямой связи между исследуемым и факторным показателями, а отрицательное – об обратной. Чем ближе значение коэффициента корреляции к единице, тем теснее связь. Близкий к нулю коэффициент корреляции говорит об отсутствии линейной связи переменных, но не свидетельствует об отсутствии их связи вообще. В случае равенства нулю показателя корреляции нельзя однозначно утверждать, что исследуемые показатели независимы. В данном случае можно попытаться найти более сложную модель их связи. Значительный интерес представляют коэффициенты корреляции, характеризующие взаимосвязь факторов между собой. В корреляционную модель следует подбирать независимые между собой факторы. Если коэффициент корреляции двух факторов выше 0,8, то один из них необходимо исключить из модели.

Так как $r_{x_1x_2} = 0,4359 < 0,8$ и $r_{x_2x_3} = 0,04064 < 0,8,$ то связь между факторами x_1 и x_2, а также x_2 и x_3 достаточно слабая, и их можно включить в модель; $r_{x_1x_3} = 0,8072 > 0,8,$ поэтому связь между факторами x_1 и x_3 достаточно сильная, имеет место мультиколлинеарность. Для устранения мультиколлинеарности применим метод исключения факторов.

Проверим *значимость коэффициентов регрессии* по критерию Стьюдента. Значения t-статистики (фактические значения) для показателей x_1 и x_3 равны соответственно 0,73643 и 0,31019. Наименьшее значение t-статистики у фактора x_3. Исключим фактор x_3 (незначим, k = 3) из рассмотрения и будем искать зависимость между факторами y и x_1, x_2.

Значимость *коэффициентов корреляции* проверяется по критерию Стьюдента. Выдвигаем нулевую гипотезу H_0: коэффициент корреляции равен нулю ($r = 0$); конкурирующая гипотеза H_1: $r \neq 0$. Если расчетное значение $t_{\text{расч}}$ выше табличного, то можно сделать вывод, что величина коэффициента корреляции является значимой, следовательно, нулевая гипотеза отвергается.

При уровне значимости $\alpha = 0,1$ и с учетом того, что в примере количество степеней свободы равно:$\gamma = n - k - 1 = 9 - 1 - 1 = 7$, получим табличное значение критерия (функция СТЬЮДРАСПОБР). Теперь вычислим фактические значения: $t_{yx_1} = \left| \dfrac{r_{yx_1}\sqrt{n-1-1}}{\sqrt{1-r_{yx_1}^2}} \right|$. Поскольку фактическое значение критерия в первых двух случаях выше табличного, то связь между результативным и факторными показателями x_1 и x_2 является надежной, а величина коэффициентов корреляции – значимой. Про фактор x_3 можно сказать, что коэффициент корреляции значимым не является.

Матрица коэффициентов парной корреляции имеет вид:

$$K = \begin{bmatrix} 1 & r_{yx_1} & \dots & r_{yx_k} \\ r_{x_1 y} & 1 & \dots & r_{x_1 x_k} \\ \dots & \dots & \dots & \dots \\ r_{x_k y} & r_{x_k x_1} & \dots & 1 \end{bmatrix}.$$

По данным этой матрицы можно примерно оценить, какие факторы существенно влияют на переменную y, а какие – несущественно, а также выявить взаимосвязь между факторами.

После исключения фактора x_3 корреляционная матрица имеет вид:

$$K = \begin{bmatrix} 1 & 0,70235 & 0,83406 \\ 0,70235 & 1 & 0,4359 \\ 0,83406 & 0,4359 & 1 \end{bmatrix}.$$

Связь между оставшимися факторами достаточно слабая ($r_{x_1 x_2} = 0,4359 < 0,8$).

Коэффициент множественной корреляции принимает значения от 0 до 1, но несет в себе более универсальный смысл: чем ближе его значение к единице, тем в большей степени учтены факторы, влияющие на зависимую переменную, тем более точной выглядит построенная на основе отобранных факторов модель. Расчет коэффициента множественной корреляции производится на основе значений коэффициентов парной корреляции:

$$R = \sqrt{1 - \frac{\det K}{K_{11}}},$$

где $\det K$ – определитель корреляционной матрицы; K_{11} – алгебраическое дополнение элемента первой строки и первого столбца матрицы K.

Найдем коэффициент множественной корреляции по формуле (или воспользуемся результатами, представленными в табл. 24):

$$R = 0{,}915.$$

Если коэффициент корреляции возвести в квадрат, получим коэффициент детерминации: $D = R^2$.

Коэффициент детерминации D = 0,837. Это значит, что изменение рентабельности на 83,7% зависит от изменения исследуемых факторов, а на долю других факторов приходится 16,3% изменения результативного показателя. После исключения фактора коэффициент детерминации уменьшился несущественно (ранее D = 0,840). Значит, в корреляционную модель удалось включить наиболее существенные факторы.

Найдем уравнение регрессии для преобразованной модели. Используем команды меню Сервис → Анализ данных → Регрессия (табл. 24).

Таблица 24

Вывод итогов

Регрессионная статистика				
Множественный R		**0,915067475**		
R-квадрат		**0,837348484**		
Нормированный R-квадрат		0,783131311		
Стандартная ошибка		0,526412377		
Наблюдения		9		
Дисперсионный анализ				
	df	*SS*	*MS*	*F*
Регрессия	2	8,559562277	4,279781138	**15,4443408**
Остаток	6	1,662659946	0,277109991	
Итого	8	10,22222222		
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-значение*
Y-пересечение	**6,587630865**	2,679543464	2,45849002	0,04922022
Переменная *X*1	**0,63590539**	0,278142062	**2,286261151**	0,06226255
Переменная *X*2	**0,520938348**	0,146226794	**3,562536884**	0,01189071

Проверим значимость коэффициентов нового уравнения регрессии по критерию Стьюдента. Значения t-статистики (фактические значения) для показателей x_1 и x_2 равны соответственно 2,286 и 3,563, $t_{\text{табл}} = t_{\alpha;n-k-1} = 1{,}943$, следовательно, коэффициенты нового уравнения регрессии являются значимыми, а мультиколлинеарность устранена.

Запишем уравнение регрессии:

$$\hat{y} = 6,5876 + 0,6359x_1 + 0,5209x_2.$$

Это уравнение выражает зависимость уровня рентабельности от производительности труда и продолжительности оборота оборотных средств.

Коэффициенты уравнения показывают количественное воздействие каждого фактора на результативный показатель при неизменности других. В примере рентабельность повышается на 0,6359% при увеличении производительности труда на 1 тыс. ден. ед.; на 0,5209% – при увеличении продолжительности оборота оборотных средств на 1 день.

Определение уравнения линейной регрессии осуществляется также с помощью функции ЛИНЕЙН категории *Статистические*. Для записи результата нужно выделить область размером 5×(k + 1), где k – число переменных, а затем вызвать функцию ЛИНЕЙН. В диалоговом окне требуется задать следующие аргументы: интервал значений Y_i; блок значений X_i; константу; статистику. В полях *Константа* и *Статистика* следует задать значение *Истина* (первое – для того чтобы получить уравнение регрессии с ненулевым свободным членом, второе – для получения оценки достоверности этого уравнения регрессии). Задав аргументы, необходимо нажать Ctrl+Shift+Enter. Вывод результата осуществляется в следующем формате (табл. 25).

Таблица 25

Результат применения функции ЛИНЕЙН

b_k	b_{k-1}	…	b_1	b_0
$s\{b_k\}$	$s\{b_{k-1}\}$	…	$s\{b_1\}$	$s\{b_0\}$
R^2	$\sqrt{s^2_{\text{ад}}}$	…	–	–
$F_{\text{расч}}$	$f_{\text{ад}}$	…	–	–
$SS_{\text{регр}}$	$SS_{\text{ост}}$	…	–	–

В первой строке записываются коэффициенты уравнения регрессии в обратном порядке, во второй – их среднеквадратические отклонения. Первый элемент третьей строки – множественный коэффициент детерминации. Здесь $SS_{\text{ост}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ – остаточная сумма квадратов; $s^2_{\text{ад}} = SS_{\text{ост}} / f_{\text{ад}}$ – дисперсия адекватности; $f_{\text{ад}} = n - k - 1$ – число степеней свободы дисперсии адекватности; $SS_{\text{регр}} = \sum_{i=1}^{n}(y_i - \overline{y})^2 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$;

$$F_{\text{расч}} = \frac{SS_{\text{регр}} / k}{SS_{\text{ост}} / (n - k - 1)}$$

используется для проверки значимости множественного коэффициента регрессии.

Проверим правильность полученных результатов, используя функцию ЛИНЕЙН:

0,520938	**0,635905**	**6,587631**
0,146227	0,278142	2,679543
0,837348	0,526412	#Н/Д
15,44434	6	#Н/Д
8,559562	1,66266	#Н/Д

Уравнение регрессии и коэффициент детерминации совпадают с полученными ранее.

Коэффициенты эластичности рассчитываются по формуле:

$$Э_i = b_i \frac{\overline{x}_i}{\overline{y}}$$

и показывают, на сколько процентов в среднем изменяется функция с изменением аргумента на 1%.

Так, для переменной x_1 коэффициент эластичности равен 0,2474, а для переменной x_2 – 0,4454.

Согласно полученным данным, рентабельность возрастает на 0,247% при увеличении производительности труда на 1%; на 0,445% – при увеличении продолжительности оборота оборотных средств на 1%.

Для того чтобы убедиться в надежности уравнения связи и правомерности его использования для практических целей, необходимо дать статистическую оценку надежности показателей связи. Для этого используются критерий Фишера, средняя ошибка аппроксимации, коэффициенты множественной корреляции и детерминации.

Критерий Фишера. Значимость построенной модели проверяется следующим образом. Выдвигаем гипотезу H_0: модель незначима. Конкурирующая гипотеза H_1: модель значима. Гипотеза проверяется по критерию Фишера. Фактическая величина $F_{\text{расч}}$ сопоставляется с табличной и делается заключение о надежности связи. Если $F_{\text{расч}} \geq F_{\text{табл}}$ при заданном уровне значимости α, то линейную модель можно считать адекватной (нулевая гипотеза отвергается).

Найдем фактическое значение критерия из результатов применения команд меню Сервис $\to$ Анализ данных $\to$ Регрессия (или функции ЛИНЕЙН):

$$F_{\text{расч}} = 15{,}444.$$

Найдем табличное значение критерия: при уровне значимости $\alpha = 0{,}1$ и с учетом того, что в примере количество степеней свободы равно: $\gamma_1 = k = 2$ и $\gamma_2 = n - k - 1 = 9 - 2 - 1 = 6$, получим: $F_{\text{табл}} = 3{,}4633$ (функция FРАСПОБР). Так как $F_{\text{расч}} \geq F_{\text{табл}}$, то построенную модель можно считать адекватной.

Для статистической оценки точности уравнения связи используется также *средняя относительная ошибка аппроксимации*:

$$E_{\text{отн}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\varepsilon_i}{y_i}\right| \cdot 100\%.$$

Чем меньше теоретическая линия регрессии (рассчитанная по уравнению) отклоняется от фактической (эмпирической), тем меньше средняя относительная ошибка аппроксимации.

$$E_{\text{отн}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\varepsilon_i}{y_i}\right| \cdot 100\% = \frac{0{,}159}{9} \cdot 100\% = 1{,}76\%.$$

Средняя ошибка мала, что также свидетельствует об адекватности модели.

Следовательно, данное уравнение можно использовать для практических целей: оценки результатов хозяйственной деятельности; расчета влияния факторов на прирост результативного показателя; подсчета резервов повышения уровня исследуемого показателя; планирования и прогнозирования его величины.

Индивидуальные задания

Задача. Зависимость уровня рентабельности y (%) от производительности труда x_1 (тыс. ден. ед.), продолжительности оборота оборотных средств предприятия x_2 (дней) и материалоотдачи x_3 (тыс. ден. ед.) приведена в табл. 26.

1. Считая, что между результативным и факторными признаками имеет место линейная связь, найти линейное уравнение регрессии.

2. Найти парные коэффициенты корреляции и составить корреляционную матрицу. По полученным данным сделать вывод о тесноте связи между рассматриваемыми переменными. Проверить значимость коэффициентов корреляции и проанализировать полученные данные. Сделать вывод о наличии либо отсутствии мультиколлинеарности и при необходимости устранить мультиколлинеарность.

3. Найти линейное уравнение регрессии для преобразованной модели. Для полученной линейной модели определить коэффициенты эластичности. Сделать выводы.

4. Проверить адекватность полученной модели по критерию Фишера и определить среднюю относительную ошибку аппроксимации. Уровень значимости $\alpha = 0{,}1$.

Таблица 26

Исходные данные

Вариант 1				Вариант 2				Вариант 3			
y	x_1	x_2	x_3	y	x_1	x_2	x_3	y	x_1	x_2	x_3
54	23	290	58	28	45	5	58	36	19,7	39	58
75	21	248	54	28	43	4,6	54	35	16,3	40	54
60	24	300	60	30	43	4,5	60	33	15,5	39	60
64	24	261	61	31	36	4,5	61	34	32,3	40	61
63	25	262	69	30	41	4,4	69	33	33,7	40	69
42	30	297	75	29	43	4,4	75	32	39,1	45	75
70	22	279	72	30	43	4,6	72	32	40,9	47	72
34	36	316	53	31	36	4,2	53	31	36,5	47	53
43	30	304	76	32	29	3,3	76	31	38,7	49	76
60	24	299	62	31	36	4,2	62	34	15,5	38	62
46	33	304	80	31	29	3,2	80	32	40	45	80
35	38	319	87	32	33	4	87	30	30	42	87

Продолжение табл. 26

Вариант 4				Вариант 5				Вариант 6			
y	x_1	x_2	x_3	y	x_1	x_2	x_3	y	x_1	x_2	x_3
20	32	33	68	48	76	41	120	90	18	33	100
24	30	31	65	45	69	40	112	55	36	35	92
28	36	41	72	43	67	44	109	30	60	35	89
30	40	39	79	41	61	39	105	75	25	45	85
31	41	46	80	40	55	35	98	42	36	60	78
33	47	43	85	38	60	45	106	20	60	68	86
34	56	34	87	37	54	42	87	45	55	69	67
37	54	38	86	34	56	43	89	38	48	70	69
34	55	34	90	33	47	43	72	70	38	72	52
38	60	42	92	30	41	46	65	38	50	78	45
40	55	35	86	–	–	–	–	–	–	–	–
Вариант 7				Вариант 8				Вариант 9			
y	x_1	x_2	x_3	y	x_1	x_2	x_3	y	x_1	x_2	x_3
28	8	41	120	28	67	31	119	10	77	6	120
28	11	44	101	28	68	29	104	5,5	32	18	112
30	10	44	98	30	59	24	98	3	40	30	109
31	5	48	94	31	62	13	94	7,5	80	6	105
30	14	48	87	30	68	10	87	4,2	80	18	98
30	23	48	95	30	68	10	95	2	60	30	106
31	5	49	74	31	70	7	74	7,1	65	12	87
31	18	49	77	31	70	4	77	3,8	54	24	89
31	21	50	64	31	70	2	64	6,3	70	12	72
32	29	50	58	32	70	1	58	3,4	45	24	65
Вариант 10				Вариант 11				Вариант 12			
y	x_1	x_2	x_3	y	x_1	x_2	x_3	y	x_1	x_2	x_3
83	7	39	121	36	26	43	130	90	6	40	119
56	18	42	105	35	24	42	11	55	18	42	104
32	30	41	98	33	23	42	119	30	30	41	98
75	6	54	94	34	21	41	114	75	6	54	94
42	18	68	87	33	22	41	97	42	18	68	87
20	30	80	45	32	23	40	106	20	30	80	45
71	12	50	91	32	20	40	85	71	12	50	91
38	24	60	77	31	21	40	88	38	24	60	77
63	12	70	65	31	22	39	76	63	12	70	64
34	24	75	58	30	19	37	69	34	24	75	58

Окончание табл. 26

Вариант 13				Вариант 14				Вариант 15			
y	x_1	x_2	x_3	y	x_1	x_2	x_3	y	x_1	x_2	x_3
90	6	40	78	578	6	80	125	28	31	40	67
55	18	42	74	552	5,8	70	91	28	29	41	64
30	30	41	80	729	7,7	41	73	30	24	43	72
75	6	54	81	586	6,1	54	81	31	13	43,5	80
42	18	68	89	460	4,9	68	89	30	10	43	80
20	30	80	103	569	6,3	80	120	30	10	42,5	85
71	15	54	92	561	5,9	50	92	31	7	43	87
38	24	60	73	665	7	60	73	31	4	41	86
63	12	70	96	465	5,2	68	96	31	2	45	90
73	6	54	82	578	6	70	89	30	12	43	92
71	12	50	80	558	6	70	93	32	1	44	86
34	24	75	107	552	5,6	75	107	–	–	–	–

5. ТЕОРЕТИЧЕСКИЙ И ПРАКТИЧЕСКИЙ МИНИМУМ

5.1. ТЕОРЕТИЧЕСКИЙ МИНИМУМ

Вопросы

1. Основные понятия эконометрики.
2. Элементы корреляционно-регрессионного анализа. Основные понятия корреляционного анализа.
3. Понятие о регрессионной модели.
4. Задачи корреляционно-регрессионного анализа.
5. Линейная парная регрессия.
6. Метод наименьших квадратов.
7. Модель множественной регрессии.
8. Коэффициент корреляции. Коэффициент детерминации.
9. Нелинейная эмпирическая регрессия.
10. Эконометрический анализ при нарушении классических предположений. Основные проблемы при нарушении классических предположений.
11. Мультиколлинеарность.
12. Автокорреляция.
13. Гетероскедастичность.
14. Временные ряды.
15. Системы одновременных уравнений.

Основные понятия эконометрики. *Эконометрика* – наука, объединяющая совокупность математико-статистических методов моделирования и количественного анализа экономических явлений и процессов.

Эконометрика позволяет найти количественное подтверждение (либо неподтверждение) того или иного экономического закона или гипотезы. Одним из важнейших направлений эконометрики является построение прогнозов по различным экономическим показателям.

Задачи эконометрики:

• спецификация модели – построение эконометрических моделей для эмпирического анализа;

• параметризация модели – оценка параметров модели;

• верификация модели – проверка качества параметров модели и самой модели в целом;

• прогнозирование модели – составление прогноза и рекомендаций для конкретных экономических явлений по результатам моделирования.

Эконометрическая модель – математическое описание соотношений между входными (объясняющими, независимыми, экзогенными) и выходными (объясняемыми, зависимыми, эндогенными) переменными изучаемого экономического явления или процесса, основанное на реальных статистических данных.

Эконометрическое моделирование – исследование экономических процессов посредством их эконометрических моделей.

Эконометрические модели условно делят на три класса.

1. *Регрессионные модели с одним уравнением.* Результативный признак представлен в виде функции факторных признаков $Y = f(X) + \varepsilon = f(X_1, X_2, \ldots, X_k) + \varepsilon$, где Y – *наблюдаемое значение* объясняемой (*эндогенной*) переменной, которое зависит от значений объясняющих (*экзогенных*) переменных (факторов); ε – *случайная ошибка* (возмущение).

Объясняемая переменная Y – случайная величина (СВ) при заданных значениях объясняющих переменных X_i, $i = \overline{1, k}$. Объясняющие переменные в модели могут также носить случайный характер. Например, модель зависимости цены от объема поставки, модель зависимости спроса от цены на отдельный товар, модель зависимости спроса от реальных доходов потребителей, модель зависимости объема производства от производственных факторов.

2. *Системы одновременных уравнений.* Они состоят из уравнений, в которые наряду с факторными признаками включены и результативные признаки, т. е. одни и те же переменные могут одновременно рассматриваться как зависимые переменные в одних уравнениях и как независимые – в других.

3. *Модели временных рядов.* Результативный признак является функцией времени или переменных, относящихся к другим моментам времени.

В эконометрическом моделировании рассматриваются следующие типы данных:

• *пространственные данные* – набор сведений по разным объектам, взятым за один и тот же период времени (объем производства предприятий региона, численность сотрудников институтов и т. д.);

• *временные данные* – набор сведений, характеризующий один и тот же объект за разные периоды времени (индекс потребительских цен и др.).

Элементы корреляционно-регрессионного анализа. Основные понятия корреляционного анализа. *Корреляционный анализ* – раздел математической статистики, изучающий силу (тесноту) связи между признаками (между двумя признаками при парной связи и между результативным и множеством факторных признаков при многофакторной связи). *Регрессионный анализ* – раздел математической статистики, изучающий форму связи между признаками.

Различают следующие типы зависимостей между явлениями и их признаками:

1. *Функциональная зависимость* – связь, при которой каждому значению независимой переменной X соответствует точно определенное значение зависимой переменной Y (зависимость выработки продукции на одного рабочего от объема выпущенной продукции и численности рабочих).

2. *Статистическая зависимость* – связь, при которой каждому значению независимой переменной X соответствует множество значений зависимой переменной Y и изменение которой происходит в условиях неопределенности, имеющей, как правило, случайный характер (зависимость всхожести семян некоторых культур от количества микроэлементов при их обработке, зависимость производительности труда на предприятии от его энерговооруженности и т. д.).

3. *Корреляционная зависимость* – частный случай статистический зависимости – связь, при которой каждому значению независимой переменной X соответствует определенное математическое ожидание (среднее значение) зависимой переменной Y.

Условным математическим ожиданием (*условной средней*) $M_x(Y) = M(Y|X = x) = \overline{Y}_x$ называется математическое ожидание СВ Y, вычисленное в предположении, что СВ X приняла значение x.

Корреляционная зависимость бывает:

• *парная* – связь между двумя признаками (результативным Y и факторным X или двумя факторными);

• *частная* – зависимость между результативным и одним факторным признаком или между двумя факторными признаками при фиксированных значениях других факторных признаков;

• *множественная* – зависимость между результативным признаком и двумя или более факторными признаками, включенными в исследование.

Теснота связи количественно выражается величиной коэффициента корреляции.

В зависимости от количества признаков, включенных в модель, связи подразделяются на два вида.

1. *Однофакторные* – связи между одним признаком-фактором и результативным признаком (при абстрагировании от влияния других факторов).

2. *Многофакторные* – связи между несколькими факторными признаками и результативным признаком (факторы действуют комплексно, т. е. одновременно и во взаимосвязи).

Корреляционная зависимость исследуется с помощью методов корреляционного и регрессионного анализа.

Понятие о регрессионной модели. *Теоретическим уравнением* (или просто *уравнением*) *регрессии* Y на X называется уравнение $M(Y \mid X = x) = \overline{Y}_x = f(x)$. Функция $f(x)$ называется *теоретической регрессией* (или просто *регрессией*) Y на X, а ее график – *линией регрессии* СВ Y на СВ X. При этом X является независимой (объясняющей) переменной, Y – зависимой (объясняемой) переменной. При рассмотрении зависимости двух СВ говорят о *парной* регрессии.

Зависимость нескольких переменных, выражаемую функцией

$$M(Y \mid x_1, x_2, ..., x_k) = f(x_1, x_2, ..., x_k),$$

где $M(Y \mid x_1, x_2, ..., x_k)$ – условное математическое ожидание (математическое ожидание СВ Y при условии, что СВ X_1, X_2, ..., X_k приняли значения x_1, x_2, ..., x_k соответственно), называют *множественной регрессией*.

Поскольку реальные значения зависимой переменной не всегда совпадают с ее условными математическими ожиданиями и могут быть различными при одном и том же значении объясняющей переменной (наборе объясняющих переменных), фактическая зависимость должна учитывать ошибку (погрешность) ε, которая также является СВ. Таким образом, связи между зависимой и объясняющей(ими) переменными можно описать соотношениями:

$$Y = M(Y \mid x) + \varepsilon;$$

$$Y = M(Y \mid x_1, x_2, ..., x_k) + \varepsilon.$$

Задачи корреляционно-регрессионного анализа. Основные задачи корреляционно-регрессионного анализа:

1. Установление формы корреляционной связи, т. е. установление вида функции регрессии (линейная, квадратичная, показательная и т. д.).

2. Оценка тесноты корреляционной связи Y и X (степени рассеяния значений СВ Y около $\overline{Y}_x$). Большое рассеяние означает слабую зависимость Y от X либо вообще ее отсутствие. Малое рассеяние указывает на существование достаточно сильной зависимости Y от X.

3. Оценка неизвестных параметров регрессионной модели, проверка гипотез об их значимости и адекватности модели рассматриваемому экономическому объекту.

Выбор формулы связи переменных называется *спецификацией* уравнения регрессии. В случае парной регрессии выбор формулы обычно осуществляется по графическому изображению реальных статистических данных.

Пример 1. Для анализа зависимости инвестиций (Y) предприятия от объемов (X) производства исследуются данные 12 однотипных предприятий, которые приведены в табл. 27.

Таблица 27

Зависимость инвестиций предприятия от объемов производства

Показатель	Предприятие											
	1	2	3	4	5	6	7	8	9	10	11	12
Инвестиции y_i, тыс. ден. ед.	12	15	18	20	22	17	19	25	31	30	40	44
Объем производства x_i, млрд. шт.	17	19	20	21	28	30	31	38	42	48	45	50

Необходимо построить корреляционное поле.

Решение. Построим корреляционное поле (рис. 17).

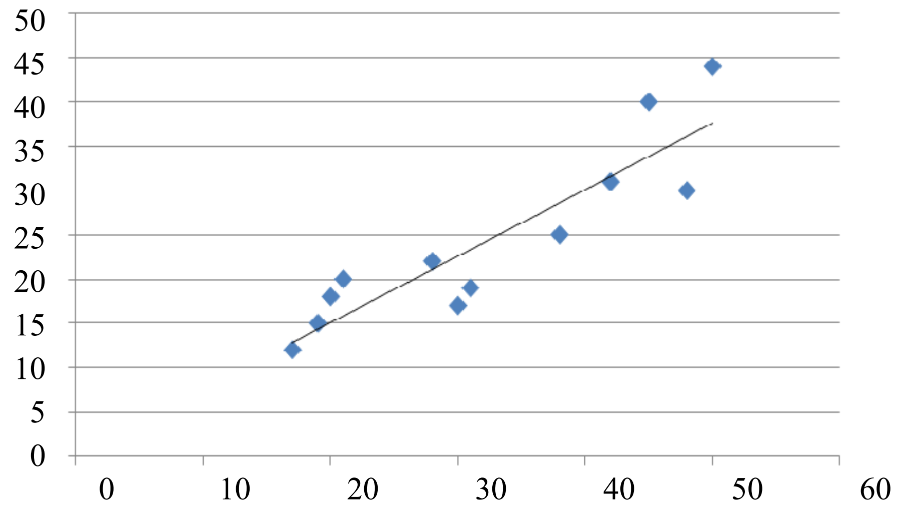


Рис. 17. Корреляционное поле

По расположению точек на корреляционном поле полагаем, что зависимость между X и Y близка к линейной.

Линейная парная регрессия. По выборке ограниченного объема можно искать (приближенно) регрессионную зависимость в определенном виде, например в виде линейной зависимости (эмпирическое линейное уравнение регрессии):

$$\hat{y}_i = b_0 + b_1 x_i + e_i, \qquad (*)$$

где $\hat{y}_i$ – *оценка* условного математического ожидания $M(Y \mid X = x_i)$; b_0 и b_1 – *оценки* неизвестных параметров, называемые *эмпирическими коэффициентами линейной регрессии*, отклонение e_i – оценка теоретического случайного отклонения ε_i.

Модель линейной регрессии (линейное уравнение) является наиболее распространенным (и простым) видом зависимости между экономическими переменными. Кроме того, построенное линейное уравнение может служить начальным этапом эконометрического анализа.

Задачи линейного регрессионного анализа (см. пример 2):

1. По имеющимся статистическим данным (x_i; y_i), $i = \overline{1, n}$, получить наилучшие оценки неизвестных параметров.

2. Проверить статистические гипотезы о параметрах модели.

3. Проверить, достаточно ли хорошо модель согласуется со статистическими данными (адекватность модели данным наблюдений).

Метод наименьших квадратов. Различные выборки из одной и той же генеральной совокупности обычно приводят к определению отличающихся друг от друга оценок. Требуется по конкретной выборке $(x_i;\, y_i)$, $i=\overline{1,\, n}$, найти оценки b_0 и b_1 неизвестных параметров уравнения (*) так, чтобы соответствующая линия регрессии (прямая) являлась бы наилучшей в некотором смысле среди всех других прямых. Другими словами, построенная прямая должна быть «ближайшей» к совокупности точек наблюдений. Мерами качества найденных оценок могут служить определенные функции отклонений (невязок) $e_i = y_i - \hat{y}_i$, $i=\overline{1,\, n}$ (рис. 18).

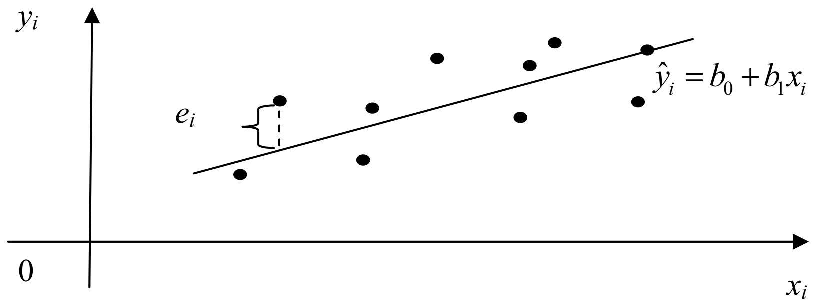


Рис. 18. Линейная регрессия

Самым распространенным методом нахождения коэффициентов (оценок) b_0 и b_1 уравнения эмпирической линейной регрессии является метод наименьших квадратов (МНК). Согласно МНК, эти коэффициенты выбираются таким образом, чтобы минимизировать функцию (сумму квадратов отклонений)

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

Необходимым условием минимума данной функции является равенство нулю ее частных производных по параметрам b_0 и b_1, откуда для определения параметров линейной регрессии получаем линейную систему алгебраических уравнений:

$$\begin{cases} nb_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases}$$

Коэффициент b_1 называется *выборочным коэффициентом регрессии* Y на X. Он показывает, на сколько единиц в среднем изменяется переменная Y при увеличении переменной X на одну единицу.

Коэффициент b_1 нельзя непосредственно использовать для оценки влияния факторного признака x на результативный признак y из-за различия единиц измерения исследуемых показателей. Для этих целей применяется коэффициент эластичности. Для эмпирической линейной регрессии коэффициент эластичности

$$Э_{yx} = b_1 \frac{\overline{x}}{\overline{y}},$$

где $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ – средние значения независимой и зависимой переменных.

Напомним, что в общем случае коэффициент эластичности определяется по формуле:

$$Э_{yx} = y' \frac{x}{y}.$$

Коэффициент эластичности показывает, на сколько процентов изменяется результативный признак y при изменении факторного признака x на один процент.

Для проверки гипотез о статистической значимости коэффициента регрессии, т. е. гипотезы H_0: $b_1 = 0$ и конкурирующей (альтернативной) гипотезы H_1: $b_1 \neq 0$ используется t-статистика:

$$t = \frac{b_1}{S_{b_1}},$$

которая, при выполнении исходных предпосылок модели, имеет распределение Стьюдента с числом степеней свободы $\gamma = n - k - 1$, где n – число наблюдений, k – число независимых переменных в уравнении регрессии, а S_{b_1} – стандартная ошибка коэффициента регрессии. Если речь идет о парной линейной регрессии, то $k = 1$.

Гипотеза H_0 отклоняется, если $\left|t_{\text{расч}}\right| = \left|\frac{b_1}{S_{b_1}}\right| \geq t_{\text{табл}} = t_{\alpha;n-k-1}$, где α – требуемый уровень значимости.

Если гипотеза H_0 принимается, то есть основание полагать, что величина Y не зависит от X. В этом случае говорят, что коэффициент b_1 статистически незначим. При отклонении гипотезы H_0 коэффициент b_1 считается статистически значимым, что указывает на наличие линейной зависимости между Y и X.

Для парной регрессии более важным является анализ статистической значимости коэффициента b_1, т. к. именно он позволяет оценить влияние объясняющей переменной X на зависимую переменную Y.

Воздействие неучтенных факторов и ошибок наблюдений определяется с помощью дисперсии случайных отклонений $D(\varepsilon_i)$. Несмещенной оценкой этой дисперсии является выборочная остаточная дисперсия.

Прогнозируемое значение переменной y вычисляется по формуле:

$$y^*_{\text{прогн}} = b_0 + b_1 x_{\text{прогн}}.$$

Данный прогноз является *точечным*.

Пример 2. Для данных примера 1 найти уравнение регрессии Y по X, а также вычислить коэффициент эластичности. Сделать статистические выводы.

Решение. Найдем уравнение регрессии Y по X. Будем искать уравнение регрессии в виде $\hat{y}_i = b_0 + b_1 x_i,\ i = \overline{1,\ n}$.

Вычисления по МНК удобно выполнять, используя данные табл. 28.

Таблица 28

Вспомогательные вычисления

i	x_i	y_i	x_i^2	$x_i y_i$
1	2	3	4	5
1	17	12	289	204
2	19	15	361	285
3	20	18	400	360
4	21	20	441	420
5	28	22	784	616

Окончание табл. 28

1	2	3	4	5
6	30	17	900	510
7	31	19	961	589
8	38	25	1 444	950
9	42	31	1 764	1 302
10	48	30	2 304	1 440
11	45	40	2 025	1 800
12	50	44	2 500	2 200
Сумма	389	293	14 045	10 952
Среднее	32,42	24,42	–	–

Согласно МНК, имеем:

$$\begin{cases} 12b_0 + 389b_1 = 293, \\ 389b_0 + 14\ 045b_1 = 10\ 952 \end{cases} \Rightarrow \begin{cases} b_0 = -0,015, \\ b_1 = 0,754. \end{cases}$$

Таким образом, эмпирическое уравнение парной линейной регрессии имеет вид:

$$\hat{y}_i = -0,015 + 0,754x_i.$$

Изобразим данную прямую на корреляционном поле (см. рис. 17). Построим ее, например, по следующим двум точкам: $(\overline{x};\ \overline{y}) = (32,42;\ 24,42)$ и $(\overline{x};\ \overline{y}) = (20;\ 15,07)$.

Коэффициент b_1 = 0,754 показывает, на какую величину изменятся инвестиции в данное предприятие, если его объем производства возрастет на одну единицу.

Коэффициент эластичности $Э_{yx} = b_1 \dfrac{\overline{x}}{\overline{y}} = 1,001$ (или $Э_{yx} = f'(x)\dfrac{x}{y}$) показывает, на сколько процентов в среднем изменяются инвестиции с изменением объема производства на 1%.

Модель множественной регрессии. *Уравнение множественной эмпирической линейной регрессии* имеет вид:

$$y_i = b_0 + b_1x_{i1} + \ldots + b_kx_{ik} + \varepsilon_i,$$

где y_i $(i = \overline{1,\ n})$ – i-е наблюдение зависимой переменной; $x_{i1}, x_{i2}, \ldots, x_{ik}$ $(i = \overline{1,\ n})$ – i-е наблюдения независимых переменных

$x_1, x_2, \ldots, x_k$; n – количество наблюдений (объем выборки); k – количество независимых переменных в уравнении.

Оценка параметров регрессии обычно осуществляется по методу наименьших квадратов:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_{i1} + \ldots + b_k x_{ik}))^2 \to \min.$$

Метод наименьших квадратов предусматривает нахождение параметров $b_0, b_1, b_2, \ldots, b_k$ из условия минимума суммы квадратов отклонений. Используя необходимое условие экстремума, получим систему уравнений для определения неизвестных коэффициентов $b_0, b_1, b_2, \ldots, b_k$:

$$\begin{cases} nb_0 + b_1\sum_{i=1}^{n}x_{i1} + \ldots + b_k\sum_{i=1}^{n}x_{ik} = \sum_{i=1}^{n}y_i, \\ b_0\sum_{i=1}^{n}x_{i1} + b_1\sum_{i=1}^{n}x_{i1}^2 + \ldots + b_k\sum_{i=1}^{n}x_{i1}x_{ik} = \sum_{i=1}^{n}x_{i1}y_i, \\ \ldots \\ b_0\sum_{i=1}^{n}x_{ik} + b_1\sum_{i=1}^{n}x_{i1}x_{ik} + \ldots + b_k\sum_{i=1}^{n}x_{ik}^2 = \sum_{i=1}^{n}x_{ik}y_i. \end{cases}$$

Оценку параметров модели можно провести в матричной форме. Уравнение линейной множественной регрессии *в матричной форме* имеет вид:

$$Y = XB + \varepsilon,$$

где $Y = (y_1, y_2, ..., y_n)'$ – вектор значений зависимой переменной размерности $n \times 1$; $X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix}$ – матрица значений независимых переменных $x_1, x_2, ..., x_k$; $B = (b_0, b_1, ..., b_k)'$ – подлежащий оценке вектор неизвестных параметров; $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ – вектор случайных отклонений.

Тогда формула для вычисления параметров регрессионного уравнения по методу наименьших квадратов имеет вид:

$$B = (X'X)^{-1} \cdot X'Y,$$

где X' – транспонированная матрица X; $(X'X)^{-1}$ – обратная матрица.

В частном случае *для двухфакторной модели* получаем *матричное уравнение*

$$(X'X)B = X'Y,$$

$$\text{где } X'X = \begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i2}x_{i1} & \sum x_{i2}^2 \end{bmatrix}; \ X'Y = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{bmatrix}; \ B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}.$$

Коэффициенты b_1, b_2, ..., b_k показывают количественное воздействие каждого фактора на результативный показатель при неизменности значений других факторов.

Коэффициенты эластичности рассчитываются по формуле:

$$Э_i = b_i \frac{\overline{x}_i}{\overline{y}}.$$

Коэффициенты эластичности показывают, на сколько процентов в среднем изменяется функция с изменением аргумента на 1%.

Коэффициент корреляции. Коэффициент детерминации. *Коэффициент парной корреляции* используется в качестве меры, характеризующей степень линейной связи двух переменных. Он представляет собой ковариацию двух наборов данных, деленную на произведение их стандартных отклонений:

$$r_{yx_j} = \frac{n\sum_{i=1}^{n} x_{ij}y_i - \sum_{i=1}^{n} x_{ij} \cdot \sum_{m=1}^{n} y_m}{\sqrt{n\sum_{i=1}^{n} x_{ij}^2 - \left(\sum_{i=1}^{n} x_{ij}\right)^2} \cdot \sqrt{n\sum_{m=1}^{n} y_m^2 - \left(\sum_{m=1}^{n} y_m\right)^2}};$$

$$r_{x_jx_l} = \frac{n\sum_{i=1}^{n} x_{ij}x_{il} - \sum_{i=1}^{n} x_{ij} \cdot \sum_{m=1}^{n} x_{ml}}{\sqrt{n\sum_{i=1}^{n} x_{ij}^2 - \left(\sum_{i=1}^{n} x_{ij}\right)^2} \cdot \sqrt{n\sum_{m=1}^{n} x_{ml}^2 - \left(\sum_{m=1}^{n} x_{ml}\right)^2}}.$$

Коэффициент корреляции принимает значения от –1 до +1. Если $r > 0$, то полагаем, что корреляционная связь между переменными является прямой, если $r < 0$ – обратной.

Если $r = \pm 1$, корреляционная связь представляется линейной функциональной зависимостью. При $r = 0$ линейная корреляционная связь отсутствует.

Множественная корреляция возникает от взаимодействия нескольких факторов с результативным показателем.

Значительный интерес представляют коэффициенты корреляции, характеризующие взаимосвязь факторов между собой. В корреляционную модель следует подбирать независимые между собой факторы. Если коэффициент корреляции двух факторов выше 0,8, то один из этих факторов рекомендуется исключить из модели.

Качественные характеристики связи представлены в табл. 29.

Таблица 29

Качественные характеристики связи

Значение r	Характер связи
От 0 до $\lvert\pm 0{,}3\rvert$	Практически отсутствует
От $\lvert\pm 0{,}3\rvert$ до $\lvert\pm 0{,}5\rvert$	Слабая
От $\lvert\pm 0{,}5\rvert$ до $\lvert\pm 0{,}7\rvert$	Умеренная
От $\lvert\pm 0{,}7\rvert$ до $\lvert\pm 1\rvert$	Сильная (тесная)

В случае парной линейной регрессии коэффициент корреляции также можно вычислить по формуле:

$$r_{xy} = b_1 \frac{\sigma_x}{\sigma_y},$$

где σ_x, σ_y – средние квадратические отклонения случайных величин x и y.

Матрица коэффициентов парной корреляции (корреляционная матрица) имеет вид:

$$K = \begin{bmatrix} 1 & r_{yx_1} & \ldots & r_{yx_k} \\ r_{x_1y} & 1 & \ldots & r_{x_1x_k} \\ \ldots & \ldots & \ldots & \ldots \\ r_{x_ky} & r_{x_kx_1} & \ldots & 1 \end{bmatrix}.$$

По данным этой матрицы можно примерно оценить, какие факторы существенно влияют на переменную y, а какие – несущественно, а также выявить взаимосвязь между факторами.

Для линейной множественной регрессии *коэффициент множественной корреляции* определяется по формуле:

$$R = \sqrt{1 - \frac{\det K}{K_{11}}},$$

где $\det K$ – определитель корреляционной матрицы; K_{11} – алгебраическое дополнение элемента первой строки и первого столбца матрицы K.

Коэффициент множественной корреляции принимает значения от 0 до 1. Чем ближе его значение к единице, тем в большей степени учтены факторы, влияющие на зависимую переменную, тем более точной является построенная на основе отобранных факторов модель.

Индекс корреляции (*коэффициент множественной корреляции*) вычисляется по формуле:

$$R = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2}}.$$

Чем выше значение R, тем вероятнее близость расчетных значений результативного признака к фактическим. Данный показатель используется при любой форме связи переменных.

Долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака характеризует *коэффициент детерминации* $D = R^2$, получаемый возведением в квадрат коэффициента множественной корреляции.

О полноте связи можно судить по величине коэффициентов множественной корреляции и детерминации. Например, если $R = 0{,}92$, а $D = 0{,}85$, то это значит, что вариация результативного признака на 85% зависит от изменения исследуемых факторов, а на долю других факторов приходится 15% вариации результативного показателя. Следовательно, в корреляционную модель удалось включить наиболее существенные факторы.

Для проверки гипотез о статистической значимости коэффициента корреляции, т.е. гипотез

$$H_0\colon r = 0; \quad H_1\colon r \neq 0,$$

при заданном уровне значимости α и объеме выборки n используется t-статистика:

$$t_{\text{расч}} = \frac{|r|}{\sqrt{1 - r^2}} \sqrt{n - 1 - 1}.$$

По таблице критических точек распределения Стьюдента по заданному уровню значимости α и числу степеней свободы $\gamma = n - 2$ находят $t_{кр} = t_{табл} = t_{\alpha;n-2}$. Если $t_{расч} \le t_{кр}$, нет оснований отвергнуть гипотезу H_0. Если $t_{расч} > t_{кр}$, то гипотезу H_0 о равенстве коэффициента корреляции нулю отвергают. Другими словами, r значительно отличается от нуля, т. е. СВ X и СВ Y коррелированны.

Значимость построенной модели проверяется следующим образом. Выдвигаем гипотезу H_0: модель незначима. Конкурирующая гипотеза H_1: модель значима. Гипотеза проверяется по критерию Фишера. Фактическая величина

$$F_{расч} = \frac{R^2}{1-R^2}\frac{(n-k-1)}{k}$$

сопоставляется с табличной и делается заключение о надежности связи. Здесь k – количество независимых переменных в уравнении связи. Если $F_{расч} \ge F_{табл} = F_{\alpha;\gamma_1;\gamma_2}$ со степенями свободы $\gamma_1 = k$; $\gamma_2 = n - k - 1$ при заданном уровне значимости α, то линейную модель можно считать адекватной, гипотеза о случайной природе зависимости между оцениваемыми характеристиками (нулевая гипотеза) отклоняется и признается статистическая значимость и надежность модели.

Определение меры точности модели производится с помощью расчета *средней относительной ошибки аппроксимации*

$$E_{отн} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_i}{y_i}\right| \cdot 100\%.$$

Допустимый предел значений $E_{отн}$ составляет не более 8–15%.

Графическое представление поведения остаточного члена e:

$$e_i = y_i - \hat{y}_i,\ i = \overline{1,\ n},$$

позволяет проанализировать наличие автокорреляции и гетероскедастичности (непостоянства дисперсий отклонений). С помощью графического представления отклонений может быть обнаружена неправильная спецификация уравнения.

Пример 3. По данным примеров 1 и 2 оценить тесноту и направление связи между переменными с помощью коэффициента корреляции; оценить значимость полученного коэффициента корреляции по критерию Стьюдента (уровень значимости α = 0,1);

проверить адекватность полученной модели по критерию Фишера. Сделать выводы.

Решение. Оценим тесноту и направление связи между переменными с помощью коэффициента корреляции $r_{xy} = 0{,}9$. Поскольку коэффициент корреляции положительный, связь прямая. Так как коэффициент корреляции близок к единице, то связь сильная.

Для проверки значимости коэффициента корреляции используется t-критерий Стьюдента

$$t_{\text{расч}} = r_{xy}\sqrt{\frac{n-1-1}{1-r_{xy}^2}} = 5{,}88.$$

При уровне значимости $\alpha = 0{,}1$ и с учетом того, что в примере количество степеней свободы равно: $\gamma = n - k - 1 = 12 - 1 - 1 = 10$, получим: $t_{\text{табл}} = t_{0{,}1;10} = 1{,}81$. Так как $t_{\text{расч}} > t_{\text{табл}}$, то значение коэффициента корреляции признается значимым. Парный коэффициент детерминации $D = r_{xy}^2 = 0{,}81$. Это значит, что изменение y на 81% зависит от изменения исследуемых факторов, а на долю других факторов приходится 19% изменения результативного показателя.

Для проверки адекватности модели используется F-статистика (критерий Фишера)

$$F_{\text{расч}} = \frac{R^2}{1-R^2}\frac{(n-k-1)}{k} = 43{,}63.$$

При заданном уровне значимости $\alpha = 0{,}1$ и с учетом того, что количество степеней свободы равно: $\gamma_1 = k = 1$, $\gamma_2 = n - k - 1 = 10$, получим: $F_{\text{табл}} = 3{,}285$. Расчетное значение критерия больше табличного, поэтому модель можно считать значимой, гипотеза о случайной природе оцениваемых характеристик отклоняется, признается их статистическая значимость.

Нелинейная эмпирическая регрессия. Многие экономические зависимости не являются линейными по своей сути, и поэтому их моделирование линейными уравнениями регрессии не дает положительного результата.

Нелинейность может проявляться как относительно переменных, так и относительно входящих в функцию коэффициентов.

Оценка параметров регрессии, нелинейной по переменным, включенным в анализ, но линейной по оцениваемым параметрам, проводится с помощью МНК путем решения системы линейных алгебраических уравнений.

1. *Степенные модели* вида $y = bx^a$, где a, b – параметры модели (рис. 19).

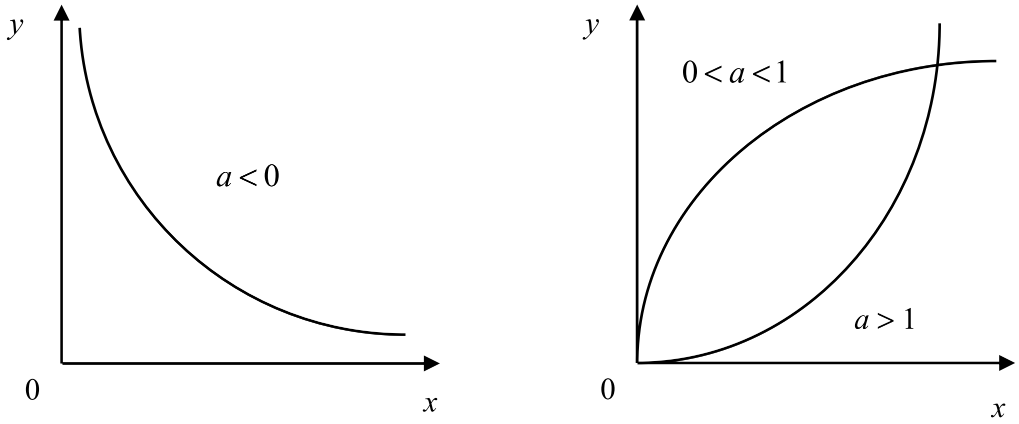


Рис. 19. Зависимость $y = bx^a$, $b > 0$

Прологарифмируем выражение $y = bx^a$: $\ln y = \ln b + a \ln x$. Выполним замену: $Y = \ln y$; $X = \ln x$; $b_0 = \ln b$; $b_1 = a$. Тогда получим: $Y = b_0 + b_1 X$. Коэффициент b_1 определяет эластичность переменной y по переменной x и является константой.

Данная модель легко обобщается на большее число переменных. Например, $\ln y = b_0 + b_1 \ln x_1 + b_2 \ln x_2 + \varepsilon$. Здесь коэффициенты b_1 и b_2 являются эластичностями переменной y по переменным x_1 и x_2 соответственно.

2. *Показательная модель* $y = be^{ax}$, $b > 0$ (рис. 20).

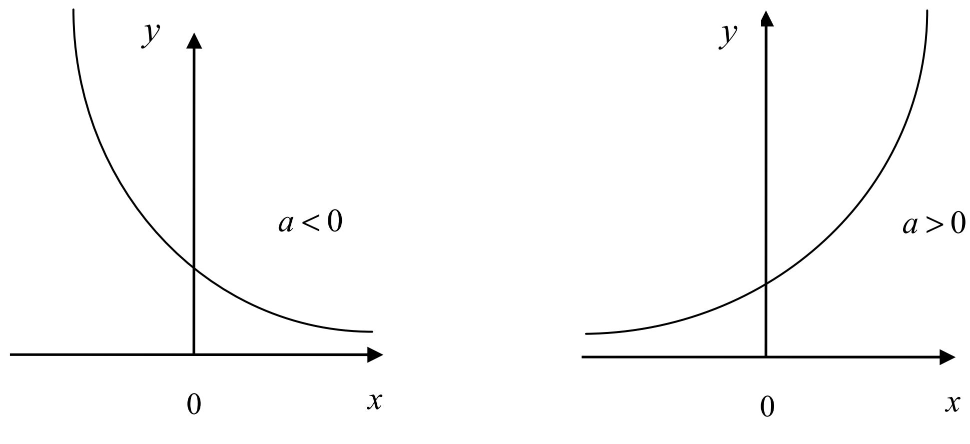


Рис. 20. Зависимость $y = be^{ax}$, $b > 0$

Прологарифмируем выражение $y = be^{ax}$: $\ln y = \ln b + ax$.

Выполним замену: $Y = \ln y$; $X = x$; $b_0 = \ln b$; $b_1 = a$. Получим линейную модель $Y = b_0 + b_1 X$.

3. *Логарифмические модели* – это модели вида $y = a\ln x + b$ (рис. 21). Они сводятся к линейной модели заменой $Y = y$; $X = \ln x$.

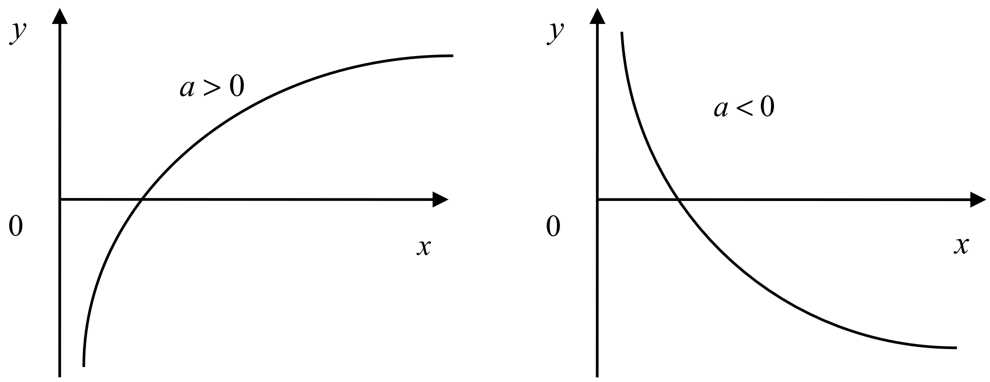


Рис. 21. Зависимость $y = a\ln x + b$

4. *Обратная модель.* Модель вида $y = \dfrac{1}{ax+b}$ (рис. 22) приводится к линейной модели $Y = aX + b$ заменой $Y = \dfrac{1}{y}$; $X = x$, а модель вида $y = \dfrac{x}{ax+b}$ (рис. 23) – заменой $Y = \dfrac{x}{y}$; $X = x$.

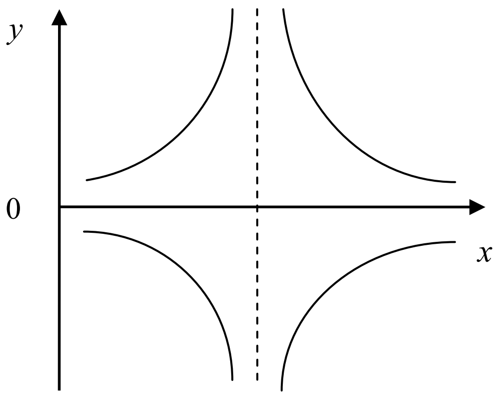


Рис. 22. Зависимость $y = \dfrac{1}{ax+b}$

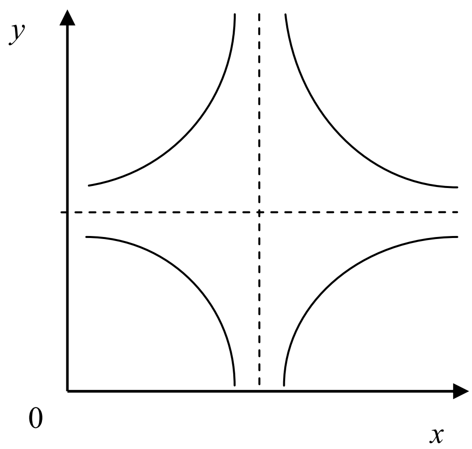


Рис. 23. Зависимость $y = \dfrac{x}{ax+b}$

Пример 4. Для 10 районов за год известны доля расходов на покупку продовольственных товаров K (%) в общих расходах и средненедельная зарплата t (ден. ед.) одного работающего. Получены следующие экспериментальные данные зависимости между K и t, представленные в табл. 30.

Таблица 30

Исходные данные

t, ден. ед.	5,0	5,6	6,0	6,4	6,8	7,2	7,6	8,0	8,4	8,8
K, %	10,4	14,4	17,1	22,5	25,9	33,1	40,4	50,0	59,2	74,1

Найти эмпирическую функциональную зависимость $K = f(t)$.

Решение. На плоскости переменных t и K построим точки $M_i(t_i; K_i)$ и соединим их плавной кривой (рис. 24).

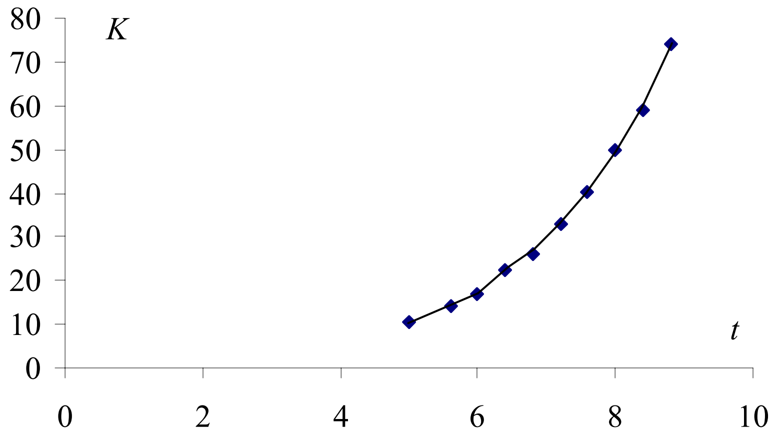


Рис. 24. Диаграмма исходных данных

По виду полученной диаграммы предполагаем, что для данного случая можно использовать зависимости $y = \dfrac{1}{ax+b}$ или $y = be^{ax}$.

Рассмотрим зависимость

$$K = \frac{1}{at+b}.$$

Используя преобразование $X = t$; $Y = \dfrac{1}{K}$, зависимость $K = \dfrac{1}{at+b}$ преобразуем в линейную: $Y = aX + b$.

Найдем значения новых переменных X и Y и занесем результаты расчетов в табл. 31.

Таблица 31

Преобразованные данные

$X = t$	5,0	5,6	6,0	6,4	6,8	7,2	7,6	8,0	8,4	8,8
$Y = \frac{1}{K}$	0,096	0,069	0,058	0,044	0,039	0,030	0,024	0,020	0,016	0,013

Построив на плоскости OXY точки $N_i(X_i;\ Y_i)$, $i = \overline{1, n}$ (рис. 25), видим, что они расположены вдоль некоторой кривой, а не прямой линии.

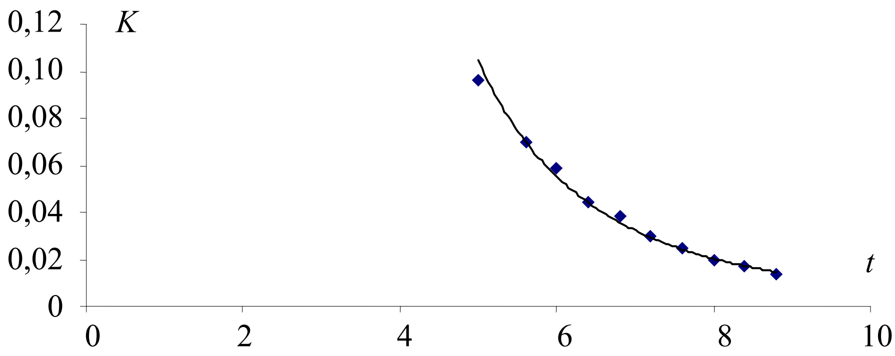


Рис. 25. Корреляционное поле для исходных данных

Предположим теперь, что зависимость описывается формулой $K = be^{at}$. Используя преобразование $Y = \ln K$; $X = t$, получим линейную зависимость $Y = \ln b + aX$.

Найдем значения новых переменных X и Y по формулам $X_i = t$; $Y_i = \ln K_i$ и запишем в табл. 32.

Таблица 32

Преобразованные данные

$X = t$	5,0	5,6	6,0	6,4	6,8	7,2	7,6	8,0	8,4	8,8
$Y = \ln K$	2,34	2,67	2,84	3,11	3,25	3,50	3,70	3,91	4,08	4,31

На плоскости OXY построим точки $P_i(X_i;\ Y_i)$, $i = \overline{1, n}$. Как видно из рис. 26, они расположены вдоль некоторой прямой линии, следовательно, выбранная зависимость лучше соответствует исходным данным.

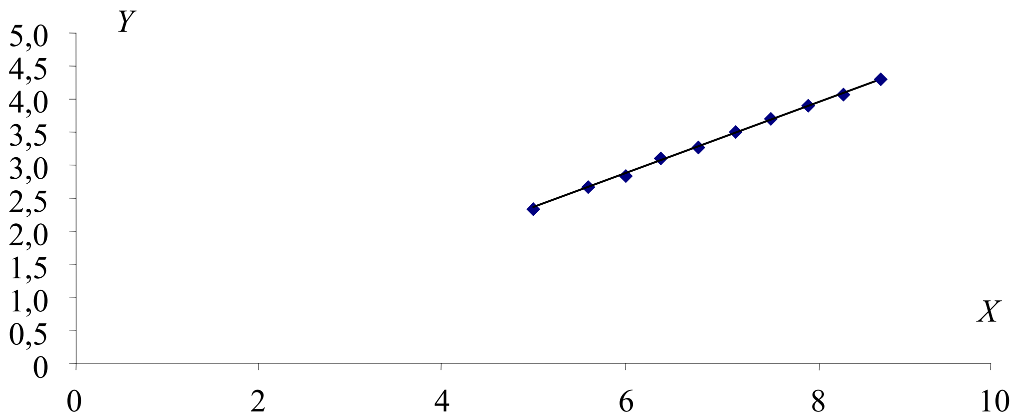


Рис. 26. Корреляционное поле для преобразованных данных

Параметры ln$b$ и a найдем с помощью МНК. Для вычисления коэффициентов системы составим табл. 33.

Таблица 33

Вспомогательные вычисления

t_i	K_i	$X_i = t_i$	$Y_i = \ln K_i$	X_i^2	X_iY_i
5,0	10,4	5,0	2,3418	25,00	11,709
5,6	14,4	5,6	2,6672	31,36	14,936
6,0	17,1	6,0	2,8391	36,00	17,034
6,4	22,5	6,4	3,1135	40,96	19,926
6,8	25,9	6,8	3,2542	46,24	22,129
7,2	33,1	7,2	3,4995	51,84	25,197
7,6	40,4	7,6	3,6988	57,76	28,111
8,0	50,0	8,0	3,9120	64,00	31,296
8,4	59,2	8,4	4,0809	70,56	34,280
8,8	74,1	8,8	4,3054	77,44	37,888
Сумма		69,8	33,7130	501,16	242,510

Составим нормальную систему уравнений:

$$\begin{cases} 10\ln b + 69{,}8a = 33{,}713, \\ 69{,}8\ln b + 501{,}16a = 242{,}51. \end{cases}$$

Решая ее, находим ln$b$ = –0,22702381 и a = 0,51547619. Отсюда получаем значение параметра b = 0,796901807. Таким образом, исходную зависимость можно описать функцией $K = 0{,}8e^{0{,}515t}$.

Эконометрический анализ при нарушении классических предположений. Основные проблемы при нарушении классических предположений. При моделировании реальных экономических процессов возникают ситуации, в которых условия классической модели регрессии оказываются нарушенными, а при их нарушении МНК может давать оценки с плохими статистическими свойствами.

1. Если имеется линейная связь экзогенных переменных, например $x_2 = b_0 + b_1 x_1$, то МНК-оценки не будут существовать. Такая ситуация в эконометрике носит название проблемы *мультиколлинеарности*.

2. Если нарушается гипотеза о взаимной независимости случайных отклонений: $M(\varepsilon_i \varepsilon_j) \neq 0$, то возникает проблема *автокорреляции*.

3. Одной из ключевых предпосылок МНК является *условие постоянства дисперсий случайных отклонений*. Выполнимость данной предпосылки называется *гомоскедастичностью*. Невыполнимость данной предпосылки называется *гетероскедастичностью* (непостоянством дисперсий отклонений).

Мультиколлинеарность. Если в модель включаются два или более тесно взаимосвязанных фактора, то наряду с уравнением изучаемой регрессии появляется и другая зависимость. *Мультиколлинеарность* – тесная зависимость между факторными признаками, включенными в модель. Она искажает величину коэффициентов регрессии и затрудняет их экономическую интерпретацию. Мультиколлинеарность возникает лишь в случае множественной регрессии.

В решении проблемы мультиколлинеарности можно выделить несколько этапов:

1. Установление наличия мультиколлинеарности.
2. Определение причин возникновения мультиколлинеарности.
3. Разработка мер по устранению мультиколлинеарности.

Способы определения наличия мультиколлинеарности:

– *анализ матрицы коэффициентов парной корреляции.* Факторы x_i и x_j могут быть признаны коллинеарными, если $\left|r_{x_i x_j}\right| > 0,8$;

– *исследование матрицы X'X.* Если определитель матрицы $X'X$ близок к нулю, то это свидетельствует о возможности наличия мультиколлинеарности;

– выявление статистически незначимых коэффициентов регрессии (т. е. имеющих низкие t-статистики) при достаточно высоком коэффициенте детерминации R^2;

Выделяют следующие методы устранения или уменьшения мультиколлинеарности:

1. *Сравнение значений линейных коэффициентов корреляции.* При отборе факторов предпочтение отдается тому, который более тесно, чем другие факторы, связан с результативным признаком, причем желательно, чтобы связь данного факторного признака с y была выше, чем его связь с другими факторными признаками.

2. *Метод включения факторов.* В модель включаются факторы по одному в определенной последовательности. После включения каждого фактора в модель рассчитывают ее характеристики и проверяют модель на достоверность.

3. *Метод исключения факторов.* В модель включаются все факторы. После построения уравнения регрессии из модели исключают фактор, коэффициент при котором незначим и имеет наименьшее значение t-критерия. Процесс исключения факторов продолжается до тех пор, пока все коэффициенты регрессии не будут значимы.

4. *Получение дополнительных данных или новой выборки.*

5. *Изменение спецификации модели.*

6. *Использование предварительной информации о некоторых параметрах.*

Автокорреляция. *Автокорреляция* (*последовательная корреляция*) определяется как корреляция между наблюдаемыми показателями, упорядоченными во времени (временные ряды) или в пространстве (перекрестные данные).

Автокорреляция остатков (отклонений) обычно встречается в регрессионном анализе при использовании данных временных рядов.

Методы определения автокорреляции:

1. *Графический метод.* По оси абсцисс откладываются либо время (момент) получения статистических данных, либо порядковый номер наблюдения, а по оси ординат – отклонения (или оценки отклонений). По графику предполагают, имеются ли определенные связи между отклонениями, т. е. автокорреляция. Отсутствие зависимости, скорее всего, свидетельствует об отсутствии автокорреляции. Данный график можно также дополнить графиком зависимости e_t от e_{t-1}.

2. *Тест Дарбина – Уотсона.*

Гетероскедастичность. Одной из ключевых предпосылок МНК является *условие постоянства дисперсий случайных отклонений*. Выполнимость данной предпосылки называется *гомоскедастичностью*, а невыполнимость – *гетероскедастичностью* (непостоянством дисперсий отклонений).

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

Не существует однозначного метода определения гетероскедастичности. Однако для проверки разработано много тестов и критериев. Наиболее популярные и наглядные: графический анализ отклонений, тест ранговой корреляции Спирмена, тест Парка, тест Глейзера, тест Голдфелда – Квандта.

Использование графического представления отклонений позволяет определиться с наличием гетероскедастичности. В этом случае по оси абсцисс откладываются значения объясняющей переменной X (либо линейной комбинации объясняющих переменных), а по оси ординат – либо отклонения, либо их квадраты.

Если все отклонения находятся внутри полосы постоянной ширины, параллельной оси абсцисс, то это говорит о независимости дисперсий от значений переменной X и их постоянстве, т. е. в этом случае выполняются условия гомоскедастичности.

Если наблюдаются некоторые систематические изменения в соотношениях между значениями переменной X и квадратами отклонений (линейная, квадратичная, гиперболическая и другие зависимости), то такие ситуации отражают большую вероятность наличия гетероскедастичности для рассматриваемых статистических данных.

Временные ряды. Для характеристики и анализа различных социально-экономических явлений за определенный период применяют показатели и методы, характеризующие эти процессы во времени (динамике). Под *временным рядом* в экономике понимается последовательность наблюдений некоторого признака (случайной величины) Y в последовательные моменты времени. Отдельные наблюдения, которые называются *уровнями* ряда, будем обозначать y_t, $t = \overline{1, n}$ (где n – число уровней). Последовательно расположенные во времени числовые показатели характеризуют уровень состояния и изменения явления или процесса.

Классификация временных рядов:

1. *В зависимости от показателя времени* временные ряды бывают *моментные* (на определенную дату) и *интервальные* (за определенный период).

2. *По форме представления* уровни во временном ряду могут быть представлены *абсолютными*, *средними* и *относительными* величинами.

3. *По расстоянию между уровнями* временные ряды подразделяются на ряды с *равноотстоящими* и *неравноотстоящими* уровнями по времени. В равноотстоящих рядах даты регистрации периода следуют друг за другом с равными интервалами, в неравноотстоящих рядах равные интервалы не соблюдаются.

4. *По содержанию* временные ряды подразделяют на состоящие из *частных* и *агрегированных* показателей. Частные показатели характеризуют явления изолированно, односторонне (например, динамика показателей среднесуточного объема потребленной воды). Агрегированные показатели являются производными от частных показателей и характеризуют изучаемое явление комплексно (например, динамика показателей экономической конъюнктуры).

Каждый уровень временного ряда формируется из трендовой, циклической, сезонной и случайной компонент. В случае относительно коротких временных рядов (например, 3–5 лет) циклическая компонента отдельно не выделяется, а происходит объединение циклической компоненты с трендом.

В общем виде модель экономического временного ряда может быть представлена следующим образом:

$$y_t = u_t + v_t + \varepsilon_t,$$

где u_t – *тренд*, плавно меняющаяся компонента, описывающая чистое влияние долговременных факторов, т. е. длительную тенденцию изменения признака (например, рост населения, экономическое развитие, изменение структуры потребления и т. п.); v_t – *сезонная компонента*, отражающая повторяемость экономических процессов в течение не очень длительного периода: года, иногда месяца, недели и т. д. (например, объем продаж товаров или перевозок пассажиров в различные времена года); ε_t – *случайная компонента*, отражающая влияние не поддающихся учету и регистрации случайных факторов.

Следует обратить внимание на то, что в отличие от компоненты ε_t составляющие u_t, v_t являются закономерными, неслучайными.

Модели, в которых временной ряд представлен как сумма перечисленных компонент, называются аддитивными; как произведение – мультипликативными моделями временного ряда.

Аддитивная модель имеет вид: $y_t = u_t + v_t + \varepsilon_t$.

Мультипликативная модель имеет вид: $y_t = u_t \cdot v_t \cdot \varepsilon_t$. Такую модель применяют в случае, если происходят существенные сезонные изменения.

Среди наиболее распространенных методов анализа временных рядов выделим корреляционный анализ, модели авторегрессии и скользящей средней.

Важную роль в анализе временных рядов играют стационарные временные ряды, вероятностные свойства которых не изменяются во времени. Стационарные временные ряды применяются, в частности, при описании случайных составляющих анализируемых рядов. Временной ряд y_t называется *стационарным*, если совместное распределение вероятностей p наблюдений $y_1, y_2, \ldots, y_p$ такое же, как и распределение вероятностей p наблюдений $y_{1+\tau}, y_{2+\tau}, \ldots, y_{p+\tau}$ при любых p и τ. Иначе говоря, свойства стационарных рядов y_t не зависят от момента t, т. е. закон распределения и его числовые характеристики не зависят от t. Поэтому математическое ожидание и среднее квадратическое отклонение могут быть оценены по наблюдениям y_t, $t=\overline{1,\,n}$, по формулам:

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n}; \quad s^2 = \frac{\sum_{t=1}^{n} (y_t - \bar{y})^2}{n}.$$

Степень тесноты связи между компонентами временного ряда y_t (сдвинутых относительно друг друга на τ единиц, или, как говорят, с *лагом* τ) может быть определена с помощью коэффициента корреляции.

Системы одновременных уравнений. Для изучения комплексных экономических явлений средствами эконометрики, как правило, применяют не отдельные уравнения регрессии, а системы уравнений.

Виды систем эконометрических уравнений:

1. *Система независимых уравнений.* Каждый результативный признак (объясняемая переменная) $y_j,\ j=\overline{1,\ n},$ является функцией одной и той же совокупности факторов (объясняющих переменных) $x_i,\ i=\overline{1,\ m}.$ Набор факторов в каждом уравнении системы может варьироваться в зависимости от изучаемого явления.

2. *Система рекурсивных уравнений.* Результативный признак y_j, $j=\overline{1,\ n},$ одного уравнения системы в каждом последующем уравнении является фактором наряду с одной и той же совокупностью факторов $x_i,\ i=\overline{1,\ m}.$

3. *Система одновременных уравнений.* Результативный признак $y_j,\ j=\overline{1,\ n},$ одного уравнения системы входит во все другие уравнения системы в качестве фактора наряду с одной и той же совокупностью факторов $x_i,\ i=\overline{1,\ m}.$ Такие системы эффективны в эконометрических исследованиях и наиболее широко применяются в макроэкономике.

Параметры системы независимых или рекурсивных уравнений определяют с помощью МНК. Для исследования системы одновременных уравнений требуются другие, отличные от МНК методы.

Система одновременных уравнений может быть представлена:

• в виде структурной формы модели;

• в виде приведенной формы модели.

Основными составляющими обеих форм записи являются эндогенные и экзогенные переменные. *Эндогенные переменные* (y) определяются внутри модели и являются зависимыми переменными. *Экзогенные переменные* (x) определяются вне системы и являются независимыми переменными. Предполагается, что экзогенные переменные не коррелируют с ошибкой в соответствующем уравнении. Под *предопределенными переменными* системы одновременных уравнений понимают экзогенные и лаговые (за предыдущие моменты времени) эндогенные переменные этой системы.

Структурная форма модели имеет вид:

$$\begin{cases} y_1 = c_{10} + b_{12}y_2 + b_{13}y_3 + \ldots + b_{1n}y_n + a_{11}x_1 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = c_{20} + b_{21}y_1 + b_{23}y_3 + \ldots + b_{2n}y_n + a_{21}x_1 + \ldots + a_{2m}x_m + \varepsilon_2, \\ \ldots \\ y_n = c_{n0} + b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + \ldots + a_{nm}x_m + \varepsilon_n, \end{cases}$$

где c_{i0}, $i=\overline{1,\ n}$, – свободный член уравнения модели; b_{ij}, $i=\overline{1,\ n}$; $j=\overline{1,\ m}$, – коэффициент при эндогенной переменной модели; a_{ij}, $i=\overline{1,\ n}$; $j=\overline{1,\ m}$, – коэффициент при экзогенной переменной; ε_i, $i=\overline{1,\ n}$, – случайная составляющая (ошибка) i-го уравнения структурной формы модели.

Наряду с регрессионными уравнениями в модели могут быть записаны и тождества.

Таким образом, структурные уравнения модели подразделяются на два класса:

1. *Поведенческие уравнения.* Описывают взаимодействие между экзогенными и эндогенными переменными.

2. *Тождества.* Устанавливают соотношения между эндогенными переменными, не содержат случайных составляющих и структурных коэффициентов модели.

Структурная форма модели может быть преобразована в приведенную форму:

$$\begin{cases} y_1 = \alpha_{10} + \alpha_{11}x_1 + \ldots + \alpha_{1m}x_m + \eta_1, \\ y_2 = \alpha_{20} + \alpha_{21}x_1 + \ldots + \alpha_{2m}x_m + \eta_2, \\ \ldots \\ y_n = \alpha_{n0} + \alpha_{n1}x_1 + \ldots + \alpha_{nm}x_m + \eta_n, \end{cases}$$

где α_{i0}, $i=\overline{1,\ n}$, – свободный член уравнения модели; α_{ij}, $i=\overline{1,\ n}$; $j=\overline{1,\ m}$, – коэффициент при предопределенной переменной, являющийся функцией коэффициентов структурной формы модели; η_i, $i=\overline{1,\ n}$, – случайная составляющая (ошибка) i-го уравнения приведенной формы модели.

Идентификация – это установление соответствия между приведенной и структурной формами модели. Единственность соответствия между приведенной и структурной формами модели составляет задачу идентификации.

Классы структурных моделей с точки зрения задачи идентификации:

1. *Идентифицируемая.* Все структурные коэффициенты однозначно определяются через приведенные коэффициенты.

2. *Неидентифицируемая.* Структурные коэффициенты невозможно найти по приведенным коэффициентам.

3. *Сверхидентифицируемая.* Структурные коэффициенты, выраженные через приведенные коэффициенты, имеют два и более числовых значения.

Необходимое условие идентифицируемости уравнений системы: если уравнение модели идентифицируемо, то количество эндогенных переменных (n) этого уравнения на единицу больше количества предопределенных переменных (p) системы, не входящих в данное уравнение: $n = p + 1$. Если $n < p + 1$, то уравнение сверхидентифицируемо; если $n > p + 1$, то уравнение неидентифицируемо.

Достаточное условие идентифицируемости уравнений системы: если определитель (Δ^*) матрицы коэффициентов (A) при переменных системы, не входящих в данное уравнение, не равен нулю ($\Delta^* \neq 0$) и количество эндогенных переменных системы без единицы равно рангу этой матрицы ($\operatorname{rank} A = n - 1$), то уравнение модели идентифицируемо.

Проверка структурной модели на идентифицируемость позволяет установить степень возможности оценки коэффициентов структурных уравнений по коэффициентам приведенных уравнений.

Модель идентифицируема тогда и только тогда, когда идентифицируемо каждое ее уравнение. Идентификация не применяется для тождеств модели.

Для получения качественных оценок параметров системы одновременных уравнений пользуются косвенным МНК, алгоритм которого следующий:

1. Структурная форма модели преобразуется в приведенную форму.

2. С помощью МНК оцениваются параметры приведенной формы.

3. Приведенная форма преобразуется в структурную форму.

Область применения косвенного МНК ограничивается идентифицируемыми системами одновременных уравнений.

Пример 5. Проверить, идентифицируемы ли уравнения (1) и (2) модели предложения и спроса кейнсианского типа:

$$\begin{cases} y_{1t} = b_{13}y_{3t} + a_{11}x_{1t} + c_{10} + \varepsilon_1, & (1) \\ y_{2t} = b_{23}y_{3t} + a_{23}y_{3,t-1} + c_{20} + \varepsilon_2, & (2) \\ y_{1t} = y_{2t}, & (3) \end{cases}$$

где y_{1t} – спрос на товар в момент времени t; y_{2t} – предложение товара в момент t; y_{3t} – цена товара в момент t; $y_{3,t-1}$ – цена товара в момент $(t-1)$; x_{1t} – доход в момент t; t – текущий период; $(t-1)$ – предыдущий период.

Решение. Запишем систему в виде:

$$\begin{cases} 0 = -y_{1t} + b_{13}y_{3t} + a_{11}x_{1t} + c_{10} + \varepsilon_1, & (1) \\ 0 = -y_{2t} + b_{23}y_{3t} + a_{23}y_{3,t-1} + c_{20} + \varepsilon_2, & (2) \\ 0 = -y_{1t} + y_{2t}. & (3) \end{cases}$$

Запишем коэффициенты последней системы в виде табл. 34.

Таблица 34

Матрица коэффициентов

Уравнения	Переменные				
	эндогенные			предопределенные	
	y_{1t}	y_{2t}	y_{3t}	$y_{3,t-1}$	x_{1t}
1	–1	0	b_{13}	0	a_{11}
2	0	–1	b_{23}	a_{23}	0
3	–1	1	0	0	0

Уравнение (1):

а) необходимое условие: эндогенных переменных две (y_{1t}, y_{3t}), отсутствующих экзогенных – одна ($y_{3,t-1}$). Таким образом, $n = 2$, $p = 1$ и выполняется необходимое условие идентификации ($n = p + 1$): $2 = 1 + 1$;

б) достаточное условие. В первом уравнении отсутствуют переменные y_{2t} и $y_{3,t-1}$. Запишем матрицу коэффициентов при этих переменных в других уравнениях системы (табл. 35).

Таблица 35

Матрица коэффициентов при отсутствующих переменных

Уравнения	Переменные	
	эндогенные	предопределенные
	y_{2t}	$y_{3,t-1}$
2	–1	a_{23}
3	1	0

Матрица коэффициентов при переменных системы, не входящих в уравнение, $A = \begin{bmatrix} -1 & a_{23} \\ 1 & 0 \end{bmatrix}$. Ранг этой матрицы rank$A$ = 2 (равен количеству эндогенных переменных модели минус один). Причем $\Delta^* = |A| = -1 \cdot 0 - 1 \cdot a_{23} = -a_{23} \neq 0$. Достаточное условие идентифицируемости также выполняется. Следовательно, можно сделать вывод, что уравнение (1) идентифицируемо.

Уравнение (2):

а) $n = 2$, $p = 1$. Выполняется необходимое условие идентификации ($n = p + 1$): 2 = 1 + 1;

б) матрица коэффициентов при переменных системы, не входящих в уравнение, $A = \begin{bmatrix} -1 & a_{11} \\ -1 & 0 \end{bmatrix}$. Ранг этой матрицы $\operatorname{rank} A = 2$ (равен количеству эндогенных переменных модели минус один). Причем $\Delta^* = |A| = a_{11} \neq 0$. Достаточное условие идентифицируемости также выполняется. Следовательно, можно сделать вывод, что уравнение (2) идентифицируемо.

Вопросы для самоконтроля

Основные понятия эконометрики

1. Перечислите задачи эконометрики.

2. Что понимается под эконометрическим моделированием?

3. Определите, к какому классу относится каждая из следующих моделей:

а) $Y = f(X_1, X_2, \ldots, X_k) + \varepsilon$;

б) $y_t = u_t + v_t + \varepsilon_t$;

в) $\begin{cases} y_{1t} = c_{10} + b_{13} y_{3t} + \varepsilon_1, \\ y_{2t} = c_{20} + b_{23} y_{3t} + \varepsilon_2, \\ y_{1t} = y_{2t}. \end{cases}$

4. Какие типы данных используются в эконометрическом моделировании?

Элементы корреляционно-регрессионного анализа

1. В чем заключается корреляционный анализ?

2. В чем заключается регрессионный анализ?

3. При каком значении линейного коэффициента корреляции связь между признаками можно считать тесной?

4. Какой критерий используют для оценки значимости коэффициента корреляции?

5. Чему равен коэффициент детерминации в модели вида $y = b_0 + b_1x$, если коэффициент корреляции между признаками x и y принимает значение 0,687?

6. Какой коэффициент определяет среднее изменение результативного признака при изменении факторного признака на 1%?

7. Какой критерий используют для оценки значимости уравнения регрессии?

8. Какие виды нелинейных эконометрических моделей Вы знаете?

Эконометрический анализ при нарушении классических предположений. Временные ряды

1. Объясните значение термина «мультиколлинеарность».
2. Каковы основные последствия мультиколлинеарности?
3. Как можно обнаружить мультиколлинеарность?
4. Как оценивается коррелированность между двумя объясняющими переменными?
5. Перечислите основные методы устранения мультиколлинеарности.
6. Что такое автокорреляция?
7. Перечислите методы обнаружения автокорреляции.
8. В чем суть гетероскедастичности?
9. Что называется временным рядом?
10. Какие составляющие временного ряда Вы знаете?
11. Какой временной ряд называется стационарным?

Системы одновременных уравнений

1. Перечислите виды систем эконометрических уравнений.
2. Какие виды переменных встречаются в системах одновременных уравнений?
3. Какой вид имеет структурная форма модели?
4. Какой вид имеет приведенная форма модели?
5. Перечислите классы структурных уравнений модели. Объясните, в чем их отличие.
6. Что такое идентификация модели?
7. Какие структурные модели выделяются с точки зрения задачи идентификации?
8. Сформулируйте необходимое условие идентифицируемости уравнений системы.

9. Сформулируйте достаточное условие идентифицируемости уравнений системы.

10. Опишите алгоритм косвенного метода наименьших квадратов.

5.2. ПРАКТИЧЕСКИЙ МИНИМУМ

Примеры решения задач

Элементы корреляционно-регрессионного анализа

Пример 6. Для анализа зависимости инвестиций (y) предприятия от объемов (x) производства исследуются данные 12 однотипных предприятий (табл. 36).

Таблица 36

Зависимость инвестиций предприятия от объемов производства

Показатель	Предприятие											
	1	2	3	4	5	6	7	8	9	10	11	12
Инвестиции y_i, тыс. ден. ед.	12	15	18	20	22	17	19	25	31	30	40	44
Объем производства x_i, млрд. шт.	17	19	20	21	28	30	31	38	42	48	45	50

Необходимо:

1) построить корреляционное поле;

2) оценить тесноту и направление связи между переменными с помощью коэффициента корреляции;

3) оценить значимость полученного коэффициента корреляции по критерию Стьюдента (уровень значимости $\alpha = 0{,}1$);

4) вычислить коэффициент детерминации;

5) найти уравнение регрессии Y по X;

6) вычислить коэффициент эластичности;

7) проверить адекватность полученной модели по критерию Фишера;

8) сделать выводы.

Решение.

Пункт 1) см. пример 1.

Пункты 2)–5) см. пример 2.

Пункты 6), 7) см. пример 3.

Отметим, что уравнение отражает лишь общую тенденцию в поведении рассматриваемых переменных. Индивидуальные значения переменных в силу различных причин могут отклоняться от модельных значений. Однако при определенных условиях уравнение регрессии служит основным качественным инструментом анализа и прогнозирования.

Пример 7. Для трех видов продукции (*A*, *B* и *C*) модели зависимости удельных постоянных расходов от объема выпускаемой продукции выглядят следующим образом: $y_A = 600$; $y_B = 80 + 0,7x$; $y_C = 40x^{0,5}$.

1. Определите коэффициенты эластичности по каждому виду продукции и поясните их смысл.
2. Сравните эластичность затрат для продукции *B* и *C* при $x = 100$.
3. Определите, каким должен быть объем выпускаемой продукции, чтобы коэффициенты эластичности для продукции *B* и *C* были равны.

Решение.

1. Коэффициент эластичности определяется по формуле $Э_{yx} = f'(x)\dfrac{x}{y}$. Поскольку $y'_A = 600' = 0$; $y'_B = (80)' + (0,7x)' = 0,7$; $y'_C = (40x^{0,5})' = 20x^{-0,5}$, то коэффициент эластичности $Э_A = 0$; $Э_B = 0,7x/(80 + 0,7x)$; $Э_C = 0,5$. Коэффициент эластичности показывает, что с изменением объема выпуска продукции на 1% удельные постоянные расходы для продукции *A* не изменяются, для продукции *B* изменяются на $0,7x/(80 + 0,7x)\%$ (в зависимости от объема выпускаемой продукции *x*); для продукции *C* в среднем изменяются на 0,5%.

2. При $x = 100$ эластичность затрат для продукции *B* равна: $Э_B = 0,7 \cdot 100/(80 + 0,7 \cdot 100) = 70/150 = 7/15 \approx 0,467$; для продукции *C* равна 0,5, следовательно, эластичность затрат для продукции *B* ниже, чем для продукции *C*.

3. Чтобы коэффициенты эластичности для продукции *B* и *C* были равны, должно выполняться равенство $0,7x/(80 + 0,7x) = 0,5$, следовательно, объем выпускаемой продукции $x = 800/7$.

Эконометрический анализ при нарушении классических предположений

Пример 8. Зависимость уровня рентабельности *y* (ед.) от производительности труда x_1 (тыс. ден. ед.), продолжительности оборота оборотных средств предприятия x_2 (дней) и материалоотдачи x_3 (тыс. ден. ед.) приведена в табл. 37.

Таблица 37

Зависимость уровня рентабельности от производительности труда, продолжительности оборота оборотных средств предприятия и материалоотдачи

y, ед.	x_1, тыс. ден. ед.	x_2, дней	x_3, тыс. ден. ед.
21	7,9	18	66
23	10,0	19	88
20	7,9	17	55
21	8,5	18	59
23	9,0	21	71
21	7,8	17	62
22	7,7	20	33
20	8,0	17	59
22	8,3	18	60
21	7,0	17	39

Найти парные коэффициенты корреляции и составить корреляционную матрицу. По полученным данным сделать вывод о тесноте связи между рассматриваемыми переменными. Проверить значимость коэффициентов корреляции и проанализировать полученные данные. Выяснить, имеет ли место мультиколлинеарность, и устранить ее при необходимости.

Решение. Найдем средние значения: $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 21,4;$ $\overline{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{i1} = 8,21;$ $\overline{x}_2 = \frac{1}{n}\sum_{i=1}^{n} x_{i2} = 18,2;$ $\overline{x}_3 = \frac{1}{n}\sum_{i=1}^{n} x_{i3} = 592,$ $n = 10$, а также среднеквадратические отклонения: $\sigma_y = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}{n}} = 1,019;$ $\sigma_{x_1} = 0,777;$ $\sigma_{x_2} = 1,327;$ $\sigma_{x_3} = 14,613.$

Найдем коэффициенты корреляции по следующей формуле: $r_{x_i y} = \frac{\sum_{j=1}^{n}(x_{ij} - \overline{x})(y_j - \overline{y})}{n\sigma_x\sigma_y}$. Коэффициенты $r_{yx_1} = 0,663;$ $r_{yx_2} = 0,828;$ $r_{yx_3} = 0,404$ показывают связь между результативным признаком y

и факторами x_1, x_2, x_3 соответственно; а коэффициенты $r_{x_1x_2} = 0{,}525$; $r_{x_1x_3} = 0{,}851$; $r_{x_2x_3} = 0{,}173$ показывают связь между факторными признаками.

Тогда корреляционная матрица имеет вид:

$$K = \begin{bmatrix} 1 & 0{,}663 & 0{,}828 & 0{,}404 \\ 0{,}663 & 1 & 0{,}512 & 0{,}851 \\ 0{,}828 & 0{,}512 & 1 & 0{,}173 \\ 0{,}404 & 0{,}851 & 0{,}173 & 1 \end{bmatrix}.$$

Так как $r_{x_1x_2} = 0{,}525 < 0{,}8$ и $r_{x_2x_3} = 0{,}173 < 0{,}8$, то связь между факторами x_1 и x_2, а также между x_2 и x_3 достаточно слабая, и эти факторы можно включить в модель; $r_{x_1x_3} = 0{,}851 > 0{,}8$, поэтому связь между факторами x_1 и x_3 достаточно сильная и есть основания полагать, что имеет место мультиколлинеарность.

Проверим значимость коэффициентов корреляции по критерию Стьюдента. При уровне значимости $\alpha = 0{,}1$ и с учетом того, что в примере количество степеней свободы равно: $\gamma = n - k - 1 = 10 - 1 - 1 = 8$, получим табличное значение критерия: $t_{\text{табл}} = 1{,}86$. Теперь вычислим фактические значения. Для переменной x_1 расчетное значение t равно 3,208, для переменной x_2 – 7,96, а для переменной x_3 – 1,448.

Поскольку фактическое значение t в первых двух случаях выше табличного, то связь между результативным и факторными показателями x_1 и x_2 является надежной, а величина коэффициентов корреляции – значимой. Про фактор x_3 можно сказать, что его следует исключить из модели, т. к. имеет место тесная связь между факторами x_1, x_3 и коэффициент корреляции значимым не является. В результате исключения фактора x_3 из рассмотрения мультиколлинеарность устраняется.

Пример 9. По данным примера 8 найти линейное уравнение связи (регрессии), считая, что между результативным и факторными признаками имеет место линейная связь. Для полученной линейной модели определить коэффициенты эластичности. Сделать выводы.

Решение. Найдем уравнение регрессии. Используя метод наименьших квадратов, получим систему уравнений, которая в матричном виде (более удобном для расчетов в нашем случае) может быть представлена следующим образом:

$$(X'X)B = X'Y,$$

где $X'X = \begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i2}x_{i1} & \sum x_{i2}^2 \end{bmatrix}$; $X'Y = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{bmatrix}$; $B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}$.

Составим следующие матрицы:

$$X'X = \begin{bmatrix} 10 & 82{,}1 & 182 \\ 82{,}1 & 680{,}09 & 1499{,}5 \\ 182 & 1499{,}5 & 3330 \end{bmatrix}; \ X'Y = \begin{bmatrix} 214 \\ 1762{,}2 \\ 3906 \end{bmatrix}.$$

Тогда по формуле $B = (X'X)^{-1} \cdot (X'Y)$ найдем вектор коэффициентов регрессии. Сначала вычислим обратную матрицу:

$$(X'X)^{-1} = \begin{bmatrix} 20{,}614 & -0{,}616 & -0{,}849 \\ -0{,}616 & 0{,}224 & -0{,}067 \\ -0{,}849 & -0{,}067 & 0{,}077 \end{bmatrix},$$

а затем вектор B:

$$B = (X'X)^{-1} \cdot (X'Y) = \begin{bmatrix} 8{,}648 \\ 0{,}426 \\ 0{,}509 \end{bmatrix}.$$

Запишем уравнение регрессии: $\hat{y} = 8{,}648 + 0{,}426x_1 + 0{,}509x_2$.

Это уравнение выражает зависимость уровня рентабельности от производительности труда и продолжительности оборота оборотных средств. Коэффициенты уравнения показывают количественное воздействие каждого фактора на результативный показатель при неизменности других. В примере рентабельность в среднем повышается на 0,426 ед. при увеличении производительности труда на 1 тыс. ден. ед. и на 0,509 ед. – при увеличении продолжительности оборота оборотных средств на 1 день.

Коэффициенты эластичности найдем по формуле $Э_i = b_i \dfrac{\overline{x_i}}{\overline{y}}$.

Для переменной x_1 коэффициент эластичности равен 0,163, а для переменной x_2 – 0,433.

Согласно полученным данным, рентабельность возрастает на 0,16% при увеличении производительности труда на 1%; на 0,43% – при увеличении продолжительности оборота оборотных средств на 1%.

Пример 10. Проверить адекватность полученной в примере 9 модели по критерию Фишера и определить среднюю ошибку аппроксимации. Уровень значимости $\alpha = 0{,}1$.

Решение. Для того чтобы убедиться в надежности уравнения связи и правомерности его использования для практических целей, необходимо дать статистическую оценку надежности показателей связи.

По корреляционной матрице

$$K = \begin{bmatrix} 1 & 0{,}663 & 0{,}828 \\ 0{,}663 & 1 & 0{,}512 \\ 0{,}828 & 0{,}512 & 1 \end{bmatrix}$$

найдем коэффициент множественной корреляции

$$R = \sqrt{1 - \frac{\det K}{K_{11}}},$$

где $\det K$ – определитель корреляционной матрицы; K_{11} – алгебраическое дополнение элемента первой строки и первого столбца матрицы K.

Фактическое значение критерия $F_{\text{расч}} = \dfrac{R^2}{1-R^2}\dfrac{(n-k-1)}{k} = 11{,}27.$

При уровне значимости $\alpha = 0{,}1$ и с учетом того, что в примере количество степеней свободы равно: $\gamma_1 = k = 2$ и $\gamma_2 = n - k - 1 = 10 - 2 - 1 = 7$, получим табличное значение критерия: $F_{\text{табл}} = 3{,}26$. Так как $F_{\text{расч}} \geq F_{\text{табл}}$, то построенную модель можно считать адекватной.

Найдем среднюю относительную ошибку аппроксимации:

$$E_{\text{отн}} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_i}{y_i}\right| \cdot 100\% = 2{,}09\%.$$

Заметим, что $\hat{y}_i$ вычисляется следующим образом: $\hat{y}_i = 8{,}648 + 0{,}426x_{i1} + 0{,}509x_{i2}$, где данные x_{i1} и x_{i2}, $i = \overline{1,\ 10}$, берутся из условия задачи.

Средняя ошибка мала, что также свидетельствует об адекватности модели.

Следовательно, данное уравнение можно использовать для различных практических целей: а) оценки результатов хозяйственной деятельности; б) расчета влияния факторов на прирост результативного показателя; в) подсчета резервов повышения уровня исследуемого показателя; г) планирования и прогнозирования его величины.

Системы одновременных уравнений

Пример 11. Проверить, идентифицируема ли модель предложения и спроса кейнсианского типа. Выписать приведенную форму модели. Указать метод оценки параметров модели:

$$\begin{cases} y_{1t} = b_{13} y_{3t} + a_{11} x_{1t} + c_{10} + \varepsilon_1, \\ y_{2t} = b_{23} y_{3t} + a_{23} y_{3,t-1} + c_{20} + \varepsilon_2, \\ y_{1t} = y_{2t}, \end{cases}$$

где y_{1t} – спрос на товар в момент времени t; y_{2t} – предложение товара в момент t; y_{3t} – цена товара в момент t; $y_{3,t-1}$ – цена товара в момент $(t - 1)$; x_{1t} – доход в момент t; t – текущий период; $(t - 1)$ – предыдущий период.

Решение. См. пример 5.

Таким образом, система одновременных уравнений идентифицируемая в силу идентифицируемости первого и второго уравнений. Для оценки параметров системы можно применять как косвенный, так и двухшаговый МНК.

Приведенная форма модели имеет вид:

$$\begin{cases} y_{1t} = \alpha_{10} + \alpha_{11} x_{1t} + \alpha_{12} y_{3,t-1} + \eta_1, \\ y_{2t} = \alpha_{20} + \alpha_{21} x_{1t} + \alpha_{22} y_{3,t-1} + \eta_2, \\ y_{3t} = \alpha_{30} + \alpha_{31} x_{1t} + \alpha_{32} y_{3,t-1} + \eta_3. \end{cases}$$

Используя соответствующие статистические данные, с помощью косвенного МНК можно найти несмещенные и состоятельные оценки структурной формы, смоделировав тем самым реальную экономическую ситуацию изучения спроса-предложения с учетом дохода в текущий период и цены товара в предыдущий период.

Отметим, что каждое уравнение системы оценивают тогда и только тогда, когда установлена его идентифицируемость. Идентификация не применяется для тождеств модели.

Пример 12. Предложение денег и спрос на деньги представлены в виде модели:

$$\begin{cases} y_{1t} = c_{10} + b_{12} y_{2t} + a_{11} x_{1t} + \varepsilon_1, & (1) \\ y_{2t} = c_{20} + b_{21} y_{1t} + \varepsilon_2, & (2) \end{cases}$$

где y_{1t} – процентные ставки в период t; y_{2t} – ВВП в период t; x_{1t} – денежная масса в период t.

Проверить, идентифицируемы ли уравнения.

Решение. Запишем коэффициенты системы в виде табл. 38.

Таблица 38

Матрица коэффициентов

Уравнения	Переменные		
	эндогенные		предопределенные
	y_{1t}	y_{2t}	x_{1t}
1	–1	b_{12}	a_{11}
2	b_{21}	–1	0

Уравнение (1):

Необходимое условие: $n = 2$ (y_{1t}, y_{2t}), $p = 0$ (т. к. x_{1t} является единственной предопределенной переменной, которая входит в первое уравнение системы), следовательно, $n > p + 1$.

Уравнение неидентифицируемо, поэтому неидентифицируема вся система.

В этом случае модель изменяют так, чтобы она, с одной стороны, содержала основные эндогенные и экзогенные переменные, которые определяют предложение и спрос на деньги, а с другой – была эконометрически разрешима.

Пример 13. Оценить параметры идентифицируемой структурной модели:

$$\begin{cases} y_1 = c_{10} + b_{12} y_2 + a_{11} x_1 + \varepsilon_1, \\ y_2 = c_{20} + b_{21} y_1 + a_{22} x_2 + \varepsilon_2, \end{cases}$$

где y_1 и y_2 – эндогенные переменные системы; x_1 и x_2 – экзогенные переменные этой системы.

Исходные данные приведены в табл. 39.

Таблица 39

Исходные данные

y_1	y_2	x_1	x_2
5	2	7	2
4	3	8	2
3	1	6	4
2	5	4	6
7	4	2	4
8	3	7	8
9	6	4	9
10	8	5	7

Решение.

1. От структурной формы перейдем к приведенной форме модели:

$$\begin{cases} y_1 = \alpha_{10} + \alpha_{11}x_1 + \alpha_{12}x_2 + \eta_1, \\ y_2 = \alpha_{20} + \alpha_{21}x_1 + \alpha_{22}x_2 + \eta_2. \end{cases}$$

2. Применим МНК для оценки коэффициентов уравнений модели.

Оценки коэффициентов можно найти, составив систему нормальных уравнений и решив ее. Можно также использовать электронные таблицы Excel (функция ЛИНЕЙН или команды меню Сервис $\to$ Анализ данных $\to$ Регрессия).

Рассмотрим сначала первое уравнение системы. Нам нужно определить коэффициенты α_{10}, α_{11}, α_{12}. Для этого используем данные табл. 39 (столбцы y_1, x_1, x_2).

Получаем уравнение: $y_1 = 2,64 - 0,02x_1 + 0,66x_2 + \eta_1$.

Аналогичным образом определяются коэффициенты второго уравнения (используются данные столбцов y_2, x_1, x_2).

Получаем уравнение: $y_2 = 3,51 - 0,34x_1 + 0,44x_2 + \eta_2$.

Запишем приведенную форму модели:

$$\begin{cases} y_1 = 2,64 - 0,02x_1 + 0,66x_2 + \eta_1, & (*) \\ y_2 = 3,51 - 0,34x_1 + 0,44x_2 + \eta_2. & (**) \end{cases}$$

3. От приведенной формы переходим к структурной форме модели. Из уравнения (*) выражаем $x_1 = (2,64 - y_1 + 0,66x_2)/0,02$ и подставляем правую часть этого равенства в уравнение (**):

$$y_2 = 3,51 - 0,34 \cdot \frac{2,64 - y_1 + 0,66x_2}{0,02} + 0,44x_2,$$

или

$$y_2 = -41,37 + 17y_1 - 10,78x_2 + \varepsilon_2.$$

Из уравнения (**) выражаем $x_2 = (y_2 - 3,51 + 0,34x_1)/0,44$ и подставляем в правую часть уравнения (*). После преобразований получаем:

$$y_1 = -2,63 + 1,5y_2 + 0,49x_1 + \varepsilon_1.$$

Таким образом, структурная форма модели имеет вид:

$$\begin{cases} y_1 = -2,63 + 1,5y_2 + 0,49x_1 + \varepsilon_1, \\ y_2 = -41,37 + 17y_1 - 10,78x_2 + \varepsilon_2. \end{cases}$$

5.3. МИНИМУМ ДЛЯ АУДИТОРНОЙ РАБОТЫ

Занятие 1: 3А1; 3А2; 3А3.
Занятие 2: 3А9; 3А11.
Занятие 3: 3А19; 3А20.

6. ПРИЛОЖЕНИЯ

6.1. ТРЕНИРОВОЧНАЯ КОНТРОЛЬНАЯ РАБОТА

Вариант 1

1А. Уравнение множественной регрессии имеет вид: $y = -27{,}16 + 1{,}37\ x_1 - 0{,}29\ x_2$. Определить коэффициенты эластичности для данной модели, если известно, что $\overline{x}_1 = 15$; $\overline{x}_2 = 10$; $\overline{y} = 30$. Указать их экономический смысл.

2А+Б. Зависимость прибыли предприятия Y (млн. ден. ед.) от расходов на рекламу X (млн. ден. ед.) за 10 лет представлена в табл. 40.

Таблица 40

Зависимость прибыли предприятия от расходов на рекламу

X, млн. ден. ед.	0,8	1,8	2,0	2,5	4,0	5,7	7,5	8,2	8,7	8,8
Y, млн. ден. ед.	5	7	13	15	17	21	22	24	23	25

Построить корреляционное поле и оценить по МНК коэффициенты предполагаемого уравнения регрессии.

3А+Б. Рассматривается система уравнений вида

$$\begin{cases} y_1 = 5x + 3y_2 + \varepsilon_1, \\ y_2 = 4y_1 + \varepsilon_2. \end{cases}$$

1. Какие из переменных данной модели являются экзогенными, эндогенными, предопределенными?

2. Проверить, является ли данная система идентифицируемой.

3. Изменится ли ответ, если во второе уравнение включить: а) константу; б) переменную x?

4Б. Приведите примеры использования логарифмических регрессионных моделей.

Вариант 2

1А+Б. Дайте определение мультиколлинеарности, поясните причины ее возникновения и методы устранения.

2А+Б. Зависимость объема продаж y (тыс. ден. ед.) от расходов на рекламу x (тыс. ден. ед.) для 12 предприятий концерна характеризуется следующим образом:

– уравнение регрессии: $y = 12{,}9 + 1{,}3x$;
– среднее квадратическое отклонение x: $\sigma_x = 3{,}5$;
– среднее квадратическое отклонение y: $\sigma_y = 4{,}8$.

Определите коэффициент корреляции. Оцените значимость коэффициента корреляции с помощью t-критерия Стьюдента.

3А+Б. Дана зависимость вида $Y = aX^b$. Для преобразованных (в логарифмах) переменных получены следующие данные:

$$\sum xy = 25{,}2;\ \sum x = 5;\ \sum x^2 = 12{,}5;\ \sum y = 4;\ \sum (y_i - \hat{y}_i)^2 = 0{,}0014.$$

Найдите параметр b. Определите показатель корреляции, считая $\sigma_Y = 0{,}08$. Оцените его значимость, если известно, что $n = 10$.

4Б. Оценить параметры идентифицируемой структурной модели

$$\begin{cases} y_1 = c_{10} + b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = c_{20} + b_{21}y_1 + a_{22}x_2 + \varepsilon_2, \end{cases}$$

если $\sum y_1 = 134$; $\sum y_2 = 38$; $\sum x_1 = 24$; $\sum x_2 = 38$; $\sum y_1x_1 = 342$; $\sum y_1x_2 = 862$; $\sum y_2x_1 = 182$; $\sum y_2x_2 = 26$; $\sum x_1x_2 = 108$; $\sum x_1^2 = 82$; $\sum x_2^2 = 222$; $n = 8$.

6.2. ВАРИАНТЫ ТЕСТОВЫХ ЗАДАНИЙ

Вариант 1

Вопрос	Варианты ответов
1. Какова цель эконометрики:	а) представить экономические данные в наглядном виде; б) разработать способы моделирования и количественного анализа реальных экономических объектов; в) определить способы сбора и группировки статистических данных; г) изучить качественные аспекты экономических явлений?
2. Набор сведений о разных объектах, взятых за один период времени, называется:	а) временными данными; б) пространственными данными; в) независимыми данными; г) системой алгебраических уравнений.

3. Какое значение не может принимать парный коэффициент корреляции:	а) –0,875; б) 0,045; в) 1,524; г) –0,165?
4. Согласно методу наименьших квадратов, минимизируется следующее выражение:	а) $\sum_{i=1}^{n}(y_i - \hat{y}_i)$; б) $\sum_{i=1}^{n}(y_i - \overline{y})^2$; в) $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$; г) $\sum_{i=1}^{n}(y_i^2 - \hat{y}_i^2)$.
5. Множественный линейный коэффициент корреляции равен 0,75. Какой процент вариации зависимой переменной y учтен в модели и обусловлен влиянием факторов:	а) 56,2; б) 75; в) 37,5; г) 25?
6. Рассмотрите модель зависимости общей величины расходов на питание (y) от располагаемого личного дохода (x) и цены продуктов питания (p): $y = a_0 + a_1x + a_2p + \varepsilon$. Определите класс модели и вид переменных:	а) регрессионная модель с одним уравнением; эндогенная переменная – расходы на питание, экзогенная переменная – располагаемый личный доход, лаговая переменная – цена продуктов питания; б) регрессионная модель с одним уравнением; эндогенная переменная – расходы на питание, экзогенные переменные – располагаемый личный доход и цена продуктов питания; в) модель временного ряда; эндогенная переменная – расходы на питание, лаговые переменные – располагаемый личный доход и цена продуктов питания.
7. Уравнение тренда $y = 35{,}5 - 4{,}6t$. Насколько в среднем за единицу времени в исследуемом периоде изменяется признак:	а) увеличивается на 35,5; б) увеличивается на 4,6; в) уменьшается на 4,6; г) уменьшается на 35,5?
8. Какой коэффициент указывает, на сколько процентов в среднем изменяется результативный показатель y при увеличении аргумента x на 1%:	а) коэффициент детерминации; б) коэффициент регрессии; в) коэффициент эластичности; г) бета-коэффициент?

9. Имеются следующие данные: коэффициент регрессии $b_1 = 4{,}225$; среднее квадратическое отклонение коэффициента регрессии $S_{b_1} = 0{,}25$. Определите t-критерий Стьюдента и оцените значимость коэффициента регрессии b_1, если $t_{табл} = 2{,}11$ при уровне значимости $\alpha = 0{,}05$:	а) 0,059; коэффициент незначим; б) 16,9; коэффициент значим; б) 4,225; коэффициент значим; в) –16,9; коэффициент незначим.

Вариант 2

Вопрос	Варианты ответов
1. Какое из определений соответствует понятию «эконометрика»:	а) это наука, предметом изучения которой является количественная сторона массовых социально-экономических явлений и процессов в конкретных условиях места и времени; б) это наука, предметом изучения которой является количественное выражение взаимосвязей экономических явлений и процессов; в) это наука, предметом изучения которой являются общие закономерности случайных явлений и методы количественной оценки влияния случайных факторов?
2. Верификация модели – это:	а) определение вида экономической модели, выражение в математической форме взаимосвязи между ее переменными; б) определение исходных предпосылок и ограничений модели; в) проверка качества как самой модели в целом, так и ее параметров; г) анализ изучаемого экономического явления.

3. Какие значения не может принимать парный коэффициент корреляции:	а) –1,975; б) 0,045; в) 0,524; г) –0,1?
4. Выберите аналоги понятия «независимая переменная»:	а) эндогенная переменная; б) фактор; в) результат; г) экзогенная переменная.
5. Множественный линейный коэффициент корреляции равен 0,85. Какой процент вариации зависимой переменной y учтен в модели и обусловлен влиянием факторов:	а) 85; б) 72,25; в) 70,5; г) 15?
6. Оценки параметров регрессии (свойства оценок метода наименьших квадратов) должны быть:	а) несмещенными; б) гетероскедастичными; в) эффективными.
7. Какой коэффициент указывает, на сколько процентов в среднем изменяется результативный показатель y при увеличении аргумента x на 1%:	а) коэффициент детерминации; б) коэффициент регрессии; в) коэффициент эластичности; г) бета-коэффициент?
8. Если структурные коэффициенты модели выражены через приведенные коэффициенты и имеют более одного числового значения, то такая модель:	а) сверхидентифицируемая; б) неидентифицируемая; в) идентифицируемая; г) нелинейная.
9. Случайная составляющая в модели вида $y_t = u_t + v_t + \varepsilon_t$ обозначена:	а) u_t; б) ε_t; в) y_t; г) v_t.

6.3. РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА (t-РАСПРЕДЕЛЕНИЕ)

γ	Вероятность												
	0,9	**0,8**	**0,7**	**0,6**	**0,5**	**0,4**	**0,3**	**0,2**	**0,1**	**0,05**	**0,02**	**0,01**	**0,001**
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,327	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,583
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015

γ	Вероятность												
	0,9	**0,8**	**0,7**	**0,6**	**0,5**	**0,4**	**0,3**	**0,2**	**0,1**	**0,05**	**0,02**	**0,01**	**0,001**
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,833
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,868	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,402	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

6.4. РАСПРЕДЕЛЕНИЕ ФИШЕРА – СНЕДЕКОРА (*F*-РАСПРЕДЕЛЕНИЕ)

γ_2	γ_1										
	1	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**24**	**60**	**∞**
	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	253,3	12,71
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366	63,66
	406523	500016	536700	562527	576449	585953	598149	610598	623432	636535	636,2
	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50	4,30
2	98,49	99,01	99,17	99,25	99,30	99,13	99,36	99,42	99,46	99,50	9,92
	998,46	999,00	999,20	999,20	999,20	999,20	999,40	999,60	999,40	999,40	31,00
	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53	3,18
3	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,60	26,12	5,84
	67,47	148,51	141,10	137,10	134,60	132,90	130,60	128,30	125,90	123,50	12,94
	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63	2,78
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	13,93	13,46	4,60
	74,13	61,24	56,18	5343	51,71	50,52	49,00	47,41	45,77	44,05	8,61
	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36	2,57
5	16,26	13,27	12,06	11,39	10,97	10,67	10,27	9,89	9,47	9,02	4,03
	47,04	36,61	33,20	31,09	20,75	28,83	27,64	26,42	25,14	23,78	6,86
	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67	2,45
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31	6,88	3,71
	35,51	26,99	23,70	21,90	20,81	20,03	19,03	17,99	16,89	15,75	5,96

γ_2	γ_1										
	1	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**24**	**60**	**∞**
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23	2,36
	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65	3,50
	29,22	21,69	18,77	17,19	16,21	15,52	14,63	13,71	12,73	11,70	5,40
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,99	2,31
	11,26	8,65	7,59	7,10	6,63	6,37	6,03	5,67	5,28	4,86	3,36
	25,42	18,49	15,83	14,39	13,49	12,86	12,04	11,19	10,30	9,35	5,04
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71	2,26
	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31	3,25
	22,86	16,39	13,90	12,56	11,71	11,13	10,37	9,57	8,72	7,81	4,78
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54	2,23
	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91	3,17
	21,04	14,91	12,55	11,28	10,48	9,92	9,20	8,45	7,64	6,77	4,59
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40	2,20
	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,02	3,60	3,11
	19,69	13,81	11,56	10,35	9,58	9,05	8,35	7,62	6,85	6,00	4,49
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30	2,18
	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36	3,06
	18,64	12,98	10,81	9,63	8,89	8,38	7,71	7,00	6,25	5,42	4,32

γ_2	γ_1										
	1	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**24**	**60**	**∞**
	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21	2,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,59	3,16	3,01
	17,81	12,31	10,21	9,07	8,35	7,86	7,21	6,52	5,78	4,97	4,12
	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13	2,14
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,43	3,00	2,98
	17,14	11,78	9,73	8,62	7,92	7,44	6,80	6,13	5,41	4,60	4,14
	4,45	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07	2,13
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87	2,95
	16,59	11,34	9,34	8,25	7,57	7,09	6,47	5,81	5,10	4,31	4,07
	4,41	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01	2,12
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75	2,92
	16,12	10,97	9,01	7,94	7,27	6,80	6,20	5,55	4,85	4,06	4,02
	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96	2,11
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,08	2,65	2,90
	15,72	10,66	8,73	7,68	7,02	6,56	5,96	5,32	4,63	3,85	3,96
	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92	2,10
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,01	2,57	2,88
	15,38	10,39	8,49	7,46	6,81	6,35	5,76	5,13	4,45	3,67	3,92

γ_2	γ_1										
	1	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**24**	**60**	**∞**
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88	2,09
	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	2,92	2,49	2,86
	15,08	10,16	8,28	7,26	6,61	6,18	5,59	4,97	4,29	3,52	3,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84	2,09
	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42	2,84
	14,82	9,95	8,10	7,10	6,46	6,02	5,44	4,82	4,15	3,38	3,85
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,82	2,08
	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36	2,83
	14,62	9,77	7,94	6,95	6,32	5,88	5,31	4,70	4,03	3,26	3,82
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78	2,07
	7,94	5,72	4,82	4,31	3,99	3,75	3,45	3,12	2,75	2,30	2,82
	14,38	9,61	7,80	6,87	6,19	5,76	5,19	4,58	3,92	3,15	3,79
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76	2,07
	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26	2,81
	14,19	9,46	7,67	6,70	6,08	5,56	5,09	4,48	3,82	3,05	3,77
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73	2,06
	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21	2,80
	14,03	9,34	7,55	6,59	5,98	5,55	4,99	4,39	3,84	2,97	3,75

γ_2	γ_1										
	1	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**24**	**60**	**∞**
5	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71	2,06
	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,62	2,17	2,79
	13,88	9,22	7,45	6,49	5,89	5,46	4,91	4,31	3,66	2,89	3,72
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69	2,06
	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13	2,78
	13,74	9,12	7,36	6,41	5,80	5,38	4,83	4,24	3,59	2,82	3,71
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67	2,05
	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10	2,77
	13,61	9,02	7,27	6,33	5,73	5,31	4,76	4,17	3,52	2,76	3,69
28	4,19	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65	2,05
	7,64	5,45	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06	2,76
	13,50	8,93	7,18	6,25	5,66	5,24	4,69	4,11	3,46	2,70	3,67
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64	2,05
	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03	2,76
	13,39	8,85	7,12	6,19	5,59	5,18	4,65	4,05	3,41	2,64	3,66
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62	2,04
	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,47	2,01	2,75
	13,29	8,77	7,05	6,12	5,53	5,12	4,58	4,00	3,36	2,59	3,64

γ_2	γ_1										
	1	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**24**	**60**	**∞**
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39	2,00
	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60	2,66
	11,97	7,76	6,17	5,31	4,76	4,37	3,87	3,31	2,76	1,90	3,36
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,03	1,96
	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,79	1,04	2,58
	10,83	6,91	5,42	4,62	4,10	3,74	3,27	2,74	2,13	1,05	3,29

Примечание. Первое значение соответствует вероятности 0,05; второе – вероятности 0,01 и третье – вероятности 0,001; γ_1 – число степеней свободы числителя; γ_2 – знаменателя.

ЛИТЕРАТУРА

Основная

1. Айвазян, С. А. Прикладная статистика и основы эконометрики: учебник для вузов / С. А. Айвазян, В. С. Мхитарян. – М.: ЮНИТИ, 1998. – 1022 с.

2. Бородич, С. А. Эконометрика: учеб. пособие / С. А. Бородич. – 2-е изд., испр. – Минск: Новое знание, 2004. – 416 с.

3. Доугерти, К. Введение в эконометрику: учебник для вузов / К. Доугерти. – М.: ИНФРА-М, 1999. – 402 с.

4. Кремер, Н. Ш. Эконометрика: учебник для студентов вузов / Н. Ш. Кремер, Б. А. Путко; под ред. Н. Ш. Кремера. – 2-е изд., стереотип. – М.: ЮНИТИ-ДАНА, 2008. – 311 с.

5. Марченко, В. М. Методы оптимизации и статистической обработки результатов измерений: учеб. пособие / В. М. Марченко, Т. Б. Копейкина. – Минск: БГТУ, 2007. – 140 с.

6. Хацкевич, Г. А. Эконометрика: учеб.-метод. комплекс для студентов экономических специальностей / Г. А. Хацкевич, А. Б. Гедранович. – Минск: Изд-во МИУ, 2005. – 252 с.

Дополнительная

7. Герасимович, А. И. Математическая статистика: учеб. пособие / А. И. Герасимович. – 2-е изд., перераб. и доп. – Минск: Вышэйшая школа, 1983. – 279 с.

8. Гмурман, В. Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов / В. Е. Гмурман. – 12-е изд., перераб. – М.: Высшее образование, 2006. – 479 с.

9. Гурский, Е. И. Сборник задач по теории вероятностей и математической статистике / Е. И. Гурский. – Минск: Вышэйшая школа, 1984. – 223 с.

10. Кремер, Н. Ш. Теория вероятностей и математическая статистика: учебник для вузов / Н. Ш. Кремер. – М.: ЮНИТИ-ДАНА, 2001. – 543 с.

11. Мацкевич, И. П. Высшая математика: теория вероятностей и математическая статистика: учебник / И. П. Мацкевич, Г. П. Свирид. – Минск: Вышэйшая школа, 1993. – 272 с.

12. Минюк, С. А. Математические методы и модели в экономике: учеб. пособие / С. А. Минюк, Е. А. Ровба, К. К. Кузьмич.– Минск: ТетраСистемс, 2002. – 432 с.

13. Письменный, Д. Т. Конспект лекций по теории вероятностей и математической статистике / Д. Т. Письменный. – М.: Айрис-пресс, 2004. – 256 с.

14. Янович, В. И. Экономико-математические методы и модели: лабораторный практикум для студентов специальностей 1-25 01 07 «Экономика и управление на предприятии», 1-25 01 08 «Бухгалтерский учет, анализ и аудит», 1-26 02 02 «Маркетинг», 1-26 02 03 «Менеджмент» / В. И. Янович, Н. В. Балашевич. – Минск: БГТУ, 2002. – 66 с.

15. Экономико-математические методы и модели: учеб.-метод. пособие по выполнению расчетных заданий с использованием табличного процессора Excel для студентов экономических специальностей / авт.-сост. Е. А. Шинкевич. – Минск: БГТУ, 2005. – 72 с.

16. Харин, Ю. С. Эконометрическое моделирование: учеб. пособие / Ю. С. Харин, В. И. Малюгин, А. Ю. Харин. – Минск: БГУ, 2003. – 313 с.

ОГЛАВЛЕНИЕ

Учебное издание

Марченко Владимир Матвеевич
Можей Наталья Павловна
Шинкевич Елена Алексеевна

ЭКОНОМЕТРИКА И ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ

В 2-х частях

Часть 1. Эконометрика

Учебное пособие

Редактор *О. А. Семенец*
Компьютерная верстка *О. А. Семенец*
Корректор *О. А. Семенец*

Подписано в печать 30.12.2011. Формат $60\times84^{1}/_{16}$.
Бумага офсетная. Гарнитура Таймс. Печать офсетная.
Усл. печ. л. 9,1. Уч.-изд. л. 9,4.
Тираж 350 экз. Заказ .

Издатель и полиграфическое исполнение:
«Белорусский государственный технологический университет».
ЛИ № 02330/0549423 от 08.04.2009.
ЛП № 02330/0150477 от 16.01.2009.
Ул. Свердлова, 13а, 220006, г. Минск.