ИНСТИТУТ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

Сборник
учебно-методических материалов

Н.В. Колачева, С.Ш. Палферова

ЭКОНОМЕТРИКА

www.tltsu.ru
Тольятти
2011

Министерство образования и науки Российской Федерации
Тольяттинский государственный университет
Факультет математики и информатики
Кафедра «Высшая математика и математическое моделирование»
Институт динстанционного обучения

Н.В. Колачева, С.Ш. Палферова

ЭКОНОМЕТРИКА

Сборник учебно-методических материалов

Тольятти
ТГУ
2011

УДК 330.43
ББК 65в631.8
К60

Рецензенты:
д.п.н., профессор Тольяттинского филиала
Московского государственного университета пищевых производств *А.Н. Ярыгин*;
д.т.н., профессор Тольяттинского государственного университета *П.Ф. Зибров*.

К60 Колачева, Н.В. Эконометрика : сборник учеб.-метод. материалов / Н.В. Колачева, С.Ш. Палферова. – Тольятти : ТГУ, 2011. – 172 с. : обл.

Сборник включает учебно-методическое пособие, курс лекций, практикум по дисциплине, материалы для контроля знаний, задания к контрольной работе с методическими рекомендациями для их выполнения.

Предназначен для студентов, обучающихся по специальностям 080105.65 «Финансы и кредит», 080109.65 «Бухгалтерский учет, анализ и аудит» и направлению подготовки 080100.62 «Экономика», изучающих дисциплину «Эконометрика».

Адресован студентам заочной формы обучения и может быть полезен для студентов очной формы обучения.

УДК 330.43
ББК 65в631.8

Рекомендовано к изданию научно-методическим советом Тольяттинского государственного университета.

ВВЕДЕНИЕ

Постоянно усложняющиеся экономические процессы потребовали создания и совершенствования особых методов изучения и анализа. Широкое распространение получило использование моделирования и количественного анализа. На этом этапе выделилось и сформировалось одно из направлений экономических исследований – *эконометрика*.

В переводе с греческого языка «эконометрика» означает «измерения в экономике». Однако область исследований данной дисциплины гораздо шире. Эконометрика – это наука, в которой на базе реальных статистических данных строятся, анализируются и совершенствуются математические модели реальных экономических явлений. Эконометрика позволяет найти количественное подтверждение либо опровержение того или иного экономического закона либо гипотезы. Одним из важнейших направлений эконометрики является построение прогнозов по различным экономическим показателям.

Эконометрика как научная дисциплина зародилась и получила развитие на основе слияния экономической теории, математической экономики, экономической и математической статистики.

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

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

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

Мощным инструментом эконометрических исследований является аппарат математической статистики. Действительно, большинство экономических показателей носят характер случайных величин, предсказать точные значения которых практически невозможно. Например, весьма сложно предвидеть доход или потребление какого-либо индивидуума, объемы экспорта и импорта страны в течение следующего года и т. д. Связи между экономическими показателями на практике никогда не носят строгий функциональный характер, а допускают наличие каких-либо случайных отклонений (особенно это касается макроэкономических данных). Вследствие этого использование методов математической статистики в эконометрике естественно и обоснованно. Однако в силу специфики получения статистических данных в экономике (например, из-за невозможности проведения управляемого эксперимента) эконометристам приходится использовать свои собственные наработки и специальные приемы анализа, которые в математической статистике не встречаются.

Настоящий сборник содержит разделы курса «Эконометрика», соответствующие программе для студентов экономических специальностей высших учебных заведений.

В учебно-методическом пособии изложены цели, задачи, теоретические вопросы, рекомендуемая литература и методические рекомендации.

В издании также представлены курс лекций, практикум с примерами решения задач, контрольными заданиями для самостоятельного решения, тестами.

Каждый студент должен последовательно изучить материал разделов, акцентировать внимание на основных понятиях и терминах. Необходимо решить контрольные задания по каждой теме в соответствии с вариантом, согласованным с преподавателем, после чего ответить на вопросы теста. В результате студент должен научиться решать задачи, аналогичные разобранным в данном сборнике, выполнить контрольную работу и тест итогового контроля.

1. УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

Изучение дисциплины «Эконометрика» для заочного отделения согласно с учебным планом предусматривает следующее распределение часов по видам учебных занятий.

Наименование специальности (направления)	6 семестр			
	Лекции (час.)	Самостояельная работа (час.)	Практические занятия (час.)	Формы контроля
080100.62 «Экономика»	12	148	10	контрольная работа, зачет
Итого	170 часов			
080105.65 «Финансы и кредит», 080109.65 «Бухгалтерский учет, анализ и аудит»	6	88	14	контрольная работа, зачет
Итого	108 часов			

1.1. Цели и задачи дисциплины

Основными целями и задачами изучения эконометрики являются:

– формирование у студентов навыков анализа связей между экономическими факторами и показателями на основе статистических данных с использованием аппарата теории вероятностей и математической статистики;
– обучение студентов применению экономической теории, экономической статистики, экономических измерений и математико-статистического инструментария;
– развитие навыков прогнозирования социально-экономических показателей, характеризующих состояние и развитие анализируемой системы;
– развитие у студентов возможностей имитации различных сценариев социально-экономического развития анализируемой системы.

1.2. Методические рекомендации по изучению дисциплины

Тема 1. Виды эконометрических моделей. Введение в регрессионный анализ

Цели: обзор основных положений и методов построения эконометрических моделей; усвоение классификации этих моделей; изучение вероятностных характеристик случайных величин.

Учебные вопросы

1. Предмет и задачи эконометрики.
2. Характеристики случайных величин.
3. Типы эконометрических моделей.
4. Типы данных при эконометрическом моделировании.
5. Основные положения регрессионного анализа.

Изучив данную тему, студент должен:

иметь представление об эконометрике как науке, классификации эконометрических моделей и данных, основных положениях регрессионного анализа;

знать понятия эконометрической модели, регрессионной модели, модели временных рядов, системы одновременных уравнений, объясняемых и объясняющих переменных, поля корреляции, среднего значения, математического ожидания, выборочной дисперсии, стандартного отклонения, выборочного коэффициента корреляции, выборочной ковариации случайных величин, стандартных предположений регрессионного анализа, гомоскедастичности и гетероскедастичности дисперсии ошибок; эндогенных и экзогенных переменных;

уметь вычислять вероятностные характеристики случайных величин, строить поле корреляции исходя из эмпирических данных;

владеть навыками анализа статистических данных.

При освоении темы необходимо:

1. Изучить материал по данной теме.
2. Акцентировать внимание на функциональных взаимосвязях реальных экономических факторов; типах эконометрических данных и моделей; стандартных предположениях регрессионного анализа; вероятностных характеристиках случайных величин и их свойствах.
3. Выполнить задание, назначенное преподавателем.
4. Выполнить тест по пройденной теме.
5. Ответить на контрольные вопросы.

Тема 2. Парная линейная регрессия

Цели: изучение и применение метода наименьших квадратов (МНК) расчета параметров уравнения парной линейной регрессии, использование интервальных оценок для прогнозирования по уравнению регрессии, изучение методов проверки качества полученной модели.

Учебные вопросы

1. Метод наименьших квадратов (МНК).
2. Использование оцененной модели для прогноза.
3. Интервальная оценка функции регрессии и ее параметров.
4. Оценка значимости уравнения регрессии (адекватности имеющимся статистическим данным).
5. Проверка выполнения стандартных предположений об ошибках в линейной модели наблюдений графическим методом.

Изучив данную тему, студент должен:

иметь представление о модели парной линейной регрессии и ее составляющих, МНК оценки параметров этой модели, методах прогнозирования по данной модели, методах оценки качества уравнения регрессии;

знать понятия объясняющей и объясняемой переменных, случайной составляющей модели, поля корреляции, МНК, доверительных интервалов для математического ожидания, индивидуального значения объясняемой переменной и параметров регрессионной модели, составляющих дисперсии объясняемой переменной, средней ошибки аппроксимации, критерия Фишера, коэффициента детерминации, графиков стандартизированных остатков регрессии;

уметь записывать уравнение парной линейной регрессии, рассчитывать параметры этого уравнения регрессии, прогнозировать средние и индивидуальные значения объясняемой переменной по полученному уравнению, оценивать качество уравнения и соблюдение стандартных предположений регрессионного анализа;

владеть навыками построения полей корреляции, графиков стандартизированных остатков.

При освоении темы необходимо:

1. Изучить материал по данной теме.
2. Акцентировать внимание на составляющих модели парной линейной регрессии; подборе формы теоретической модели по виду облака рассеяния на поле корреляции; применении МНК оценки параметров линейного уравнения парной регрессии; **построении доверительных интервалов для среднего значения у полученного по функции регрессии $M^{\hat{y}}(x)$, индивидуальных значений зависимой переменной, параметров регрессионной модели**; основной идее дисперсионного анализа; **процедуре проверки значимости линейной связи между переменными**; свойствах коэффициента детерминации; проверке выполнения стандартных предположений об ошибках в линейной модели наблюдений графическим методом.
3. Выполнить задание, назначенное преподавателем.
4. Выполнить тест по пройденной теме.
5. Ответить на контрольные вопросы.

Тема 3. Нелинейные регрессионные модели

Цели: изучение различных форм нелинейных моделей, освоение и применение метода линеаризации нелинейных уравнений регрессии, итерационных методов подбора нелинейных моделей, методов проверки качества таких моделей.

Учебные вопросы

1. Нелинейные модели с двумя переменными, приводимые к линейной форме.
2. Итерационные методы подбора нелинейных моделей.
3. Нелинейные модели множественной регрессии.
4. Проверка статистических гипотез о значениях отдельных коэффициентов.

Изучив данную тему, студент должен:

иметь представление о степенной форме эконометрической модели; методах приведения степенной модели к линейной форме; оценки параметров модели и ее качества; степенной модели с мультипликативными возмущениями; логарифмической, правой и левой полулогарифмических моделей; экспоненциальной модели; обратной модели; полиномиальной модели; интерактивного уравнения; степенной модели с аддитивными возмущениями; проверке статистических гипотез о значениях отдельных коэффициентов;

знать понятия парной и множественной нелинейной регрессии, предельной склонности и эластичности функции, условий постоянства и изменения предельной склонности и эластичности, итерационных методов подбора нелинейных моделей, статистической гипотезы, статистического критерия, уровня значимости критерия, мощности критерия, t-критерия, критического множества;

уметь приводить нелинейные модели к линейной форме, рассчитывать оценки параметров модели и ее качества, проверять статистические гипотезы о значениях отдельных коэффициентов;

владеть навыками линеаризации нелинейных моделей, прогнозирования по нелинейным моделям, оценки качества нелинейных моделей, определения эластичности функции.

При освоении темы необходимо:

1. Изучить материал по данной теме.
2. Акцентировать внимание на формах нелинейных моделей; методах линеаризации нелинейных уравнений; применении методов оценки качества к нелинейным регресси-

онным моделям; понятии предельной склонности и эластичности функции; условиях постоянства и изменения предельной склонности и эластичности функции; нелинейных моделях, которые не могут быть приведены к линейной форме; итерационных методах подбора нелинейных моделей;проверке статистических гипотез о значениях отдельных коэффициентов.
3. Выполнить задание, назначенное преподавателем.
4. Выполнить тест по пройденной теме.
5. Ответить на контрольные вопросы.

Тема 4. Множественная линейная регрессия и корреляция

Цели: овладение навыками ранжирования факторов для включения в модель множественной регрессии, изучение и применение метода наименьших квадратов (МНК) расчета параметров уравнения множественной линейной регрессии, использование интервальных оценок для прогнозирования по уравнению регрессии, изучение методов проверки качества полученной модели.

Учебные вопросы

1. Отбор факторов для модели множественной регрессии.
2. Оценивание параметров множественной линейной регрессии МНК.
3. Проверка значимости уравнения множественной линейной регрессии.
4. Метод взвешенных наименьших квадратов (обобщенный МНК).
5. Фиктивные переменные.

Изучив данную тему, студент должен:

иметь представление о методах построения уравнения множественной регрессии, модели множественной линейной регрессии и ее составляющих, методе наименьших квадратов (МНК) оценки параметров этой модели, методах прогнозирования по данной модели, методах оценки качества уравнения регрессии, фиктивных переменных, моделях дисперсионного анализа, ковариационного анализа, ловушке фиктивной переменной;

знать понятия коэффициента интеркорреляции, коллинеарных переменных, мультиколлинеарности факторов, уравнения множественной регрессии в стандартизированном масштабе, стандартизованные коэффициенты регрессии, частных коэффициентов эластичности, коэффициента множественной корреляции, коэффициента множественной детерминации, скорректированного коэффициента множественной корреляции, скорректированного коэффициента множественной детерминации, частных коэффициентов корреляции, общего и частных *F*-критериев, обобщенного МНК;

уметь применять правила отбора факторов в модель множественной регрессии, рассчитывать параметры этого уравнения с помощью МНК, прогнозировать значения объясняемой переменной по полученному уравнению, оценивать качество уравнения и соблюдение стандартных предположений регрессионного анализа;

владеть навыками расчета матрицы показателей линейной корреляции, отбора факторов в модель множественной регрессии, оценки параметров уравнения множественной регрессии, преобразования уравнения множественной регрессии в случае смещенности и неэффективности оценок параметров модели.

При освоении темы необходимо:

1. Изучить материал по данной теме.
2. Акцентировать внимание на экономических процессах, описываемых с помощью уравнений множественной регрессии; анализе факторов, включаемых в модель множест-

венной регрессии; **методах построения уравнения множественной регрессии**; **методе наименьших квадратов (МНК) оценивания параметров множественной линейной регрессии**; применении МНК для стандартизированного уравнения множественной линейной регрессии; **частных коэффициентах эластичности**; свойствах коэффициента детерминации; коэффициентах множественной корреляции и детерминации; частных и общем коэффициентах корреляции; проверке значимости уравнения линейной множественной регрессии с помощью критериев Фишера и Стьюдента; методе взвешенных наименьших квадратов (обобщенном МНК); необходимости использования фиктивных переменных; моделях, содержащих только качественные объясняющие переменные; моделях, в которых объясняющие переменные носят как количественный, так и качественный характер.
3. Выполнить задание, назначенное преподавателем.
4. Выполнить тест по пройденной теме.
5. Ответить на контрольные вопросы.

Тема 5. Временные ряды

Цель: изучение теоретических основ, методологических принципов и конкретных подходов постановки, решения и анализа задач моделирования временных рядов.

Учебные вопросы

1. Составляющие временных рядов.
2. Коэффициент автокорреляции. Автокорреляционная функция.
3. Моделирование тенденции временного ряда.
4. Моделирование сезонных колебаний.
5. Автокорреляция в остатках. Критерий Дарбина-Уотсона.

Изучив данную тему, студент должен:

иметь представление о группах факторов, влияющих на формирование временного ряда, стационарных и нестационарных временных рядах, аддитивной и мультипликативной моделях временных рядов, свойствах коэффициента автокорреляции;

знать понятия тенденции, сезонной и случайной компонент временного ряда, коэффициента автокорреляции, автокорреляционной функции, лага;

уметь строить коррелограмму, анализировать автокорреляционную функцию, применять метод скользящей средней;

владеть навыками аналитического выравнивания временного ряда, определения типа тенденции, моделирования тенденции и сезонных колебаний, расчета критерия Дарбина-Уотсона.

При освоении темы необходимо:

1. Изучить материал поданной теме.
2. Акцентировать внимание на факторах, формирующих тенденцию ряда, циклические колебания ряда и случайных факторах; стационарных и нестационарных временных рядах; аддитивной и мультипликативной моделях временных рядов; свойствах коэффициента автокорреляции; анализе автокорреляционной функции; аналитическом выравнивании временного ряда; методе скользящей средней выравнивания исходного ряда; проверке модели на автокорреляцию в остатках.
3. Выполнить задание, назначенное преподавателем.
4. Выполнить тест по пройденной теме.
5. Ответить на контрольные вопросы.

Тема 6. **Системы эконометрических уравнений**

Цель: изучение теоретических основ, методологических принципов и конкретных подходов постановки, решения и анализа задач моделирования систем эконометрических уравнений.

Учебные вопросы

1. Классификация систем регрессионных уравнений.
2. Оценка параметров систем одновременных уравнений.
3. Проблема идентификации структурных моделей.
4. Методы оценки параметров структурной модели.

Изучив данную тему, студент должен:

иметь представление о системах эконометрических уравнений, их классификации, особенностях использования МНК для оценивания структурных коэффициентов модели по коэффициентам приведенной модели, проблеме идентификации структурных моделей, косвенном МНК, двухшаговом МНК, трехшаговом МНК, методе максимального правдоподобия с полной информацией, методе максимального правдоподобия при ограниченной информации;

знать понятия системы независимых уравнений, системы рекурсивных уравнений, системы взаимозависимых (совместных, одновременных) уравнений, структурной формы модели, эндогенных и экзогенных переменных, лаговых переменных, структурных коэффициентов модели, приведенной формы модели, идентифицируемой, неидентифицируемой и сверхидентифицируемой моделей;

уметь выражать коэффициенты приведенной формы модели через коэффициенты структурной формы и наоборот, анализировать условие идентифицируемости системы;

владеть навыками оценки параметров системы независимых уравнений, системы рекурсивных уравнений и системы взаимозависимых уравнений.

При освоении темы необходимо:

1. Изучить материал по данной теме.
2. Акцентировать внимание на классификации систем эконометрических уравнений; проблеме идентификации эконометрических уравнений; особенностях применения МНК для различных видов и форм систем эконометрических уравнений.
3. Выполнить задание, назначенное преподавателем.
4. Выполнить тест по пройденной теме.
5. Ответить на контрольные вопросы.

1.3. Вопросы для итогового контроля

1. Понятие и задачи эконометрики как науки.
2. Эконометрическая модель и ее составляющие.
3. Поле корреляции.
4. Характеристики случайных величин: математическое ожидание, среднее значение, выборочная дисперсия, стандартное отклонение.
5. Выборочный корреляционный момент (выборочная ковариация), коэффициент корреляции (r) и его свойства при большом объеме выборки.
6. Виды эконометрических моделей.
7. Понятие регрессионной модели.
8. Типы данных при эконометрическом моделировании.

9. Стандартные предположения регрессионного анализа.
10. Понятия гомоскедастичности и гетероскедастичности дисперсии ошибок.
11. Модель парной линейной регрессии.
12. МНК оценки параметров парной регрессионной модели, их статистические свойства.
13. Использование модели парной линейной регрессии для прогноза.
14. Геометрический смысл регрессионной модели, составляющие дисперсии.
15. Доверительный интервал для функции регрессии, т. е. для $M_x(y)$.
16. Доверительный интервал для индивидуальных значений зависимой переменной.
17. Доверительный интервал для параметра   регрессионной модели.
18. Доверительный интервал для параметра ^{2} регрессионной модели.
19. Основная идея дисперсионного анализа.
20. Процедура проверки значимости линейной связи между переменными, использование F-критерия (критерия Фишера-Снедекора).
21. Коэффициент детерминации (R^2) и его свойства.
22. Графический метод проверки стандартных предположений регрессионного анализа.
23. Понятие предельной склонности потребления в модели доход–потребление.
24. Приведение степенной модели к линейной форме модели, оценка параметров модели и ее качества.
25. Понятие предельной склонности и эластичности функции. Условия постоянства предельной склонности и эластичности функции.
26. Обратно пропорциональная зависимость, модели с убывающей эластичностью.
27. Итерационные методы подбора нелинейных моделей.
28. Нелинейные модели множественной регрессии, приводимые к линейной форме.
29. Проверка статистических гипотез о значениях отдельных коэффициентов.
30. Определение и примеры моделей множественной линейной регрессии.
31. Отбор факторов в модель множественной регрессии.
32. Отбор факторов в модель линейной множественной регрессии.
33. Методы построения уравнения множественной регрессии.
34. МНК оценивания параметров множественной линейной регрессии.
35. Уравнение множественной регрессии в стандартизированном масштабе.
36. Понятие частных коэффициентов эластичности.
37. Частные и общий коэффициенты корреляции.
38. Проверка значимости уравнения множественной линейной регрессии с помощью критериев Фишера и Стьюдента.
39. Сущность метода взвешенных наименьших квадратов (обобщенного МНК).
40. Понятие фиктивных переменных.
41. Модели, содержащие только качественные объясняющие переменные.
42. Модели, в которых объясняющие переменные носят как количественный, так и качественный характер.
43. Виды моделей временных рядов, составляющие временного ряда.
44. Стационарные и нестационарные временные ряды.
45. Аддитивная и мультипликативная модели временных рядов.
46. Коэффициент автокорреляции, его свойства.
47. Автокорреляционная функция, коррелограмма, их анализ.
48. Моделирование тенденции временного ряда.

49. Моделирование сезонных колебаний временного ряда.
50. Автокорреляция в остатках. Критерий Дарбина-Уотсона.
51. Классификация систем регрессионных уравнений.
52. Оценка параметров систем одновременных уравнений.
53. Проблема идентификации структурных моделей.
54. Методы оценки параметров структурной модели.

2. КУРС ЛЕКЦИЙ

2.1. Виды эконометрических моделей. Введение в регрессионный анализ

2.1.1. Предмет и задачи эконометрики

Эконометрика – это совокупность методов анализа количественных связей между экономическими факторами и показателями на основании реальных статистических данных с использованием аппарата теории вероятностей и математической статистики.

Анализируя характер имеющихся статистических данных методами эконометрики, исследователь должен сделать определенные заключения о возможной форме подходящей теоретической экономической модели. Статистические данные указывают на то, в каком направлении нужно искать теоретические модели. Построение окончательной модели производится с учетом представлений экономической теории и с учетом информации, содержащейся в эмпирических данных.

Обобщенный вид эконометрической модели:

$$y = f(x_1, x_2, ..., x_n) + \varepsilon, \tag{2.1}$$

где y – наблюдаемое значение зависимой переменной (объясняемая переменная); $x_1, x_2, ..., x^n$ – объясняющие переменные; $f(x_1, x_2,..., x^n)$ – объясненная часть, зависящая от значений объясняющих переменных; ε – случайная составляющая.

Рассмотрим связь между годовым располагаемым доходом x и годовыми расходами на личное потребление y (в 1999 году, в условных единицах) 20 домашних хозяйств (табл. 2.1).

Таблица 2.1

i	x	y
1	2508	2406
2	2572	2464
3	2408	2336
4	2522	2281
5	2700	2641
6	2531	2385
7	2390	2297
8	2595	2416
9	2524	2460
10	2685	2549

i	x	y
11	2435	2311
12	2354	2278
13	2404	2240
14	2381	2183
15	2581	2408
16	2529	2379
17	2562	2378
18	2624	2554
19	2407	2232
20	2448	2356

Известный психолог Кейнс отметил как фундаментальный закон психологии склонность людей увеличивать расходы на личное потребление по мере возрастания их доходов, но не в той степени, в какой возрастает доход, то есть $y = f(x)$, где обе переменные измерены в одних единицах и функция $f(x)$ должна быть возрастающей, скорость изменения этой функции должна быть меньше 1.

Для того чтобы установить форму функциональной связи, строят диаграмму рассеяния или поле корреляции (рис. 2.1).

Простейшей моделью связи является линейная модель (модель наблюдений)

$$y = \alpha + \beta x + \varepsilon, \tag{2.2}$$

где β – некоторая постоянная величина, $0 < \beta < 1$, характеризующая в данном круге домашних хозяйств их склонность к потреблению, связанную с традиционными привычками; α – постоянное потребление; $\varepsilon = y - (a + \beta x)$ – это отклонение реально наблюдаемых расходов на потребление y_i от значения $y = a + \beta x$, предсказываемого гипотетической линейной моделью связи для i-го домашнего хозяйства.

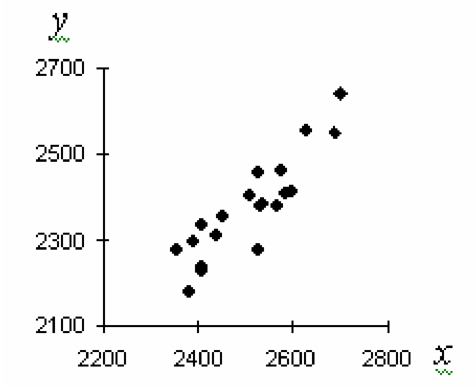

Рис. 2.1

В связи с наличием случайной составляющей ε точки не лежат на одной прямой, а образуют облако рассеяния.

Предложив для описания имеющихся статистических данных модель, учитывающую указанное отклонение от теоретической модели, мы должны оценить с их помощью величину параметров α и β. Затем, используя соответствующие критерии, вынести на основании этих данных суждение о пригодности выбранной модели.

2.1.2. Характеристики случайных величин

Если в рассмотренном ранее примере обозначить $x_1, x_2, \dots, x^n$ – последовательно располагаемые доходы домашних хозяйств; $y_1, y_2, \dots, y^n$ – расходы домашних хозяйств на потребление, мы сможем говорить о наблюдаемых значениях двух этих переменных. Всего мы имеем здесь $n = 20$ наблюдаемых пар значений переменных x и y: $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$.

Наиболее простыми показателями, характеризующими последовательности $x_1, x_2, \dots, x_n$ и $y_1, y_2, \dots, y_n$ являются *средние значения* этих дискретных величин.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i ; \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i . \tag{2.3}$$

В рассматриваемом примере

$$\bar{x} = \frac{2508 + 2572 + \dots + 2448}{20} = 2508; \quad \bar{y} = \frac{2406 + 2572 + \dots + 2356}{20} = 2377{,}7 .$$

Математическое ожидание дискретных случайных величин – это сумма произведений всех значений дискретной величины на их вероятности, оно приближенно равно их средним значениям.

$$M(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{i=1}^{n} x_i p_i, \qquad M(X) = y_1 p_1 + y_2 p_2 + \dots + y_i p_n = \sum_{i=1}^{n} y_i p_i. \tag{2.4}$$

Выборочные дисперсии (вариации) характеризуют степень разброса значений $x_1, x_2, \dots, x_n$ $(y_1, y_2, \dots, y^n)$ вокруг своего среднего значения $\bar{x}(\bar{y})$

$$\mathrm{var}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2, \quad \mathrm{var}(y) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 . \tag{2.5}$$

Стандартное отклонение (среднеквадратическое отклонение) более удобно для характеристики рассеяния дискретной случайной величины, так как измеряется в тех же единицах, что и сама величина.

$$S(x)=\sqrt{\operatorname{var}(x)}, \qquad S(y)=\sqrt{\operatorname{var}(y)}. \tag{2.6}$$

Удобным графическим средством анализа данных является диаграмма рассеяния (поле корреляции), на которой в прямоугольной системе координат располагаются точки (x_i, y_i), $I = 1, 2, ..., n$. Для того чтобы по форме облака рассеяния делать выводы о характере зависимости между величинами, при построении диаграммы желательно выбирать масштабы и интервалы изменения переменных таким образом, чтобы окно диаграммы имело вид квадрата, и чтобы на диаграмме имелись точки, достаточно близко расположенные к каждой из четырех границ этого квадрата.

Степень выраженности линейной связи между произвольными переменными x и y, принимающими значения x_i, y_i, $i = 1, 2, ..., n$, оценивается посредством *выборочного коэффициента корреляции*

$$r_{xy}=\frac{\operatorname{cov}(x,y)}{S(x)S(y)}=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{S_xS_y}; \tag{2.7}$$

$$\operatorname{cov}(x,y)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\overline{xy}-\overline{x}\cdot\overline{y}. \tag{2.8}$$

Величина $cov(x, y)$ называется *выборочной ковариацией* (*выборочным корреляционным моментом*). Характеризует степень зависимости двух случайных величин и степень их рассеяния.

Для расчета r_{xy} можно использовать формулу

$$r_{xy}=\frac{n\sum_{i=1}^{n}x_iy_i-\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)}{\sqrt{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}\sqrt{n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2}}. \tag{2.9}$$

Здесь r_{xy} находится с использованием непосредственных данных и на его значении не скажутся округления данных, связанные с расчетом средних значений.

Свойства выборочного коэффициента корреляции r_{xy} (при достаточно большом объеме выборки n).

1. Коэффициент корреляции по абсолютной величине не превосходит единицу ($-1 \le r_{xy} \le 1$). Чем ближе $|r_{xy}|$ к единице, тем теснее связь (рис. 2.2).

2. При $r_{xy} = \pm1$ все наблюдаемые значения лежат на одной прямой. Корреляционная связь представляет линейную функциональную зависимость (рис. 2.3, 2.4).

3. При $r_{xy} = 0$ линейная связь отсутствует (рис. 2.5, 2.6, 2.7).

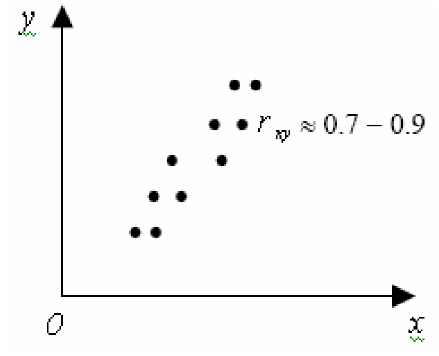

Рис. 2.2

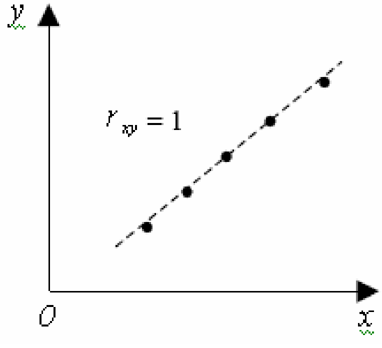

Рис. 2.3

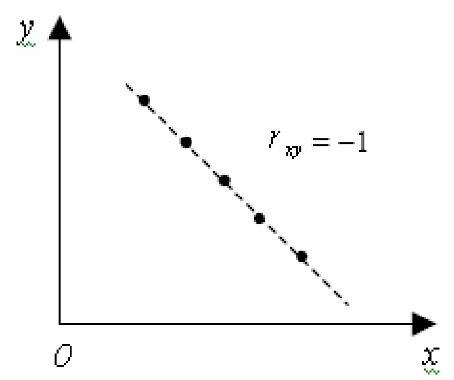

Рис. 2.4

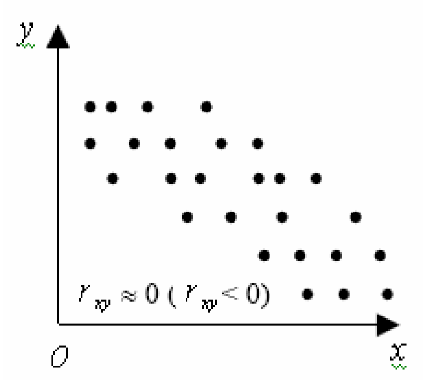

Рис. 2.5

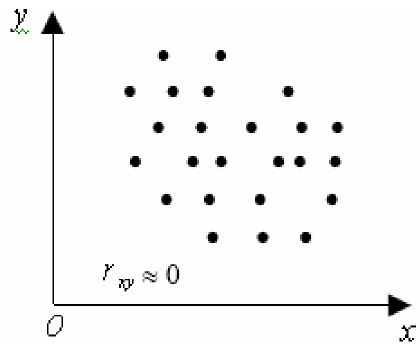

Рис. 2.6

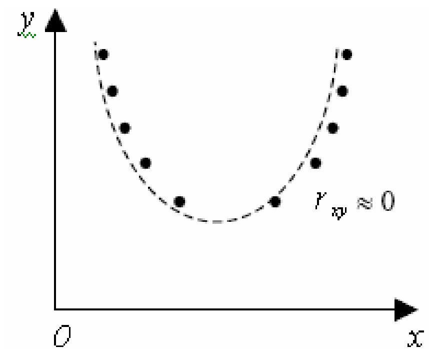

Рис. 2.7

2.1.3. Типы эконометрических моделей

Модели временных рядов

К ним относятся:

Модели тренда

$$y(t) = T(t) + \varepsilon_t, \tag{2.10}$$

где $T(t)$ – временной тренд заданного параметрического вида (например, линейный $T(t) = a + bt$), плавно меняющаяся компонента, описывающая чистое влияние долговременных факторов – рост населения, экономическое развитие, изменение структуры потребления и т. д.; ε_t – случайная (стохастическая) компонента.

Модели сезонности

$$y(t) = S(t) + \varepsilon_t, \tag{2.11}$$

где $S(t)$ – периодическая (сезонная) компонента, отражающая повторяемость экономических процессов в течении не очень длительного времени – объем продаж товаров, перевозок пассажиров в различные времена года и т. д.; ε_t – случайная (стохастическая) компонента.

Модели тренда и сезонности

– аддитивная

$$y(t) = T(t) + S(t) + \varepsilon_t; \tag{2.12}$$

– мультипликативная

$$y(t) = T(t) \cdot S(t) \cdot \varepsilon_t, \tag{2.13}$$

где $T(t)$ – временной тренд заданного параметрического вида; $S(t)$ – периодическая (сезонная) компонента; ε_t – случайная (стохастическая) компонента.

К моделям временных рядов относится множество более сложных моделей, таких как модели адаптивного прогноза, модели авторегрессии, скользящего среднего и др. Их общей чертой является то, что они объясняют поведение временного ряда, исходя только из его предыдущих значений.

Такие модели могут применяться, например, для изучения и прогнозирования объема продаж авиабилетов, спроса на мороженое, краткосрочного прогноза процентных ставок и т. п.

Регрессионные модели с одним уравнением

Зависимость, при которой изменение одной из величин влечет изменение распределения другой, называется *статистической*. Если статистическая зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение (условное математическое ожидание) другой, то она называется *корреляционной*.

В регрессионных моделях зависимая (объясняемая) переменная y представляется в виде функции

$$y = M_y(x_1, \ldots, x_p) + \varepsilon, \tag{2.14}$$

где $M_y(x_1, \ldots, x_p)$ – условное математическое ожидание величины y, полученное при данном наборе значений объясняющих переменных, или так называемая функция регрессии; ε – случайная составляющая.

По виду функции различают линейные и нелинейные регрессионные модели.

Например, можно исследовать спрос на мороженое как функцию от времени года, температуры воздуха, среднего уровня доходов или установить зависимость зарплаты от возраста, пола, уровня образования, стажа работы и т. п.

Область применения таких моделей, даже линейных, значительно шире, чем моделей временных рядов. Эта тема – основная в эконометрике.

Системы одновременных уравнений

Эти модели описываются системами уравнений. Системы могут состоять из тождеств и регрессионных уравнений, каждое из которых может, кроме объясняющих переменных, включать в себя также объясняемые переменные из других уравнений системы. Таким образом, мы имеем здесь набор объясняемых переменных, связанных через уравнения системы. Системы одновременных уравнений требуют относительно более сложный математический аппарат. Используются в макроэкономике.

Рассмотрим, например, модель спроса и предложения. Пусть Q_t^D – спрос на товар в момент времени t, Q_t^S – предложение товара в момент времени t, P_t – цена товара в момент времени t, Y_t – доход в момент времени t.

Составим следующую систему уравнений «спрос-предложение»:

– предложение

$$Q_t^S = \alpha_1 + \alpha_2 + \alpha_3\, P_t + \alpha_3\, P_{t-1} + \varepsilon_t;$$

– спрос

$$Q_t^D = \beta_1 + \beta_2 + \beta_3\, P_{\mathrm{t}} + \beta_3\, Y_{\mathrm{t}} + u_t;$$

– равновесие

$$Q_t^S = Q_t^D.$$

Цена на товар P_t и спрос на товар Q_t^S, равный предложению Q_t^D, определяются из уравнений модели, они являются эндогенными переменными. Предопределенными (экзогенными) переменными в данной модели являются доход Y_t и значение цены товара в предыдущий момент времени P_{t-1}.

Взаимосвязанные переменные, описывающие экономический объект и формирующиеся внутри функционирования объекта, называются *эндогенными*, а задаваемые извне – *экзогенными*. *Лаговыми* называются переменные, взятые в предыдущий момент времени.

2.1.4. Типы данных при эконометрическом моделировании

Пространственные данные

Данные в определенный момент времени называются *пространственными данными*. Например:

– набор сведений (объем производства, количество работников, доход и др.) по разным фирмам в один и тот же момент времени (пространственный срез);
– данные по курсам покупки – продажи наличной валюты в какой-то день по городу.

Временные ряды

Данные через определенные отрезки времени называются *временными рядами*. Например:

– ежеквартальные данные по инфляции, средней заработной плате, национальному доходу, денежной эмиссии за последние годы;
– ежедневный курс доллара США, цены фьючерсных контрактов на поставку доллара США, котировки акций за ряд последних лет.

Отличительной чертой временных рядов является то, что они естественным образом упорядочены по времени, кроме того, наблюдения в близкие моменты времени часто бывают зависимыми.

2.1.5. Основные положения регрессионного анализа

В регрессионном анализе объясняемая переменная y представляется в виде функции

$$y = M_y\,(x_1, x_2, ..., x_n) + \varepsilon, \tag{2.15}$$

где $M_y\,(x_1, x_2, ..., x_n)$ – функция, значение которой является условным математическим ожиданием величины y, полученным при данном наборе значений объясняющих переменных (функция регрессии); ε – случайная составляющая.

В случае парной регрессии

$$y = M_y\,(x) + \varepsilon. \tag{2.16}$$

Для линейной парной регрессии вид модели:

$$y_i = \alpha + \beta x_i + \varepsilon, \tag{2.17}$$

где $(x_i; y_i)$ – элементы выборки, содержащей n пар значений переменных.

Стандартные предположения регрессионного анализа (предположения о процессе порождения данных):

1. В модели не все значения $x_1, x_2, ..., x_n$ совпадают между собой (условие идентифицируемости), тогда можно вычислить величину $\sum_{i=1}^{n}(x_i - \bar{x})^2 \neq 0$, входящую в формулы числовых характеристик величины x.

2. Значения y_1, y_2, ..., y_n получаются наложением на значения $(\alpha + \beta x_i)$ случайных ошибок ε_i, то есть значения $(\alpha + \beta x_i)$ рассматриваются как некоторые постоянные, хотя и неизвестные наблюдателю. А значения $y_1, y_2,..., y_n$ носят случайный характер, определяемый случайным характером ε_i. Это условие предполагает отсутствие автокорреляционной зависимости остатков от номера наблюдения.

3. Ошибки ε_1, ε_2, ..., ε_n – независимые случайные величины, имеющие нормальное распределение. Соответственно и y_1, y_2, ..., y_n – независимые случайные величины. Математическое ожидание ошибок $M\varepsilon_i = 0$.

Некоторые сведения из теории вероятностей

Предположим, что задана функция распределения случайной величины ε_i, *то есть для каждого* $-\infty \le x \le \infty$ *определена вероятность* $F(x)$ *того, что наблюдаемое значение отклонения* ε_i *не превзойдет величину x, причем эта вероятность не зависит от номера наблюдения* $i = 1, 2, ..., n$

$$F(x) = P(\varepsilon_i\ x). \tag{2.18}$$

Тогда плотность распределения вероятностей ε_i, *то есть закон распределения*

$$f(x) = F'(x). \tag{2.19}$$

Нормальным называют распределение вероятностей непрерывной случайной величины, которое описывается плотностью

$$f(x) = \frac{1}{\delta\sqrt{2\pi}} e^{-(x-m)^2/2\delta^2}, \tag{2.20}$$

где m *– математическое ожидание*; δ *– среднеквадратическое (стандартное) отклонение.*

График плотности нормального распределения называют нормальной кривой (кривой Гаусса), он представлен на рис. 2.8.

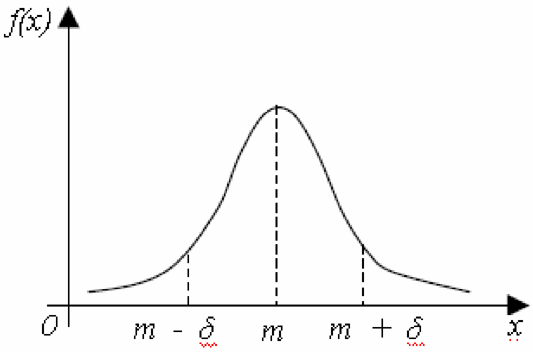

Рис. 2.8

График симметричен прямой $x = m$, *точки* $m - \delta$ и $m + \delta$ *являются точками перегиба. Площадь фигуры между графиком и осью ох равна 1.*

Нормальное распределение часто приемлемо для описания случайных ошибок в моделях наблюдений, где результирующая ошибка является следствием сложения большого количества независимых случайных ошибок, каждая из которых достаточно мала.

4. Дисперсия ошибок ε_i (соответственно, и величины y_i) постоянна для любого i.

$$\operatorname{var}(\varepsilon_i) = \delta^2(\varepsilon_i) = const, \quad \operatorname{var}(y_i) = \delta^2(y_i) = const. \tag{2.21}$$

Это предположение называют условием *равноизменчивости (гомоскедастичности)* ошибок ε_i и, соответственно, объясняемой переменной y_i. Под *гетероскедастичностью* ошибок ε_i и объясняемой переменной y_i понимают условие

$$\operatorname{var}(\varepsilon_i) = \delta^2(\varepsilon_i) \ne const, \quad \operatorname{var}(y_i) = \delta^2(y_i) \ne const. \tag{2.22}$$

2.2. Парная линейная регрессия

2.2.1. Метод наименьших квадратов

Ранее отмечалось, что если между переменными x и y существует теоретическая линейная связь в виде

$$y = \alpha + \beta x, \tag{2.23}$$

то наблюдаемые значения x_i, y_i, $i = 1, 2, ..., n$ этих переменных связаны линейной моделью наблюдений

$$y_i = (\alpha + \beta x_i) + \varepsilon_i\,,\ i = 1, 2, ..., n \tag{2.24}$$

Если α и β – истинные значения параметров линейной модели связи, то величина $\varepsilon_i = y_i - (\alpha + \beta x_i)$ представляет собой ошибку в i-м наблюдении.

Поиск коэффициентов α и β осуществляется таким образом, чтобы величина ε_i стремилась к минимуму (в идеале к нулю). Если $\varepsilon_i = 0$, то все точки лежат на одной прямой. В результате получают подобранную модель линейной связи

$$\hat{y}_i = a + bx_i\,. \tag{2.25}$$

В подобранной модели наблюдаемому значению x_i переменной x сопоставляется значение $\hat{y}_i$ переменной y. Значения подобранного $\hat{y}_i$ и реального наблюдаемоего y обычно отличаются. Разность

$$e_i = y_i - \hat{y}_i = y_i - (a + bx_i) \tag{2.26}$$

называется остатком в i-м наблюдении.

Для реальных данных, как правило, все остатки отличны от нуля, так что часть из них имеет положительный знак, а остальные – отрицательный. При этом необходимо соблюдение *принципа наименьших квадратов*

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \to \min; \quad \sum_{i-1}^{n} (y_i - a - bx_i)^2 \to \min. \tag{2.27}$$

Получаемые при этом оценки a и b называются оценками наименьших квадратов. Свойством оценок наименьших квадратов является то, что соответствующая им прямая проходит через точку $(\bar{x}, \bar{y})$. Поиск пары чисел a и b с помощью метода наименьших квадратов (МНК) сводятся к математической задаче поиска точки минимума функции двух переменных. В результате получаем коэффициенты в подобранной модели

$$b = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)}; \tag{2.28}$$

$$a = \bar{y} - b\bar{x}\,, \tag{2.29}$$

где $\operatorname{cov}(x, y) = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n} = \overline{xy} - \bar{x} \cdot \bar{y}, \quad \operatorname{var}(x) = \dfrac{\sum_{i=1}^{n} (x - \bar{x})^2}{n}.$

При подстановке в формулу (2.3) выражения (2.7) получаем оценку уравнения парной линейной регрессии (функция регрессии)

$$\hat{y}_i = \bar{y} - b\bar{x} + bx_i\,. \tag{2.30}$$

При выполнении стандартных предположений регрессионного анализа (предпосылок МНК) МНК-оценки параметров уравнения регрессии будут обладать следующими *статистическими свойствами*:

1. Несмещенность. Статистическая оценка некоторого параметра называется *несмещенной*, если ее математическое ожидание равно истинному значению этого параметра. В случае парной линейной регрессии: $M(a) = \alpha$, $M(b) = \beta$.

*2. Состоятельность.*При неограниченном возрастании объема выборки значение оценки должно стремиться по вероятности к истинному значению параметра, а дисперсии оценок параметров должны уменьшаться и в пределе стремиться к 0: $\mathrm{var}(a) \to 0$, $\mathrm{var(b)} \to 0$ при $\mathrm{n} \to \infty$.

*3. Эффективность.*Оценка называется эффективной, если она имеет минимальную дисперсию по сравнению с другими оценками заданного класса.

2.2.2. Использование оцененной модели для прогноза

Пусть мы имеем модель наблюдений в виде

$$y_i = \alpha + \beta x_i + \varepsilon_i,\ i = 1, 2, \ldots, n \tag{2.31}$$

и хотим дать прогноз, каким будет значение объясняемой переменной y^n при некотором выбранном (фиксированном) значении x^n объясняющей переменной x, если мы будем продолжать наблюдения.

Мы умеем оценивать коэффициенты α и β методом наименьших квадратов и получать подобранную модель

$$\hat{y} = a + bx, \tag{2.32}$$

где $\hat{y}$ – прогнозируемое значение объясняемой переменной.

Вопрос: насколько надежным является выбор такого значения в качестве прогнозируемого?

Выбирая в качестве прогноза в $(n+1)$-м наблюдении для y_0 значение $\hat{y}_0 = a + bx_0$, мы тем самым допускаем ошибку прогноза

$$\hat{y}_0 - y_0 = (a + bx_0) - (\alpha + \beta_0 x + \varepsilon_0) = (a - \alpha) + (b - \beta)x_0 + \varepsilon_0. \tag{2.33}$$

Эта ошибка является следствием:

- неопределенности, связанной с отклонением вычисленных значений случайных величин a и b от истинных значений параметров α и β;
- неопределенности, связанной со случайной ошибкой ε_0 в $(n+1)$-м наблюдении.

При наших стандартных предположениях о линейной модели ошибка прогноза является случайной величиной с математическим ожиданием

$$M(\hat{y} - y_0) = \mathrm{M}(a - \alpha) + x_0 M(b - \beta) + M(\varepsilon_0) = 0;$$
$$[M(a) = \alpha;\ M(b) = \beta;\ M(\varepsilon_0) = 0]. \tag{2.34}$$

Теоретическая точность такого прогноза характеризуется дисперсией ошибки прогноза

$$\mathrm{var}(\hat{y} - y_0) = \mathrm{var}(\varepsilon_0) = 0. \tag{2.35}$$

Далее мы будем рассматривать оценку этой дисперсии $S\varepsilon_i$, то есть оценку дисперсии ошибки прогноза (дисперсию остатков)

$$S^2{}_{\varepsilon_i} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-m} == \frac{\sum_{i=1}^{n} e_i^2}{n-m}, \tag{2.36}$$

где n – количество наблюдений; m – количество параметров уравнения регрессии.

2.2.3. Интервальная оценка функции регрессии и ее параметров

Доверительный интервал для среднего значения y, полученного по функции регрессии $M_{\hat{y}}(x)$

Строится с использованием *t*-статистики (распределения Стьюдента).

Задается надежность (доверительная вероятность) $\gamma = 1 - \alpha$ (0,95–0,99), с которой значение, полученное по уравнению регрессии, должно находиться в доверительном интервале.

Уравнение регрессии

$$\hat{y} = a + bx; \quad \hat{y} = \bar{y} - b\bar{x} + bx, \tag{2.37}$$

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

Остаток в i-м наблюдении

$$e_i = y_i - \hat{y}_i = y_i - \bar{y} - b(x - \bar{x}) = y_i - (a + bx_i). \tag{2.38}$$

На рис. 2.9 линия регрессии изображена графически. Для произвольного наблюдения значения y_i выделены его составляющие: среднее значение – $\bar{y}$, приращение $b(x_i - \bar{x})$, образующие расчетное значение $\hat{y}_i$ и остаток e_i регрессии, $tg\,\alpha = b$.

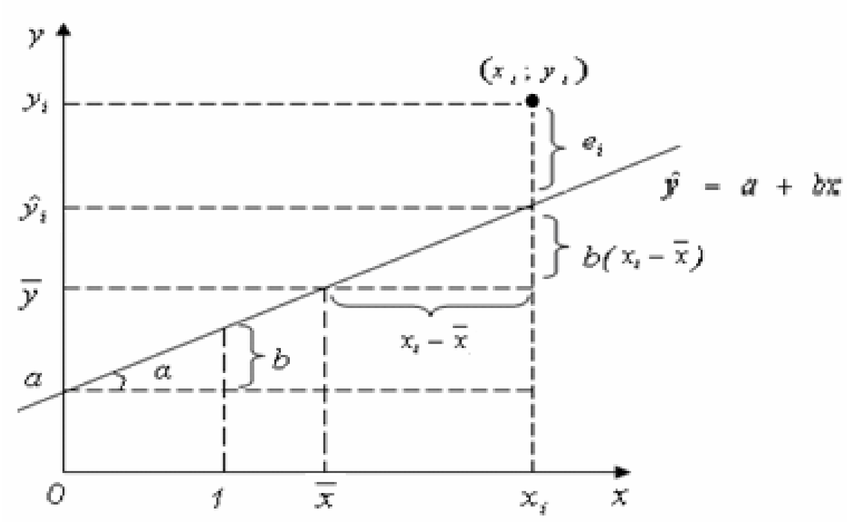

Рис. 2.9

Вариация (дисперсия) группового среднего значения $\hat{y}$ равна сумме дисперсий двух слагаемых выражения

$$\hat{y} = \bar{y} + b(x - \bar{x}); \tag{2.39}$$

$$\mathrm{var}(y) = \mathrm{var}(\bar{y}) + \mathrm{var}[b(x - \bar{x}]. \tag{2.40}$$

Запишем выражение (2.40) через стандартное отклонение и вынесем неслучайную величину $(x - \bar{x})$ за знак дисперсии, возведя ее в квадрат:

$$\delta_{\hat{y}}{}^2 = \delta_{\bar{y}}^2 + \delta_b^2 (x - x_i)^2. \tag{2.41}$$

Следовательно, оценка дисперсии значения $\hat{y}$, найденного по уравнению регрессии, складывается из дисперсии среднего значения $\bar{y}$ и дисперсии *b*.

$$S_{\hat{y}}^2 = S_{\varepsilon_i}^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right). \tag{2.42}$$

Доверительный интервал для математического ожидания $\hat{y}_i$, найденного по уравнению регрессии

$$\hat{y} - t_{1-\alpha;k} S_{\hat{y}} \le M_{\hat{y}}(x) \le \hat{y} + t_{1-\alpha;k} S_{\hat{y}} \ , \tag{2.43}$$

где $t = \dfrac{\hat{y} - M_{\hat{y}}(x)}{S_{\hat{y}}}$ – статистика, имеющая t – распределение Стьюдента с $k = n$-2 степенями свободы; n – объем выборки.

Величина доверительного интервала зависит от объясняющей переменной x (рис. 2.2). При $x = \bar{x}$ она стремиться к минимуму, по мере удаления x от $\bar{x}$ увеличивается.

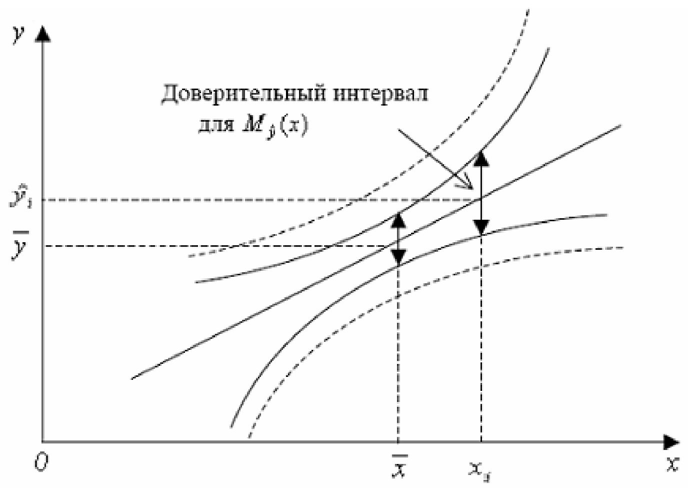

Рис. 2.10

Таким образом, прогноз значений объясняемой переменной y по уравнению регрессии оправдан, если значение x объясняющей переменной не выходит за диапазон наблюдаемых значений x_0.

Доверительный интервал для индивидуальных значений зависимой переменной

Построенная на рис. 2.10 доверительная область для $M_{\hat{y}}(x)$ определяет местоположение условного математического ожидания или среднего значения зависимой переменной (модельной линии регрессии), но не отдельных возможных значений зависимой переменной, которые отклоняются от средней.

Поэтому при определении доверительного интервала для индивидуальных значений y_0 зависимой переменной учитывают еще один источник вариации – рассеяния вокруг линии регрессии, то есть в оценку суммарной дисперсии $S^2_{\hat{y}}$ включается величина $S^2_{\varepsilon_i}$. В результате оценка дисперсии индивидуальных значений y_0 при $x = x_0$ равна

$$S^2_{\hat{y}_0} = S^2_{\varepsilon_i}\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\right]. \tag{2.44}$$

Соответствующий доверительный интервал для прогнозов индивидуальных значений y_0 определяется по формуле:

$$\hat{y}_0 - t_{1-\alpha;\ n-2} S_{\hat{y}0} \le y_0 \le \hat{y}_0 + t_{1-\alpha;\ n-2} S_{\hat{y}0}. \tag{2.45}$$

На рис. 2.10 данный доверительный интервал показан пунктиром.

Доверительные интервалы для параметров регрессионной модели

Наряду с интервальными оценками функции регрессии иногда представляет интерес построение доверительных интервалов для параметров регрессионной модели ($y_i = \alpha + \beta x_i + \varepsilon_i$), в частности для α, β и $\delta^2 = \mathrm{var}(\varepsilon_i)$ (дисперсии возмущения ε_i или зависимой переменной y_i).

При построении доверительного интервала параметра β исходят из того, что статистика $t = \frac{b-\beta}{S}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ имеет t – распределение Стьюдента с $k = n - 2$ степенями свободы. Интервальная оценка параметра β на уровне значимости α имеет вид:

$$b - t_{1-\alpha;\ n-2}\frac{S_{\varepsilon_i}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \le \beta \le b + t_{1-\alpha;\ n-2}\frac{S_{\varepsilon_i}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}. \qquad (2.46)$$

Поскольку знак коэффициента регрессии указывает на рост результативного признака y при увеличении признака-фактора x ($b > 0$), уменьшение результативного признака при увеличении признака-фактора ($b < 0$) или его независимость от независимой переменной ($b = 0$) (рис. 2.11), то границы доверительного интервала для коэффициента регрессии не должны содержать противоречивых результатов, например, $-1,5 \le b \le 0,8$. Такого рода запись указывает на то, что истинное значение коэффициента регрессии одновременно содержит положительные и отрицательные величины и даже ноль, чего не может быть.

Интервальная оценка параметра α на уровне значимости α имеет вид:

$$\alpha - t_{1-\alpha;\ n-2}S_{\varepsilon_i}\sqrt{\frac{\sum_{i=1}^{n}(x_i)^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}} \le \alpha \le \alpha + t_{1-\alpha;\ n-2}S_{\varepsilon_i}\sqrt{\frac{\sum_{i=1}^{n}(x_i)^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}}. \qquad (2.47)$$

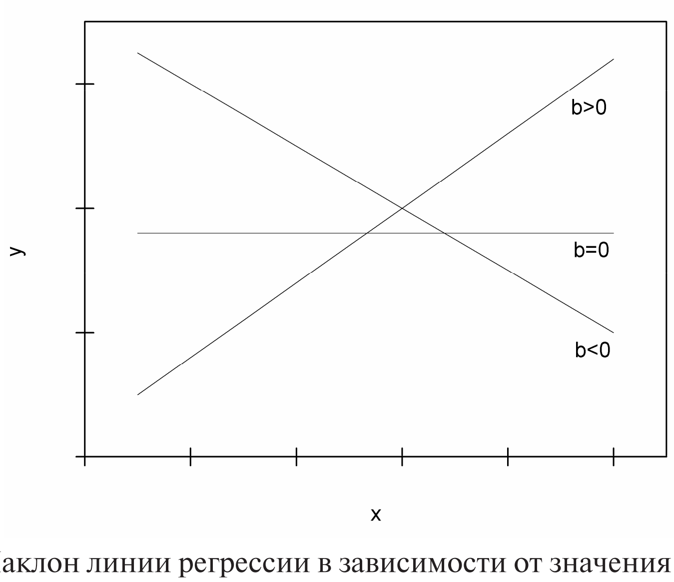

Рис. 2.11. Наклон линии регрессии в зависимости от значения параметра b

При построении доверительного интервала для параметра δ^2 исходят из того, что статистика nS^2/δ^2 имеет χ^2-распределение с $k = n - 2$ степенями свободы. Поэтому интервальная оценка для δ^2 на уровне значимости α имеет вид:

$$\frac{nS^2_{\varepsilon_i}}{\chi^2_{\frac{\alpha}{2};\ n-2}} \le \delta^2 \le \frac{nS^2_{\varepsilon_i}}{\chi^2_{\frac{1-\alpha}{2};\ n-2}}\ . \qquad (2.48)$$

Доверительный интервал выбирается таким образом, чтобы вероятность

$$P\left(\chi^2 \langle \chi^2_{\frac{1-\alpha}{2};n-2}\right) = P\left(\chi^2 > \chi^2_{\frac{1-\alpha}{2};n-2}\right) = \frac{\alpha}{2}. \qquad (2.49)$$

2.2.4. Оценка значимости уравнения регрессии (адекватно имеющимся статистическим данным)

Основная идея дисперсионного анализа

Для того чтобы проверить значимость уравнения регрессии, необходимо установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

На рис. 2.12 изображены наблюдаемые значения переменных x_i, y_i, соответствующая этим значениям линия регрессии, обозначены составляющие регрессии: $AB = y_i - \bar{y}$; $AC = y_i - \hat{y}$; $CB = \hat{y}_i - \bar{y}$.

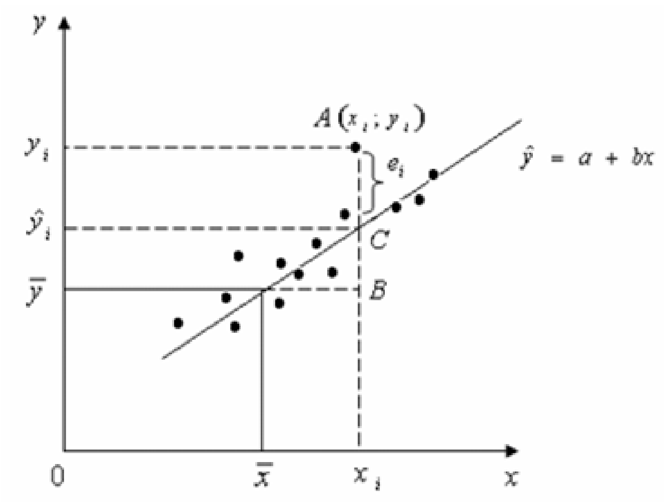

Рис. 2.12

По рисунку $AB = CB + AC$ или $y_i - \bar{y} = (y_i - \hat{y}) + (\hat{y}_i - \bar{y})$, где $\hat{y}_i = a + bx_i$ – ордината точки прямой, соответствующей уравнению регрессии, имеющей абсциссу x_i.

Возведя обе части в квадрат и просуммировав выражение для каждого i-го случая, имеем

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \approx$$
$$\approx \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \qquad (2.50)$$

где $Q = \sum_{i=1}^{n}(y_i - \bar{y})^2$ – полная сумма квадратов; $Q_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ – сумма квадратов, объясненная моделью; $Q_\ell = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$ – остаточная сумма квадратов.

Если поделить выражение на *n,* то получим

$$\mathrm{var}(y) = \mathrm{var}(\hat{y}) + \mathrm{var}(e), \tag{2.51}$$

то есть дисперсия переменной *y* частично объясняется изменчивостью $\hat{y}$, а частично изменчивостью остатка регрессии.

Оценка этой изменчивости

$$S_R^2 = \frac{Q_R}{m-1}, \tag{2.52}$$

где S_R^2 – обусловлена уравнением регрессии или объясняющей переменной.

$$S_{\varepsilon_i}^2 = \frac{Q_e}{n-m}, \tag{2.53}$$

здесь $S_{\varepsilon_i}^2$ – воздействием неучтенных случайных факторов и ошибок.

Процедура проверки значимости линейной связи между переменными

Эта процедура будет иметь смысл при соблюдении стандартных предположений о модели.

Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации:

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\% = \sum_{i=1}^{n}\left|\frac{e_i}{y_i}\right| . \tag{2.54}$$

Средняя ошибка аппроксимации не должна превышать 8–10%.

Оценка значимости уравнения регрессии в целом производится на основе *F*-критерия Фишера, которому предшествует дисперсионный анализ. Согласно основной идее дисперсионного анализа, общая сумма квадратов отклонений переменной у от среднего значения $\bar{y}$ раскладывается на две части – «объясненную» (S_R^2) и «необъясненную» ($S_{\varepsilon_i}^2$).

Случайные величины $S_R^2 = \dfrac{Q_R}{m-1}$ и $S_{\varepsilon_i}^2 = \dfrac{Q_e}{n-m}$ имеют χ^2-распределение соответственно с (m-1) и (n-m) степенями свободы. Поэтому уравнение регрессии значимо на уровне α, если фактически наблюдаемое значение статистики

$$F = \frac{Q_R(n-m)}{Q_\ell(m-1)} = \frac{S_R^2}{S_{\varepsilon_i}^2} > F_{\alpha;\,k_1;\,k_2}, \tag{2.55}$$

где $Q_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ – сумма квадратов, объясненная моделью; $Q_\ell = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}e_i^2$ – остаточная сумма квадратов; m – число оцениваемых параметров уравнения регрессии; n – число наблюдений; $F_{\alpha;\,k_1;\,k_2}$ – табличное значение критерия.

В случае линейной парной регрессии $m = 2$ и уравнение регрессии значимо на уровне значимости α, если

$$F = \frac{Q_R(n-2)}{Q_e} > F_{\alpha:1:\;n-2}. \tag{2.56}$$

Одной из наиболее эффективных оценок значимости уравнения регрессии является *коэффициент детерминации*. Он характеризует степень выраженности связи между переменными. Определяется по формуле

$$R^2 = \frac{Q_R}{Q} = 1 - \frac{Q_\ell}{Q}. \tag{2.57}$$

Величина R^2 показывает, какая доля вариации зависящей переменной обусловлена вариацией объясняющей переменной. Так как $0 \le Q_R \le Q$, то $0 \le R^2 \le 1$.

Свойства коэффициента детерминации:

1. При $R^2 \to 1$ регрессия хорошо отражает эмпирические данные, наблюдения примыкают к линии регрессии.

2. При $R^2 = 1$ точки $(x_i; y_i)$ лежат на линии регрессии и между переменными существует линейная функциональная зависимость.

3. При $R^2 = 0$ вариация зависимой переменной полностью обусловлена воздействием неучтенных в модели переменных и линия регрессии параллельна оси OX.

Если известен коэффициент детерминации R^2, то критерий значимости уравнения регрессии

$$F = \frac{R^2(n-m)}{(1-R^2)(m-1)} > F_{\alpha;\ k_1;\ k_2}. \tag{2.58}$$

В случае линейной регрессионной модели коэффициент детерминации равен квадрату коэффициента корреляции.

$$R^2 = r_{xy}^{\ 2}. \tag{2.59}$$

Оценка статистической значимости коэффициентов парной линейной регрессии и корреляции

В парной линейной регрессии оценивается значимость не только уравнения в целом, но и отдельных его параметров. С этой целью по каждому из параметров определяется его стандартная ошибка: S_b и S_a.

Стандартная ошибка коэффициента регрессии определяется по формуле

$$S_b = \frac{S_{\varepsilon_i}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}, \tag{2.60}$$

где $S^2{}_{\varepsilon_i} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-m} == \frac{\sum_{i=1}^{n} e_i^2}{n-m}$ – остаточная дисперсия.

Величина стандартной ошибки совместно с t-распределением Стьюдента при $(n - 2)$ степенях свободы применяется для проверки существенности коэффициента регрессии и для расчета его доверительного интервала.

Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой. Определяется фактическое значение t-критерия Стьюдента $t_b = \frac{b}{S_b}$, которое затем сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы $(n - 2)$. Доверительный интервал для коэффициента регрессии определяется как $b \pm t_{1-\alpha;\ n-2} \cdot S_b$, этот процесс был описан в п. 2.3.3.

Стандартная ошибка параметра α определяется по формуле

$$S_a = S_{\varepsilon_i} \sqrt{\frac{\sum_{i=1}^{n}(x_i)^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}}. \tag{2.61}$$

Процедура оценивания существенности данного параметра не отличается от рассмотренной выше для коэффициента регрессии. Вычисляется t-критерий: $t_a = \frac{a}{S_a}$, его величина сравнивается с табличным значением при $n-2$ степенях свободы.

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции S_r:

$$S_r = \sqrt{\frac{1-r^2}{n-2}}. \tag{2.62}$$

Фактическое значение t-критерия Стьюдента определяется как $t_r = \frac{r}{S_r}$.

Существует связь между t-критерием Стьюдента и F-критерием Фишера:

$$t_b = t_r = \sqrt{F}\,. \tag{2.63}$$

2.2.5. Проверка выполнения стандартных предположений об ошибках в линейной модели наблюдений графическим методом

Перед тем как построить модель наблюдений (см. п. 1.5), мы сделали четыре стандартных предположения о процессе порождения данных. Первое из этих предположений (в модели не все значения $x_1, x_2, \ldots, x_n$ совпадают между собой) проверяется перед построением регрессионной модели. Остальные три можно проверить только тогда, когда уравнение регрессии уже составлено. Без соблюдения этих условий построенная модель теряет смысл.

Оцененная модель проверяется на отсутствие автокорреляционной зависимости остатков от номера наблюдения, на независимость случайных ошибок $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$, математическое ожидание которых должно стремиться к нулю ($M\varepsilon_i = 0$), на постоянство или гомоскедастичность дисперсии ошибок [$\mathrm{var}(\varepsilon_i) = \delta^2(\varepsilon_i) = \mathrm{const}$]. Анализ соблюдения перечисленных условий (дисперсионный анализ) проводят, используя графики стандартизированных остатков

$$C_i = \frac{y_i - \hat{y}_i}{S_{\varepsilon_i}} = \frac{e_i}{S_{\varepsilon_i}} (i = 1, 2, \ldots, n), \tag{2.64}$$

где $S^2_{\varepsilon_i}$ – оценка дисперсии остатков,

$$S^2_{\varepsilon_i} = \frac{Q_e}{n-m} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-m}. \tag{2.65}$$

Графики позволяют выявлять типичные отклонения от стандартных предположений о модели наблюдений по характеру поведения остатков. При большом количестве наблюдений поведение стандартизированных остатков имитирует поведение ошибок ε_i.

Наиболее часто используют графики зависимости стандартизированных остатков (как ординат):

- от оцененных значений y (по оси абсцисс);
- отдельных объясняющих переменных;
- номера наблюдения, если наблюдения производятся в последовательные моменты времени равными интервалами.

График зависимости C_i от $\hat{y}_i$ позволяет выявить три довольно распространенных нарушения стандартных предположений о модели наблюдений:

1. Выделяющиеся наблюдения – наличие отдельных наблюдений, для которых либо математическое ожидание ошибок отлично от нуля $M(\varepsilon_i) \neq 0$, либо дисперсия ошибки $\mathrm{var}(\varepsilon_i)$ существенно превышает величину дисперсий остальных ошибок. Они могут обнаруживать себя на графике как наблюдения со слишком большими по величине остатками (рис. 2.12).

2. Неоднородность дисперсии (гетероскедастичность), например,в форме той или иной функциональной зависимости $\mathrm{var}(\varepsilon_i)$ от величины $\hat{y}_i$. Если рассматриваемый график имеет вид, изображенный на рис. 2.12, то это, скорее всего, отражает рост дисперсий ошибок с ростом значений $\hat{y}_i$.

3. Неправильная спецификация модели в отношении множества объясняющих переменных, приводящих к нарушению условия $M(\varepsilon_i) = 0$ (рис. 2.13).

График зависимости c_i *от значений* x_{ij}(j = 1, 2, ..., n) j-й объясняющей переменной помогает выявить нелинейную зависимость y от j-й объясняющей переменной в случае множественной регрессионной модели (рис. 2.14, 2.15).

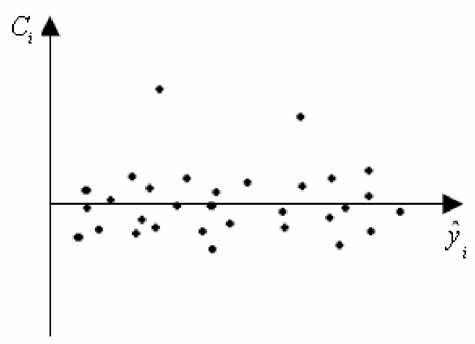

Рис. 2.13

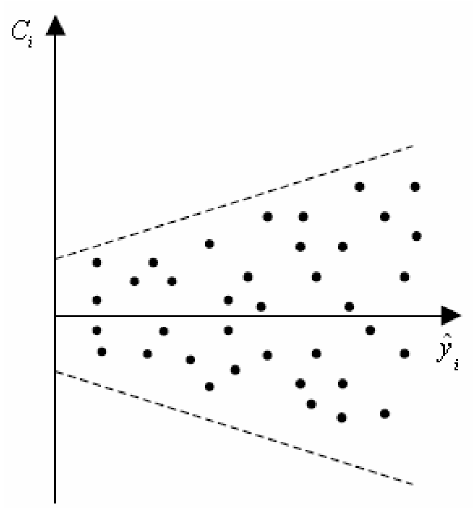

Рис. 2.14

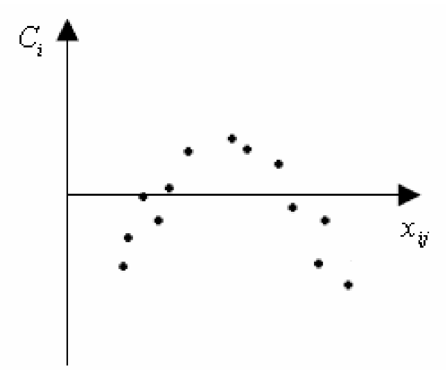

Рис. 2.15

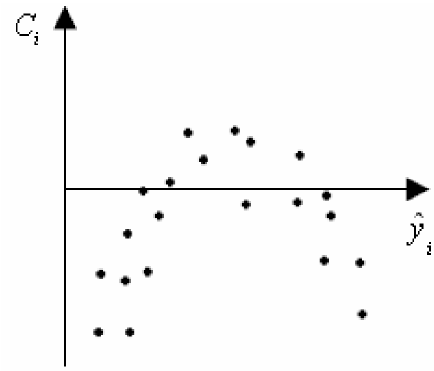

Рис. 2.16

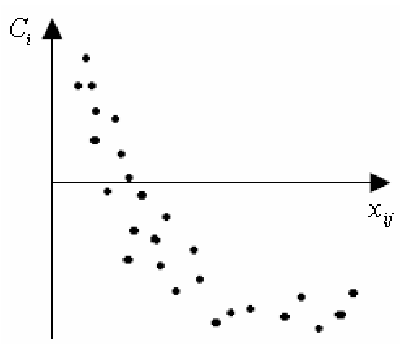

Рис. 2.17

График зависимости остатков от номера наблюдения полезен в случае, когда наблюдения производятся последовательно во времени (через равные интервалы времени). По такому графику можно обнаружить:

1. Изменение дисперсии ошибок var(ε_i) с течением времени.

2. Невключение в модель переменных, зависящих от времени и существенно влияющих на объясняемую переменную.

3. Невыполнение условия независимости в совокупности случайных ошибок ε_i в форме их автокоррелированности (определенной функциональной зависимости). График остатков в случае положительной автокоррелированности приведен на рис. 2.18 и в случае отрицательной – на рис. 2.19.

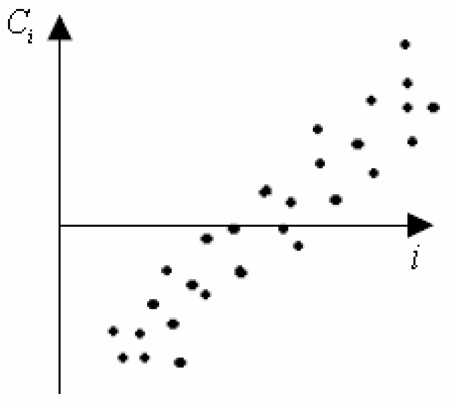

Рис. 2.18

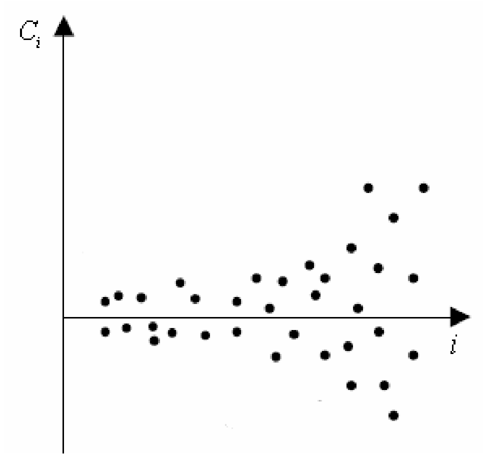

Рис. 2.19

В первом случае проявляется тенденция сохранения знака остатка при переходе к следующему наблюдению (за положительным остатком скорее следует также положительный, а за отрицательным – отрицательный). Во втором случае проявляется тенденция смены знака остатка при переходе к следующему наблюдению (за положительным остатком скорее следует отрицательный, а за отрицательным – положительный).

Помимо графических, существует довольно много процедур, предназначенных для проверки стандартных предположений о линейной модели наблюдений, использующих статистические критерии.

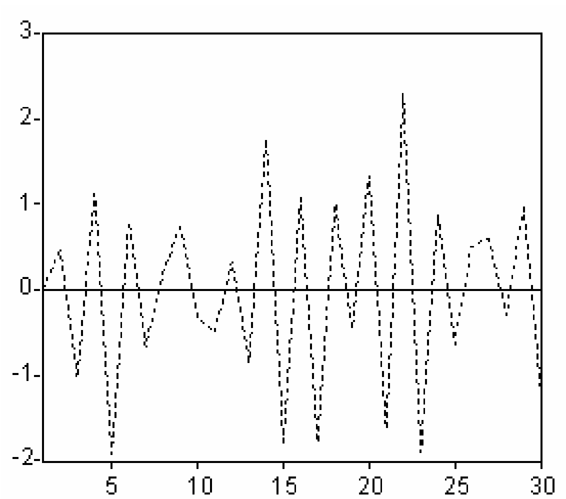

Рис. 2.20

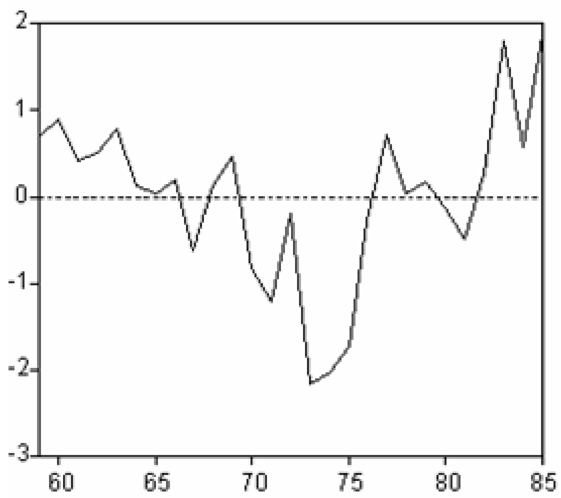

Рис. 2.21

2.3. Нелинейные регрессионные модели

2.3.1. Нелинейные модели с двумя переменными, приводимые к линейной форме

Степенная форма эконометрической модели

Связь между конкретными экономическими факторами не обязательно линейная. Например, если рассматривать зависимость от располагаемого дохода x не всех затрат на личное потребление y, а лишь затрат V на некоторый продукт питания (или группу продуктов питания), например, на куриные яйца, то уже по чисто физиологическим причинам функция связи $V = V(x)$, скорее всего, должна замедлять свой рост при возрастании x. Возможный график этой функции имеет вид, приведенный на рис. 2.22.

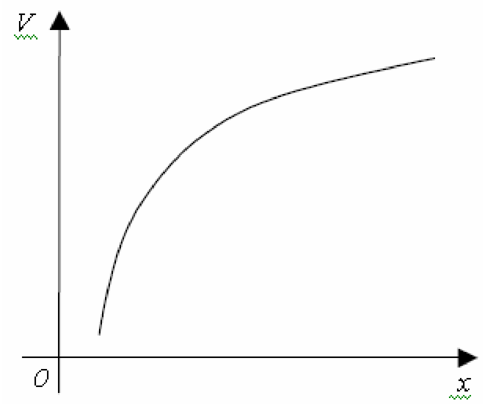

Рис. 2.22

В данной ситуации нельзя говорить о склонности к потреблению данного продукта как о постоянной величине. Вводят понятие предельной склонности к потреблению.

Определение. *Предельной склонностью к потреблению (нормой потребления)* называют величину $C(x)$, которая для заданной величины располагаемого дохода x определяется формулой

$$C(x) = \lim_{\Delta x \to 0} \frac{V(x+\Delta x) - V(x)}{\Delta x} = \frac{dV}{dx} = V'(x). \tag{2.66}$$

Замедление скорости роста функции $V(x)$ соответствует убыванию $C(x)$ при возрастании располагаемого дохода x. Уточняя предположение о поведении нормы потребления $C(x)$, можно получить ту или иную формулу связи между переменными V и x.

Возможной формой связи затрат V на некоторый продукт питания от располагаемого дохода x может быть степенная связь

$$V = V(x) = \alpha \cdot x^{\beta}, \tag{2.67}$$

где $\alpha > 0$, $0 < \beta < 1$.

Для такой связи производная (норма потребления) $C(x) = \alpha \cdot \beta \cdot x^{\beta-1}$, то есть эта склонность к потреблению монотонно убывает с ростом располагаемого дохода x.

Приведение степенной модели к линейной форме модели.
Оценка параметров модели и ее качества

Если вместо уровней дохода и расходов на потребление рассмотреть логарифмы уровней по одному и тому же основанию (натуральные или десятичные log), то степенную форму связи можно привести к линейной.

$$V = \alpha \cdot x^{\beta} \quad \to \quad \ln V = \ln \alpha x^{\beta}.$$

$$\ln V = \ln\alpha + \ln x^{\beta};$$

$$\ln V = \ln\alpha + \beta \ln x. \qquad (2.68)$$

Если обозначить $\ln V = V'$, $\ln \alpha = \alpha'$, $\ln x = x'$, то получим

$$V' = \alpha' + \beta \cdot x'. \qquad (2.69)$$

Линейной форме модели в логарифмах соответствует линейная модель наблюдений

$$V_i' = \alpha' + \beta x_i' + \varepsilon_i \quad i = 1, 2, \ ..., \ n, \qquad (2.70)$$

где $\beta = \dfrac{d(\ln V)}{d(\ln x)}$.

Модель $y_i = \alpha \cdot x_i^{\beta} \cdot \varepsilon_i$, линеаризуемая путем логарифмирования, называется *степенной моделью с мультипликативными возмущениями*.

Оценить параметры такой модели можно с помощью МНК, применив данный метод к уравнению (2.68) и, далее, перейдя от логарифмов к исходным показателям. Качество непосредственно степенной модели (2.67) оценивают графическим методом (сравнивая облако рассеяния с полученной по теоретической модели кривой) с помощью средней ошибки аппроксимации и коэффициента детерминации. Соблюдение стандартных предположений регрессионного анализа проще всего проследить, используя графики стандартизированных остатков (см. п. 2.2.5). О качестве исходной модели (2.67) также можно судить по модели (2.68), применив методы, изложенные в п. 2.2.

Понятие предельной склонности и эластичности функции

Определение. Если имеется связь между какими-то переменными экономическими факторами x и y в виде $Y = f(x)$, то функция

$$C(x) = \frac{dy}{dx} = f'(x) = \lim_{\Delta x \to 0} \frac{y(x + \Delta x) - y(x)}{\Delta x} \qquad (2.71)$$

является *предельной склонностью* величины y по отношению к величине x.

Определение. Величина

$$\eta(x) = \lim_{\Delta x \to 0} \frac{\dfrac{f(x + \Delta x) - f(x)}{f(x)} \cdot 100\%}{\dfrac{\Delta x}{x} \cdot 100\%} == \frac{x}{y} \cdot \frac{dy}{dx}, \qquad (2.72)$$

$$\eta(x) = \frac{x}{y} C(x) \qquad (2.73)$$

в экономической теории называется *эластичностью (функцией эластичности)*.

Эластичность показывает, на какое количество процентов изменится y при изменении x на 1%.

Если $|\eta(x)| > 1$, то фактор y эластичен по отношению к фактору x.

Если же $|\eta(x)| < 1$, то фактор y неэластичен по отношению к фактору x. Отдельно выделяют пограничные случаи $\eta(x) = \pm 1$.

Найдем отношение дифференциалов для логарифмов уровней факторов y и x.

$$\frac{d \ln f(x)}{d \ln x} = \frac{\dfrac{df(x)}{f(x)}}{\dfrac{dx}{x}} = \frac{x}{y} \frac{dy}{dx} = \eta(x). \qquad (2.74)$$

Значение предельной склонности $C(x_0)$ равно угловому коэффициенту наклона касательной к графику функции $y = f(x)$ при $x = x_0$.

Значение эластичности $\eta(x_0)$ равно угловому коэффициенту касательной к графику зависимости $\ln y$ от $\ln x$ при $x = x_0$.

Следствие

1. Условие постоянства предельной склонности $C(x_0) = \beta$ означает линейную связь между уровнями факторов x и y:

$$y = \alpha + \beta x. \tag{2.75}$$

2. Условие постоянства эластичности $\eta(x_0) = \beta$ означает линейную связь между логарифмами уровней

$$\ln y = \alpha' + \beta \ln x, \tag{2.76}$$

соответствующую степенной связи между уровнями факторов. Преобразовав выражение (2.76), можно получить формулу (2.67):

$$e^{\ln y} = e^{(\alpha' + \beta \ln x)}.$$

$$y = e^{\alpha'}(e^{\ln x})^{\beta} \rightarrow y = e^{\alpha'} \cdot x^{\beta} \rightarrow y = \alpha x^{\beta} \ (\alpha = e^{\alpha'}).$$

Выражает степенное возрастание (при $\beta > 0$) или убывание (при $\beta < 0$) уровней фактора y при возрастании уровней фактора x.

Если $\eta(x) = \beta$, то коэффициент β можно трактовать как процентное изменение уровня фактора y при изменении фактора x на 1%.

Если $|\beta| > 1$, то фактор y эластичен по отношению к фактору x.

Если $|\beta| < 1$, то фактор y не эластичен по отношению к фактору x.

Пограничные случаи, когда $\beta = \pm 1$ соответствуют единичной эластичности.

Другие виды эконометрических моделей, приводимые к линейной форме

Если y – объем плановых инвестиций, а z – норма банковского процента, то между ними существует связь (рис. 2.23), которая часто выражается в форме *обратно пропорциональной зависимости:*

$$y = \alpha + \frac{\beta}{z} \ (\alpha > 0, \beta > 0). \tag{2.77}$$

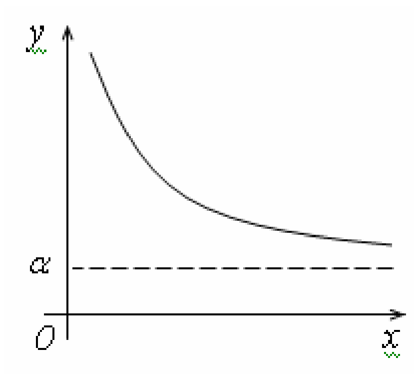

Рис. 2.23

Эта модель приводится к линейной заменой переменной $x = \frac{1}{z}$:

$$y = \alpha + \beta x.$$

В этой модели эластичность y по z отрицательна и меньше единицы по абсолютной величине

$$\eta(z) = \frac{dy}{dz} \cdot \frac{z}{y} = -\frac{\beta}{z^2} \cdot \frac{z}{\alpha + \frac{\beta}{z}} = -\frac{\beta}{\beta + \alpha z}, \tag{2.78}$$

следовательно, объем плановых инвестиций неэластичен по отношению к норме процента (проценты растут быстрее, чем инвестиции).

В моделях «доход-потребление», относящихся к потреблению продуктов питания, линейная модель в логарифмах уровней, выражающая уменьшение нормы потребления $C(x)$ с ростом доходов x, не всегда удовлетворительна, поскольку эластичность в такой модели $\eta = \text{const}$.

Часто более подходящей является *модель связи с убывающей эластичностью* (рис. 2.24). Например:

$$y = \alpha + \beta \ln x. \tag{2.79}$$

Тогда

$$\eta(z) = \frac{dy}{dz} \cdot \frac{z}{y} = \frac{\beta}{z} \cdot \frac{z}{\alpha + \beta \ln z}; \tag{2.80}$$

$$\lim_{z \to \infty}(z) = \lim_{z \to \infty} \frac{\beta}{(\alpha + \beta \mathrm{h}\ z)} = 0. \tag{2.81}$$

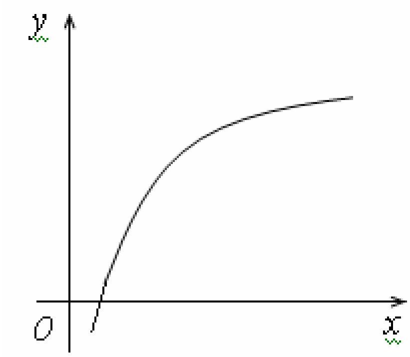

Рис. 2.24

В этой модели возникают проблемы с отрицательными значениями y при малых значениях z.

Такого недостатка нет в модели

$$\ln y = \alpha - \frac{\beta}{z} \quad (\beta > 0), \tag{2.82}$$

то есть

$$y = e^{(\alpha - \frac{\beta}{z})}. \tag{2.83}$$

Здесь

$$\eta(z) = \frac{dy}{dz} \cdot \frac{z}{y} = \frac{e^{(\alpha - \frac{\beta}{z})} * \beta * z}{z^2 e^{(\alpha - \frac{\beta}{z})}};$$

$$\eta(z) = \frac{\beta}{z}. \tag{2.84}$$

Последнее соотношение выражает *закон Энгеля* – убывания эластичности потребления продуктов питания по доходу. Значения y в этой модели ограничены сверху значением e^{α}.

Модели 2.79 и 2.82 сводятся к линейной форме путем перехода от уровней переменных к их логарифмам или обратным величинам.

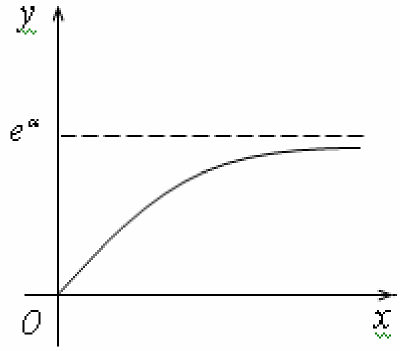

Рис. 2.25

2.3.2. Итерационные методы подбора нелинейных моделей

Если исследователь принимает модель наблюдений

$$\ln y_i = \alpha' + \beta \ln x_i + \varepsilon_i , \tag{2.85}$$

то тем самым он соглашается с видом модели

$$y_i = e^{\alpha} \cdot x_i^{\beta} \cdot e^{\varepsilon_i} \text{ или } y_i = \alpha \cdot x_i^{\beta} \cdot v_i , \tag{2.86}$$

то есть соглашается с мультипликативным вхождением ошибок v_i в нелинейное уравнение y_i.

Однако не исключено, что, по существу, модель должна иметь вид

$$y_i = \alpha \cdot x_i^{\beta} + v_i , \tag{2.87}$$

то есть имеет аддитивные ошибки.

В этой модели взятие логарифмов от обеих частей не приводит к линейной модели наблюдений.

Для получения оценок наименьших квадратов параметров α и β сумму квадратов

$$\sum_{i=1}^{n} (y_i - \alpha x_i^{\beta})^2 = Q(a,b) \tag{2.88}$$

минимизируют, используя итерационные методы, в процессе реализации которых сначала задаются некоторые «стартовые значения» оцениваемых параметров, а затем производится последовательное приближение значений (a и b) оценок параметров α и β, минимизирующих $Q(a, b)$.

2.3.3. Нелинейные модели множественной регрессии

Ранее нами рассматривались примеры некоторых регрессионных моделей с двумя переменными, приводимых к линейной форме путем логарифмирования или замены переменной. Регрессионные модели, имеющие более двух переменных, называются *моделями множественной регрессии*. Приведем перечень нелинейных моделей множественной регрессии, линеаризуемых изложенными методами:

1. Степенная модель с мультипликативными возмущениями

$$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \ldots \cdot x_n^{b_n} \cdot \varepsilon, \tag{2.89}$$

линеаризуется с помощью логарифмирования

$$\ln y = \ln a + b_1 \ln x_1 + b_2 \ln x_2 + \ldots + b_n \ln x_n + \ln \varepsilon. \tag{2.90}$$

2. Для логарифмической, правой и левой полулогарифмических моделей

$$\ln y = a + b_1 \ln x_1 + b_2 \ln x_2 + \ldots + b_n \ln x_n + \varepsilon ; \tag{2.91}$$

$$y = a + b_1 \ln x_1 + b_2 \ln x_2 + \ldots + b_n \ln x_n + \varepsilon; \tag{2.92}$$

$$\ln y = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + \varepsilon \tag{2.93}$$

целесообразно заменить логарифмы уровней.

$$y' = a + b_1 x_1' + b_2 x_2' + \ldots + b_n x_n' + \varepsilon; \tag{2.94}$$

$$y = a + b_1 x_1' + b_2 x_2' + \ldots + b_n x_n' + \varepsilon; \tag{2.95}$$

$$y' = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + \varepsilon. \tag{2.96}$$

3. Экспоненциальная модель

$$y = e^{a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + \varepsilon} \tag{2.97}$$

приводится к виду (2.93) с помощью логарифмирования.

4. Обратная модель

$$\frac{1}{y} = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + \varepsilon \tag{2.98}$$

линеаризуется заменой переменной $\frac{1}{y} = y'$:

$$y' = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + \varepsilon. \tag{2.99}$$

5. Для полиномиальной модели

$$y = a + b_1 x_1 + b_2 x_2^{\,2} + b_3 x_3^{\,3} + \ldots + b_n x_n^{\,n} + \varepsilon \tag{2.100}$$

применяется замена переменных $x_2^{\,2} = x_2'$; $x_3^{\,3} = x_3'$; ... ; $x_n^{\,n} = x_n'$.

$$y = a + bx_1 + b_2 x_2' + b_3 x_3' + \ldots + b_n x_n' + \varepsilon. \tag{2.101}$$

6. Степенную модель

$$y = a + b_1 x + b_2 x^2 + b_3 x^3 + \ldots + b_n x^n + \varepsilon \tag{2.102}$$

преобразуют заменой $x^2 = x_2'$; $x^3 = x_3'$; ... ; $x^n = x_n'$,

$$y = a + b_1 x_1 + b_2 x_2' + b_3 x_3' + \ldots + b_n x_n' + \varepsilon. \tag{2.103}$$

7. При линеаризации интерактивного уравнения

$$y = a + bx_1 + cx_2 + dx_1 x_2 + \ldots + \varepsilon, \tag{2.104}$$

$x_1 x_2 = x'$, ..., тогда

$$y = a + bx_1 + cx_2 + dx' + \ldots + \varepsilon. \tag{2.105}$$

После приведения перечисленных моделей к линейной форме их параметры и качество оцениваются с помощью методов, применяемых для линейных множественных регрессионных моделей, рассмотренных ниже.

Степенная модель с аддитивными возмущениями

$$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \ldots \cdot x_n^{b_n} + \varepsilon \tag{2.106}$$

не может быть приведена к линейной форме. Для оценки ее параметров используются итерационные методы подбора нелинейных моделей.

2.3.4. Проверка статистических гипотез о значениях отдельных коэффициентов

Ранее мы говорили о способе построения доверительного интервала на уровне значимости α.

$$b - t_{1-\alpha;\, n-2} \frac{S_{\varepsilon_i}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \le \beta \le b + t_{1-\alpha;\, n-2} \frac{S_{\varepsilon_i}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}, \tag{2.107}$$

где $S_{\varepsilon_i}^2$ – оценка дисперсии ошибки прогноза $Var(\varepsilon_i)$.

$$S^2_{\varepsilon_i} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n - m}, \qquad (2.108)$$

S^2_b – среднеквадратическое (стандартное) отклонение для b

$$S_b = \frac{S_{\varepsilon_i}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}. \qquad (2.109)$$

Полученный результат показывает, что при любом истинном значении параметра β вероятность накрытия его доверительным интервалом равна $(1 - \alpha)$.

Предположим, мы взяли значение β_0, не принадлежащее данному интервалу. Вероятность такого события будет очень мала, меньше чем значение $(1 - \alpha)$. Таким образом, факт ненакрытия взятого значения β_0 представляет собой осуществление редкого события, имеющего малую вероятность, и это дает нам основание сомневаться в том, что значение параметра $\beta = \beta_0$.

Априорные предположения о значениях параметров модели называют *статистическими гипотезами*.

О проверяемой гипотезе говорят как об исходной «нулевой» гипотезе и обозначают ее H_o, в нашем случае H_o: $\beta = \beta_0$.

В соответствии со сказанным выше, такую гипотезу следует отвергать, если значение β_0 не принадлежит $(1 - \alpha)$-процентному доверительному интервалу. β_0 не будет принадлежать этому интервалу в том случае, если наблюдаемое значение отношения больше табличного по абсолютной величине

$$\left| \frac{b - \beta_0}{S_b} \right| > t_{1-\alpha;\ n-2}. \qquad (2.110)$$

Это означает слишком большое отклонение оценки b от гипотетического значения β_0 параметра β в сравнении с оценкой S_b стандартного отклонения этого параметра.

Правило решения вопроса об отклонении или неотклонении статистической гипотезы H_o называется *статистическим критерием* проверки гипотезы H_o, а выбранное при формулировании этого правила значение α называется *уровнем значимости критерия*.

В практических исследованиях чаще всего используют $\alpha = 0{,}05$, хотя иногда $\alpha = 0{,}01$, $\alpha = 0{,}1$ и другие. Выбор большего или меньшего значения α определяется степенью значимости для исследования исходной гипотезы H_o. Если мы выбираем при исследовании меньшее значение α, то уменьшаем вероятность ошибки и отвержения верной гипотезы. Такие вероятности называют *мощностью критерия*.

В реальных ситуациях статистические критерии имеют довольно низкую мощность, так что рассматриваемая H_o отвергается редко, поэтому правильнее говорить о неотвержении гипотезы, а не о ее принятии.

Всякий статистический критерий основывается на использовании той или иной статистики, то есть случайной величины, значения которой могут быть вычислены приближенно на основании имеющихся статистических данных.

В нашем случае критерий проверки гипотезы H_o: $\beta = \beta_0$ основан на использовании t-статистики $\frac{b - \beta_0}{S_b}$, значение которой можно вычислить по данным наблюдений. Критерии, основанные на использовании t-статистики (распределения) Стьюдента называют *t-критериями*. Каждому статистическому критерию соответствует критическое множество R значений статистики критерия, при которых гипотеза H_o отвергается в соответствии с принятыми правилами (то есть множество значений t-статистики, превышающих по абсолютной величине $t_{1-\alpha;\ n-m}$).

Таким образом, статистический критерий определяется заданием:
– статистической гипотезы $H_о$;
– уровня значимости α;
– статистики критерия (t-статистики, χ^2-статистики, F-статистики);
– критического множества R.

2.4. Множественная линейная регрессия и корреляция

2.4.1. Отбор факторов для модели множественной регрессии

Экономические процессы, описываемые с помощью уравнений множественной регрессии

Парная регрессия может дать хороший результат при моделировании, если влиянием других факторов, воздействующих на объект исследования, можно пренебречь. Если же этим влиянием пренебречь нельзя, то в этом случае следует попытаться выявить влияние других факторов, введя их в модель, т. е. построить уравнение множественной регрессии, где y – зависимая переменная (результативный признак), x_i – независимые, или объясняющие, переменные (признаки-факторы).

Множественная регрессия широко используется в решении проблем спроса, доходности акций, при изучении функции издержек производства, в макроэкономических расчетах и целом ряде других вопросов эконометрики. Например, потребление отдельного товара на душу населения зависит от располагаемого дохода на душу населения, цены данного товара, цен на сопутствующие товары, привлекательности товара и других факторов.

В настоящее время множественная регрессия – один из наиболее распространенных методов в эконометрике. Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

Анализ факторов, включаемых в модель множественной регрессии

Включаемые во множественную регрессию факторы должны объяснить вариацию зависимой переменной. Если строится модель с набором m факторов, то для нее рассчитывается *коэффициент детерминации* R^2, который фиксирует долю объясненной вариации результативного признака за счет рассматриваемых в регрессии m факторов. Влияние других, не учтенных в модели факторов, оценивается как $1 - R^2$ с соответствующей *остаточной дисперсией* $S^2_{\varepsilon_i}$.

При дополнительном включении в регрессию $m+1$ фактора коэффициент детерминации должен возрастать, а остаточная дисперсия $S^2_{\varepsilon_i}$ уменьшаться:

$$R^2_{m+1} \geq R^2_m \quad \text{и} \quad S^2_{m+1} \leq S^2_m.$$

Если же этого не происходит и данные показатели практически не отличаются друг от друга, то включаемый в анализ фактор x_{m+1} не улучшает модель и практически является лишним.

Хотя теоретически регрессионная модель позволяет учесть любое число факторов, практически в этом нет необходимости. Отбор факторов производится на основе качественного теоретико-экономического анализа. Отбор факторов в *модель линейной регрессии* обычно осуществляется в две стадии: на первой подбираются факторы исходя из сущности проблемы; на второй – на основе *матрицы показателей линейной корреляции* определяют статистики для параметров регрессии.

Коэффициенты интеркорреляции (т. е. линейной корреляции между объясняющими переменными) позволяют исключать из модели дублирующие факторы. Считается, что две переменные явно *коллинеарны*, т. е. находятся между собой в линейной зависимости, если $r_{x_i x_j} \geq 0{,}7$. Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из регрессии. Предпочтение при этом отдается не фактору, более тесно связанному с результатом, а тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами. В этом требовании проявляется специфика множественной регрессии как метода исследования комплексного воздействия факторов в условиях их независимости друг от друга.

Пусть, например, при изучении некоторой зависимости матрица парных коэффициентов корреляции оказалась следующей (табл. 2.2).

Таблица 2.2

	y	x_1	x_2	x_3
y	1	0,8	0,7	0,6
x_1	0,8	1	0,8	0,5
x_2	0,7	0,8	1	0,2
x_3	0,6	0,5	0,2	1

Очевидно, что факторы x_1 и x_2 $\left(r_{x_1 x_2} = 0{,}8\right)$ дублируют друг друга. В анализ целесообразно включить фактор x_2, а не x_1, хотя корреляция x_2 с результатом y слабее, чем корреляция фактора x_1 с y $\left(r_{yx_2} = 0{,}7 < r_{yx_1} = 0{,}8\right)$, но зато значительно слабее межфакторная корреляция $r_{x_2 x_3} = 0{,}2 < r_{x_1 x_3} = 0{,}5$. Поэтому в данном случае в уравнение множественной регрессии включаются факторы x_2, x_2.

Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии *мультиколлинеарности* факторов, когда более чем два фактора связаны между собой линейной зависимостью, т. е. имеет место совокупное воздействие факторов друг на друга. В результате вариация в исходных данных перестает быть полностью независимой и нельзя оценить воздействие каждого фактора в отдельности.

Включение в модель мультиколлинеарных факторов нежелательно в силу следующих последствий: 1) затрудняется интерпретация параметров множественной регрессии как характеристик действия факторов в «чистом» виде, ибо факторы коррелированы; параметры линейной регрессии теряют экономический смысл; 2) оценки параметров ненадежны, обнаруживают большие стандартные ошибки и меняются с изменением объема наблюдений (не только по величине, но и по знаку), что делает модель непригодной для анализа и прогнозирования.

Для оценки мультиколлинеарности факторов может использоваться определитель матрицы парных коэффициентов линейной корреляции между факторами.

Если бы факторы не коррелировали между собой, то матрица парных линейных коэффициентов корреляции между факторами была бы единичной, поскольку все недиагональные элементы $r_{x_i x_j}$ $(i \neq j)$были бы равны нулю. Так, для уравнения линейной регрессии, включающего три объясняющих переменных

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + b_3 x_3 \tag{2.111}$$

матрица коэффициентов корреляции между факторами имела бы определитель, равный единице:

$$\operatorname{Det}\mathbf{R}=\begin{vmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} \\ r_{x_3x_1} & r_{x_3x_2} & r_{x_3x_3} \end{vmatrix}=\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}=1.$$

Если же, наоборот, между факторами существует полная линейная зависимость и все коэффициенты корреляции равны единице, то определитель такой матрицы равен нулю:

$$\operatorname{Det}\mathbf{R}=\begin{vmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} \\ r_{x_3x_1} & r_{x_3x_2} & r_{x_3x_3} \end{vmatrix}=\begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix}=0.$$

Чем ближе к нулю определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И, наоборот, чем ближе к единице определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Существует ряд подходов преодоления сильной межфакторной корреляции. Самый простой путь устранения мультиколлинеарности состоит в исключении из модели одного или нескольких факторов. Другой подход связан с преобразованием факторов, при котором уменьшается корреляция между ними.

Одним из путей учета внутренней корреляции факторов является переход к совмещенным уравнениям регрессии, т. е. к уравнениям, которые отражают не только влияние факторов, но и их взаимодействие. Так, если $y = f(x_1, x_2, x_3)$, то возможно построение следующего совмещенного уравнения:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + \varepsilon. \qquad (2.112)$$

Рассматриваемое уравнение включает взаимодействие первого порядка (взаимодействие двух факторов). Возможно включение в модель и взаимодействий более высокого порядка, если будет доказана их статистическая значимость по F-критерию Фишера, но, как правило, взаимодействия третьего и более высоких порядков оказываются статистически незначимыми.

Методы построения уравнения множественной регрессии

Отбор факторов, включаемых в регрессию, является одним из важнейших этапов практического использования методов регрессии. Подходы к отбору факторов на основе показателей корреляции могут быть разные. Они приводят, соответственно, к разным методикам построения уравнения множественной регрессии. В зависимости от того, какая методика построения принята, меняется алгоритм ее решения на ЭВМ.

Наиболее широкое применение получили следующие *методы построения уравнения множественной регрессии*: 1) метод исключения – отсев факторов из полного его набора; 2) метод включения – дополнительное введение фактора; 3) шаговый регрессионный анализ – исключение ранее введенного фактора.

При отборе факторов также рекомендуется пользоваться следующим *правилом*: число включаемых факторов обычно в 6–7 раз меньше объема совокупности, по которой строится регрессия. Если это соотношение нарушено, то число степеней свободы остаточной дисперсии очень мало. Это приводит к тому, что параметры уравнения регрессии оказываются статистически незначимыми, а F-критерий меньше табличного значения.

Возможны разные виды уравнений множественной регрессии: *линейные и нелинейные*.

Ввиду четкой интерпретации параметров наиболее широко используется линейная функция. В *линейной множественной регрессии* $\hat{y}_x = a + b_1x_1 + b_2x_2 + ... + b_mx_m$ параметры при x называются коэффициентами «чистой» регрессии. Они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при сохраненном значении других факторов, закрепленных на среднем уровне.

2.4.2. Оценивание параметров множественной линейной регрессии методом наименьших квадратов

Метод наименьших квадратов

Рассмотрим линейную модель множественной регрессии

$$y = a + b_1x_1 + b_2x_2 + ... + b_mx_m + \varepsilon . \qquad (2.113)$$

Классический подход к оцениванию параметров линейной модели множественной регрессии основан на методе наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от расчетных $\hat{y}$ минимальна:

$$\sum_u \left(y_i - \hat{y}_{x_i}\right)^2 \to \min . \qquad (2.114)$$

Как известно из курса математического анализа, для того чтобы найти экстремум функции нескольких переменных, надо вычислить частные производные первого порядка по каждому из параметров и приравнять их к нулю.

Итак. Имеем функцию m + 1 аргумента:

$$S(a,\ b_1,\ b_2,\ ...,\ b_m) = \sum \left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right)^2 . \qquad (2.115)$$

Находим частные производные первого порядка:

$$\begin{cases} \dfrac{\partial S}{\partial a} = -2\sum \left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right) = 0; \\ \dfrac{\partial S}{\partial b_1} = -2\sum x_1 \left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right) = 0; \\ .... \\ \dfrac{\partial S}{\partial b_m} = -2\sum x_m \left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right) = 0. \end{cases} \qquad (2.116)$$

После элементарных преобразований приходим к системе линейных нормальных уравнений для нахождения параметров линейного уравнения множественной регрессии (2.68):

$$\begin{cases} na + b_1\sum x_1 + b_2\sum x_2 + ... + b_m\sum x_m = \sum y, \\ a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 + ... + b_m\sum x_1x_m = \sum yx_1, \\ .... \\ a\sum x_m + b_1\sum x_1x_m + b_2\sum x_2x_m + ... + b_m\sum x_m^2 = \sum yx_m. \end{cases} \qquad (2.117)$$

Для двухфакторной модели данная система будет иметь вид:

$$\begin{cases} na + b_1\sum x_1 + b_2\sum x_2 = \sum y, \\ a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 = \sum yx_1, \\ a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2 = \sum yx_2. \end{cases} \qquad (2.118)$$

Применение метода наименьших квадратов для стандартизированного уравнения множественной линейной регрессии

Метод наименьших квадратов применим и к *уравнению множественной регрессии в стандартизированном масштабе*:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \ldots + \beta_m t_{x_m} + \varepsilon, \tag{2.119}$$

где t_y, t_{x_1}, ..., t_{x_m} – стандартизированные переменные: $t_y = \dfrac{y - \overline{y}}{\sigma_y}$, $t_{x_i} = \dfrac{x_i - \overline{x_i}}{\sigma_{x_i}}$, для которых среднее значение равно нулю: $\overline{t_y} = \overline{t_{x_i}} = 0$, а среднее квадратическое отклонение равно единице: $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$; β_i – стандартизированные коэффициенты регрессии.

Стандартизованные коэффициенты регрессии показывают, на сколько единиц изменится в среднем результат, если соответствующий фактор x_i изменится на одну единицу при неизменном среднем уровне других факторов. В силу того что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии β_i можно сравнивать между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов «чистой» регрессии, которые несравнимы между собой.

Частные коэффициенты эластичности

В отличие от парной регрессии частные уравнения регрессии характеризуют изолированное влияние фактора на результат, ибо другие факторы закреплены на неизменном уровне. Эффекты влияния других факторов присоединены в них к свободному члену уравнения множественной регрессии. Это позволяет на основе частных уравнений регрессии определять *частные коэффициенты эластичности*:

$$Э_{y_{x_i}} = b_i \cdot \frac{x_i}{\hat{y}_{x_i \cdot x_1, x_2, \ldots x_{i-1}, x_{i+1}, \ldots, x_m}}, \tag{2.120}$$

где b_i – коэффициент регрессии для фактора x_i в уравнении множественной регрессии; $\hat{y}_{x_i \cdot x_1, x_2, \ldots x_{i-1}, x_{i+1}, \ldots, x_m}$ – частное уравнение регрессии.

Наряду с частными коэффициентами эластичности могут быть найдены средние по совокупности показатели эластичности:

$$\overline{Э}_i = b_i \cdot \frac{\overline{x}_i}{\overline{y}_{x_i}}, \tag{2.121}$$

которые показывают, на сколько процентов в среднем изменится результат при изменении соответствующего фактора на 1%. Средние показатели эластичности можно сравнивать друг с другом и, соответственно, ранжировать факторы по силе их воздействия на результат.

2.4.3. Проверка значимости уравнения множественной линейной регрессии

Коэффициенты множественной корреляции и детерминации

Коэффициент множественной корреляции характеризует тесноту линейной связи рассматриваемого набора факторов с исследуемым признаком, или, иначе, оценивает тесноту совместного влияния факторов на результат.

Независимо от формы связи коэффициент множественной корреляции может быть найден следующим образом:

$$R_{yx_1x_2...x_m} = \sqrt{1 - \frac{S^2_{\varepsilon_i}}{S^2_y}}, \tag{2.122}$$

где S^2_y – оценка общей дисперсии результативного признака; $S^2_{\varepsilon_i}$ – оценка остаточной дисперсии.

Границы изменения индекса множественной корреляции от 0 до 1. Чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов. Величина коэффициента множественной корреляции должна быть больше или равна максимальному парному коэффициенту корреляции:

$$R_{yx_1x_2...x_m} \ge r_{yx_i(\max)} \quad \left(i = \overline{1,m}\right). \tag{2.123}$$

При правильном включении факторов в регрессионную модель величина коэффициента множественной корреляции будет существенно отличаться от коэффициента корреляции парной зависимости. Если же дополнительно включенные в уравнение множественной регрессии факторы третьестепенны, то коэффициент множественной корреляции может практически совпадать с коэффициентом парной корреляции (различия в третьем, четвертом знаках). Отсюда ясно, что, сравнивая коэффициенты множественной и парной корреляции, можно сделать вывод о целесообразности включения в уравнение регрессии того или иного фактора.

Расчет коэффициента множественной корреляции предполагает определение уравнения множественной регрессии и на его основе остаточной дисперсии:

$$S^2_{\varepsilon_i} = \frac{1}{n}\sum\left(y_i - \hat{y}_{x_1x_2...x_m}\right)^2. \tag{2.124}$$

Можно пользоваться следующей формулой *коэффициента множественной детерминации*:

$$R^2_{yx_1x_2...x_m} = 1 - \frac{\sum\left(y_i - \hat{y}_{x_1x_2...x_m}\right)^2}{\sum\left(y_i - \bar{y}\right)^2}. \tag{2.125}$$

При линейной зависимости признаков формула *коэффициента множественной корреляции* может быть представлена следующим выражением:

$$R_{yx_1x_2...x_m} = \sqrt{\sum \beta_i \cdot r_{yx_i}}, \tag{2.126}$$

где β_i – стандартизованные коэффициенты регрессии; r_{yx_i} – парные коэффициенты корреляции результата с каждым фактором.

Формула коэффициента множественной корреляции для линейной регрессии получила название *линейного коэффициента множественной корреляции*, или, что то же самое, *совокупного коэффициента корреляции*.

Возможно также при линейной зависимости определение совокупного коэффициента корреляции через матрицу парных коэффициентов корреляции:

$$R_{yx_1x_2,...,x_p} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}, \tag{2.127}$$

где

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & ... & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & ... & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & ... & r_{x_2x_p} \\ ... & ... & ... & ... & ... \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & ... & 1 \end{vmatrix} \tag{2.128}$$

– определитель матрицы парных коэффициентов корреляции;

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix} \qquad (2.129)$$

– определитель матрицы межфакторной корреляции.

Как видно, величина множественного коэффициента корреляции зависит не только от корреляции результата с каждым из факторов, но и от межфакторной корреляции. Рассмотренная формула позволяет определять совокупный коэффициент корреляции, не обращаясь при этом к уравнению множественной регрессии, а используя лишь парные коэффициенты корреляции.

Скорректированный коэффициент множественной корреляции содержит поправку на число степеней свободы, а именно остаточная сумма квадратов $\sum\left(y_i - \hat{y}_{x_1x_2\dots x_m}\right)^2$ делится на число степеней свободы остаточной вариации ($n - m$), а общая сумма квадратов отклонений $\sum\left(y_i - \bar{y}\right)^2$ на число степеней свободы в целом по совокупности ($n - 1$).

Формула *скорректированного коэффициента множественной детерминации* имеет вид:

$$R^2_{ск} = 1 - \frac{\sum(y_i - \hat{y}_i)^2/(n-m)}{\sum(y_i - \bar{y})^2/(n-1)}, \qquad (2.130)$$

где m – число оцениваемых параметров уравнения регрессии; n – число наблюдений.

Поскольку $\dfrac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - R^2$ то величину скорректированного индекса детерминации можно представить в виде:

$$R^2_{ск} = 1 - (1 - R^2)\cdot\frac{n-1}{n-m}. \qquad (2.131)$$

Чем больше величина m, тем сильнее различия $R^2_{ск}$ и R^2.

Как было показано выше, ранжирование факторов, участвующих во множественной линейной регрессии, может быть проведено через стандартизованные коэффициенты регрессии (β-коэффициенты). Эта же цель может быть достигнута с помощью частных коэффициентов корреляции (для линейных связей). Кроме того, частные показатели корреляции широко используются при решении проблемы отбора факторов: целесообразность включения того или иного фактора в модель можно доказать величиной показателя частной корреляции.

Частные и общий коэффициенты корреляции

Частные коэффициенты корреляции характеризуют тесноту линейной связи между результатом и соответствующим фактором при элиминировании (устранении влияния) других факторов, включенных в уравнение регрессии.

Показатели частной корреляции представляют собой отношение сокращения остаточной дисперсии за счет дополнительного включения в анализ нового фактора к остаточной дисперсии, имевшей место до введения его в модель.

В общем виде при наличии m факторов для уравнения

$$y = a + b_1x_1 + b_2x_2 \dots + b_mx_m + \varepsilon \qquad (2.132)$$

коэффициент частной корреляции, измеряющий влияние на y фактора x_i, при неизменном уровне других факторов, можно определить по формуле:

$$r_{yx_i \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_m} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2 \ldots x_i \ldots x_m}}{1 - R^2_{yx_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_m}}}, \quad (2.133)$$

где $R^2_{yx_1 x_2 \ldots x_i \ldots x_m}$ – множественный коэффициент детерминации всех m факторов с результатом; $R^2_{yx_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_m}$ – тот же показатель детерминации, но без введения в модель фактора x_i.

При двух факторах формула (2.133) примет вид:

$$r_{yx_1 \cdot x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2}}{1 - r^2_{yx_2}}}; \quad r_{yx_2 \cdot x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2}}{1 - r^2_{yx_1}}}. \quad (2.134)$$

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. Например, $r_{yx_1 \cdot x_2 x_3}$– коэффициент частной линейной корреляции первого порядка. Соответственно, коэффициенты парной линейной корреляции называются коэффициентами нулевого порядка. Коэффициенты частной корреляции более высоких порядков можно определить через коэффициенты частной корреляции более низких порядков по рекуррентной формуле

$$r_{yx_i \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_m} = \frac{r_{yx_i \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_{m-1}} - r_{yx_m \cdot x_1 x_2 \ldots x_{m-1}} \cdot r_{x_i x_m \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_{m-1}}}{\sqrt{\left(1 - r^2_{yx_m \cdot x_1 x_2 \ldots x_{m-1}}\right) \cdot \left(1 - r^2_{x_i x_m \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_{m-1}}\right)}}. \quad (2.135)$$

При двух факторах данная формула примет вид:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right) \cdot \left(1 - r^2_{x_1 x_2}\right)}}; \quad r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{\left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{x_1 x_2}\right)}}. \quad (2.136)$$

Для уравнения регрессии с тремя факторами частные линейные коэффициенты корреляции второго порядка определяются на основе частных коэффициентов корреляции первого порядка. Так, по уравнению $y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + \varepsilon$ возможно исчисление трех частных коэффициентов корреляции второго порядка: $r_{yx_1 \cdot x_2 x_3}$, $r_{yx_2 \cdot x_1 x_3}$, $r_{yx_3 \cdot x_1 x_2}$, каждый из которых определяется по рекуррентной формуле. Например, при имеем формулу для расчета $r_{yx_1 \cdot x_2 x_3}$:

$$r_{yx_1 \cdot x_2 x_3} = \frac{r_{yx_1 \cdot x_2} - r_{yx_3 \cdot x_2} \cdot r_{x_1 x_3 \cdot x_2}}{\sqrt{\left(1 - r^2_{yx_3 \cdot x_2}\right)\left(1 - r^2_{x_1 x_3 \cdot x_2}\right)}}. \quad (2.137)$$

Рассчитанные по рекуррентной формуле частные коэффициенты линейной корреляции изменяются в пределах от -1 до +1, а по формулам через множественные коэффициенты детерминации – от 0 до 1. Сравнение их друг с другом позволяет ранжировать факторы по тесноте их связи с результатом. Частные коэффициенты линейной корреляции дают меру тесноты связи каждого фактора с результатом в чистом виде. Если из стандартизованного уравнения регрессии $t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \beta_3 t_{x_3} + \varepsilon$ следует, что $\beta_1 > \beta_2 > \beta_3$, т. е. по силе влияния на результат порядок факторов таков: x_1, x_2, x_3, то этот же порядок факторов определяется и по соотношению частных коэффициентов корреляции, $r_{yx_1 \cdot x_2 x_3} > r_{yx_2 \cdot x_1 x_3} > r_{yx_3 \cdot x_1 x_2}$.

В эконометрике частные коэффициенты линейной корреляции обычно не имеют самостоятельного значения. Их используют на стадии формирования модели. Так, строя многофакторную модель, на первом шаге определяется уравнение регрессии с полным набором факторов и рассчитывается матрица частных коэффициентов корреляции. На втором шаге отбирается фактор с наименьшей и несущественной по t-критерию Стьюдента величиной показателя частной корреляции. Исключив его из модели, строится новое уравнение регрессии. Процедура продолжается до тех пор, пока не окажется, что все частные коэффициенты корреляции существенно отличаются от нуля. Если исключен несущественный фактор, то множественные коэффициенты детерминации на двух смежных шагах построения регрессионной модели почти не отличаются друг от друга, $R^2_{m+1} \approx R^2_m$, где m – число факторов.

Из приведенных выше формул частных коэффициентов корреляции видна связь этих показателей с совокупным коэффициентом корреляции. Зная частные коэффициенты корреляции (последовательно первого, второго и более высокого порядка), можно определить совокупный по формуле

$$R_{yx_1x_2\ldots x_m} = \sqrt{1-\left(1-r^2_{yx_1}\right)\cdot\left(1-r^2_{yx_2\cdot x_1}\right)\cdot\left(1-r^2_{yx_3\cdot x_1x_2}\right)\cdot\ldots\cdot\left(1-r^2_{yx_m\cdot x_1x_2\ldots x_{m-1}}\right)}. \quad (2.138)$$

В частности, для двухфакторного уравнения формула (2.16) принимает вид:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1-\left(1-r^2_{yx_1}\right)\cdot\left(1-r^2_{yx_2\cdot x_1}\right)}. \quad (2.139)$$

При полной зависимости результативного признака от исследуемых факторов коэффициент совокупного их влияния равен единице. Из единицы вычитается доля остаточной вариации результативного признака$(1 - r^2)$, обусловленная последовательно включенными в анализ факторами. В результате подкоренное выражение характеризует совокупное действие всех исследуемых факторов.

Проверка значимости уравнений множественной регрессии с помощью критериев Фишера и Стьюдента

Значимость уравнения множественной регрессии в целом, так же как и в парной регрессии, оценивается с помощью F-критерия Фишера:

$$F = \frac{Q_R}{Q_e} = \frac{R^2}{1-R^2}\cdot\frac{n-m}{m-1}, \quad (2.140)$$

где Q_R – факторная сумма квадратов на одну степень свободы; Q_e – остаточная сумма квадратов на одну степень свободы; R^2 – коэффициент (индекс) множественной детерминации; m – число оцениваемых параметров уравнения регрессии; n – число наблюдений.

Оценивается значимость не только уравнения в целом, но и фактора, дополнительно включенного в регрессионную модель. Необходимость такой оценки связана с тем, что не каждый фактор, вошедший в модель, может существенно увеличивать долю объясненной вариации результативного признака. Кроме того, при наличии нескольких факторов они могут вводиться в модель в разной последовательности. Ввиду корреляции между факторами значимость одного и того же фактора может быть разной в зависимости от последовательности его введения в модель. Мерой для оценки включения фактора в модель служит частный F-критерий, т. е. F_{x_i}.

Частный F-критерий построен на сравнении прироста факторной дисперсии, обусловленного влиянием дополнительно включенного фактора, с остаточной дисперсией на одну степень свободы по регрессионной модели в целом. В общем виде для фактора x_i частный F-критерий определится как

$$F_{\text{част},x_1} = \frac{R^2{}_{yx_1\ldots x_i\ldots x_m} - R^2{}_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_m}}{1-R^2{}_{yx_1\ldots x_i\ldots x_{m1}}}\cdot\frac{n-m}{m-1}, \quad (2.141)$$

где $R^2_{yx_1\ldots x_i\ldots x_m}$ – коэффициент множественной детерминации для модели с полным набором факторов, $R^2_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_m}$ – тот же показатель, но без включения в модель фактора x_i, n – число наблюдений, m – число параметров в модели.

Фактическое значение частного F-критерия сравнивается с табличным при уровне значимости α и числе степеней свободы: 1 и $(n - m)$. Если фактическое значение F_{x_i} превышает $F_{табл}(\alpha, k_1, k_2)$, то дополнительное включение фактора x_i в модель статистически оправданно и коэффициент чистой регрессии b_i при факторе x_i статистически значим. Если же фактическое значение F_{x_i} меньше табличного, то дополнительное включение в модель фактора x_i не увеличивает существенно долю объясненной вариации признака y, следовательно, нецелесообразно его включение в модель; коэффициент регрессии при данном факторе в этом случае статистически незначим.

Для двухфакторного уравнения частные F-критерии имеют вид:

$$F_{\text{част, x}_1} = \frac{R^2_{yx_1x_2} - R^2_{yx_2}}{1 - R^2_{yx_1}} \cdot \frac{n-3}{2};$$
$$F_{част, x_2} = \frac{R^2_{yx_1x_2} - R^2_{yx_1}}{1 - R^2_{yx_2}} \cdot \frac{n-3}{2}. \tag{2.142}$$

С помощью частного F-критерия можно проверить значимость всех коэффициентов регрессии в предположении, что каждый соответствующий фактор x_i вводился в уравнение множественной регрессии последним.

Частный F-критерий оценивает значимость коэффициентов чистой регрессии. Зная величину F_{x_i}, можно определить и t-критерий для коэффициента регрессии при i-м факторе, t_{b_i}, а именно:

$$t_{b_i} = \sqrt{F_{x_i}}\,. \tag{2.143}$$

Оценка значимости коэффициентов чистой регрессии по t-критерию Стьюдента может быть проведена и без расчета частных F-критериев. В этом случае, как и в парной регрессии, для каждого фактора используется формула:

$$t_{b_i} = \frac{b_i}{S_{b_i}}, \tag{2.144}$$

где b_i – коэффициент чистой регрессии при факторе x_i; S_{b_i} – среднее квадратическое (стандартное) отклонение коэффициента регрессии b_i.

Для уравнения множественной регрессии $\hat{y} = a + b_1x_1 + b_2x_2 + ... + b_mx_m$ среднее квадратическое отклонение коэффициента регрессии может быть определено по следующей формуле:

$$S_{b_i} = \frac{S_y\sqrt{1 - R^2_{yx_1...x_m}}}{S_{x_i}\sqrt{1 - R^2_{x_ix_1...x_m}}} \cdot \frac{1}{\sqrt{n-m}}, \tag{2.145}$$

где S_y – среднее квадратическое отклонение для признака y, S_{x_i} – среднее квадратическое отклонение для признака x_i, $R^2_{yx_1...x_m}$ – коэффициент детерминации для уравнения множественной регрессии, $R^2_{x_ix_1...x_m}$ – коэффициент детерминации для зависимости фактора x_i со всеми другими факторами уравнения множественной регрессии; $(n - m)$ – число степеней свободы для остаточной суммы квадратов отклонений.

Как видим, чтобы воспользоваться данной формулой, необходимы матрица межфакторной корреляции и расчет по ней соответствующих коэффициентов детерминации $R^2_{x_ix_1...x_m}$. Так, для уравнения $\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3$ оценка значимости коэффициентов регрессии b_1, b_2, b_3 предполагает расчет трех межфакторных коэффициентов детерминации: $R^2_{x_1 \cdot x_2x_3}$, $R^2_{x_2 \cdot x_1x_3}$, $R^2_{x_3 \cdot x_1x_2}$.

Взаимосвязь показателей частного коэффициента корреляции, частного F-критерия и t-критерия Стьюдента для коэффициентов чистой регрессии может использоваться в процедуре отбора факторов.

2.4.4. Метод взвешенных наименьших квадратов (обобщенный МНК)

Данный метод применяется, когда построенная модель множественной регрессии не отвечает требованию гомоскедастичности возмущений. Найденные параметры такой модели не обладают свойствами несмещенности и эффективности, о чем можно судить по гетероскедастичности остатков. От таких нежелательных свойств модели можно избавиться с помощью следующих преобразований.

Согласно обобщенному (взвешенному) МНК оценки параметров определяются из условия минимума суммы взвешенных квадратов ошибок:

$$\frac{\sum_{i=1}^{n} e_i^2}{S_{\varepsilon_i}} = \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{S_{\varepsilon_i}} \to \min. \tag{2.146}$$

Если полученное после оценки параметров модели

$$y_i = \alpha + \beta_1 x_{i1} + \ldots + \beta_j x_{ij} + \ldots + \beta_k x_{ik} + \varepsilon_i$$

обычным МНК уравнение имеет вид

$$y_i = a + b_1 x_{i1} + \ldots + b_j x_{ij} + \ldots + b_k x_{ik}, \tag{2.147}$$

то чтобы избавиться от нежелательных свойств ошибок его делят на оценку стандартного отклонения ошибок.

$$\frac{y_i}{S_{\varepsilon_i}} = \frac{a}{S_{\varepsilon_i}} + \frac{b_1 x_{i1}}{S_{\varepsilon_i}} + \ldots + \frac{b_j x_{ij}}{S_{\varepsilon_i}} + \ldots + \frac{b_k x_{ik}}{S_{\varepsilon_i}}. \tag{2.148}$$

Такая модель будет гомоскедастичной и для ее оценивания можно применить обычный МНК.

Осуществленное преобразование приводит к модификации критерия метода наименьших квадратов. Теперь ошибки (остатки) модели не просто складываются, а берутся с «весами», причем наблюдениям с меньшими дисперсиями (более точным) придаются большие веса. Это позволяет получить эффективные, несмещенные оценки параметров модели. Обобщенный МНК для моделей с гетероскедастичностью называют методом взвешенных наименьших квадратов. Данный метод может применяться и для моделей с автокоррелированными остатками.

2.4.5. Фиктивные переменные

Необходимость использования фиктивных переменных

Зачастую в регрессионных моделях в качестве объясняющих переменных приходится использовать не только количественные (определяемые численно), но и *качественные переменные*. Например, спрос на некоторое материальное благо может определяться ценой данного блага, ценой на заменители данного блага, ценой дополняющих благ, доходом потребителей и т. д. (эти показатели определяются количественно). Но спрос может также зависеть от вкусов потребителей, их ожиданий, национальных и религиозных особенностей и т. д. А эти показатели представить в численном виде нельзя. Возникает проблема отражения в модели влияния таких переменных на исследуемую величину. Это достаточно сложная задача.

Обычно в моделях влияние качественного фактора выражается в виде *фиктивной (искусственной) переменной*, которая отражает два противоположных состояния качественного фактора. Например, «фактор действует» – «фактор не действует», «курс валюты фиксированный» – «курс валюты плавающий», «сезон летний» – «сезон зимний» и т. д. В этом случае фиктивная переменная может выражаться в двоичной форме:

$$d = \begin{cases} 0, & \textit{фактор не действует}, \\ 1, & \textit{фактор действует}. \end{cases}$$

В зависимости от целей исследования используются различные фиктивные переменные:

$$d = \begin{cases} 0, & \textit{если в обществе имеются инфляционные ожидания}, \\ 1, & \textit{если инфляционных ожиданий нет}; \end{cases}$$

$$d = \begin{cases} 0, & \textit{если рассматривается первый квартал}, \\ 1, & \textit{в противоположном случае}; \end{cases}$$

$$d = \begin{cases} 0, & \textit{субъект не имеет работы}, \\ 1, & \textit{субъект имеет работу}; \end{cases}$$

$$d = \begin{cases} 0, & \textit{экзамен не сдан с первой попытки}, \\ 1, & \textit{экзамен сдан с первой попытки}; \end{cases}$$

$$d = \begin{cases} 0, & \textit{в процессе обучения не используются компьютеры}, \\ 1, & \textit{в противоположном случае}\ldots \end{cases}$$

Переменная d называется *фиктивной (искусственной, двоичной) переменной (индикатором)*.

Таким образом, кроме моделей, содержащих только количественные объясняющие переменные (обозначаемые x_i), в регрессионном анализе рассматриваются также модели, содержащие лишь качественные переменные (обозначаемые d_i), либо и те и другие одновременно. Искусственные переменные могут входить в состав регрессионных моделей в качестве как объясняющих, так и объясняемых переменных.

Методы оценки параметров и качества моделей, содержащих фиктивные переменные, аналогичны рассмотренным ранее методам для моделей, содержащих только количественные переменные. Рассмотрим примеры простейших моделей с фиктивными переменными.

Модели, содержащие только качественные объясняющие переменные

Регрессионные модели, содержащие лишь качественные объясняющие переменные, называются *моделями дисперсионного анализа (ANOVA-моделями)*.

Например, пусть y – начальная заработная плата.

$$d = \begin{cases} 0, & \textit{если претендент не имеет высшего образования}, \\ 1, & \textit{если претендент имеет высшее образование}. \end{cases}$$

Тогда зависимость можно выразить моделью парной регрессии

$$y = \alpha + \beta d + \varepsilon. \tag{2.149}$$

Очевидно, что $M_y(d = 0) = \alpha + \beta \cdot 0 = 0$, $M_y(d = 1) = \alpha + \beta \cdot 1 = \alpha + \beta$.

При этом коэффициент α определяет среднюю начальную заработную плату при отсутствии высшего образования. Коэффициент β указывает, на какую величину отличаются средние начальные заработные платы при наличии или отсутствии высшего образования у претендента. Проверяя статистическую значимость коэффициента β с помощью t-статистики либо значимость коэффициента детерминации R^2 с помощью F-статистики, можно определить, влияет или нет наличие высшего образования на начальную заработную плату.

Нетрудно заметить, что *ANOVA*-модели представляют собой кусочно-постоянные функции. Однако такие модели в экономике крайне редки. Гораздо чаще встречаются модели, содержащие как качественные, так и количественные переменные.

Модели, в которых объясняющие переменные носят как количественный, так и качественный характер

Они называются *моделями ковариационного анализа (ANCOVA-моделями).*

Рассмотрим простейшую *ANCOVA*-модель с одной количественной и одной качественной переменной, имеющей два альтернативных состояния:

$$y = \alpha + \beta_1 x + \beta_2 d + \varepsilon. \tag{2.150}$$

Пусть, например, y – заработная плата сотрудника фирмы, x – стаж сотрудника, d – пол сотрудника, т. е.

$$d = \begin{cases} 0, & \text{если сотрудник – женщина}, \\ 1, & \text{если сотрудник – мужчина}. \end{cases}$$

Тогда ожидаемое значение заработной платы сотрудников при x годах трудового стажа будет:

– для женщины

$$M_y(x, d = 0) = \alpha + \beta_1 x + \beta_2 \cdot 0 = \alpha + \beta_1 x\,; \tag{2.151}$$

– для мужчины

$$M_y(x, d = 1) = \alpha + \beta_1 x + \beta_2 \cdot 1 = (\alpha + \beta_2) + \beta_1 x\,. \tag{2.152}$$

Заработная плата в данном случае является линейной функцией от стажа работы. Причем и для мужчин, и для женщин заработная плата меняется с одним и тем же коэффициентом пропорциональности β_1. А вот свободные члены в моделях (2.151), (2.152) отличаются на величину β_2. Проверив с помощью t-статистики статистические значимости коэффициентов α и $(\alpha + \beta_2)$, можно определить: имеет ли место в фирме дискриминация по половому признаку. Если эти коэффициенты окажутся статистически значимыми, то, очевидно, дискриминация есть. Более того, при $\beta_2 > 0$ она будет в пользу мужчин, а при $\beta_2 < 0$ – в пользу женщин.

В данном случае пол сотрудников имеет два альтернативных значения, и в модели это отражается одной фиктивной переменной. Возникает вопрос, нельзя ли с помощью большего числа фиктивных переменных обрисовать более сложные комбинации? Например, пусть

$$y = \alpha + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \varepsilon. \tag{2.153}$$

$$d_1 = \begin{cases} 0, & \text{если сотрудник – мужчина}, \\ 1, & \text{если сотрудник – женщина}; \end{cases}$$

$$d_2 = \begin{cases} 0, & \text{если сотрудник – женщина}, \\ 1, & \text{если сотрудник – мужчина}. \end{cases}$$

Но в этой ситуации между переменными d_1 и d_2 существует строгая линейная зависимость: $d_2 = 1 - d_1$. Мы попадаем в ситуацию совершенной мультиколлинеарности, при которой коэффициенты β_2 и β_3 однозначно определены быть не могут. Простейшим способом преодоления данной проблемы является отбрасывание одной из фиктивных переменных и использование для рассматриваемой задачи модели (2.150). Применяя аналогичные выкладки, можно получить следующее общее правило:

Если качественная переменная имеет к альтернативных значении, то при моделировании используются только (к – 1) фиктивных переменных.

Если не следовать данному правилу, то при моделировании исследователь попадает в ситуацию совершенной мультиколлинеарности, или так называемую *ловушку фиктивной переменной*.

Значения фиктивной переменной можно изменять на противоположные. Суть модели от этого не изменится. Например, в модели (2.150) можно положить, что

$$d = \begin{cases} 0, & \text{если сотрудник – мужчина,} \\ 1, & \text{если сотрудник – женщина.} \end{cases}$$

Однако при этом знак коэффициента β_1 изменится на противоположный.

Значение качественной переменной, для которого принимается $d = 0$, называется *базовым* или *сравнительным*. Выбор базового значения обычно диктуется целями исследования, но может быть и произвольным.

Коэффициент β_1 в модели (2.150) иногда называется *дифференциальным коэффициентом свободного члена,* т. к. он показывает, на какую величину отличается свободный член модели при значении фиктивной переменной, равном единице, от свободного члена модели при базовом значении фиктивной переменной.

Пусть рассматривается модель с двумя объясняющими переменными, одна из которых количественная, а другая – качественная. Причем качественная переменная имеет три альтернативы. Например, расходы на содержание ребенка могут быть связаны с доходами домохозяйств и возрастом ребенка: дошкольный, младший школьный и старший школьный. Так как качественная переменная связана с тремя альтернативами, то по общему правилу моделирования необходимо использовать две качественные переменные. Таким образом, модель может быть представлена в виде:

$$y = \alpha + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \varepsilon, \quad (1.154)$$

где y – расходы; x – доходы домохозяйств.

$$d_1 = \begin{cases} 0, & \text{если в семье ребенок дошкольного возраста,} \\ 1, & \text{если в семье ребенок другого возраста;} \end{cases}$$

$$d_2 = \begin{cases} 0, & \text{если в семье младший школьник,} \\ 1, & \text{в противоположном случае.} \end{cases}$$

Таким образом, получаются следующие зависимости.

Средний расход на дошкольника:

$$M_y(x, d_1 = 0, d_2 = 0) = \alpha + \beta_1 x. \quad (2.155)$$

Средний расход на младшего школьника:

$$M_y(x, d_1 = 1, d_2 = 0) = \alpha + \beta_1 x + \beta_2 \cdot 1 + \beta_3 \cdot 0 = (\alpha + \beta_2) + \beta_1 x. \quad (2.156)$$

Средний расход на старшего школьника:

$$M_y(x, d_1 = 1, d_2 = 1) = \alpha + \beta_1 x + \beta_2 \cdot 1 + \beta_3 \cdot 1 = (\alpha + \beta_2 + \beta_3) + \beta_1 x. \quad (2.157)$$

Здесь β_2, β_3 – дифференциальные свободные члены. Базовым значением качественной переменной является значение «дошкольник». После определения коэффициентов регрессии (2.154) определяется статистическая значимость коэффициентов β_2, β_3 на основе t-статистики. Если коэффициенты оказываются статистически незначимыми, то можно сделать вывод, что возраст ребенка не оказывает существенного влияния на его содержание.

2.5. Временные ряды

2.5.1. Составляющие временных рядов

Группы факторов, влияющие на формирование временного ряда

При построении эконометрической модели используются два типа данных:
1) данные, характеризующие совокупность различных объектов в определенный момент времени;
2) данные, характеризующие один объект за ряд последовательных моментов времени.

Модели, построенные по данным первого типа, называются *пространственными моделями*. Модели, построенные на основе второго типа данных, называются *моделями временных рядов*.

Временной ряд (ряд динамики) – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени. Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:
1) факторы, формирующие *тенденцию* ряда (тренд T);
2) факторы, формирующие *циклические колебания* ряда (сезонная компонента S);
3) *случайные* факторы (E).

Рассмотрим воздействие каждого фактора на временной ряд в отдельности.

Большинство временных рядов экономических показателей имеют *тенденцию*, характеризующую совокупное долговременное воздействие множества факторов на динамику изучаемого показателя. Все эти факторы, взятые в отдельности, могут оказывать разнонаправленное воздействие на исследуемый показатель. Однако в совокупности они формируют его возрастающую или убывающую тенденцию. На рис. 2.26 показан гипотетический временной ряд, содержащий возрастающую тенденцию.

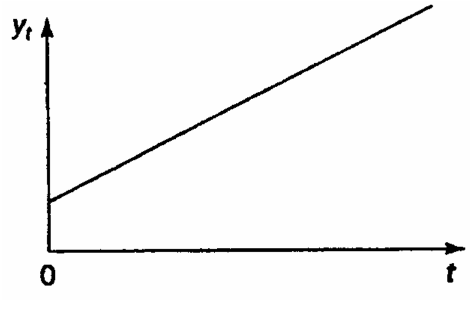

Рис. 2.26

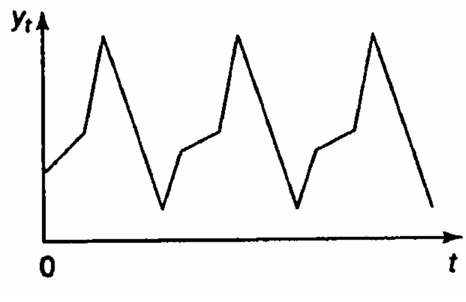

Рис. 2.27

Также изучаемый показатель может быть подвержен *циклическим колебаниям*. Эти колебания могут носить сезонный характер, поскольку экономическая деятельность ряда отраслей экономики зависит от времени года (например, цены на сельскохозяйственную продукцию в летний период выше, чем в зимний; уровень безработицы в курортных городах в зимний период выше по сравнению с летним). При наличии больших массивов данных за длительные промежутки времени можно выявить циклические колебания, связанные с общей динамикой конъюнктуры рынка. На рис. 2.27 представлен гипотетический временной ряд, содержащий только сезонную компоненту.

Стационарные и нестационарные временные ряды

Некоторые временные ряды не содержат тенденции и циклической компоненты, а каждый следующий их уровень образуется как сумма среднего уровня ряда и некоторой (положительной или отрицательной) случайной компоненты. Такие ряды называются *стационарными*. Ряды, содержащие в своем составе две или три компоненты, называются *нестационарными*.

Важными характеристиками случайного процесса, то есть стационарного ряда являются математическое ожидание и дисперсия. *Математическим ожиданием* процесса $X(t)$ является неслучайная функция $M_x(t)$, значение которой в момент времени t равно математическому ожиданию. *Дисперсией* случайного процесса является неслучайная функция $D_x(t)$, значение которой также равно дисперсии в каждый момент времени t.

Временной ряд *стационарен*, если порождающий его механизм не меняется при сдвиге во времени, а соответствующий случайный процесс достиг статистического равновесия. Формально стационарный временной ряд определяется как такой случайный процесс, для которого математическое ожиданиие, дисперсия и ковариации между отдельными членами ряда случайно варьируют вокруг постоянного, не зависящего от времени уровня (так называемая «стационарность» в широком смысле, которая только и рассматривается для временных рядов):

$$M_x(t) = \text{const}; \; D_x(t) = \text{const}.$$

Простейшим примером стационарного временного ряда является «белый шум» – чисто случайный процесс, значения которого в различные моменты времени независимы и одинаково распределены (рис. 2.28).

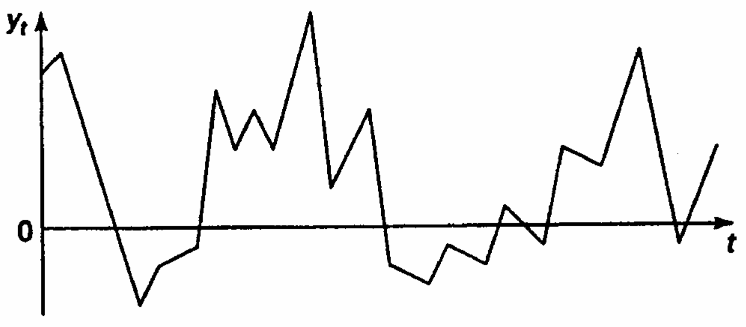

Рис. 2.28

Аддитивная и мультипликативная модели временных рядов

Очевидно, что реальные данные не следуют целиком и полностью из каких-либо описанных выше моделей. Чаще всего они содержат все три компоненты. Каждый их уровень формируется под воздействием тенденции, сезонных колебаний и случайной компоненты.

В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется *аддитивной моделью* временного ряда. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется *мультипликативной моделью* временного ряда. Основная задача эконометрического исследования отдельного временного ряда – выявление и придание количественного выражения каждой из перечисленных выше компонент с тем, чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

2.5.2. Коэффициент автокорреляции. Автокорреляционная функция

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущего. Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда.

Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Формула для расчета коэффициента автокорреляции имеет вид:

$$r_1 = \frac{\sum_{t=2}^{n}\left(y_t - \overline{y}_1\right)\left(y_{t-1} - \overline{y}_2\right)}{\sqrt{\sum_{t=2}^{n}\left(y_t - \overline{y}_1\right)^2 \sum_{t=2}^{n}\left(y_{t-1} - \overline{y}_2\right)^2}}, \tag{2.158}$$

где $\overline{y}_1 = \frac{1}{n-1}\sum_{t=2}^{n} y_t$, $\quad \overline{y}_2 = \frac{1}{n-1}\sum_{t=2}^{n} y_{t-1}$.

Эту величину называют *коэффициентом автокорреляции* уровней ряда первого порядка, так как он измеряет зависимость между соседними уровнями ряда *t* и y_{t-1}.

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями *t* и y_{t-2} и определяется по формуле

$$r_2 = \frac{\sum_{t=3}^{n}\left(y_t - \overline{y}_3\right)\left(y_{t-2} - \overline{y}_4\right)}{\sqrt{\sum_{t=3}^{n}\left(y_t - \overline{y}_3\right)^2 \sum_{t=3}^{n}\left(y_{t-2} - \overline{y}_4\right)^2}}, \tag{2.159}$$

где $\overline{y}_3 = \frac{1}{n-2}\sum_{t=3}^{n} y_t$, $\quad \overline{y}_4 = \frac{1}{n-2}\sum_{t=3}^{n} y_{t-2}$.

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют *лагом*(τ). С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Считается целесообразным для обеспечения статистической достоверности коэффициентов автокорреляции использовать *правило* – максимальный лаг должен быть не больше $n/4$.

Свойства коэффициента автокорреляции

1. Он строится по аналогии с линейным коэффициентом корреляции и таким образом характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Поэтому по коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию (например, параболу второго порядка или экспоненту), коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.

2. По знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержат положительную автокорреляцию уровней, однако при этом могут иметь убывающую тенденцию.

Последовательность коэффициентов автокорреляции уровней первого, второго и т. д. порядков называют *автокорреляционной функцией* временного ряда. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется *коррелограммой*.

Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором автокорреляция наиболее высокая, а следовательно, и лаг, при котором связь между текущим и предыдущими уровнями ряда наиболее тесная, т. е. при помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ, то ряд содержит циклические колебания с периодичностью в τ моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать одно из двух предположений относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ. Поэтому коэффициент автокорреляции уровней и автокорреляционную функцию целесообразно использовать для выявления во временном ряде наличия или отсутствия трендовой компоненты и циклической (сезонной) компоненты.

2.5.3. Моделирование тенденции временного ряда

Распространенным способом моделирования тенденции временного ряда является построение аналитической функции, характеризующей зависимость уровней ряда от времени или тренда. Этот способ называют *аналитическим выравниванием временного ряда*.

Поскольку зависимость от времени может принимать разные формы, для ее формализации можно использовать различные виды функций. Для построения трендов чаще всего применяются следующие функции:

линейный тренд: $\hat{y}_t = a + b \cdot t$;

гипербола: $\hat{y}_t = a + b/t$;

экспоненциальный тренд: $\hat{y}_t = \mathrm{e}^{a+b\cdot t}$ (или $\hat{y}_t = a \cdot b^t$);

степенная функция: $\hat{y}_t = a \cdot t^b$;

полиномы различных степеней: $\hat{y}_t = a + b_1 \cdot t + b_2 \cdot t^2 + \ldots + b_m \cdot t^m$.

Параметры каждого из перечисленных выше трендов можно определить обычным МНК, используя в качестве независимой переменной время $t = 1, 2, \ldots, n$, а в качестве зависимой переменной – фактические уровни временного ряда y_t. Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

Существует несколько *способов определения типа тенденции*. К числу наиболее распространенных способов относятся качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени. В этих же целях можно использовать и коэффициенты автокорреляции уровней ряда. Тип тенденции можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным уровням ряда. Если временной ряд имеет *линейную тенденцию*, то его соседние уровни y_t и y_{t-1} тесно коррелируют. В этом случае коэффициент автокорреляции первого порядка уровней исходного ряда должен быть высоким. Если временной ряд содержит *нелинейную тенденцию*, например, в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда. Чем сильнее выражена нелинейная тенденция в изучаемом временном ряде, тем в большей степени будут различаться значения указанных коэффициентов.

Выбор наилучшего уравнения, в случае когда ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации и средней ошибки аппроксимации. Этот метод легко реализуется при компьютерной обработке данных.

2.5.4. Моделирование сезонных колебаний

Простейший подход к моделированию сезонных колебаний – это расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Общий вид *аддитивной модели* следующий:

$$Y = T + S + E. \quad (2.160)$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой (T), сезонной (S) и случайной (E) компонент.

Общий вид *мультипликативной модели* выглядит так:

$$Y = T \cdot S \cdot E\,. \quad (2.161)$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой (T), сезонной (S) и случайной (E) компонент.

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету значений T, S и E для каждого уровня ряда.

Процесс построения модели включает в себя следующие шаги.

1. Выравнивание исходного ряда методом скользящей средней.
2. Расчет значений сезонной компоненты S.
3. Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных ($T + E$) в аддитивной или ($T \cdot E$) в мультипликативной модели.
4. Аналитическое выравнивание уровней ($T + E$) или ($T \cdot E$) и расчет значений T с использованием полученного уравнения тренда.
5. Расчет полученных по модели значений ($T + E$) или ($T \cdot E$).
6. Расчет абсолютных и/или относительных ошибок. Если полученные значения ошибок не содержат автокорреляции, ими можно заменить исходные уровни ряда и в дальнейшем использовать временной ряд ошибок E для анализа взаимосвязи исходного ряда и других временных рядов.

2.5.5. Автокорреляция в остатках. Критерий Дарбина-Уотсона

Автокорреляция в остатках может быть вызвана несколькими причинами, имеющими различную природу.

1. Она может быть связана с исходными данными и вызвана наличием ошибок измерения в значениях результативного признака.

2. В ряде случаев автокорреляция может быть следствием неправильной спецификации модели. Модель может не включать фактор, который оказывает существенное воздействие на результат и влияние которого отражается в остатках, вследствие чего последние могут оказаться автокоррелированными. Очень часто этим фактором является фактор времени t.

От истинной автокорреляции остатков следует отличать ситуации, когда причина автокорреляции заключается в *неправильной спецификации функциональной формы модели*. В этом случае следует изменить форму модели, а не использовать специальные методы расчета параметров уравнения регрессии при наличии автокорреляции в остатках.

Один из более распространенных методов определения автокорреляции в остатках – это *расчет критерия Дарбина-Уотсона*:

$$d = \frac{\sum_{t=2}^{n}\left(\varepsilon_t - \varepsilon_{t-1}\right)^2}{\sum_{t=1}^{n}\varepsilon_t^2}. \tag{2.162}$$

То есть величина d – отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии.

Можно показать, что при больших значениях n существует следующее соотношение между критерием Дарбина-Уотсона d и коэффициентом автокорреляции остатков первого порядка r_1:

$$d = 2 \cdot (1 - r_1). \tag{2.163}$$

Таким образом, если в остатках существует полная положительная автокорреляция и $r_1 = 1$, то $d = 0$. Если в остатках полная отрицательная автокорреляция, то $r_1 = -1$ и, следовательно, $d = 4$. Если автокорреляция остатков отсутствует, то $r_1 = 0$ и $d = 2$, то есть $0 \le d \le 4$.

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина-Уотсона следующий. Выдвигается гипотеза H_0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H_1 и H_1^* состоят, соответственно, в наличии положительной или отрицательной автокорреляции в остатках. Далее по специальным таблицам (прил. 3) определяются критические значения критерия Дарбина-Уотсона d_L и d_U для заданного числа наблюдений n, числа независимых переменных модели m и уровня значимости α. По этим значениям числовой промежуток [0; 4] разбивают на пять отрезков. Принятие или отклонение каждой из гипотез с вероятностью $1 - \alpha$ осуществляется следующим образом:

$0 < d < d_L$ – есть положительная автокорреляция остатков, H_0 отклоняется, с вероятностью $P = 1 - \alpha$ принимается H_1;

$d_L < d < d_U$ – зона неопределенности;

$d_U < d < 4 - d_U$ – нет оснований отклонять H_0, т. е. автокорреляция остатков отсутствует;

$4 - d_U < d < 4 - d_L$ – зона неопределенности;

$4 - d_L < d < 4$ – есть отрицательная автокорреляция остатков, H_0 отклоняется, с вероятностью $P = 1 - \alpha$ принимается H_1^*.

Если фактическое значение критерия Дарбина-Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу H_0. Существует несколько ограничений на применение критерия Дарбина-Уотсона.

1. Он неприменим к моделям, включающим в качестве независимых переменных лаговые значения результативного признака.

2. Методика расчета и использования критерия Дарбина-Уотсона направлена только на выявление автокорреляции остатков первого порядка.

3. Критерий Дарбина-Уотсона дает достоверные результаты только для больших выборок.

2.6. Системы эконометрических уравнений

2.6.1. Классификация систем регрессионных уравнений

При использовании отдельных уравнений регрессии, например для экономических расчетов, в большинстве случаев предполагается, что аргументы (факторы) можно изменять независимо друг от друга. Однако это предположение является очень грубым: изменение одной переменной, как правило, не может происходить при абсолютной неизменности других. Ее изменение повлечет за собой изменения во всей системе взаимосвязанных признаков. Следовательно, отдельно взятое уравнение множественной регрессии не может характеризовать истинные влияния отдельных признаков на вариацию результирующей переменной.

Система уравнений в эконометрических исследованиях может быть построена поразному.

Возможна *система независимых уравнений,* когда каждая зависимая переменная y рассматривается как функция одного и того же набора факторов x:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m. \end{cases} \tag{2.164}$$

Набор факторов x_j в каждом уравнении может варьироваться. Каждое уравнение системы независимых уравнений может рассматриваться самостоятельно. Для нахождения его параметров используется метод наименьших квадратов. По существу, каждое уравнение этой системы является уравнением регрессии. Так как фактические значения зависимой переменной отличаются от теоретических на величину случайной ошибки, то в каждом уравнении присутствует величина случайной ошибки ε_i.

Если зависимая переменная y одного уравнения выступает в виде фактора x в другом уравнении, то исследователь может строить модель в виде *системы рекурсивных уравнений:*

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = b_{m1}y_1 + \ldots + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m. \end{cases} \tag{1.165}$$

В данной системе зависимая переменная у включает в каждое последующее уравнение в качестве факторов все зависимые переменные предшествующих уравнений наряду с набором собственно факторов *x*. Каждое уравнение этой системы может рассматриваться самостоятельно, и его параметры определяются методом наименьших квадратов (МНК).

Наибольшее распространение в эконометрических исследованиях получила *система взаимозависимых уравнений*. В ней одни и те же зависимые переменные в одних уравнениях входят в левую, а в других — в правую часть системы:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \ldots + b_{1m}y_m \qquad + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2m}y_m \qquad + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + \ldots + b_{3m}y_m \qquad + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = b_{m1}y_1 + b_{m2}y_2 + \ldots + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_n. \end{cases} \tag{2.166}$$

Система взаимозависимых уравнений получила название *системы совместных, одновременных уравнений*. Тем самым подчеркивается, что в системе одни и те же переменные одновременно рассматриваются как зависимые в одних уравнениях и как независимые в других. В эконометрике эта система уравнений называется также *структурной формой модели*. В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим. С этой целью используются специальные приемы оценивания.

2.6.2. Оценка параметров систем одновременных уравнений

Система совместных, одновременных уравнений (или структурная форма модели) обычно содержит эндогенные и экзогенные переменные.

Эндогенные переменные — это зависимые переменные, число которых равно числу уравнений в системе и которые обозначаются через *y*.

Экзогенные переменные — это предопределенные переменные, влияющие на эндогенные переменные, но не зависящие от них. Обозначаются через *x*.

Классификация переменных на эндогенные и экзогенные зависит от теоретической концепции принятой модели. Экономические переменные могут выступать в одних моделях как эндогенные, а в других как экзогенные. Внеэкономические переменные (например, климатические условия, социальное положение, пол, возрастная категория) входят в систему только как экзогенные переменные. В качестве экзогенных переменных могут рассматриваться значения эндогенных переменных за предшествующий период времени (*лаговые переменные*).

Структурная форма модели позволяет увидеть влияние изменений любой экзогенной переменной на значения эндогенной. Целесообразно в качестве экзогенных выбирать такие переменные, которые могут быть объектом регулирования. Меняя их и управляя ими, можно заранее иметь целевые значения эндогенных переменных.

Структурная форма модели в правой части содержит при эндогенных переменных коэффициенты b_{ik} и экзогенных переменных — коэффициенты a_{ij}, которые называются *структурными коэффициентами* модели. Все переменные в модели выражены в отклонениях от среднего уровня, т. е. под *x* подразумевается $x - \bar{x}$, а под *y* — соответственно $y - \bar{y}$. Поэтому свободный член в каждом уравнении системы (6.3) отсутствует.

Использование МНК для оценивания структурных коэффициентов модели дает, как принято считать в теории, смещенные и несостоятельные оценки. Поэтому обычно для определения структурных коэффициентов модели структурная форма модели преобразуется в *приведенную форму* модели.

Приведенная форма модели представляет собой систему линейных функций эндогенных переменных от экзогенных:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + \ldots + \delta_{1n}x_n + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + \ldots + \delta_{2n}x_n + u_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = \delta_{m1}x_1 + \delta_{m2}x_2 + \ldots + \delta_{mn}x_n + u_m, \end{cases} \tag{2.167}$$

где δ_{ij} – коэффициенты приведенной формы модели, u_i – остаточная величина для приведенной формы.

По своему виду приведенная форма модели ничем не отличается от системы независимых уравнений, параметры которой оцениваются традиционным МНК. Применяя МНК, можно оценить δ_{ij}, а затем оценить значения эндогенных переменных через экзогенные.

Коэффициенты приведенной формы модели представляют собой нелинейные функции коэффициентов структурной формы модели. Рассмотрим это положение на примере простейшей структурной модели, выразив коэффициенты приведенной формы модели через коэффициенты структурной.

Для структурной модели вида

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2 \end{cases} \tag{1.168}$$

приведенная форма модели имеет вид

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + u_2. \end{cases} \tag{2.169}$$

Из первого уравнения (6.5) можно выразить y_2 следующим образом (ради упрощения опускаем случайную величину):

$$y_2 = \frac{y_1 - a_{11}x_1}{b_{12}}.$$

Подставляя во второе уравнение (6.5), имеем

$$\frac{y_1 - a_{11}x_1}{b_{12}} = b_{21}y_1 + a_{22}x_2,$$

откуда

$$y_1 = \frac{a_{11}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}b_{12}}{1 - b_{12}b_{21}}x_2.$$

Поступая аналогично со вторым уравнением системы (3.5), получим

$$y_2 = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}}{1 - b_{12}b_{21}}x_2,$$

т. е. система (6.5) принимает вид

$$\begin{cases} y_1 = \dfrac{a_{11}}{1 - b_{12}b_{21}}x_1 + \dfrac{a_{22}b_{12}}{1 - b_{12}b_{21}}x_2, \\ y_2 = \dfrac{a_{11}b_{21}}{1 - b_{12}b_{21}}x_1 + \dfrac{a_{22}}{1 - b_{12}b_{21}}x_2. \end{cases}$$

Отсюда, можно сделать вывод о том, что коэффициенты приведенной формы модели будут выражаться через коэффициенты структурной формы следующим образом:

$$\delta_{11} = \frac{a_{11}}{1 - b_{12}b_{21}}, \quad \delta_{12} = \frac{a_{22}b_{12}}{1 - b_{12}b_{21}},$$

$$\delta_{21} = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}, \quad \delta_{22} = \frac{a_{22}}{1 - b_{12}b_{21}}.$$

Следует заметить, что приведенная форма модели, хотя и позволяет получить значения эндогенной переменной через значения экзогенных переменных, но аналитически уступает структурной форме модели, так как в ней отсутствуют оценки взаимосвязи между эндогенными переменными.

2.6.3. Проблема идентификации структурных моделей

При переходе от приведенной формы модели к структурной эконометрист сталкивается с проблемой идентификации. *Идентификация* – это единственность соответствия между приведенной и структурной формами модели.

Структурная модель (2.165) в полном виде содержит $m \cdot (\mathrm{m} + n - 1)$ параметров, а приведенная форма модели в полном виде содержит $m \cdot n$ параметров. То есть в полном виде структурная модель содержит большее число параметров, чем приведенная форма модели. Соответственно, $m \cdot (m + n - 1)$ параметров структурной модели не могут быть однозначно определены из $m \cdot n$ параметров приведенной формы модели.

Чтобы получить единственно возможное решение для структурной модели, необходимо предположить, что некоторые из структурных коэффициентов модели, ввиду слабой взаимосвязи признаков с эндогенной переменной из левой части системы, равны нулю. Тем самым уменьшится число структурных коэффициентов модели. Уменьшение числа структурных коэффициентов модели возможно и другим путем, например, за счет приравнивания некоторых коэффициентов друг к другу, т. е. предположения, что их воздействие на формируемую эндогенную переменную одинаково. На структурные коэффициенты могут накладываться, например, ограничения вида $b_{ik} + a_{ij} = 0$.

С позиции идентифицируемости структурные модели можно подразделить на три вида: идентифицируемые; неидентифицируемые; сверхидентифицируемые.

Модель *идентифицируема*, если все структурные ее коэффициенты определяются однозначно единственным образом по коэффициентам приведенной формы модели, т. е. если число параметров структурной модели равно числу параметров приведенной формы модели. В этом случае структурные коэффициенты модели оцениваются через параметры приведенной формы модели и модель идентифицируема.

Модель *неидентифицируема*, если число приведенных коэффициентов меньше числа структурных коэффициентов и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели.

Модель *сверхидентифицируема*, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. В этой модели число структурных коэффициентов меньше числа коэффициентов приведенной формы. Сверхидентифицируемая модель, в отличие от неидентифицируемой, практически решаема, но требует для этого специальных методов исчисления параметров.

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых требуется проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение.

Выполнение *условия идентифицируемости модели* проверяется для каждого уравнения системы. Чтобы уравнение было идентифицируемо, необходимо, чтобы число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного.

Если обозначить число эндогенных переменных в i-м уравнении системы через H, а число экзогенных (предопределенных) переменных, которые содержатся в системе, но не входят в данное уравнение, – через D, то условие идентифицируемости модели может быть записано в виде следующего счетного правила (табл. 2.3).

Таблица 2.3

$D + 1 = H$	уравнение идентифицируемо
$D + 1 < H$	уравнение неидентифицируемо
$D + 1 > H$	уравнение сверхидентифицируемо

Для оценки параметров структурной модели система должна быть идентифицируема или сверхидентифицируема.

Рассмотренное счетное правило отражает необходимое, но недостаточное условие идентификации. Более точно условия идентификации определяются, если накладывать ограничения на коэффициенты матриц параметров структурной модели. Уравнение идентифицируемо, если по отсутствующим в нем переменным (эндогенным и экзогенным) можно из коэффициентов при них в других уравнениях системы получить матрицу, определитель которой не равен нулю, а ранг матрицы не меньше, чем число эндогенных переменных в системе без одного.

В эконометрических моделях часто наряду с уравнениями, параметры которых должны быть статистически оценены, используются балансовые тождества переменных, коэффициенты при которых равны ± 1. В этом случае, хотя само тождество и не требует проверки на идентификацию, ибо коэффициенты при переменных в тождестве известны, в проверке на идентификацию собственно структурных уравнений системы тождества участвуют.

2.6.4. Методы оценки параметров структурной модели

Коэффициенты структурной модели могут быть оценены разными способами в зависимости от вида системы одновременных уравнений. Наибольшее распространение в литературе получили следующие методы:

1) косвенный метод наименьших квадратов;
2) двухшаговый метод наименьших квадратов;
3) трехшаговый метод наименьших квадратов;
4) метод максимального правдоподобия с полной информацией;
5) метод максимального правдоподобия при ограниченной информации.

Рассмотрим вкратце сущность каждого из этих методов.

Косвенный метод наименьших квадратов (КМНК) применяется в случае точно идентифицируемой структурной модели. Процедура применения КМНК предполагает выполнение следующих этапов работы.

1. Структурная модель преобразовывается в приведенную форму модели.

2. Для каждого уравнения приведенной формы модели обычным МНК оцениваются приведенные коэффициенты δ_{ij}.

3. Коэффициенты приведенной формы модели трансформируются в параметры структурной модели.

Если система сверхидентифицируема, то КМНК не используется, ибо он не дает однозначных оценок для параметров структурной модели. В этом случае могут использоваться разные методы оценивания, среди которых наиболее распространенным и простым является двухшаговый метод наименьших квадратов (ДМНК).

Основная идея ДМНК – на основе приведенной формы модели получить для сверхидентифицируемого уравнения теоретические значения эндогенных переменных, содержащихся в правой части уравнения.

Далее, подставив их вместо фактических значений, можно применить обычный МНК к структурной форме сверхидентифицируемого уравнения. Метод получил название двухшагового МНК, ибо дважды используется МНК: на первом шаге при определении приведенной формы модели и нахождении на ее основе оценок теоретических значений эндогенной переменной $\widehat{y}_i = \delta_{i1}x_1 + \delta_{i2}x_2 + \ldots + \delta_{in}x_n$ и на втором шаге применительно к структурному сверхидентифицируемому уравнению при определении структурных коэффициентов модели по данным теоретических (расчетных) значений эндогенных переменных.

Сверхидентифицируемая структурная модель может быть двух типов:

1) все уравнения системы сверхидентифицируемы;
2) система содержит наряду со сверхидентифицируемыми точно идентифицируемые уравнения.

Если все уравнения системы сверхидентифицируемые, то для оценки структурных коэффициентов каждого из них используется ДМНК. Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

3. ПРАКТИКУМ

Практикум позволяет студенту получить навыки решения задач по дисциплине «Эконометрика». Перед выполнением практических заданий по текущей теме необходимо внимательно изучить лекции по этой теме, при необходимости обращаясь к источникам, указанным в библиографическом списке. Решение практических задач позволит закрепить в памяти теоретические положения дисциплины. Далее, студент должен разобрать приведенные в практикуме примеры решения задач по изучаемой теме. После этого необходимо приступать к выполнению задач для самостоятельного решения и тестовых заданий с использованием примеров, курса лекций и литературы. С помощью неоднократного выполнения заданий практикума студент должен научиться решать задачи и выполнять тестовые задания без применения первоисточников. Успешное самостоятельное выполнение всех заданий практикума без использования примеров, курса лекций и литературы означает, что студент получил достаточно знаний и умений по данной теме.

Практическая работа 3.1
Виды эконометрических моделей. Введение в регрессионный анализ

Цели: закрепление основных положений и методов построения эконометрических моделей; усвоение классификации этих моделей; вычисление вероятностных характеристик случайных величин.

Примеры решения задач

По десяти шахтам известны следующие данные, характеризующие процесс добычи угля.

Таблица 3.1

Номер шахты, i	1	2	3	4	5	6	7	8	9	10
Мощность пласта, x (м)	8	11	12	9	8	8	9	9	8	12
Сменная добыча угля на одного рабочего, y (т)	5	10	10	7	5	6	6	5	6	8

1. Постройте поле корреляции и по виду облака рассеяния сделайте предположение о форме связи между переменными x и y.

2. Рассчитайте характеристики случайных величин:

1) средние значения $\bar{x}$, $\bar{y}$;
2) выборочные дисперсии (вариации) $var(x)$, $var(y)$;
3) стандартные (среднеквадратические) отклонения $S(x)$, $S(y)$;
4) выборочную ковариацию (выборочный корреляционный момент) $cov(x,y)$;
5) выборочный коэффициент корреляции r_{xy}.

3. С помощью этих характеристик оцените степень рассеяния случайных величин вокруг средних значений и тесноту связи переменных.

Решение. Для удобства вычислений составим таблицу.

Таблица 3.2

i	x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$x_i \cdot y_i$
1	2	3	4	5	6	7	8
1	8	5	-1,4	1,96	-1,8	3,24	40
2	11	10	1,6	2,56	3,2	10,24	110

i	x_i	y_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$x_i \cdot y_i$
3	12	10	2,6	6,76	3,2	10,24	120
4	9	7	-0,4	0,16	0,2	0,04	63
5	8	5	-1,4	1,96	-1,8	3,24	40
6	8	6	-1,4	1,96	-0,8	0,64	48
7	9	6	-0,4	0,16	-0,8	0,64	54
8	9	5	-0,4	0,16	-1,8	3,24	45
9	8	6	-1,4	1,96	-0,8	0,64	48
10	12	8	2,6	6,76	1,2	1,44	96
Итого	94	68		24,4		33,6	664
Среднее значение	9,4	6,8		2,44		3,36	66,4

1. Построим поле корреляции (диаграмму рассеяния) по данным 2 и 3 столбцов таблицы (рис. 3.1).

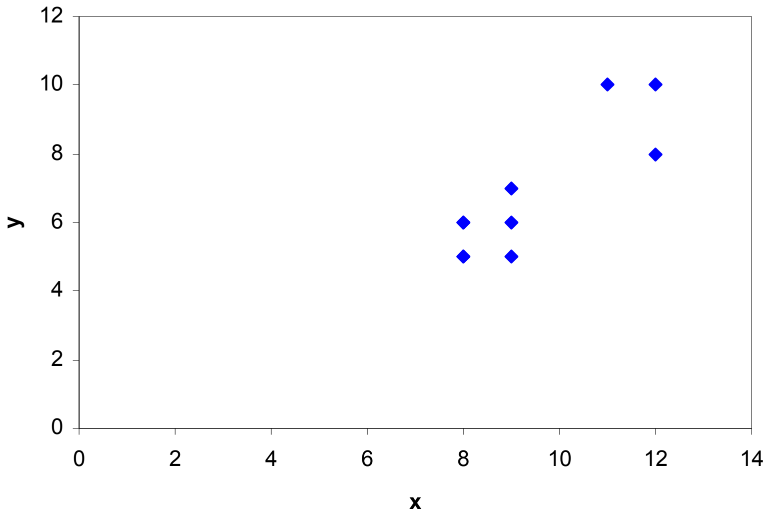

Рис. 3.1. Поле корреляции

По виду облака рассеяния предположим линейную форму связи между переменными x и y.

2. а) По формулам (2.3) рассчитаем средние значения $\bar{x}$, $\bar{y}$:

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\ \bar{x}=\frac{8+11+12+...+12}{10}=9{,}4 \text{ (см. столбец 2 таблицы);}$$

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i,\ \bar{y}=\frac{5+10+10+...+8}{10}=6{,}8 \text{ (см. столбец 3 таблицы);}$$

б) для вычисления выборочных дисперсий (вариаций) $var(x)$, $var(y)$ используем формулы (2.5):

$$\operatorname{var}(x)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=2{,}44 \text{ (столбцы 4 и 5 таблицы);}$$

$$\operatorname{var}(y)=\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2=3{,}36 \text{ (столбцы 6 и 7 таблицы);}$$

в) стандартные (среднеквадратические) отклонения $S(x)$, $S(y)$ найдем по формулам (2.6):

$$S(x)=\sqrt{\operatorname{var}(x)}=\sqrt{2{,}44}=1{,}56;\quad S(x)=\sqrt{\operatorname{var}(x)}=\sqrt{3{,}36}=1{,}83;$$

г) выборочная ковариация (выборочный корреляционный момент) $cov(x, y)$ находится по формуле (2.8). Среднее значение произведений вычисляется в 8 столбце таблицы, а из него вычитается произведение средних:

$$\mathrm{cov}(x,y)=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\overline{xy}-\bar{x}\cdot\bar{y}=66{,}4-9{,}4\cdot 6{,}8=2{,}48;$$

д) вычислим выборочный коэффициент корреляции r_{xy} по формуле (1.7):

$$r_{xy}=\frac{\mathrm{cov}(x,y)}{S(x)S(y)}=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{S_xS_y}=\frac{2{,}48}{1{,}56\cdot 1{,}83}=0{,}87.$$

3. Степень рассеяния величин x и y вокруг средних значений $\bar{x}=9{,}4$ (м), $\bar{y}=6{,}8$ (т) хорошо характеризуют стандартные отклонения $S(x)$ = 1,56 (м) и $S(y)$ = 1,83 (т). Рассеяние x находится в пределах 17%, а рассеяние y – 27%. По выборочной ковариации можно судить о зависимости двух случайных величин, их небольшом рассеянии относительно среднего значения. Коэффициент корреляции r_{xy} = 0,87, величина которого близка к единице, говорит о наличии сильной линейной зависимости между переменными x и y.

Задачи для самостоятельного решения

В табл. 3.3 приведены данные по годовым доходностям акций двух компаний, относящихся к одной отрасли.

Таблица 3.3

Номер наблюдения, i	Доходность акций А, x (%)	Доходность акций Б, y (%)
1	-2,54	-5,31
2	26,65	16,84
3	4,44	0,07
4	17,12	10,03
5	10,19	4,98
6	13,88	7,52
7	4,55	0,23
8	10,28	5,3
9	11,76	5,94
10	11,89	6,09
11	5,14	0,93
12	7,7	3,22
13	7,17	2,08
14	7,57	2,81
15	17,46	10,73

1. Постройте поле корреляции и по виду облака рассеяния сделайте предположение о форме связи между переменными x и y.

2. Рассчитайте характеристики случайных величин:

1) средние значения $\bar{x}$, $\bar{y}$;
2) выборочные дисперсии (вариации) $var(x)$, $var(y)$;
3) стандартные (среднеквадратические) отклонения $S(x)$, $S(y)$;
4) выборочную ковариацию (выборочный корреляционный момент) $cov(x, y)$;
5) выборочный коэффициент корреляции r_{xy}.

3. С помощью этих характеристик оцените степень рассеяния случайных величин вокруг средних значений и тесноту связи переменных.

Тесты

1. Если коэффициент корреляции равен -1, то между переменными...
a) связь нелинейна;
b) связь линейна, функциональна;
c) связь слабая;
d) связь отсутствует.

2. Если коэффициент корреляции равен 1, то между переменными...
a) связь линейна, функциональна;
b) связь нелинейна;
c) связь слабая;
d) связь отсутствует.

3. Значение коэффициента корреляции находится в промежутке...
a) [0; 1];
b) [-1; 1];
c) [-1; 0];
d) (-1; 1).

4. Значение линейного коэффициента корреляции характеризует тесноту ________ связи
a) множественной;
b) нелинейной;
c) линейной;
d) случайной.

5. Эндогенными переменными являются...
a) зависимые переменные;
b) переменные, значение которых определяется вне системы;
c) случайные переменные;
d) независимые переменные.

6. Основной задачей эконометрики является...
a) установление связей между различными процессами в обществе и техническим процессом;
b) исследование взаимосвязей экономических явлений и процессов;
c) отражение особенностей социального развития общества;
d) анализ технического прогресса на примере социально-экономических показателей.

7. Гетероскедастичность остатков подразумевает...
a) независимость математического ожидания остатков от значения фактора;
b) зависимость дисперсии остатков от значения фактора;
c) зависимость математического ожидания остатков от значения фактора;
d) постоянство дисперсии остатков независимо от значения фактора.

8. Предположением регрессионного анализа является...
a) отсутствие корреляции между результатом и фактором;
b) присутствие автокорреляции между результатом и фактором;
c) присутствие автокорреляции в остатках;
d) отсутствие автокорреляции в остатках.

9. Экзогенными переменными являются...
a) переменные, значение которых определяется внутри системы;
b) зависимые переменные;
c) случайные переменные;
d) независимые переменные.

10. Если значение коэффициента корреляции равно единице, то связь между результатом и фактором...
a) отсутствует;
b) вероятностная;
c) функциональная;
d) стохастическая.

11. Графическое изображение наблюдений на декартовой плоскости координат называется полем...
a) автокорреляции;
b) случайных воздействий;
c) регрессии;
d) корреляции.

12. Отсутствие автокорреляции в остатках предполагает, что значения ___________ не зависят друг от друга.
a) фактора;
b) результата;
c) независимых переменных;
d) остатков.

13. Эндогенными переменными **не являются**...
a) зависимые переменные;
b) независимые переменные;
c) переменные, значение которых определяется внутри системы;
d) переменные y в уравнениях системы вида $y = f(x)$.

14. Предпосылки регрессионного анализа исследуют поведение...
a) остаточных величин;
b) неслучайных величин;
c) переменных уравнения регрессии;
d) параметров уравнения регрессии.

15. Построена модель парной регрессии зависимости предложения от цены $y = a + bx + \varepsilon$. Влияние случайных факторов на величину предложения в этой модели учтено посредством...
a) случайной величины x;
b) случайной величины ε;
c) параметра b;
d) константы ε.

16. К ошибкам спецификации относится...
a) учет в модели случайных факторов;
b) неправильный выбор той или иной математической функции;

c) однородность выбранной совокупности;
d) учет в модели существенных факторов.

17. Выбор формы зависимости экономических показателей и определение количества факторов в модели называется эконометрической модели.
a) апробацией;
b) линеаризацией;
c) спецификацией;
d) идентификацией.

18. Если факторы входят в модель как сумма, то модель называется...
a) аддитивной;
b) суммарной;
c) мультипликативной;
d) производной.

19. Если факторы входят в модель как произведение, то модель называется...
a) мультипликативной;
b) аддитивной;
c) суммарной;
d) производной.

20. Предпосылкой регрессионного анализа является то, что...
a) при увеличении моделируемых значений результативного признака значение остатка увеличивается;
b) остаток подчиняется закону нормального распределения;
c) при уменьшении моделируемых значений результативного признака значение остатка уменьшается;
d) остаточные величины имеют неслучайный характер.

21. Предпосылкой регрессионного анализа **не является** условие...
a) гомоскедастичности остатков;
b) неслучайного характера остатков;
c) отсутствия автокорреляции в остатках;
d) случайного характера остатков.

Практическая работа 3.2
Парная линейная регрессия

Цели: применение метода наименьших квадратов (МНК) расчета параметров уравнения парной линейной регрессии; использование интервальных оценок для прогнозирования по уравнению регрессии; освоение методов проверки качества полученной модели.

Примеры решения задач

По результатам опроса восьми групп семей известны данные о связи расходов населения на продукты питания с уровнем доходов семьи (табл. 3.4).

Таблица 3.4

Расходы на продукты питания, y, тыс. руб.	0,9	1,2	1,8	2,2	2,6	2,9	3,3	3,8
Доходы семьи, x, тыс. руб.	1,2	3,1	5,3	7,4	9,6	11,8	14,5	18,7

1. Постройте линейную регрессионную модель связи переменных, где y интерпретируется как объясняемая переменная, а x – объясняющая, используя оценки наименьших квадратов.

2. Рассчитайте линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.

3. Оцените статистическую значимость параметров регрессии и корреляции на уровне значимости $\alpha = 0{,}05$ с помощью F-критерия Фишера и t-критерия Стьюдента.

4. Выполните прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума x, составляющем 110% от среднего уровня.

5. Оцените точность прогноза, рассчитав 95% доверительные интервалы для среднего и индивидуального значения объясняемой переменной при том же значении x.

6. Найдите с надежностью 0,95 интервальные оценки параметров уравнения регрессии α и β.

7. На одном графике (графике подбора) постройте исходные данные и теоретическую прямую. Сделайте вывод.

Решение

1. Предположим, что связь между доходами семьи и расходами на продукты питания линейная. Для подтверждения нашего предположения построим поле корреляции (рис. 3.2).

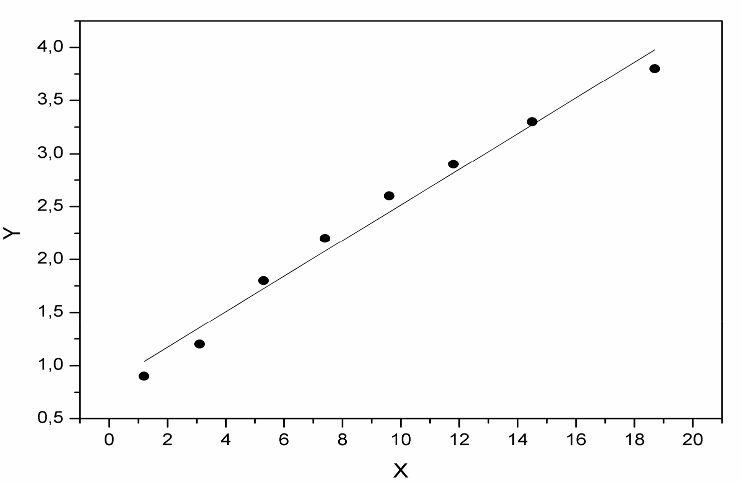

Рис. 3.2. Поле корреляции

По графику видно, что точки выстраиваются в некоторую прямую линию.

Для удобства дальнейших вычислений составим таблицы (табл. 3.5, 3.6).

Таблица 3.5

i	x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	1,2	0,9	-7,75	60,063	-1,4375	2,066
2	3,1	1,2	-5,85	34,223	-1,1375	1,294
3	5,3	1,8	-3,65	13,323	-0,5375	0,289
4	7,4	2,2	-1,55	2,403	-0,1375	0,019
5	9,6	2,6	0,65	0,423	0,2625	0,069
6	11,8	2,9	2,85	8,123	0,5625	0,316
7	14,5	3,3	5,55	30,803	0,9625	0,926
8	18,7	3,8	9,75	80,103	1,4625	2,139
Итого	71,6	18,7		244,47		7,119
Среднее значение	8,95	2,34		30,56		0,89

Таблица 3.6

	x	y	$x \cdot y$	x^2	y^2	$\hat{y}_x$	$y-\hat{y}_x$	$(y-\hat{y}_x)^2$	A_i,%
1	1,2	0,9	1,08	1,44	0,81	1,038	-0,138	0,0190	15,33
2	3,1	1,2	3,72	9,61	1,44	1,357	-0,157	0,0246	13,08
3	5,3	1,8	9,54	28,09	3,24	1,726	0,074	0,0055	4,11
4	7,4	2,2	16,28	54,76	4,84	2,079	0,121	0,0146	5,50
5	9,6	2,6	24,96	92,16	6,76	2,449	0,151	0,0228	5,81
6	11,8	2,9	34,22	139,24	8,41	2,818	0,082	0,0067	2,83
7	14,5	3,3	47,85	210,25	10,89	3,272	0,028	0,0008	0,85
8	18,7	3,8	71,06	349,69	14,44	3,978	-0,178	0,0317	4,68
Итого	71,6	18,7	208,71	885,24	50,83	18,717	-0,017	0,1257	52,19
Среднее значение	8,95	2,34	26,09	110,66	6,35	2,34	–	0,0157	6,52
S	5,53	0,943	–	–	–	–	–	–	–
var	30,56	0,89	–	–	–	–	–	–	–

Рассчитаем МНК-оценки параметров линейного уравнения парной регрессии $\hat{y}_x = a + b \cdot x$. Для этого воспользуемся формулами (2.28) и (2.29):

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)} = \frac{26{,}09 - 8{,}95 \cdot 2{,}34}{30{,}56} = 0{,}168\,;$$

$$a = \bar{y} - b \cdot \bar{x} = 2{,}34 - 0{,}168 \cdot 8{,}95 = 0{,}836\,.$$

Получили уравнение: $\hat{y}_x = 0{,}836 + 0{,}168 \cdot x$. То есть с увеличением дохода семьи на 1000 руб. расходы на питание увеличиваются на 168 руб.

2. Уравнение линейной регрессии всегда дополняется показателем тесноты связи – линейным коэффициентом корреляции r_{xy} (формула 2.7):

$$r_{xy} = \frac{\operatorname{cov}(x, y)}{S(x)S(y)} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x S_y} = \frac{26{,}09 - 8{,}95 \cdot 2{,}34}{5{,}53 \cdot 0{,}943} = 0{,}993\,.$$

Близость коэффициента корреляции к 1 указывает на тесную линейную связь между признаками.

Для модели линейной регрессии коэффициент детерминации равен квадрату коэффициента корреляции. Коэффициент детерминации $R^2 = {r_{xy}}^2 = 0{,}987$ показывает, что уравнением регрессии объясняется 98,7% дисперсии результативного признака, а на долю прочих факторов приходится лишь 1,3%. То есть 98,7% вариации расходов на продукты питания (y) объясняется вариацией фактора x – дохода семьи.

Среднюю ошибку аппроксимации находим по формуле (2.53) с помощью столбца 10 табл. 3.5. $A_i = \left|\frac{y_i - \hat{y}_{x_i}}{y_i}\right| \times 100\%$, $\bar{A} = \frac{1}{n}\sum A_i = 6{,}52\%$ говорит о хорошем качестве уравнения регрессии, т. е. свидетельствует о хорошем подборе модели к исходным данным.

3. Оценим качество уравнения регрессии в целом с помощью F-критерия Фишера. Рассчитаем фактическое значение F-критерия. Так как коэффициент детерминации уже известен, проще всего использовать формулу (2.36):

$$F = \frac{R^2(n-m)}{(1-R^2)(m-1)} = \frac{0{,}987}{1-0{,}987} \cdot 6 = 455{,}54,$$

где m – число оцениваемых параметров уравнения регрессии; n – число наблюдений.

Табличное значение F-критерия найдем по прил. 1 ($k_1 = m - 1 = 2 - 1 = 1$, $k_2 = n - 2 = 8 - 2 = 6$, $\alpha = 0{,}05$): $F_{табл} = 0{,}99$. Так как $F_{факт} > F_{табл}$, то признается статистическая значимость уравнения в целом.

Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитаем t-критерий Стьюдента каждого из показателей. Определим оценку дисперсии ошибки прогноза по формуле (2.36):

$$S^2{}_{\varepsilon_i} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-m} = \frac{0{,}1257}{8-2} = 0{,}021\ .$$

Рассчитаем случайные ошибки параметров линейной регрессии и коэффициента корреляции по формулам (2.60), (2.61) и (2.62):

$$S_b = \frac{S_{\varepsilon_i}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \frac{0.021}{\sqrt{244{,}47}} = 0{,}0093;$$

$$S_a = S_{\varepsilon_i}\sqrt{\frac{\sum_{i=1}^{n}(x_i)^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}} = 0{,}021 \cdot \sqrt{\frac{885{,}24}{8 \cdot 244{,}47}} = 0{,}0975;$$

$$S_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-0{,}987}{8-2}} = 0{,}0465.$$

Фактические значения t-статистик:

$$t_b = \frac{b}{S_b} = \frac{0{,}168}{0{,}0093} = 18{,}065; \quad t_a = \frac{b}{S_a} = \frac{0{,}836}{0{,}0975} = 8{,}574; \quad t_r = \frac{b}{S_r} = \frac{0{,}994}{0{,}0465} = 21{,}376.$$

Табличное значение t-критерия Стьюдента при $\alpha = 0{,}05$ и числе степеней свободы $n - 2 = 6$ есть $t_{табл} = 2{,}447$. Так как фактические значения t-статистики превосходят табличное значение ($t_b > t_{табл}$, $t_a > t_{табл}$ и $t_r > t_{табл}$), то признаем статистическую значимость параметров регрессии и показателя тесноты связи.

4. Полученные оценки уравнения регрессии позволяют использовать его для прогноза.

Найдем прогнозное значение результативного фактора $\hat{y}_p$ при значении признака-фактора, составляющем 110% от среднего уровня $x_p = 1{,}1 \cdot \bar{x} = 1{,}1 \cdot 8{,}95 = 9{,}845$ (тыс. руб.). Найдем предполагаемые расходы на питание при данных доходах.

$$\hat{y}_p = a + b \cdot x = 0{,}836 + 0{,}168 \cdot 9{,}845 = 2{,}490 \text{ (тыс. руб.)}.$$

Таким образом, если доходы семьи составят 9,845 тыс. руб., то расходы на питание будут 2,490 тыс. руб.

5. Оценим точность прогноза, рассчитав 95% доверительные интервалы для среднего и индивидуального значения объясняемой переменной при том же значении x.

Оценку дисперсии значения $\hat{y}$, найденного по уравнению регрессии, рассчитаем по формуле (2.42):

$$S_{\hat{y}}^2 = S_{\varepsilon_i}^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) = 0{,}021\left(\frac{1}{8} + \frac{(9{,}845 - 8{,}95)^2}{244{,}47}\right) = 0{,}0027.$$

Доверительный интервал для математического ожидания (среднего значения) $\hat{y}$, найденного по уравнению регрессии по формуле (2.43):

$$\hat{y} - t_{1-\alpha;k} S_{\hat{y}} \le M_{\hat{y}}(x) \le \hat{y} + t_{1-\alpha;k} S_{\hat{y}};$$
$$2{,}49 - 2{,}447 \cdot \sqrt{0{,}0027} \le M_{\hat{y}}(x) \le 2{,}49 + 2{,}447 \cdot \sqrt{0{,}0027};$$
$$2{,}36 \le M_{\hat{y}}(x) \le 2{,}67.$$

Таким образом, с вероятностью 0,95, семьи с доходами 9845 руб. будут расходовать на продукты питания в среднем от 2360 до 2670 руб.

При определении доверительного интервала для индивидуальных значений y_0 зависимой переменной используем формулу (2.44):

$$S_{\hat{y}_0}^2 = S_{\varepsilon_i}^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right] = 0{,}021 \left(1 + \frac{1}{8} + \frac{(9{,}845 - 8{,}95)^2}{244{,}47} \right) = 0{,}0237.$$

Найдем доверительный интервал прогноза по формуле (2.45):

$$\hat{y}_0 - t_{1-\alpha;n-2} S_{\hat{y}_0} \le y_0 \le \hat{y}_0 + t_{1-\alpha;n-2} S_{\hat{y}_0};$$
$$2{,}49 - 2{,}447 \cdot \sqrt{0{,}0237} \le y_0 \le 2{,}49 + 2{,}447 \cdot \sqrt{0{,}0237};$$
$$2{,}113 \le y_0 \le 2{,}867.$$

Таким образом, с вероятностью 0,95, семья с доходом 9845 руб. будет расходовать на продукты питания от 2113 до 2867 руб.

6. Рассчитаем доверительные интервалы для параметров регрессии a и b: $a \pm t_{1-\alpha;\,n-2} \cdot S_a$ и $b \pm t_{1-\alpha;\,n-2} \cdot S_b$. Получим, что $a \in [0{,}597;\ 1{,}075]$ и $b \in [0{,}145;\ 0{,}191]$.

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью $p = 1 - \alpha = 0{,}95$ параметры a и b, находясь в указанных границах, не принимают нулевых значений, т. е. не являются статистически незначимыми и существенно отличны от нуля.

7. Теперь на одном графике изобразим исходные данные и линию регрессии (рис. 3.3).

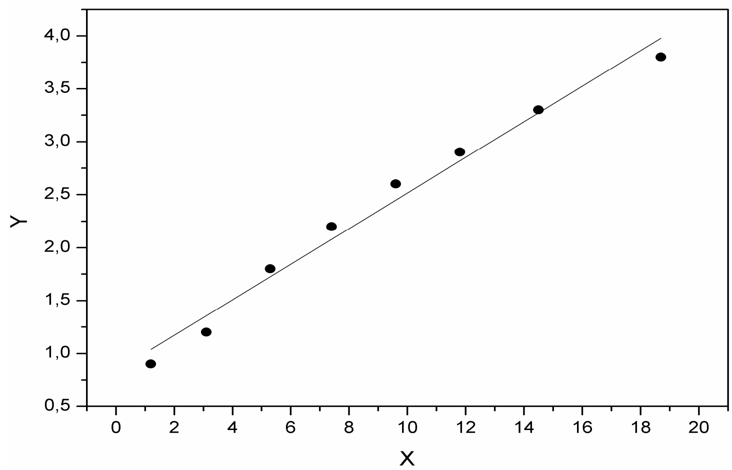

Рис. 3.3. График подбора

С помощью графика подбора можно наглядно убедиться, что построенная модель хорошо аппроксимирует исходные данные.

Задачи для самостоятельного решения

Имеются статистические данные за ряд лет о доходах домашних хозяйств и их расходах на местный автотранспорт (табл. 3.7).

Таблица 3.7

Номер наблюдения, i	Доход домашнего хозяйства, x($)	Расходы домашнего хозяйства на местный автотранспорт, y ($)
1	695,2	3,1
2	751,9	3,3
3	810,3	3,4
4	914	3,6
5	998,1	4
6	1096,2	4,4
7	1194,3	4,7
8	1313,5	5
9	1474,3	5,5
10	1650,5	6,2
11	1828,7	6,3
12	2040,9	6,2
13	2180,1	6,6
14	2333,2	6,6

1. Постройте линейную регрессионную модель связи переменных, где y интерпретируется как объясняемая переменная, а x – объясняющая, используя оценки наименьших квадратов.

2. Рассчитайте линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.

3. Оцените статистическую значимость параметров регрессии и корреляции на уровне значимости $\alpha = 0,05$ с помощью F-критерия Фишера и t-критерия Стьюдента.

4. Выполните прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума x, составляющем 110% от среднего уровня.

5. Оцените точность прогноза, рассчитав 95% доверительные интервалы для среднего и индивидуального значения объясняемой переменной при том же значении x.

6. Найдите интервальные оценки параметров уравнения регрессии α и β с надежностью 0,95.

7. На одном графике (графике подбора) постройте исходные данные и теоретическую прямую. Сделайте вывод.

Тесты

1. Значение коэффициента корреляции не характеризует...

a) корень из значения коэффициента детерминации;
b) тесноту связи;
c) силу связи;
d) статистическую значимость уравнения.

2. Если расчетное значение критерия Фишера больше табличного значения, то гипотеза о статистической значимости уравнения...

a) принимается;
b) отвергается;

c) не принимается;
d) не отвергается.

3. Если расчетное значение критерия Фишера меньше или равно табличному значению, то гипотеза о статистической незначимости уравнения...
a) не принимается;
b) не отвергается;
c) принимается;
d) отвергается.

4. Оценка значимости параметров уравнения регрессии осуществляется по критерию...
b) Фишера;
c) Стьюдента;
d) Дарбина-Уотсона;
e) Пирсона.

5. Общая дисперсия служит для оценки влияния...
a) учтенных в модели факторов;
b) случайных воздействий;
c) не учтенных в модели факторов;
d) как учтенных факторов, так и случайных воздействий.

6. Оценка значимости уравнения в целом осуществляется по критерию...
a) Стьюдента;
b) Фишера;
c) Дарбина-Уотсона;
d) Пирсона.

7. Факторaная дисперсия служит для оценки...
a) как учтенных факторов, так и случайных воздействий;
b) учтенных в модели факторов;
c) не учтенных в модели факторов;
d) случайных воздействий.

8. Минимальная дисперсия остатков характерна для оценок, обладающих свойством...
a) несостоятельности;
b) эффективности;
c) несмещенности;
d) состоятельности.

9. Остаточная дисперсия служит для оценки влияния...
a) явно учтенных в модели факторов;
b) величины постоянной составляющей в уравнении;
c) случайных воздействий;
d) как учтенных факторов, так и случайных воздействий.

10. Величина отклонений фактических значений результативного признака от его теоретических значений представляет собой...
a) показатель эластичности;
b) ошибку корреляции;

c) значение критерия Фишера;
d) ошибку аппроксимации.

11. Для модели зависимости дохода населения (руб.) от объема производства (млн руб.) получено уравнение $y = 0{,}003x + 1200 + e$. При изменении объема производства на 1 млн руб. доход в среднем изменится...
a) на 1200 млн руб.;
b) 0,003 млн руб.;
c) 0,003 руб.;
d) 1200 руб.

12. В линейном уравнении парной регрессии $y = a + bx + \varepsilon$ коэффициентом регрессии является значение:
a) переменной x;
b) параметра a;
c) параметров a и b;
d) параметра b.

13. Значение коэффициента корреляции равно 0,9. Следовательно, значение коэффициента детерминации равно...
a) 0,81;
b) 0,1;
c) 0,3;
d) 0,95.

14. Если предпосылки метода наименьших квадратов нарушены, то...
a) полученное уравнение статистически незначимо;
b) оценки параметров могут не обладать свойствами эффективности, состоятельности, несмещенности;
c) коэффициент регрессии является несущественным;
d) коэффициент корреляции является несущественным.

15. Расчетное значение критерия Фишера определяется как ____________факторной дисперсии и остаточной, рассчитанных на одну степень свободы.
a) произведение;
b) сумма;
c) отношение;
d) разность.

16. Математическое ожидание остатков равно нулю, если оценки параметров обладают свойством...
a) несмещенности;
b) эффективности;
c) смещенности;
d) состоятельности.

17. Случайный характер остатков предполагает...
a) зависимость остатков от величины предсказанных по модели значений результативного признака;

b) независимость остатков от величины предсказанных по модели значений результативного признака;
c) зависимость предсказанных по модели значений результативного признака от значений факторного признака;
d) независимость предсказанных по модели значений результативного признака от значений факторного признака;

18. При расчете коэффициента детерминации используется отношение...
a) математических ожиданий;
b) дисперсий;
c) остаточных величин;
d) параметров уравнения регрессии.

19. Если доверительный интервал для параметра проходит через точку ноль, следовательно...
a) параметр признается статистически значимым;
b) параметр является существенным;
c) параметр является несущественным;
d) значение параметра может принимать как отрицательные, так и положительные значения.

20. Смысл расчета средней ошибки аппроксимации состоит в определении среднего арифметического значения отклонений ε, выраженных в процентах...
a) от предсказанных значений случайных факторов;
b) фактических значений результативного признака;
c) наблюдаемых значений независимой переменной;
d) теоретических значений результативного признака.

21. Несмещенность оценки на практике означает...
a) невозможность перехода от точечного оценивания к интервальному;
b) что при большом числе выборочных оцениваний остатки не будут накапливаться;
c) уменьшение точности с увеличением объема выборки;
d) что найденное значение коэффициента регрессии нельзя рассматривать как среднее значение из возможно большего количества несмещенных оценок;

22. Оценки параметров, найденные при помощи метода наименьших квадратов, обладают свойствами эффективности, состоятельности и несмещенности, если выполняются __________ метода наименьших квадратов.
a) альтернативные гипотезы;
b) допустимые значения;
c) нулевые гипотезы;
d) предпосылки.

23. Если расчетное значение критерия Фишера меньше табличного значения, то гипотеза о статистической незначимости уравнения...
a) отвергается;
b) несущественна;
c) принимается;
d) незначима.

24. Для уравнения $y = 3,14 + 2x + \varepsilon$ значение коэффициента регрессии составило 2. Следовательно...
a) теснота связи в два раза сильнее, чем для функциональной связи;
b) связь функциональная;
c) при увеличении фактора на единицу значение результата увеличивается в 2 раза;
d) значение коэффициента корреляции рассчитано с ошибкой.

25. Оценки параметров уравнений регрессии при помощи метода наименьших квадратов находятся на основании решения...
a) системы нормальных уравнений;
b) уравнения регрессии;
c) системы нормальных неравенств;
d) двойственной задачи.

26. Значение коэффициента детерминации составило 0,9, следовательно...
a) уравнением регрессии объяснено 10% дисперсии результативного признака;
b) уравнением регрессии объяснено 90% дисперсии результативного признака;
c) доля дисперсии результативного признака, объясненная регрессией, в общей дисперсии результативного признака составила 0,1;
d) доля дисперсии факторного признака, объясненная регрессией, в общей дисперсии факторного признака составила 0,9.

27. Свойствами оценок МНК являются...
a) эффективность, состоятельность и смещенность;
b) эффективность, несостоятельность и несмещенность;
c) эффективность, состоятельность и несмещенность;
d) эффективность, несостоятельность и смещенность.

28. Если оценка параметра эффективна, то это означает...
a) наименьшую дисперсию остатков;
b) уменьшение точности с увеличением объема выборки;
c) максимальную дисперсию остатков;
d) равенство нулю математического ожидания остатков.

29. Случайными воздействиями обусловлено 12% результативного признака, следовательно, значение коэффициента детерминации составило...
a) 0,88;
b) 12;
c) 0,12;
d) 88.

30. В исходном соотношении МНК сумма квадратов отклонений фактических значений результативного признака от его теоретических значений...
a) приравнивается к системе нормальных уравнений;
b) минимизируется;
c) максимизируется;
d) приравнивается к нулю.

31. Качество подбора уравнения оценивает коэффициент...
a) регрессии;
b) эластичности;
c) корреляции;
d) детерминации.

32. Критические значения критерия Фишера определяются...
a) по уровню значимости и степеням свободы факторной и остаточной дисперсии;
b) уровню значимости и степени свободы общей дисперсии;
c) уровню значимости;
d) степеням свободы факторной и остаточной дисперсии;

33. Состоятельность оценки характеризуется...
a) уменьшением ее точности с увеличением объема выборки;
b) увеличением ее точности с увеличением объема выборки;
c) независимостью от объема выборки значения математического ожидания остатков;
d) зависимостью от объема выборки значения математического ожидания остатков;

34. Величина параметра a в уравнении парной линейной регрессии $y = a + bx + \varepsilon$ характеризует значение...
a) факторной переменной при нулевом значении случайного фактора;
b) факторной переменной при нулевом значении результата;
c) результирующей переменной при нулевом значении фактора;
d) результирующей переменной при нулевом значении случайной величины.

35. Доверительный интервал характеризует...
a) интервал значений параметра, куда с заданной вероятностью попадает истинное значение параметра;
b) интервал значений фактора, куда с заданной вероятностью попадает истинное значение параметра;
c) интервал значений результата, куда с заданной вероятностью попадает истинное значение параметра;
d) интервал значений коэффициента корреляции, куда с заданной вероятностью попадает истинное значение параметра.

36. При использовании метода наименьших квадратов исследуются свойства...
a) оценок переменных и параметров уравнения регрессии;
b) оценок параметров уравнения регрессии;
c) оценок случайных величин уравнения регрессии;
d) оценок переменных уравнения регрессии.

37. Переход от точечного оценивания к интервальному возможен, если оценки являются...
a) эффективными и несмещенными;
b) состоятельными и смещенными;
c) неэффективными и состоятельными;
d) эффективными и несостоятельными.

38. Предпосылкой метода наименьших квадратов является...
a) отсутствие корреляции между результатом и фактором;
b) присутствие автокорреляции между результатом и фактором;
c) присутствие автокорреляции в остатках;
d) отсутствие автокорреляции в остатках.

39. Критерий Фишера используется для оценки значимости...
a) параметров;
b) коэффициента детерминации;
c) коэффициента регрессии;
d) построенного уравнения.

40. Стандартная ошибка рассчитывается для проверки существенности...
a) коэффициента корреляции;
b) параметра;
c) случайной величины;
d) коэффициента детерминации.

41. Увеличение точности оценок с увеличением объема выборки описывает свойство _________оценки.
a) состоятельности;
b) смещенности;
c) несмещенности;
d) эффективности.

42. Расчет значений коэффициента детерминации не позволяет оценить...
a) существенность коэффициента регрессии;
b) долю остаточной дисперсии результативного признака в общей дисперсии результативного признака;
c) долю факторной дисперсии результативного признака в общей дисперсии результативного признака;
d) качество подбора уравнения регрессии.

43. Табличное значение критерия Фишера служит...
a) для проверки статистической гипотезы о равенстве факторной и остаточной дисперсии;
b) проверки статистической гипотезы о равенстве двух математических ожиданий;
c) проверки статистической гипотезы о равенстве математического ожидания некоторой гипотетической величине;
d) проверки статистической гипотезы о равенстве дисперсии некоторой гипотетической величине.

44. В качестве показателя тесноты связи для линейного уравнения парной регрессии используется...
a) множественный коэффициент линейной корреляции;
b) линейный коэффициент корреляции;
c) линейный коэффициент регрессии;
d) линейный коэффициент детерминации.

45. Расчетное значение критерия Фишера определяется как отношение...
a) математических ожиданий;
b) дисперсий;
c) результата к фактору;
d) случайных величин.

46. Предпосылки метода наименьших квадратов исследуют поведение...
a) остаточных величин;
b) неслучайных величин;
c) переменных уравнения регрессии;
d) параметров уравнения регрессии.

47. Эффективность оценки на практике характеризуется...
a) невозможностью перехода от точечного оценивания к интервальному;
b) уменьшением точности с увеличением объема выборки;
c) возможностью перехода от точечного оценивания к интервальному;
d) отсутствием накапливания значений остатков при большом числе выборочных оцениваний.

48. В основе метода наименьших квадратов лежит...
a) минимизация суммы квадратов отклонений фактических значений результативного признака от его средних значений;
b) максимизация суммы квадратов отклонений фактических значений результативного признака от его теоретических значений;
c) равенство нулю суммы квадратов отклонений фактических значений результативного признака от его теоретических значений;
d) минимизация суммы квадратов отклонений фактических значений результативного признака от его теоретических значений.

49. Значение индекса детерминации рассчитывается как отношение дисперсии результативного признака, объясненной регрессией, к ___________ дисперсии результативного признака.
a) средней;
b) остаточной;
c) факторной;
d) общей.

Практическая работа 3.3
Нелинейные регрессионные модели

Цели: закрепление знаний о различных формах нелинейных моделей, освоение и применение метода линеаризации нелинейных уравнений регрессии, итерационных методов подбора нелинейных моделей, методов проверки качества таких моделей.

Примеры решения задач

Исследуется зависимость расходов на приобретение некоторого товара (группы товаров) семейными хозяйствами от располагаемого дохода.

В течение года i-я семья, имеющая располагаемый доход x_i, затратила на приобретение этого товара V_i руб. (табл. 3.8).

Таблица 3.8

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	150537,1	3736,022
2	136570,9	3155,929
3	151518,1	4091,394
4	110318,6	3037,814
5	155144,1	3603,569
6	129398,2	3025,638
7	118036	3041,839
8	153232,6	3907,761
9	174761,2	3961,124
10	158744,2	4072,685
11	151702,4	3991,685
12	143872,3	3692,751
13	166110,4	4227,418
14	164493	3783,255
15	114337,7	3174,847
16	136811,3	3265,973
17	135744,2	3359,623
18	120100,7	2737,437
19	169115,2	3801,232
20	156830,3	3828,464

1. Подберите модель зависимости, в которой эластичность потребления рассматриваемого товара по отношению к располагаемому доходу не зависит от размера располагаемого дохода. Постоянство эластичности предполагает оценивание модели, линейной в логарифмах уровней.

2. Постройте график подбора значений регрессии. Рассчитайте среднюю ошибку аппроксимации. Сделайте выводы.

3. Проверьте значимость подобранной модели на уровне $\alpha = 0{,}05$, используя коэффициент детерминации и критерий Фишера.

4. С помощью графического метода оцените соответствие используемых для построения модели статистических данных стандартным предположениям регрессионного анализа.

5. В рамках подобранной модели проверьте гипотезы о том, что:

1) потребление данного товара эластично по отношению к располагаемому доходу. Эластичное потребление соответствует значению эластичности, большему единицы по абсолютной величине ($|\eta| = |\beta| = 1$);
2) потребление данного товара неэластично по отношению к располагаемому доходу ($|\eta| = |\beta| < 1$).

Решение

1. Предположим, что модель зависимости затрат на приобретение некоторого товара V от располагаемого дохода семейного хозяйства $\boldsymbol{x}$ имеет постоянную эластичность потребления. Постоянство эластичности означает модель вида (2.67):

$$V = V(x) = \alpha \cdot x^{\beta}.$$

Для подтверждения нашего предположения построим поле корреляции (рис. 3.4).

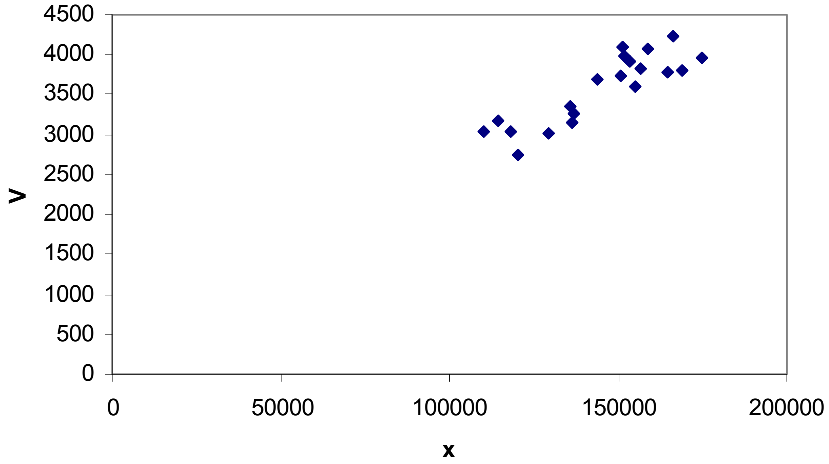

Рис. 3.4. Поле корреляции

По форме облака рассеяния видно, что наше предположение о степенной модели наблюдений (см. рис. 2.22) подтверждается. При оценивании коэффициентов такой модели метод наименьших квадратов применяют к логарифмам уровней. Степенная модель линеаризуется путем логарифмирования:

$$V = \alpha \cdot x^{\beta} \quad \to \quad \ln V = \ln \alpha x^{\beta};$$

$$\ln V = \ln \alpha + \ln x^{\beta};$$

$$\ln V = \ln \alpha + \beta \ln x.$$

Если обозначить $\ln V = V'$, $\ln \alpha = \alpha'$, $\ln x = x'$, то $V' = \alpha' + \beta \cdot x'$.

Рассчитаем логарифмы уровней и оценки наименьших квадратов для линейного уравнения парной регрессии $\hat{V}' = \alpha' + \beta \cdot x'$. Промежуточные вычисления будем заносить в табл. 3.9.

Таблица 3.9

i	x_i	V_i	$\ln x = x'$	$\ln V = V'$	$x' - \overline{x}'$	$(x' - \overline{x}')^2$	$x' \cdot V'$
1	150537,1	3736,022	11,922	8,226	0,047	0,0022	98,067
2	136570,9	3155,929	11,825	8,057	-0,05	0,0025	95,271
3	151518,1	4091,394	11,928	8,317	0,0535	0,0029	99,205
4	110318,6	3037,814	11,611	8,019	-0,264	0,0696	93,108
5	155144,1	3603,569	11,952	8,19	0,0771	0,0059	97,884
6	129398,2	3025,638	11,771	8,015	-0,104	0,0109	94,34
7	118036	3041,839	11,679	8,02	-0,196	0,0385	93,666
8	153232,6	3907,761	11,94	8,271	0,0647	0,0042	98,75
9	174761,2	3961,124	12,071	8,284	0,1962	0,0385	100
10	158744,2	4072,685	11,975	8,312	0,1	0,01	99,537
11	151702,4	3991,685	11,93	8,292	0,0547	0,003	98,921
12	143872,3	3692,751	11,877	8,214	0,0017	3E-06	97,557
13	166110,4	4227,418	12,02	8,349	0,1454	0,0211	100,36
14	164493	3783,255	12,011	8,238	0,1356	0,0184	98,948
15	114337,7	3174,847	11,647	8,063	-0,228	0,052	93,909
16	136811,3	3265,973	11,826	8,091	-0,049	0,0024	95,691
17	135744,2	3359,623	11,819	8,12	-0,056	0,0032	95,962
18	120100,7	2737,437	11,696	7,915	-0,179	0,032	92,572
19	169115,2	3801,232	12,038	8,243	0,1633	0,0267	99,233
20	156830,3	3828,464	11,963	8,25	0,0879	0,0077	98,697

i	x_i	V_i	$\ln x = x'$	$\ln V = V'$	$x' - \overline{x'}$	$(x' - \overline{x'})^2$	$x' \cdot V'$
Итого		71496,46	237,5	163,5		0,3518	1941,7
Среднее значение		3574,82	11,875	8,174		0,0176	97,084

Воспользуемся формулами (2.28) и (2.29):

$$b = \frac{\overline{xV} - \bar{x} \cdot \bar{V}}{\overline{x^2} - \bar{x}^2} = \frac{\operatorname{cov}(x, V)}{\operatorname{var}(x)} = \frac{97{,}084 - 11{,}875 \cdot 8{,}174}{0{,}0176} = 0{,}808;$$

$$a' = \bar{V} - b \cdot \bar{x} = 8{,}174 - 0{,}808 \cdot 11{,}875 = -1{,}415.$$

Получили уравнение: $\hat{V}' = -1{,}415 + 0{,}808 \cdot x'$, связывающее логарифмы уровней. Перейдем к исходной форме модели $V = \alpha \cdot x^{\beta}$. Для этого рассчитаем коэффициент $a = e^{\alpha'} = e^{-1{,}415} = 0{,}243$. Таким образом, подобранная модель с постоянной эластичностью $\eta = b = 0{,}808$ имеет вид: $\hat{V} = 0{,}243 \cdot x^{0{,}808}$.

2. Построим график подбора значений регрессии. Для этого вычислим значения $\hat{V}'$, подставляя в найденную модель наблюдаемые значения x. Эти и дальнейшие вычисления отразим в табл. 3.10.

Таблица 3.10

i	x_i	V_i	$\hat{V}'$	$V - \hat{V}'$	$(V - \hat{V}')$	A_i,%	$(V - \bar{V})^2$
1	150537,1	3736,022	3685,715	50,30694	2530,7886	1,347	25985,12
2	136570,9	3155,929	3407,013	-251,084	63043,247	7,956	175472,2
3	151518,1	4091,394	3705,099	386,2947	149223,56	9,442	266845,6
4	110318,6	3037,814	2867,513	170,3007	29002,319	5,606	288378,7
5	155144,1	3603,569	3776,54	-172,971	29918,989	4,8	826,3325
6	129398,2	3025,638	3261,766	-236,128	55756,234	7,804	301604,2
7	118036	3041,839	3028,447	13,39168	179,33707	0,44	284071,9
8	153232,6	3907,761	3738,92	168,8414	28507,433	4,321	110847,7
9	174761,2	3961,124	4157,697	-196,573	38640,771	4,963	149228,5
10	158744,2	4072,685	3847,153	225,5317	50864,562	5,538	247866,6
11	151702,4	3991,685	3708,738	282,9466	80058,797	7,088	173773,9
12	143872,3	3692,751	3553,367	139,3845	19428,035	3,775	13907,01
13	166110,4	4227,418	3990,686	236,7319	56042,005	5,6	425880,2
14	164493	3783,255	3959,277	-176,022	30983,852	4,653	43443,9
15	114337,7	3174,847	2951,586	223,2607	49845,327	7,032	159980,8
16	136811,3	3265,973	3411,855	-145,882	21281,685	4,467	95388,32
17	135744,2	3359,623	3390,349	-30,7257	944,06702	0,915	46311,04
18	120100,7	2737,437	3071,155	-333,718	111367,99	12,19	701215,3
19	169115,2	3801,232	4048,882	-247,65	61330,322	6,515	51261,04
20	156830,3	3828,464	3809,652	18,81159	353,87606	0,491	64333,76
Итого		71496,46			879303,2	104,9	3626622
Среднее значение		3574,82				5,25	

На одном графике изобразим поле корреляции и подобранную по модели кривую (рис. 3.5).

По графику видно, что подобранная модель хорошо аппроксимирует исходные данные.

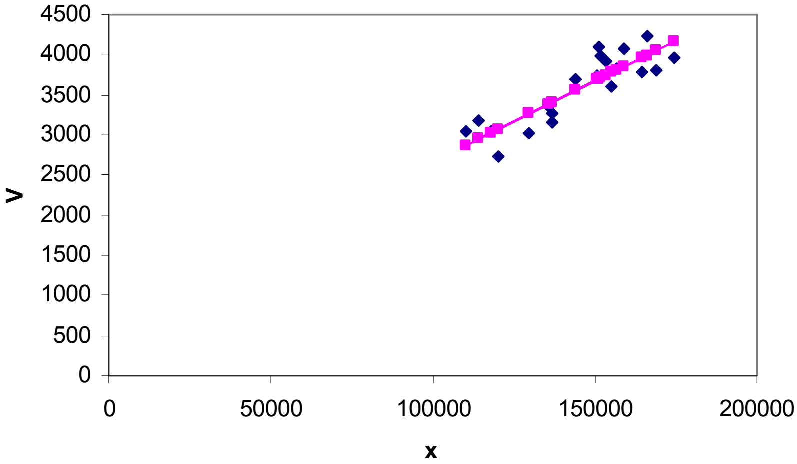

Рис. 3.5. График подбора

Среднюю ошибку аппроксимации находим по формуле (2.53) с помощью столбцов 5 и 7 табл. 3.10. $A_i = \left|\frac{V_i - \hat{V}_{x_i}}{V_i}\right| \times 100\%$, $\bar{A} = \frac{1}{n}\sum A_i = 5{,}25\%$ свидетельствует о хорошем подборе модели к исходным данным.

3. Проверим значимость подобранной модели на уровне $\alpha = 0{,}05$, используя коэффициент детерминации и критерий Фишера.

Одной из наиболее эффективных оценок значимости уравнения регрессии является коэффициент детерминации. Он характеризует степень выраженности связи между переменными. Определяется по формуле (2.56):

$$R^2 = \frac{Q_R}{Q} = 1 - \frac{Q_\ell}{Q} = 1 - \frac{879303{,}2}{3626622} = 0{,}758,$$

где $Q = \sum_{i=1}^{n}\left(V_i - \bar{V}\right)^2 = 3626622$ – полная сумма квадратов; $Q_R = \sum_{i=1}^{n}\left(\hat{V}_i - \bar{V}\right)^2$ – сумма квадратов, объясненная моделью; $Q_\ell = \sum_{i=1}^{n}\left(V_i - \hat{V}_i\right)^2 = \sum_{i=1}^{n} e_i^2 = 879303{,}2$ – остаточная сумма квадратов.

Коэффициент детерминации $R_2 = 0{,}785$ показывает, что уравнением регрессии объясняется 75,8% дисперсии результативного признака, а на долю прочих факторов приходится 24,2%. То есть 75,8% вариации расходов на приобретение некоторого товара (V) объясняется вариацией фактора x – дохода семейного хозяйства.

Оценим качество уравнения регрессии в целом с помощью F-критерия Фишера. Рассчитаем фактическое значение F-критерия. Так как коэффициент детерминации уже известен, проще всего использовать формулу (2.57):

$$F = \frac{R^2(n-m)}{(1-R^2)(m-1)} = \frac{0{,}758}{1-0{,}758} \cdot 18 = 56{,}24,$$

где m – число оцениваемых параметров уравнения регрессии; n – число наблюдений.

Табличное значение F-критерия найдем по прил. 1 ($k_1 = m - 1 = 2 = 1 = 1$, $k_2 = n - 2 = 20 - 2 = 18$): $F_{табл} = 4{,}41$. Так как $F_{факт} > F_{табл}$, то признается статистическая значимость уравнения в целом.

4. С помощью графического метода оценим соответствие используемых для построения модели статистических данных стандартным предположениям регрессионного анализа.

Подобранная модель проверяется на отсутствие автокорреляционной зависимости остатков от номера наблюдения, на независимость случайных ошибок ε_1, ε_2, ..., ε_n, математическое ожидание которых должно стремиться к нулю ($M\varepsilon_i = 0$), на постоянство или гомоскедастичность дисперсии ошибок [$\mathrm{var}(\varepsilon_i) = \delta^2(\varepsilon_i) = \mathrm{const}$]. Анализ соблюдения перечисленных условий проводят, используя графики стандартизированных остатков (формула 2.64).

$$C_i = \frac{V_i - \hat{V}_i}{S_{\varepsilon_i}} = \frac{e_i}{S_{\varepsilon_i}} (i = 1, 2, ..., n),$$

где $S^2_{\varepsilon_i}$ - оценка дисперсии остатков (формула 2.65).

$$S^2_{\varepsilon_i} = \frac{Q_e}{n-m} = \frac{\sum_{i=1}^{n}\left(V_i - \hat{V}_i\right)^2}{n-m} = \frac{879303{,}2}{48} = 18318{,}82.$$

Данные для построения графика зависимости стандартизированных остатков C_i (как ординат) от оцененных значений $\hat{V}'$ (по оси абсцисс) занесем в табл. 3.11.

Таблица 3.11

i	x_i	V_i	$\hat{V}'$	$V - \hat{V}'$	$(V - \hat{V}')$	C_i
1	150537,1	3736,022	3685,715	50,30694	2530,7886	0,372
2	136570,9	3155,929	3407,013	-251,084	63043,247	-1,86
3	151518,1	4091,394	3705,099	386,2947	149223,56	2,854
4	110318,6	3037,814	2867,513	170,3007	29002,319	1,258
5	155144,1	3603,569	3776,54	-172,971	29918,989	-1,28
6	129398,2	3025,638	3261,766	-236,128	55756,234	-1,74
7	118036	3041,839	3028,447	13,39168	179,33707	0,099
8	153232,6	3907,761	3738,92	168,8414	28507,433	1,247
9	174761,2	3961,124	4157,697	-196,573	38640,771	-1,45
10	158744,2	4072,685	3847,153	225,5317	50864,562	1,666
11	151702,4	3991,685	3708,738	282,9466	80058,797	2,091
12	143872,3	3692,751	3553,367	139,3845	19428,035	1,03
13	166110,4	4227,418	3990,686	236,7319	56042,005	1,749
14	164493	3783,255	3959,277	-176,022	30983,852	-1,3
15	114337,7	3174,847	2951,586	223,2607	49845,327	1,65
16	136811,3	3265,973	3411,855	-145,882	21281,685	-1,08
17	135744,2	3359,623	3390,349	-30,7257	944,06702	-0,23
18	120100,7	2737,437	3071,155	-333,718	111367,99	-2,47
19	169115,2	3801,232	4048,882	-247,65	61330,322	-1,83
20	156830,3	3828,464	3809,652	18,81159	353,87606	0,139
Итого		71496,46			879303,2	
Среднее значение		3574,82				
$S^2_{\varepsilon_i}$					18318,82	
S_{ε_i}					135,35	

Проанализируем полученный график (рис. 3.6).

На графике нет выделяющихся наблюдений, что могло бы указывать на отличие математического ожидания ошибок от нуля $M(\varepsilon_i) \neq 0$ либо на неоднородность дисперсии ошибок. Не наблюдается функциональной зависимости $\mathrm{var}(\varepsilon_i)$ от величины $\hat{V}'_i$, то есть дисперсия ошибок гомоскедастична. Судя по графику, условие $M(\varepsilon_i) = 0$ выполняется, то есть спецификация модели подобрана правильно.

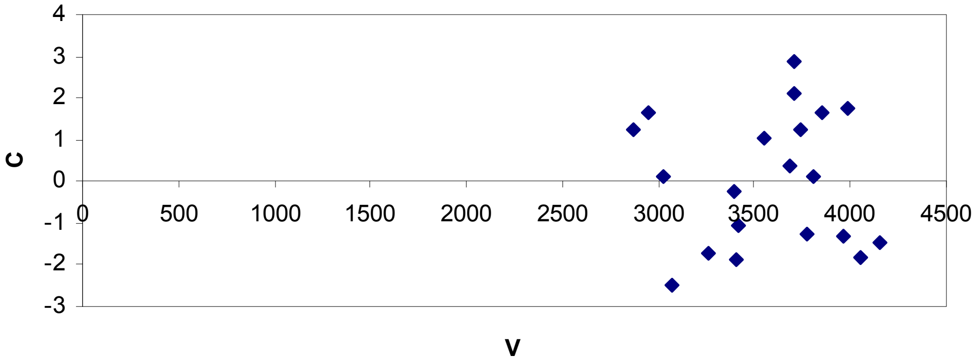

Рис. 3.6. График стандартизированных остатков

Таким образом, используемые для построения модели статистические данные соответствуют стандартным предположениям регрессионного анализа.

5. Проверим гипотезу *A* о том, что потребление данного товара эластично по отношению к располагаемому доходу. Эластичное потребление соответствует значению эластичности, большему единицы по абсолютной величине (A: $|\eta| = |\beta| > 1$).

Воспользуемся формулой (2.106):

$$\left|\frac{b-\beta_0}{S_b}\right| > t_{1-\alpha;n-2},$$

где S_b – среднеквадратическое (стандартное) отклонение параметра модели b. Для его расчета используем формулу (2.105) и табл. 3.12.

$$S_b = \frac{S_{\varepsilon_i}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \frac{135{,}35}{\sqrt{6868629912}} = \frac{135{,}35}{82877{,}2} = 0{,}00163.$$

Таблица 3.12

i	x_i	$x-\bar{x}$	$(x-\bar{x})^2$
1	150537,1	5668,175	32128207,83
2	136570,9	-8298,025	68857218,9
3	151518,1	6649,175	44211528,18
4	110318,6	-34550,325	1193724958
5	155144,1	10275,175	105579221,3
6	129398,2	-15470,725	239343332
7	118036	-26832,925	720005864,1
8	153232,6	8363,675	69951059,51
9	174761,2	29892,275	893548104,7
10	158744,2	13875,275	192523256,3
11	151702,4	6833,475	46696380,58
12	143872,3	-996,625	993261,3906
13	166110,4	21241,475	451200260,2
14	164493	19624,075	385104319,6
15	114337,7	-30531,225	932155700
16	136811,3	-8057,625	64925320,64
17	135744,2	-9124,725	83260606,33
18	120100,7	-24768,225	613464969,7
19	169115,2	24246,275	587881851,4
20	156830,3	11961,375	143074491,9
Итого	2897379		6868629912
Среднее значение	144868,9		

Значение критерия Стьюдента определим по прил. 2: $t_{1-0,05;\,18} = 2{,}1$.

$$\left|\frac{0{,}808-\beta_0}{0{,}001633}\right| > 2{,}1.$$

Нетрудно видеть, что при подстановке любого числа, большего 1 по абсолютной величине, данное неравенство выполняется. Так как наблюдаемое значение отношения больше табличного по абсолютной величине, такую гипотезу с вероятностью 0,95 следует отвергнуть. Это означает слишком большое отклонение оценки b от гипотетического значения β_0 параметра β в сравнении с оценкой S_b стандартного отклонения этого параметра.

Аналогично проверим гипотезу *Б*, состоящую в том, что потребление данного товара неэластично по отношению к располагаемому доходу (*В*: $|\eta| = |\beta| < 1$). Данная гипотеза будет выполняться только при условии подстановки в неравенство чисел из интервала $b-t_{1-\alpha;\,n-2}\cdot S_b \le \beta \le b+t_{1-\alpha;\,n-2}\cdot S_b$ или $0{,}8041 \le \beta \le 0{,}811$.

Таким образом, гипотеза *Б* (потребление товара неэластично по отношению к располагаемому доходу) с вероятностью 0,95 принимается для значений $0{,}8041 \le \beta \le 0{,}811$, и с этой же вероятностью отвергается для других значений эластичности, меньших 1.

Как видим, при значениях параметра, не принадлежащих 0,95-процентному доверительному интервалу, обе гипотезы отвергаются. Следовательно, проверить обе гипотезы, то есть осуществить прогноз по подобранной модели, можно только в рассчитанном интервале.

Задачи для самостоятельного решения

В течение года i-я семья, имеющая располагаемый доход x_i, затратила на приобретение этого товара V_i, руб. (табл. 3.13).

Таблица 3.13

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	144354,6	3932,599
2	149545,6	3886,132
3	170410,3	3352,526
4	153039,5	3295,72
5	159254,2	3900,406
6	149145,2	3777,833
7	160664,6	4602,151
8	116919,8	3422,525
9	144054,5	3700,39
10	168181,5	3845,074
11	139014,7	3869,305
12	144764	3561,304
13	158904,1	3595,007
14	136085,9	3385,927
15	126789,5	3124,223
16	142027,4	3055,907
17	148287,3	3590,945
18	151869,5	3719,061
19	170769,1	3972,955
20	112560,6	2699,635

1. Подберите модель зависимости, в которой эластичность потребления рассматриваемого товара по отношению к располагаемому доходу не зависит от размера располага-

емого дохода. Постоянство эластичности предполагает оценивание модели, линейной в логарифмах уровней.

2. Постройте график подбора значений регрессии. Рассчитайте среднюю ошибку аппроксимации. Сделайте выводы.

3. Проверьте значимость подобранной модели на уровне = 0,05, используя коэффициент детерминации и критерий Фишера.

4. С помощью графического метода оцените соответствие используемых для построения модели статистических данных стандартным предположениям регрессионного анализа.

5. В рамках подобранной модели проверьте гипотезы о том, что:

А) потребление данного товара эластично по отношению к располагаемому доходу. Эластичное потребление соответствует значению эластичности, большему единицы по абсолютной величине ($|\eta| = |\beta| > 1$);

Б) потребление данного товара неэластично по отношению к располагаемому доходу ($|\eta| = |\beta| < 1$).

Тесты

1. Требованием к уравнениям регрессии, параметры которых можно найти при помощи МНК, является...

a) равенство нулю средних значений факторного признака;
b) линейность параметров;
c) нелинейность параметров;
d) равенство нулю средних значений результативной переменной.

2. Для существенного параметра расчетное значение критерия Стьюдента по модулю...

a) больше табличного значения;
b) не меньше табличного значения;
c) меньше табличного значения;
d) не больше табличного значения.

3. При оценке статистической значимости уравнения регрессии и существенности связи осуществляется проверка...

a) существенности коэффициента корреляции;
b) существенности параметров;
c) существенности коэффициента детерминации;
d) нулевой гипотезы.

4. Статистические гипотезы используются для оценки...

a) значимости уравнения регрессии в целом;
b) тесноты связи между результатом и случайным фактором;
c) тесноты связи между результатом и фактором;
d) автокорреляции в остатках.

5. Совокупность значений критерия, при которых принимается нулевая гипотеза, называется областью __________ гипотезы.

a) принятия;
b) отрицания;

c) альтернативной;
d) нулевых значений.

6. Критическое значение критерия Стьюдента определяет...
a) максимально возможную величину, допускающую принятие гипотезы о существенности параметра;
b) минимально возможную величину, допускающую принятие гипотезы о несущественности параметра;
c) максимально возможную величину, допускающую принятие гипотезы о несущественности параметра;
d) минимально возможную величину, допускающую принятие гипотезы о равенстве нулю значения параметра.

7. Параметр является существенным, если...
a) доверительный интервал проходит через ноль;
b) доверительный интервал не проходит через ноль;
c) стандартная ошибка превышает половину самого параметра;
d) расчетное значение критерия Стьюдента равно табличному значению.

8. Критерий Стьюдента предназначен для определения значимости...
a) построенного уравнения в целом;
b) каждого коэффициента корреляции;
c) каждого коэффициента регрессии;
d) уравнения.

9. Спецификация модели «нелинейная парная регрессия» (простая) подразумевает нелинейную зависимость и...
a) одну независимую переменную;
b) две зависимых переменных;
c) пару независимых переменных;
d) пару существенных переменных.

10. Значение индекса детерминации, рассчитанное для нелинейного уравнения регрессии, характеризует...
a) долю дисперсии результативного признака, объясненную нелинейной регрессией в общей дисперсии результативного признака;
b) долю дисперсии результативного признака, объясненную линейной корреляцией в общей дисперсии результативного признака;
c) долю дисперсии результативного признака, объясненную линейной регрессией в общей дисперсии результативного признака;
d) долю дисперсии результативного признака, необъясненную нелинейной корреляцией в общей дисперсии результативного признака.

11. Основной целью линеаризации уравнения регрессии является...
a) повышение существенности связи между рассматриваемыми признаками;
b) улучшение качества модели;
c) возможность применения метода наименьших квадратов для оценки параметров;
d) получение новых нелинейных моделей.

12. При выборе спецификации нелинейная регрессия используется, если...

a) нелинейная зависимость для исследуемых экономических показателей является несущественной;
b) между экономическими показателями не обнаруживается нелинейная зависимость;
c) между экономическими показателями обнаруживается нелинейная зависимость;
d) между экономическими показателями обнаруживается линейная зависимость.

13. Нелинейным называется уравнение регрессии, если...

a) параметры и случайные факторы входят в уравнение нелинейным образом, а переменные линейны;
b) случайные факторы нелинейны;
c) независимые переменные входят в уравнение нелинейным образом;
d) параметры входят в уравнение нелинейным образом, а переменные линейны.

14. Для нелинейных уравнений метод наименьших квадратов применяется...

a) к нелинейным уравнениям;
b) преобразованным линеаризованным уравнениям;
c) непреобразованным линейным уравнениям;
d) обратным уравнениям.

15. Назовите показатель тесноты связи для нелинейных моделей регрессии.

a) линейный коэффициент корреляции;
b) индекс корреляции;
c) парный коэффициент линейной корреляции;
d) индекс детерминации.

16. Нелинейным не является уравнение...

a) $y = a + bx + cx^2 + \varepsilon$;

b) $y = a + \frac{b}{x} + \varepsilon$;

c) $y = a + bx_1 + cx_2 + \varepsilon$;

d) $y = \frac{1}{a + bx} + \varepsilon$.

17. Известно, что с увеличением объема производства себестоимость единицы продукции уменьшается за счет того, что происходит перераспределение постоянных издержек. Пусть *a* – совокупная величина постоянных издержек, а *b* – величина переменных издержек в расчете на одно изделие. Тогда зависимость себестоимости единицы продукции от объема производства можно описать с помощью модели...

a) $y = \frac{a}{x} + b + \varepsilon$;

b) $y = \frac{a}{b} x + \varepsilon$;

c) $y = \frac{a}{x} + \frac{b}{x} + \varepsilon$;

d) $y = \frac{b}{x} + ax + \varepsilon$.

18. Нелинейную модель зависимостей экономических показателей нельзя привести к линейному виду, если...
a) линейная модель является внутренне нелинейной;
b) нелинейная модель является внутренне линейной;
c) линейная модель является внутренне линейной;
d) нелинейная модель является внутренне нелинейной.

19. Замена $x_1 = x$, $x_2 = x^2$ подходит для уравнения...
a) $y = a + \frac{b}{x} + \frac{c}{x^2} + \varepsilon$;
b) $y = a + bx + cx^2 + dx^3 + \varepsilon$;
c) $y = a + bx + cx^2 + \varepsilon$;
d) $y = \frac{1}{a + bx + cx^2} + \varepsilon$.

20. Для моделирования зависимости предложения от цены **не может быть** использовано уравнение регрессии...
a) $y = a + bx + \varepsilon$;
b) $y = a + \frac{b}{x} + \varepsilon$;
c) $y = a + bx^2 + \varepsilon$;
d) $y = a + x^b + \varepsilon$.

21. Если спецификация модели отображает нелинейную форму зависимости между экономическими показателями, то нелинейно уравнение...
a) корреляции;
b) регрессии;
c) статистической зависимости;
d) системы.

22. Если спецификация модели $y = f(x) + \varepsilon$ нелинейное уравнение регрессии, то нелинейной является функция...
a) $f(x)$;
b) $f(y)$;
c) $f(\varepsilon)$;
d) $f(x, \varepsilon)$.

23. Оценить статистическую значимость нелинейного уравнения регрессии можно с помощью...
a) индекса детерминации;
b) средней ошибки аппроксимации;
c) индекса корреляции;
d) критерия Фишера.

24. Если между экономическими показателями существует нелинейная связь, то...
a) целесообразно использовать спецификацию нелинейного уравнения регрессии;
b) целесообразно использовать нелинейное уравнение парной регрессии;

c) необходимо включить в модель другие факторы и использовать линейное уравнение множественной регрессии;
d) нецелесообразно использовать спецификацию нелинейного уравнения регрессии.

25. Парабола второй степени может быть использована для зависимостей экономических показателей, если...
a) характер связи зависит от случайных факторов;
b) исходные данные не обнаруживают изменения направления связи;
c) для определенного интервала значений фактора меняется характер связи рассматриваемых показателей: прямая связь меняется на обратную или обратная на прямую;
d) для определенного интервала значений фактора меняется скорость изменений значений результата, то есть возрастает динамика роста или спада.

26. Значение индекса корреляции, рассчитанное для нелинейного уравнения регрессии, характеризует...
a) тесноту случайной связи;
b) тесноту линейной связи;
c) тесноту нелинейной связи;
d) тесноту обратной связи.

27. Экспоненциальным не является уравнение регрессии...
a) $y = e^{a+bx} + \varepsilon$;
b) $y = e + bx + \varepsilon$;
c) $y = e^{x} + \varepsilon$;
d) $y = e^{x} \cdot \varepsilon$.

28. При помощи модели степенного уравнения регрессии вида $y = a \cdot x_1^{\;b} \cdot \varepsilon$ ($b > 1$, то есть x возрастает и y тоже возрастает) **не может быть описана** зависимость...
a) заработной платы от выработки;
b) выработки от трудоемкости;
c) выработки от уровня квалификации;
d) объема предложения от цены.

29. Линеаризация **не подразумевает** процедуру...
a) включения в модель дополнительных существенных факторов;
b) преобразования уравнений;
c) приведения нелинейного уравнения к линейному;
d) замены переменных.

30. Относительно формы зависимости различают...
a) линейную и множественную регрессию;
b) линейную и нелинейную регрессию;
c) положительную и отрицательную регрессию;
d) простую и множественную регрессию.

31. Расчет средней ошибки аппроксимации для нелинейного уравнения регрессии связан с расчетом разности между...
a) прогнозным и теоретическим значениями результативной переменной;
b) прогнозным и теоретическим значениями независимой переменной;

c) фактическим и теоретическим значениями независимой переменной;
d) фактическим и теоретическим значениями результативной переменной.

32. Линеаризация подразумевает процедуру...
b) приведения нелинейного уравнения относительно параметров к уравнению линейному относительно результата;
c) приведения уравнения множественной регрессии к парной;
d) приведения линейного уравнения к нелинейному виду;
e) приведения нелинейного уравнения к линейному виду.

33. Назовите показатель корреляции для нелинейных моделей регрессии.
a) парный коэффициент линейной корреляции;
a) линейный коэффициент корреляции;
b) индекс корреляции;
c) индекс детерминации.

34. Нелинейное уравнение регрессии означает нелинейную форму зависимости между...
a) результатом и факторами;
b) результатом и параметрами;
c) фактором и результатами;
d) фактором и случайной величиной.

35. По результатам исследования было выявлено, что рентабельность производства падает с увеличением трудоемкости. Какую спецификацию уравнения регрессии можно использовать для построения модели такой зависимости?
a) $y = e^{a+bx} + \varepsilon$;
b) $y = e + bx + cx^2 + \varepsilon$;
c) $y = \dfrac{1}{a + b\sqrt{x}} + \varepsilon$;
d) $y = a + b\sqrt{x} + \varepsilon$.

36. Результатом линеаризации полиномиальных уравнений являются…
a) нелинейные уравнения множественной регрессии;
b) линейные уравнения множественной регрессии;
c) линейные уравнения парной регрессии;
d) нелинейные уравнения парной регрессии.

37. Уравнение регрессии $y = a + \dfrac{b}{x} + \varepsilon$ характеризует __________ зависимость.
a) прямо пропорциональную;
b) обратно пропорциональную;
c) функциональную;
d) линейную.

38. Спецификацию нелинейного уравнения парной регрессии целесообразно использовать, если...
a) значение линейного коэффициента корреляции для исследуемой зависимости ближе к 1;
b) доля остаточной дисперсии результативного признака в его общей дисперсии стремится к 1;

c) значение индекса корреляции для исследуемой зависимости ближе к 0;
d) значение индекса детерминации, рассчитанное для данной модели, достаточно близко к 1.

39. Метод наименьших квадратов позволяет оценить ________уравнений регрессии.
a) параметры;
b) переменные;
c) случайные факторы;
d) факторы.

40. В нелинейной модели парной регрессии уравнение, связывающее зависимую и независимую переменную.
a) линейное;
b) частное;
c) нелинейное;
d) стандартизированное.

41. Нелинейное уравнение регрессии, которое не может быть линеаризовано, имеет вид...
a) $y = a + bx + \varepsilon$;
b) $y = ax^b \cdot \varepsilon$;
c) $y = a + bx^c + \varepsilon$;
d) $y = a + b\sqrt{x} + \varepsilon$.

42. Нелинейным является уравнение...
a) $y = a + bx + cx^2 + \varepsilon$;
b) $y = a + bx + \varepsilon$;
c) $y = a + bx_1 + cx_2 + \varepsilon$;
d) $y = ax + \varepsilon$.

43. При выборе спецификации модели парная регрессия используется в случае, если...
a) среди множества факторов, влияющих на результат, нельзя выделить доминирующий фактор;
b) среди множества факторов, влияющих на результат, можно выделить несколько факторов;
c) среди множества факторов, влияющих на результат, можно выделить лишь случайные факторы;
d) среди множества факторов, влияющих на результат, можно выделить доминирующий фактор.

Практическая работа 3.4
Множественная линейная регрессия и корреляция

Цели: овладение навыками ранжирования факторов для включения в модель множественной регрессии; применение метода наименьших квадратов (МНК) расчета параметров уравнения множественной линейной регрессии; использование интервальных оценок для прогнозирования по уравнению регрессии; освоение методов проверки качества полученной модели.

Примеры решения задач

По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%).

Таблица 3.14

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,9	10	11	9	6,0	21
2	7	3,9	14	12	11	6,4	22
3	7	3,7	15	13	9	6,8	22
4	7	4,0	16	14	11	7,2	25
5	7	3,8	17	15	12	8,0	28
6	7	4,8	19	16	12	8,2	29
7	8	5,4	19	17	12	8,1	30
8	8	4,4	20	18	12	8,5	31
9	8	5,3	20	19	14	9,6	32
10	10	6,8	20	20	14	9,0	36

1. Постройте линейную модель множественной регрессии. Запишите стандартизованное уравнение множественной регрессии. На основе стандартизованных коэффициентов регрессии и средних коэффициентов эластичности ранжируйте факторы по степени их влияния на результат.

2. Найдите коэффициенты парной, частной и множественной корреляции. Проанализируйте их.

3. Найдите скорректированный коэффициент множественной детерминации. Сравните его с нескорректированным (общим) коэффициентом детерминации.

4. С помощью F-критерия Фишера оцените статистическую надежность уравнения регрессии и коэффициента детерминации $R^2_{yx_1x_2}$.

5. С помощью частных F-критериев Фишера оцените целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1.

6. Составьте уравнение линейной парной регрессии, оставив лишь один значащий фактор.

Решение

Для удобства проведения расчетов поместим результаты промежуточных расчетов в табл. 3.15.

Таблица 3.15

№	y	x_1	x_2	yx_1	yx_2	x_1x_2	x_1^2	x_2^2	y^2
1	7,0	3,9	10,0	27,3	70,0	39,0	15,21	100,0	49,0
2	7,0	3,9	14,0	27,3	98,0	54,6	15,21	196,0	49,0
3	7,0	3,7	15,0	25,9	105,0	55,5	13,69	225,0	49,0
4	7,0	4,0	16,0	28,0	112,0	64,0	16,0	256,0	49,0
5	7,0	3,8	17,0	26,6	119,0	64,6	14,44	289,0	49,0
6	7,0	4,8	19,0	33,6	133,0	91,2	23,04	361,0	49,0

№	y	x_1	x_2	yx_1	yx_2	x_1x_2	x_1^2	x_2^2	y^2
7	8,0	5,4	19,0	43,2	152,0	102,6	29,16	361,0	64,0
8	8,0	4,4	20,0	35,2	160,0	88,0	19,36	400,0	64,0
9	8,0	5,3	20,0	42,4	160,0	106,0	28,09	400,0	64,0
10	10,0	6,8	20,0	68,0	200,0	136,0	46,24	400,0	100,0
11	9,0	6,0	21,0	54,0	189,0	126,0	36,0	441,0	81,0
12	11,0	6,4	22,0	70,4	242,0	140,8	40,96	484,0	121,0
13	9,0	6,8	22,0	61,2	198,0	149,6	46,24	484,0	81,0
14	11,0	7,2	25,0	79,2	275,0	180,0	51,84	625,0	121,0
15	12,0	8,0	28,0	96,0	336,0	224,0	64,0	784,0	144,0
16	12,0	8,2	29,0	98,4	348,0	237,8	67,24	841,0	144,0
17	12,0	8,1	30,0	97,2	360,0	243,0	65,61	900,0	144,0
18	12,0	8,5	31,0	102,0	372,0	263,5	72,25	961,0	144,0
19	14,0	9,6	32,0	134,4	448,0	307,2	92,16	1024,0	196,0
20	14,0	9,0	36,0	126,0	504,0	324,0	81,0	1296,0	196,0
Сумма	192	123,8	446	1276,3	4581	2997,4	837,74	10828,0	1958,0
Среднее значение	9,6	6,19	22,3	63,815	229,05	149,87	41,887	541,4	97,9

1. Найдем средние квадратические отклонения признаков:

$$\sigma_y = \sqrt{\overline{y^2} - \overline{y}^2} = \sqrt{97{,}9 - 9{,}6^2} = 2{,}396;$$

$$\sigma_{x_1} = \sqrt{\overline{x_1^2} - \overline{x}_1^2} = \sqrt{41{,}887 - 6{,}19^2} = 1{,}890;$$

$$\sigma_{x_2} = \sqrt{\overline{x_2^2} - \overline{x}_2^2} = \sqrt{541{,}4 - 22{,}3^2} = 6{,}642.$$

Вычисление параметров линейного уравнения множественной регрессии.

Для нахождения параметров линейного уравнения множественной регрессии $y = a + b_1x_1 + b_2x_2$ необходимо решить следующую систему линейных уравнений относительно неизвестных параметров a, b_1, b_2:

$$\begin{cases} na + b_1\sum x_1 + b_2\sum x_2 = \sum y; \\ a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 = \sum yx_1; \\ a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2 = \sum yx_2 \end{cases}$$

либо воспользоваться готовыми формулами:

$$b_1 = \frac{\sigma_y}{\sigma_{x_1}} \cdot \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}; \quad b_2 = \frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{r_{yx_2} - r_{yx_1}r_{x_1x_2}}{1 - r_{x_1x_2}^2}; \quad a = \overline{y} - b_1\overline{x}_1 - b_2\overline{x}_2.$$

Рассчитаем сначала парные коэффициенты корреляции:

$$r_{yx_1} = \frac{\operatorname{cov}(y,\, x_1)}{\sigma_y \cdot \sigma_{x_1}} = \frac{63{,}815 - 6{,}19 \cdot 9{,}6}{1{,}890 \cdot 2{,}396} = 0{,}970;$$

$$r_{yx_2} = \frac{\operatorname{cov}(y,\, x_2)}{\sigma_y \cdot \sigma_{x_2}} = \frac{229{,}05 - 22{,}3 \cdot 9{,}6}{6{,}642 \cdot 2{,}396} = 0{,}941;$$

$$r_{x_1x_2} = \frac{\operatorname{cov}(x_1,\, x_2)}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{149{,}87 - 6{,}19 \cdot 22{,}3}{1{,}890 \cdot 6{,}642} = 0{,}943.$$

Находим

$$b_1 = \frac{2{,}396}{1{,}890} \cdot \frac{0{,}970 - 0{,}941 \cdot 0{,}943}{1 - 0{,}943^2} = 0{,}946; \quad b_2 = \frac{2{,}396}{6{,}642} \cdot \frac{0{,}941 - 0{,}970 \cdot 0{,}943}{1 - 0{,}943^2} = 0{,}0856;$$

$$a = 9,6 - 0,946 \cdot 6,19 - 0,0856 \cdot 22,3 = 1,835.$$

Таким образом, получено следующее уравнение множественной регрессии:

$$\hat{y} = 1,835 + 0,946 \cdot x_1 + 0,0856 \cdot x_2.$$

Коэффициенты β_1 и β_2 стандартизованного уравнения регрессии $t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \varepsilon$, находятся по формулам:

$$\beta_1 = b_1 \frac{\sigma_{x_1}}{\sigma_y} = 0,946 \cdot \frac{1,890}{2,396} = 0,746; \quad \beta_2 = b_2 \frac{\sigma_{x_2}}{\sigma_y} = 0,0856 \cdot \frac{6,642}{2,396} = 0,237.$$

Стандартизированное уравнение будет выглядеть следующим образом:

$$\hat{t}_y = 0,746 \cdot t_{x_1} + 0,237 \cdot t_{x_2}.$$

Так как стандартизованные коэффициенты регрессии можно сравнивать между собой, то можно сказать, что ввод в действие новых основных фондов оказывает большее влияние на выработку продукции, чем удельный вес рабочих высокой квалификации.

Сравнивать влияние факторов на результат можно также при помощи средних коэффициентов эластичности:

$$\bar{Э}_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}_{x_i}}.$$

Вычисляем:

$$\bar{Э}_1 = 0,946 \cdot \frac{6,19}{9,6} = 0,61; \quad \bar{Э}_2 = 0,0856 \cdot \frac{22,3}{9,6} = 0,20.$$

Следовательно, увеличение только основных фондов (от своего среднего значения) или только удельного веса рабочих высокой квалификации на 1% увеличивает в среднем выработку продукции на 0,61 или 0,20% соответственно. Таким образом, подтверждается большее влияние на результат y фактора x_1, чем фактора x_2.

2. Коэффициенты парной корреляции мы уже нашли:

$$r_{yx_1} = 0,970; \quad r_{yx_2} = 0,941; \quad r_{x_1x_2} = 0,943.$$

Они указывают на весьма сильную связь каждого фактора с результатом, а также высокую межфакторную зависимость (факторы x_1 и x_2 явно коллинеарны, т. к. $r_{x_1x_2} = 0,943 > 0,7$). При такой сильной межфакторной зависимости рекомендуется один из факторов исключить из рассмотрения.

Частные коэффициенты корреляции характеризуют тесноту связи между результатом и соответствующим фактором при элиминировании (устранении влияния) других факторов, включенных в уравнение регрессии.

При двух факторах частные коэффициенты корреляции рассчитываются следующим образом:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r_{yx_2}^2\right) \cdot \left(1 - r_{x_1x_2}^2\right)}} = \frac{0,970 - 0,941 \cdot 0,943}{\sqrt{\left(1 - 0,941^2\right) \cdot \left(1 - 0,943^2\right)}} = 0,734;$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r_{yx_1}^2\right) \cdot \left(1 - r_{x_1x_2}^2\right)}} = \frac{0,941 - 0,970 \cdot 0,943}{\sqrt{\left(1 - 0,970^2\right) \cdot \left(1 - 0,943^2\right)}} = 0,325.$$

Если сравнить коэффициенты парной и частной корреляции, то можно увидеть, что из-за высокой межфакторной зависимости коэффициенты парной корреляции дают завышенные оценки тесноты связи. Именно по этой причине рекомендуется при наличии сильной коллинеарности (взаимосвязи) факторов исключать из исследования тот фактор, у которого теснота парной зависимости меньше, чем теснота межфакторной связи.

Коэффициент множественной корреляции определить через матрицу парных коэффициентов корреляции:

$$R_{yx_1x_2} = \sqrt{1 - \frac{\Delta_r}{\Delta_{r_{11}}}},$$

где

$$\Delta_r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1x_2} \\ r_{yx_2} & r_{x_2x_1} & 1 \end{vmatrix}$$

– определитель матрицы парных коэффициентов корреляции;

$$\Delta_{r_{11}} = \begin{vmatrix} 1 & r_{x_1x_2} \\ r_{x_2x_1} & 1 \end{vmatrix}$$

– определитель матрицы межфакторной корреляции.

$$\Delta_r = \begin{vmatrix} 1 & 0,970 & 0,941 \\ 0,970 & 1 & 0,943 \\ 0,941 & 0,943 & 1 \end{vmatrix} = 1 + 0,8607 + 0,8607 -$$

$$-0,8855 - 0,8892 - 0,9409 = 0,0058$$

$$\Delta_{r_{11}} = \begin{vmatrix} 1 & 0,943 \\ 0,943 & 1 \end{vmatrix} = 1 - 0,8892 = 0,1108.$$

Коэффициент множественной корреляции

$$R_{yx_1x_2} = \sqrt{1 - \frac{0,0058}{0,1108}} = 0,973.$$

Аналогичный результат получим при использовании других формул:

$$R_{yx_1x_2} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}} = \sqrt{1 - \frac{0,305}{5,74}} = 0,973;$$

$$R_{yx_1x_2} = \sqrt{\sum \beta_i \cdot r_{yx_i}} = \sqrt{0,746 \cdot 0,970 + 0,237 \cdot 0,941} = 0,973;$$

$$R_{yx_1x_2 \ldots x_m} = \sqrt{1 - \left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{yx_2 \cdot x_1}\right)} =$$

$$= \sqrt{1 - \left(1 - 0,970^2\right) \cdot \left(1 - 0,325^2\right)} = 0,973.$$

Коэффициент множественной корреляции показывает на весьма сильную связь всего набора факторов с результатом.

3. Нескорректированный коэффициент множественной детерминации $R^2_{yx_1x_2} = 0,947$ оценивает долю вариации результата за счет представленных в уравнении факторов в общей вариации результата. Здесь эта доля составляет 94,7% и указывает на весьма высокую степень обусловленности вариации результата вариацией факторов, иными словами, на весьма тесную связь факторов с результатом.

Скорректированный коэффициент множественной детерминации

$$\bar{R}^2 = 1 - \left(1 - R^2\right) \frac{(n-1)}{(n-m-1)} = 1 - (1 - 0,947) \frac{20-1}{20-2-1} = 0,941$$

определяет тесноту связи с учетом степеней свободы общей и остаточной дисперсий. Он дает такую оценку тесноты связи, которая не зависит от числа факторов и поэтому может сравниваться по разным моделям с разным числом факторов. Оба коэффициента указывают на весьма высокую (более 94%) детерминированность результата ***y*** в модели факторами x_1 и x_2.

4. Оценку надежности уравнения регрессии в целом и показателя тесноты связи $R_{yx_1x_2}$ дает F-критерий Фишера:

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

В нашем случае фактическое значение F-критерия Фишера:

$$F_{\text{факт}} = \frac{0{,}973^2}{1 - 0{,}973^2} \cdot \frac{20 - 2 - 1}{2} = 151{,}88.$$

Получается, что $F_{факт} > F_{табл} = 3{,}49$ (при $n = 20$), т. е. вероятность случайно получить такое значение F-критерия не превышает допустимый уровень значимости 0,05. Следовательно, полученное значение не случайно, оно сформировалось под влиянием существенных факторов. Подтверждается статистическая значимость всего уравнения и показателя тесноты связи $R^2_{yx_1x_2}$.

5. С помощью частных F-критериев Фишера оценим целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1 при помощи формул:

$$F_{\text{част},\, x_1} = \frac{R^2_{yx_1x_2} - R^2_{yx_2}}{1 - R^2_{yx_1}} \cdot \frac{n - m - 1}{m}; \quad F_{\text{част},\, x_2} = \frac{R^2_{yx_1x_2} - R^2_{yx_1}}{1 - R^2_{yx_2}} \cdot \frac{n - m - 1}{m}.$$

Найдем $R^2_{yx_1}$ и $R^2_{yx_2}$.

$$R^2_{yx_1} = r^2_{yx_1} = 0{,}970^2 = 0{,}941; \quad R^2_{yx_2} = r^2_{yx_2} = 0{,}941^2 = 0{,}885.$$

Имеем

$$F_{\text{част},\, x_1} = \frac{0{,}947 - 0{,}885}{1 - 0{,}941} \cdot \frac{20 - 2 - 1}{2} = 8{,}9322; \quad F_{\text{част},\, x_2} = \frac{0{,}947 - 0{,}941}{1 - 0{,}885} \cdot \frac{20 - 2 - 1}{2} = 0{,}4435.$$

Получается, что $F_{\text{част},\, x_2} < F_{\text{табл}} = 3{,}49$. Следовательно, включение в модель фактора x_2, после того как в модель включен фактор x_1, статистически нецелесообразно: прирост факторной дисперсии за счет дополнительного признака x_2 оказывается незначительным, несущественным; фактор x_2 включать в уравнение после фактора x_1 не следует.

Если поменять первоначальный порядок включения факторов в модель и рассмотреть вариант включения x_1 после x_2, то результат расчета частного F-критерия для x_1 будет иным. $F_{\text{част},\, x_1} > F_{\text{табл}} = 3{,}49$, т. е. вероятность его случайного формирования меньше принятого стандарта $\alpha = 0{,}05$ (5%). Следовательно, значение частного F-критерия для дополнительно включенного фактора x_{21} не случайно, является статистически значимым, надежным, достоверным. Прирост факторной дисперсии за счет дополнительного фактора x_1 является существенным. Фактор x_1 должен присутствовать в уравнении, в том числе в варианте, когда он дополнительно включается после фактора x_2.

6. Общий вывод состоит в том, что множественная модель с факторами x_1 и x_2 с $R^2_{yx_1x_2} = 0{,}947$ содержит неинформативный фактор x_2. Если исключить фактор x_2,то можно ограничиться уравнением парной регрессии:

Задачи для самостоятельного решения

По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%).

Таблица 3.16

Номер предприятия	x_2	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,6	12	11	10	7,3	23
2	7	4,1	14	12	11	7,6	25
3	8	4,3	16	13	12	7,8	26
4	7	4,4	17	14	11	7,9	28
5	7	4,5	18	15	12	8,2	30
6	8	4,8	19	16	12	8,4	31
7	8	5,3	20	17	12	8,6	32
8	8	5,6	20	18	13	8,8	32
9	9	6,7	21	19	13	9,2	33
10	10	6,9	22	20	14	9,6	34

1. Постройте линейную модель множественной регрессии. Запишите стандартизованное уравнение множественной регрессии. На основе стандартизованных коэффициентов регрессии и средних коэффициентов эластичности ранжируйте факторы по степени их влияния на результат.

2. Найдите коэффициенты парной, частной и множественной корреляции. Проанализируйте их.

3. Найдите скорректированный коэффициент множественной детерминации. Сравните его с нескорректированным (общим) коэффициентом детерминации.

4. С помощью F-критерия Фишера оцените статистическую надежность уравнения регрессии и коэффициента детерминации $R^2_{yx_1x_2}$

5. С помощью частных F-критериев Фишера оцените целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1. Составьте уравнение линейной парной регрессии, оставив лишь один значащий фактор.

Тесты

1. Величина остаточной дисперсии при включении существенного фактора в модель...

a) не изменится;
b) будет увеличиваться;
c) будет уменьшаться;
d) будет равна единице.

2. При включении фиктивных переменных в модель им присваиваются...

a) числовые метки;
b) одинаковые значения;
c) качественные метки;
d) нулевые значения.

3. Показатель, характеризующий на сколько сигм изменится в среднем результат при изменении соответствующего фактора на одну сигму и неизменном уровне других факторов, называется ____________ коэффициентом регрессии.

a) центрированным;
b) нормализованным;
c) выравненным;
d) стандартизированным.

4. Систему МНК, построенную для оценки параметров линейного уравнения множественной регрессии, можно решить...

a) методом скользящей средней;
b) методом первых разностей;
c) методом определителей;
d) симплекс-методом.

5. Основным требованием к факторам включаемой в модель множественной регрессии является...

a) наличие тесной взаимосвязи между факторами;
b) отсутствие взаимосвязи между факторами;
c) наличие линейной взаимосвязи между факторами;
d) отсутствие взаимосвязи между результатом и фактором.

6. Исследуется зависимость, которая характеризуется линейным уравнением множественной корреляции. Для уравнения рассчитано значение тесноты связи результативной переменной с набором факторов. В качестве этого показателя был использован множественный коэффициент...

a) эластичности;
b) регрессии;
c) корреляции;
d) детерминации.

7. Объем выборки должен превышать число рассчитываемых параметров при исследуемых факторах...

a) в 2–3 раза;
b) в 10–12 раз;
c) в 20–25 раз;
d) в 5–6 раз.

8. Фиктивные переменные включаются в уравнение множественной регрессии для учета действия на результат признаков...

a) несущественного характера;
b) случайного характера;
c) качественного характера;
d) количественного характера.

9. Факторные переменные уравнения множественной регрессии, преобразованные из качественных в количественные, называются...

a) парными;
b) фиктивными;
c) множественными;
d) аномальными.

10. Линейное уравнение множественной регрессии имеет вид $y = a + 2{,}5x_1 - 3{,}7x_2 + \varepsilon$. Определите, какой из факторов x_1 или x_2 оказывает более сильное влияние на y.

a) x_2, так как $2{,}5 < 3{,}7$;
b) оказывают одинаковое влияние;

c) x_1, так как 2,5 > -3,7;
d) по этому уравнению нельзя ответить на поставленный вопрос, так как коэффициенты регрессии несравнимы между собой.

11. Обобщенный метод наименьших квадратов отличается от обычного МНК тем, что при использовании ОМНК...
a) уменьшается количество наблюдений;
b) остатки приравниваются к нулю;
c) преобразуются исходные уровни переменных;
d) остатки не изменяются.

12. Величина коэффициента детерминации при включении существенного фактора в эконометрическую модель...
a) будет уменьшаться;
b) будет равна нулю;
c) будет увеличиваться;
d) существенно не изменится.

13. Включение фактора в модель целесообразно, если коэффициент регрессии при этом факторе является...
a) существенным;
b) нулевым;
c) несущественным;
d) незначимым.

14. Исследуется зависимость, которая характеризуется линейным уравнением множественной корреляции. Для уравнения рассчитано значение тесноты связи результативной переменной с набором факторов. В качестве этого показателя был использован множественный коэффициент...
a) эластичности;
b) корреляции;
c) регрессии;
d) детерминации.

15. При использовании метода наименьших остатков уменьшить гетероскедастичность остатков удается путем...
a) преобразования параметров;
b) введения дополнительных результатов в модель;
c) введения дополнительных факторов в модель;
d) преобразования переменных.

16. В стандартизированном уравнении свободный член...
a) равен коэффициенту множественной детерминации;
b) равен коэффициенту множественной корреляции;
c) равен 1;
d) отсутствует.

17. Если коэффициент регрессии является несущественным, то его значение приравнивается...
a) к табличному и соответствующий фактор не включается в модель;
b) нулю и соответствующий фактор не включается в модель;
c) нулю и соответствующий фактор включается в модель;
d) единице и не влияет на результат.

18. В матрице парных коэффициентов корреляции отображены значения парных коэффициентов линейной корреляции между...
a) параметрами и переменными;
b) переменными;
c) параметрами;
d) переменными и случайными факторами.

19. Обобщенный метод наименьших квадратов не используется для моделей с ___________ остатками.
a) гомоскедастичными;
b) гетероскедастичными;
c) автокоррелированными и гетероскедастичными;
d) автокоррелированными.

20. Матрица парных коэффициентов корреляции строится для выявления коллинеарных и мультиколлинеарных...
a) случайных факторов;
b) параметров;
c) результатов;
d) существенных факторов.

21. Дано уравнение регрессии $y = a + bx_1 - cx_2 + \varepsilon$. Определите спецификацию модели.
a) линейное уравнение множественной регрессии;
b) полиномиальное уравнение множественной регрессии;
c) линейное уравнение простой регрессии;
d) полиномиальное уравнение парной регрессии.

22. Одним из методов присвоения числовых значений фиктивным переменным является...
a) ранжирование;
b) нахождение среднего значения;
c) выравнивание числовых значений по убыванию;
d) выравнивание числовых значений по возрастанию.

23. Обобщенный метод наименьших квадратов используется для корректировки...
a) автокорреляции между независимыми переменными;
b) гетероскедастичности остатков в уравнении регрессии;
c) параметров нелинейного уравнения регрессии;
d) точности определения коэффициента множественной корреляции.

24. В стандартизированном уравнении множественной регрессии переменными являются...
a) средние значения исходных переменных;
b) исходные переменные;
c) стандартизированные переменные;
d) стандартизированные параметры.

25. Методом присвоения числовых значений фиктивным переменным **не является**...
a) присвоение количественных значений;
b) ранжирование;
c) присвоение числовых меток;
d) нахождение среднего значения.

26. Отбор факторов в модель множественной регрессии при помощи метода включения основан на сравнении остаточной дисперсии до и после включения __________ в модель.
a) случайных факторов;
b) зависимой переменной;
c) независимой переменной;
d) постоянной величины.

27. Что преобразуется при использовании обобщенного метода наименьших квадратов?
a) стандартизированные коэффициенты регрессии;
b) исходные уровни переменных;
c) дисперсия факторного признака;
d) дисперсия результативного признака.

28. Мультиколлинеарность факторов эконометрической модели подразумевает...
a) наличие линейной зависимости между двумя факторами;
b) наличие линейной зависимости между более чем двумя факторами;
c) отсутствие зависимости между факторами;
d) наличие нелинейной зависимости между двумя факторами.

29. На основании преобразования переменных при помощи обобщенного метода наименьших квадратов, получаем новое уравнение регрессии, которое представляет собой...
a) взвешенную регрессию, в которой переменные взяты с весами $1/\sqrt{K}$;
b) нелинейную регрессию, в которой переменные взяты с весами $\sqrt{K}$;
c) взвешенную регрессию, в которой переменные взяты с весами $\sqrt{K}$;
d) нелинейную регрессию, в которой переменные взяты с весами $1/\sqrt{K}$.

30. Объем выборки определяется числом параметров...
a) при случайных факторах;
b) независимых переменных;
c) независимых и зависимых переменных;
d) зависимых переменных.

31. Уравнение регрессии, которое связывает результирующий признак с одним из факторов при зафиксированном на среднем уровне значении других переменных, называется...
a) частным;
b) несущественным;

c) существенным;
d) множественным.

32. Метод оценки параметров моделей с гетероскедастичными остатками называется... методом наименьших квадратов.
a) обобщенным;
b) минимальным;
c) косвенным;
d) обычным.

33. Взаимодействие факторов эконометрической модели означает, что
a) влияние факторов на результирующий признак зависит от значений другого неколлинеарного им фактора;
b) факторы дублируют влияние друг друга на результат;
c) влияние факторов на результирующий признак усиливается, начиная с определенного уровня значений факторов;
d) влияние одного из факторов на результирующий признак не зависит от значений другого фактора.

34. Исходные значения фиктивных переменных предполагают... значения.
a) одинаковые;
b) нулевые;
c) количественно измеримые;
d) качественные.

35. Обобщенный метод наименьших квадратов рекомендуется применять в случае...
a) автокорреляции результативного признака;
b) гомоскедастичных остатков;
c) автокорреляции остатков;
d) нормально распределенных остатков.

36. В качестве фиктивных переменных в модель множественной регрессии включаются факторы, ...
a) имеющие вероятностные значения;
b) имеющие количественные значения;
c) не имеющие количественных значений;
d) не имеющие качественных значений.

37. Факторы эконометрической модели являются коллинеарными, если коэффициент...
a) корреляции между ними по модулю больше 0,7;
b) детерминации между ними по модулю больше 0,7;
c) корреляции между ними по модулю меньше 0,7;
d) детерминации между ними по модулю меньше 0,7.

38. Фиктивные переменные включаются в уравнение ___________ регрессии.
a) множественной;
b) парной;
c) косвенной;
d) случайной.

39. Из пары коллинеарных факторов в эконометрическую модель включается тот фактор, ...

a) который при отсутствии связи с результатом имеет меньшую связь с другими факторами;
b) который при достаточно тесной связи с результатом имеет меньшую связь с другими факторами;
c) который при достаточно тесной связи с результатом имеет большую связь с другими факторами;
d) который при отсутствии связи с результатом имеет оптимальную связь с другими факторами.

40. Производится исследование зависимости выработки сотрудника предприятия от ряда факторов. Примером фиктивной переменной в данной модели будет являться _____ _________ работника.

a) уровень образования;
b) стаж;
c) возраст;
d) заработная плата.

41. В стандартизированном уравнении множественной регрессии $\beta_1 = 0{,}3$, $\beta_2 = -2{,}1$. Определите, какой из факторов x_1 или x_2 оказывает более сильное влияние на y.

a) x_1, так как $0{,}3 > -2{,}1$;
b) по этому уравнению нельзя ответить на поставленный вопрос, так как стандартизированные коэффициенты регрессии несравнимы между собой;
c) по этому уравнению нельзя ответить на поставленный вопрос, так как неизвестны коэффициенты «чистой» регрессии;
d) x_2, так как $2{,}1 > 0{,}3$.

42. Оценки параметров линейного уравнения множественной регрессии можно найти при помощи метода...

a) средних квадратов;
b) наименьших квадратов;
c) нормальных квадратов;
d) наибольших квадратов.

43. Обобщенный метод наименьших квадратов подразумевает...

a) преобразование переменных;
b) переход от множественной регрессии к парной;
c) линеаризацию уравнения регрессии;
d) двухэтапное применение метода наименьших квадратов.

44. Множественная регрессия является результатом преобразования уравнения...

a) $y = a + bx + cx^2 + dx^3 + \varepsilon$;
b) $y = a + bx + cx^2 + \varepsilon$;
c) $y = a + \frac{b}{x} + \frac{1}{x} + \varepsilon$;
d) $y = a + \frac{1}{x} + \frac{1}{x^2} + \varepsilon$.

45. Относительно количества факторов, включенных в уравнение регрессии, различают...
a) линейную и нелинейную регрессию;
b) множественную и многофакторную регрессию;
c) непосредственную и косвенную регрессию;
d) простую и множественную регрессию.

Практическая работа 3.5
Временные ряды

Цель – закрепление знаний теоретических основ, методологических принципов и конкретных подходов постановки, решения и анализа задач моделирования временных рядов.

Примеры решения задач

Пусть имеются некоторые условные данные по общему количеству правонарушений на таможне одного из субъектов РФ (например, Республики Татарстан).

Таблица 3.17

Год	Квартал	t	Количество возбужденных дел, y_t
1999	I	1	375
	II	2	371
	III	3	869
	IV	4	1015
2000	I	5	357
	II	6	471
	III	7	992
	IV	8	1020
2001	I	9	390
	II	10	355
	III	11	992
	IV	12	905
2002	I	13	461
	II	14	454
	III	15	920
	IV	16	927

Построим поле корреляции (рис. 3.7).

Уже исходя из графика видно, что значения y образуют пилообразную фигуру. Рассчитаем несколько последовательных коэффициентов автокорреляции. Для этого составляем первую вспомогательную таблицу (табл. 3.18).

Таблица 3.18

t	y_t	y_{t-1}	$y_t-\overline{y}_1$	$y_{t-1}-\overline{y}_2$	$(y_t-\overline{y}_1)\times$ $\times(y_{t-1}-\overline{y}_2)$	$(y_t-\overline{y}_1)^2$	$(y_{t-1}-\overline{y}_2)^2$
1	375	–	–	–	–	–	–
2	371	375	-328,33	-288,13	94601,72	107800,59	83018,90
3	869	371	169,67	-292,13	-49565,70	28787,91	85339,94
4	1015	869	315,67	205,87	64986,98	99647,55	42382,46
5	357	1015	-342,33	351,87	-120455,66	117189,83	123812,50
6	471	357	-228,33	-306,13	69898,66	52134,59	93715,58

t	y_t	y_{t-1}	$y_t-\overline{y}_1$	$y_{t-1}-\overline{y}_2$	$(y_t-\overline{y}_1)\times$ $\times(y_{t-1}-\overline{y}_2)$	$(y_t-\overline{y}_1)^2$	$(y_{t-1}-\overline{y}_2)^2$
7	992	471	292,67	-192,13	-56230,69	85655,73	36913,94
8	1020	992	320,67	328,87	105458,74	102829,25	108155,48
9	390	1020	-309,33	356,87	-110390,60	95685,05	127356,20
10	355	390	-344,33	-273,13	94046,85	118563,15	74600,00
11	992	355	292,67	-308,13	-90180,41	85655,73	94944,10
12	905	992	205,67	328,87	67638,69	42300,15	108155,48
13	461	905	-238,33	241,87	-57644,88	56801,19	58501,10
14	454	461	-245,33	-202,13	49588,55	60186,81	40856,54
15	920	454	220,67	-209,13	-46148,72	48695,25	43735,36
16	927	920	227,67	256,87	58481,59	51833,63	65982,20
Сумма	10499	9947	9,05	0,05	74085,16	1153766,39	1187469,73
Среднее значение	699,33	663,13	–	–	–	–	–

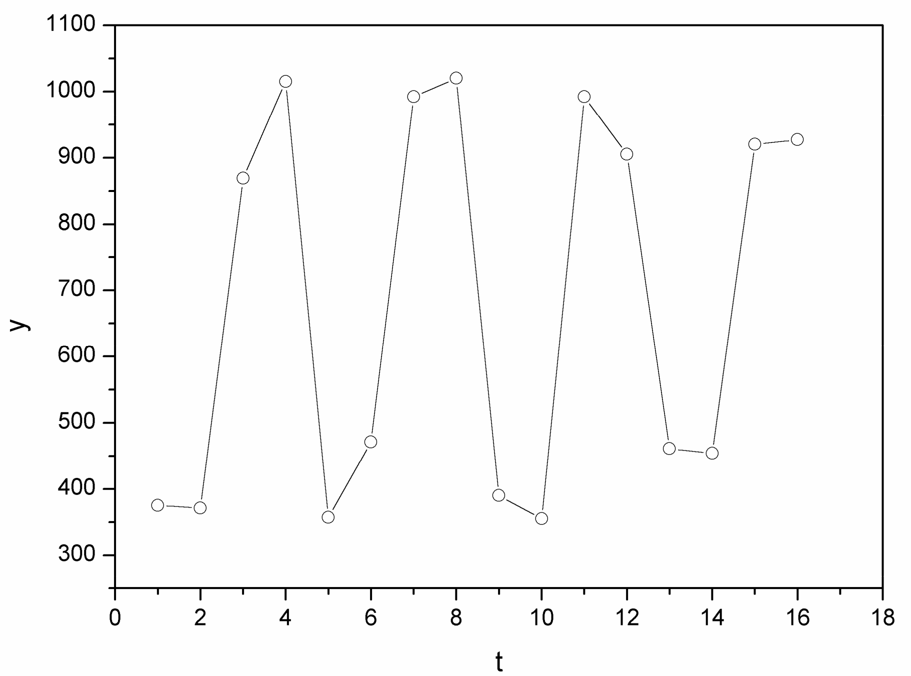

Рис. 3.7. Поле корреляции

Следует заметить, что среднее значение получается путем деления не на 16, а на 15, так как у нас теперь на одно наблюдение меньше.

Теперь вычисляем коэффициент автокорреляции первого порядка по формуле (2.157):

$$r_1 = \frac{74085,16}{\sqrt{1153756,39 \cdot 1187469,73}} = 0,063294\,.$$

Составляем вспомогательную таблицу для расчета коэффициента автокорреляции второго порядка (табл. 3.19).

Таблица 3.19

t	y_t	y_{t-2}	$y_t-\overline{y}_3$	$y_{t-2}-\overline{y}_4$	$(y_t-\overline{y}_3)\times$ $\times(y_{t-2}-\overline{y}_4)$	$(y_t-\overline{y}_3)^2$	$(y_{t-2}-\overline{y}_4)^2$
1	375	–	–	–	–	–	–
2	371	–	–	–	–	–	–
3	869	375	145,57	-269,79	-39273,33	21190,62	72786,64
4	1015	371	291,57	-273,79	-79828,95	85013,06	74960,96
5	357	869	-366,43	224,21	-82157,27	134270,94	50270,12
6	471	1015	-252,43	370,21	-93452,11	63720,90	137055,44
7	992	357	268,57	-287,79	-77291,76	72129,84	82823,08
8	1020	471	296,57	-173,79	-51540,90	87953,76	30202,96
9	390	992	-333,43	347,21	-115770,23	111175,56	120554,78
10	355	1020	-368,43	375,21	-138238,62	135740,66	140782,54
11	992	390	268,57	-254,79	-68428,95	72129,84	64917,94
12	905	355	181,57	-289,79	-52617,17	32967,66	83978,24
13	461	992	-262,43	347,21	-91118,32	68869,50	120554,78
14	454	905	-269,43	260,21	-70108,38	72592,52	67709,24
15	920	461	196,57	-183,79	-36127,60	38639,76	33778,76
16	927	454	203,57	-190,79	-38839,12	41440,74	36400,82
Сумма	10128	9027	-0,02	-0,06	-1034792,71	1037835,43	1116776,36
Среднее значение	723,43	644,79	–	–	–	–	–

Следовательно,

$$r_2 = \frac{-1034792,71}{\sqrt{1037835,43 \cdot 1116776,36}} = -0,961183.$$

Аналогично находим коэффициенты автокорреляции более высоких порядков, а все полученные значения заносим в сводную таблицу (табл. 3.20).

Таблица 3.20

Лаг (τ)	Коэффициент автокорреляции уровней
1	0,063294
2	–0,961183
3	–0,036290
4	0,964735
5	0,050594
6	–0,976516
7	–0,069444
8	0,964629
9	0,162064
10	-0,972918
11	-0,065323
12	0,985761

По данным табл. 3.20 строим график (рис. 3.8).

Анализ коррелограммы и графика исходных уровней временного ряда позволяет сделать вывод о наличии в изучаемом временном ряде сезонных колебаний периодичностью в четыре квартала.

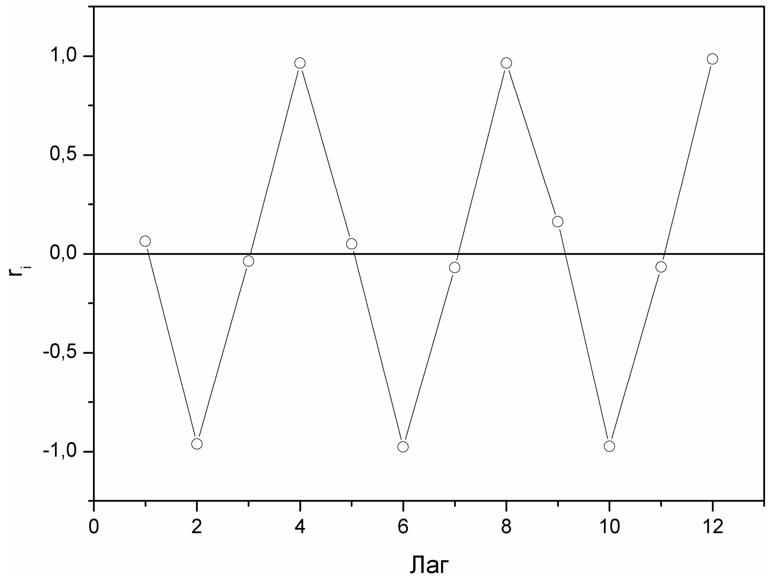

Рис. 3.8. Коррелограмма

Построение аддитивной модели временного ряда

Было показано, что данный временной ряд содержит сезонные колебания периодичностью 4, так как количество правонарушений в первый–второй кварталы ниже, чем в третий–четвертый. Рассчитаем компоненты аддитивной модели временного ряда.

Шаг 1. Проведем выравнивание исходных уровней ряда методом скользящей средней. Для этого:

1.1. Просуммируем уровни ряда последовательно за каждые четыре квартала со сдвигом на один момент времени и определим условные годовые объемы правонарушений (ст. 3 табл. 3.21).

1.2. Разделив полученные суммы на 4, найдем скользящие средние (ст. 4 табл. 3.21). Полученные таким образом выровненные значения уже не содержат сезонной компоненты.

1.3. Приведем эти значения в соответствие с фактическими моментами времени, для чего найдем средние значения из двух последовательных скользящих средних – центрированные скользящие средние (ст. 5 табл. 3.21).

Таблица 3.21

№ квартала, t	Количество правонарушений, y_t	Итого за четыре квартала	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	375	–	–	–	–
2	371	2630	657,5	–	–
3	869	2612	653	655,25	213,75
4	1015	2712	678	665,5	349,5
5	357	2835	708,75	693,75	-336,75
6	471	2840	710	709,375	-238,375
7	992	2873	718,25	714,125	277,875
8	1020	2757	689,25	703,75	316,25
9	390	2757	689,25	689,25	-299,25
10	355	2642	660,5	674,875	-319,875
11	992	2713	678,25	669,375	322,625

№ квартала, t	Количество правонарушений, y_t	Итого за четыре квартала	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
12	905	2812	703	690,625	214,375
13	461	2740	685	694	-233
14	454	2762	690,5	687,75	-233,75
15	920	–	–	–	–
16	927	–	–	–	–

Шаг 2. Найдем оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними (ст. 6 табл. 3.21). Используем эти оценки для расчета значений сезонной компоненты S (табл. 3.22). Для этого найдем средние за каждый квартал (по всем годам) оценки сезонной компоненты S_i. В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимно погашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю.

Таблица 3.22

Показатели	Год	№ квартала, i			
		I	II	III	IV
	1999	–	–	213,75	349,5
	2000	-336,75	-238,375	277,875	316,25
	2001	-299,25	-319,875	322,625	214,375
	2002	-233	-233,75	–	–
Всего за i-й квартал		-869	-792	814,25	880,125
Средняя оценка сезонной компоненты для i-го квартала, $\bar{S}_i$		-289,667	-264	271,417	293,375
Скорректированная сезонная компонента, S_i		-292,448	-266,781	268,636	290,593

Для данной модели имеем:

$$-289,667-264+271,417+293,375=11,125.$$

Корректирующий коэффициент: $k=11,125/4=2,781$.

Рассчитываем скорректированные значения сезонной компоненты ($S_i=\bar{S}_i-k$) и заносим полученные данные в табл. 3.22.

Проверим равенство нулю суммы значений сезонной компоненты:

$$-292,448 - 266,781 + 268,636 + 290,593 = 0.$$

Шаг 3. Исключим влияние сезонной компоненты, вычитая ее значение из каждого уровня исходного временного ряда. Получим величины $T + E = Y - S$ (ст. 4 табл. 3.23). Эти значения рассчитываются за каждый момент времени и содержат только тенденцию и случайную компоненту.

Таблица 3.23

t	y_t	S_i	$y_t - S_i$	T	$T + S$	$E = y_t - (T + S)$	E^2
1	375	-292,448	667,448	672,700	380,252	-5,252	27,584
2	371	-266,781	637,781	673,624	406,843	-35,843	1284,721
3	869	268,636	600,364	674,547	943,183	-74,183	5503,117
4	1015	290,593	724,407	675,470	966,063	48,937	2394,830
5	357	-292,448	649,448	676,394	383,946	-26,946	726,087

t	y_t	S_i	$y_t - S_i$	T	$T + S$	$E = y_t - (T + S)$	E^2
6	471	-266,781	737,781	677,317	410,536	60,464	3655,895
7	992	268,636	723,364	678,240	946,876	45,124	2036,175
8	1020	290,593	729,407	679,163	969,756	50,244	2524,460
9	390	-292,448	682,448	680,087	387,639	2,361	5,574
10	355	-266,781	621,781	681,010	414,229	-59,229	3508,074
11	992	268,636	723,364	681,933	950,569	41,431	1716,528
12	905	290,593	614,407	682,857	973,450	-68,450	4685,403
13	461	-292,448	753,448	683,780	391,332	69,668	4853,630
14	454	-266,781	720,781	684,703	417,922	36,078	1301,622
15	920	268,636	651,364	685,627	954,263	-34,263	1173,953
16	927	290,593	636,407	686,550	977,143	-50,143	2514,320
Сумма							37911,973

Шаг 4. Определим методом наименьших квадратов компоненту T данной модели. Для этого проведем аналитическое выравнивание ряда ($T + E$) с помощью линейного тренда. Результаты аналитического выравнивания следующие:

$$T = 671{,}777 + 0{,}9233 \cdot t.$$

Подставляя в это уравнение значения $t = 1, 2, ..., 16$, найдем уровни T для каждого момента времени (ст. 5 табл. 3.23).

Шаг 5. Найдем значения уровней ряда, полученные по аддитивной модели. Для этого прибавим к уровням T значения сезонной компоненты для соответствующих кварталов (ст. 6 табл. 3.23).

На одном графике отложим фактические значения уровней временного ряда и теоретические, полученные по аддитивной модели (рис. 3.9).

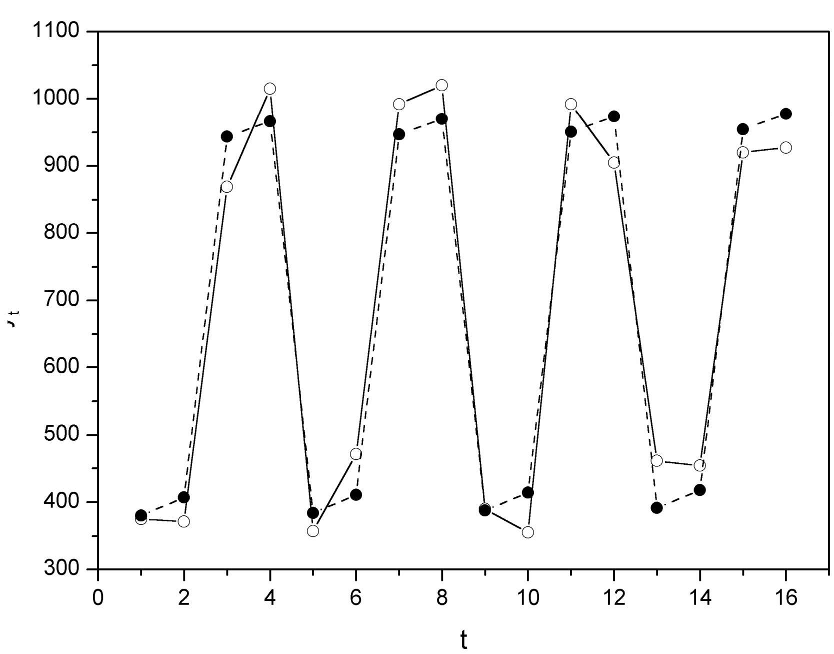

Рис. 3.9. График подбора

Для оценки качества построенной модели применим сумму квадратов полученных абсолютных ошибок. Рассчитаем коэффициент детерминации

$$R^2 = 1 - \frac{\sum E^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{37911{,}973}{1252743{,}75} = 0{,}97.$$

Следовательно, можно сказать, что аддитивная модель объясняет 97% общей вариации уровней временного ряда количества правонарушений по кварталам за 4 года.

Шаг 6. Прогнозирование по аддитивной модели. Предположим, что по нашему примеру необходимо спрогнозировать общий объем правонарушений на I и II кварталы 2003 года. Прогнозное значение F_t уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда

$$T = 671{,}777 + 0{,}9233 \cdot t.$$

Получим

$$T_{17} = 671{,}777 + 0{,}9233 \cdot 17 = 687{,}473;$$
$$T_{18} = 671{,}777 + 0{,}9233 \cdot 18 = 688{,}396.$$

Значения сезонных компонент за соответствующие кварталы равны: S_1 = -292,448 и S_2 = -266,781. Таким образом,

$$F_{17} = T_{17} + S_1 = 687{,}473 - 292{,}448 \approx 395;$$
$$F_{18} = T_{18} + S_2 = 688{,}396 - 266{,}781 \approx 422.$$

Согласно подобранной модели, в первые два квартала 2003 года следовало ожидать порядка 395 и 422 правонарушений.

Построение мультипликативной модели рассмотрим на данных предыдущего примера.

Шаг 1. Методика, применяемая на этом шаге, полностью совпадает с методикой построения аддитивной модели.

Таблица 3.24

№ квартала, t	Количество правонарушений, $\hat{y}_t$	Итого за четыре квартала	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	375	–	–	–	–
2	371	2630	657,5	–	–
3	869	2612	653	655,25	1,3262
4	1015	2712	678	665,5	1,5252
5	357	2835	708,75	693,75	0,5146
6	471	2840	710	709,375	0,6640
7	992	2873	718,25	714,125	1,3891
8	1020	2757	689,25	703,75	1,4494
9	390	2757	689,25	689,25	0,5658
10	355	2642	660,5	674,875	0,5260
11	992	2713	678,25	669,375	1,4820
12	905	2812	703	690,625	1,3104
13	461	2740	685	694	0,6643
14	454	2762	690,5	687,75	0,6601
15	920	–	–	–	–
16	927	–	–	–	–

Шаг 2. Найдем оценки сезонной компоненты как частное от деления фактических уровней ряда на центрированные скользящие средние (ст. 6 табл. 3.24). Эти оценки используются для расчета сезонной компоненты S (табл. 3.25). Для этого найдем средние за каждый квартал оценки сезонной компоненты S_i. Так же как и в аддитивной модели, считается, что сезонные воздействия за период взаимопогашаются. В мультипликативной модели это выражается в том, что сумма значений сезонной компоненты по всем

кварталам должна быть равна числу периодов в цикле. В нашем случае число периодов одного цикла равно 4.

Таблица 3.25

Показатели	Год	№ квартала, i			
		I	II	III	IV
	1999	–	–	1,3262	1,5252
	2000	0,5146	0,6640	1,3891	1,4494
	2001	0,5658	0,5260	1,4820	1,3104
	2002	0,6643	0,6601	–	–
Всего за i-й квартал		1,7447	1,8501	4,1973	4,2850
Средняя оценка сезонной компоненты для i-го квартала, $\bar{S}_i$		0,5816	0,6167	1,3991	1,4283
Скорректированная сезонная компонента, S_i		0,5779	0,6128	1,3901	1,4192

Имеем

$$0{,}5816 + 0{,}6167 + 1{,}3991 + 1{,}4283 = 4{,}0257.$$

Определяем корректирующий коэффициент:

$$k = 4/4{,}0257 = 0{,}9936.$$

Скорректированные значения сезонной компоненты S_i получаются при умножении ее средней оценки $\bar{S}_i$ на корректирующий коэффициент k.

Проверяем условие равенства 4 суммы значений сезонной компоненты:

$$0{,}5779 + 0{,}6128 + 1{,}3901 + 1{,}4192 = 4.$$

Шаг 3. Разделим каждый уровень исходного ряда на соответствующие значения сезонной компоненты. В результате получим величины $T \cdot E = Y/S$ (ст. 4 табл. ***3.***26), которые содержат только тенденцию и случайную компоненту.

Таблица 3.26

t	y_t	S_i	y_t/S_i	T	$T \cdot S$	$E = y_t/(T \cdot S)$
1	375	0,5779	648,9012	654,9173	378,4767	0,9908
2	371	0,6128	605,4178	658,1982	403,3439	0,9198
3	869	1,3901	625,1349	661,4791	919,5221	0,9451
4	1015	1,4192	715,1917	664,7600	943,4274	1,0759
5	357	0,5779	617,7539	668,0409	386,0608	0,9247
6	471	0,6128	768,6031	671,3218	411,3860	1,1449
7	992	1,3901	713,6177	674,6027	937,7652	1,0578
8	1020	1,4192	718,7148	677,8836	962,0524	1,0602
9	390	0,5779	674,8572	681,1645	393,6450	0,9907
10	355	0,6128	579,3081	684,4454	419,4281	0,8464
11	992	1,3901	713,6177	687,7263	956,0083	1,0377
12	905	1,4192	637,6832	691,0072	980,6774	0,9228
13	461	0,5779	797,7159	694,2881	401,2291	1,1490
14	454	0,6128	740,8616	697,5690	427,4703	1,0621
15	920	1,3901	661,8229	700,8499	974,2515	0,9443
16	927	1,4192	653,1849	704,1308	999,3024	0,9277

Шаг 4. Определим компоненту T в мультипликативной модели. Для этого рассчитаем параметры линейного тренда, используя уровни $T \cdot E$. В результате получим уравнение тренда:

$$T = 651{,}6364 + 3{,}2809 \cdot t.$$

Подставляя в это уравнение значения $t = 1, 2, ..., 16$, найдем уровни T для каждого момента времени (ст. 5 табл. 3.26).

Шаг 5. Найдем уровни ряда, умножив значения T на соответствующие значения сезонной компоненты (ст. 6 табл.3. 26). На одном графике откладываем фактические значения уровней временного ряда и теоретические, полученные по мультипликативной модели (рис. 3.10).

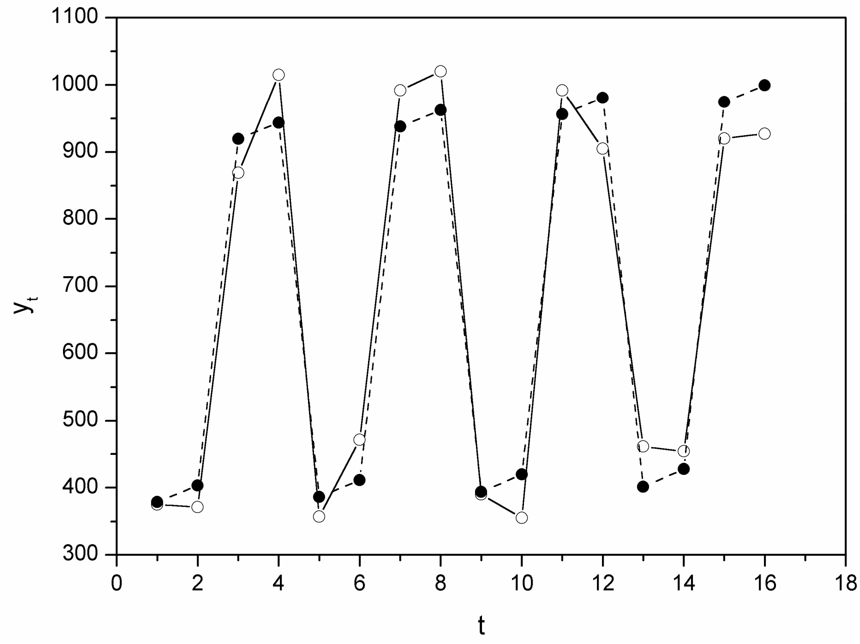

Рис. 3.10. График подбора

Расчет ошибки в мультипликативной модели производится по формуле

$$E = Y/(T \cdot S).$$

Для сравнения мультипликативной модели и других моделей временного ряда можно, по аналогии с аддитивной моделью, использовать сумму квадратов абсолютных ошибок $(y_t - T \cdot S)^2$:

$$R^2 = 1 - \frac{\sum (y_t - T \cdot S)^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{43065{,}02}{1252743{,}75} = 0{,}966.$$

Сравнивая показатели детерминации аддитивной и мультипликативной моделей, делаем вывод, что они примерно одинаково аппроксимируют исходные данные.

Шаг 6. Прогнозирование по мультипликативной модели. Если предположить, что по нашему примеру необходимо спрогнозировать общий объем правонарушений на I и II кварталы 2003 года, то прогнозное значение F_t уровня временного ряда в мультипликативной модели есть произведение трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда

$$T = 651{,}6364 + 3{,}2809 \cdot t.$$

Получим

$$T_{17} = 651{,}6364 + 3{,}2809 \cdot 17 = 707{,}4117;$$

$$T_{18} = 651{,}6364 + 3{,}2809 \cdot 18 = 710{,}6926.$$

Значения сезонных компонент за соответствующие кварталы равны:

$$S_1 = 0{,}45779 \text{ и } S_2 = 0{,}6128.$$

Таким образом,

$$F_{17} = T_{17} \cdot S_1 = 707{,}4117 \cdot 0{,}5779 \approx 409;$$
$$F_{18} = T_{18} \cdot S_2 = 710{,}6926 \cdot 0{,}6128 \approx 436.$$

Прогнозируя по модели, в первые два квартала 2003 года следовало ожидать порядка 409 и 436 правонарушений.

Таким образом, аддитивная и мультипликативная модели дают примерно одинаковый результат по прогнозу.

Проверим гипотезу о наличии автокорреляции в остатках для аддитивной модели нашего временного ряда. Исходные данные и промежуточные расчеты заносим в табл. 3.27.

Таблица 3.27

t	y_t	$\varepsilon_t = E$	ε_{t-1}	$(\varepsilon_t - \varepsilon_{t-1})^2$	ε_t^2
1	375	-5,252	–	–	27,584
2	371	-35,843	-5,252	935,8093	1284,7
3	869	-74,183	-35,843	1469,956	5503,1
4	1015	48,937	-74,183	15158,53	2394,8
5	357	-26,946	48,937	5758,23	726,09
6	471	60,464	-26,946	7640,508	3655,9
7	992	45,124	60,464	235,3156	2036,2
8	1020	50,244	45,124	26,2144	2524,5
9	390	2,361	50,244	2292,782	5,574
10	355	-59,229	2,361	3793,328	3508,1
11	992	41,431	-59,229	10132,44	1716,5
12	905	-68,450	41,431	12073,83	4685,4
13	461	69,668	-68,45	19076,58	4853,6
14	454	36,078	69,668	1128,288	1301,6
15	920	-34,263	36,078	4947,856	1174
16	927	-50,143	-34,263	252,1744	2514,3
Сумма		-0,002	50,141	84921,85	37911,97

Фактическое значение критерия Дарбина-Уотсона (формула 2.161) для данной модели составляет:

$$d = \frac{84921{,}85}{37911{,}97} = 2{,}24.$$

Сформулируем гипотезы: H_0 – в остатках нет автокорреляции; H_1 – в остатках есть положительная автокорреляция; H^*_1 – в остатках есть отрицательная автокорреляция. Зададим уровень значимости $\alpha = 0{,}05$. По таблице значений критерия Дарбина-Уотсона определим для числа наблюдений $n = 16$ и числа независимых параметров модели $k = 1$ (мы рассматриваем только зависимость от времени t) критические значения $d_L = 1{,}10$ и $d_U = 1{,}37$. Фактическое значение d-критерия Дарбина-Уотсона попадает в интервал $d_U < d < 4 - d_U$ ($1{,}37 < 2{,}24 < 2{,}63$). Следовательно, нет основания отклонять гипотезу H_0 об отсутствии автокорреляции в остатках.

Задачи для самостоятельного решения

Имеются условные данные по объемам потребления электроэнергии (y_t) жителями региона за 16 кварталов (табл. 3.28).

Таблица 3.28

t	y_t	t	y_t
1	5,6	9	8,2
2	4,7	10	5,5
3	5,2	11	6,4
4	9,1	12	10,8
5	7,0	13	9,1
6	5,1	14	6,7
7	6,0	15	7,5
8	10,2	16	11,3

Требуется:

1. Построить автокорреляционную функцию и сделать вывод о наличии сезонных колебаний.

2. Построить аддитивную модель временного ряда (для нечетных вариантов) или мультипликативную модель временного ряда (для четных вариантов).

3. Сделать прогноз на 2 квартала вперед.

Тесты

1. Структуру временного ряда можно выявить с помощью коэффициента...
a) детерминации уровней ряда;
b) регрессии уровней ряда;
c) автокорреляции уровней ряда;
d) авторегрессии уровней ряда.

2. Временной ряд – это совокупность значений экономических показателей...
a) за несколько последовательных моментов (периодов) времени;
b) не зависящих от времени;
c) по однотипным объектам;
d) за несколько непоследовательных моментов (периодов) времени.

3. Параметры уравнения тренда определяются ____________ методом наименьших квадратов.
a) двухшаговым;
b) обычным;
c) косвенным;
d) обобщенным.

4. Уровнем временного ряда является...
a) совокупность значений временного ряда;
b) значение временного ряда в конкретный момент (период) времени;
c) среднее значение временного ряда;
d) значение конкретного момента (периода) времени.

5. Если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит...
a) тенденцию;
b) случайную компоненту;
c) циклические колебания;
d) сильную нелинейную тенденцию.

6. Моделирование тенденции осуществляется на основе построения уравнения регрессии зависимости...

a) случайной компоненты от времени;
b) уровня ряда от времени;
c) сезонной компоненты от времени;
d) трендовой компоненты от времени.

7. Под стационарным процессом можно понимать...

a) функциональный процесс;
b) стохастический процесс, для которого среднее значение и дисперсия независимо от рассматриваемого периода имеют постоянное значение;
c) процесс с постоянной тенденцией;
d) процесс с возрастающей тенденцией.

8. Значение коэффициента автокорреляции второго порядка характеризует связь между...

a) исходными уровнями и уровнями другого временного ряда;
b) двумя временными рядами;
c) исходными уровнями и уровнями другого ряда, сдвинутыми на 2 момента времени;
d) исходными уровнями и уровнями этого же ряда, сдвинутыми на 2 момента времени.

9. Стационарность временного ряда означает отсутствие...

a) значений уровней ряда;
b) тренда;
c) временной характеристики;
d) наблюдений по уровням временного ряда.

10. Известны значения аддитивной модели временного ряда: Y_t – значение уровня ряда, $Y_t = 30$, T – значение тренда, $T = 15$, E – значение компоненты случайных факторов, $E = 2$, определите значение сезонной компоненты S.

a) $S = 13$;
b) $S = 0$;
c) $S = 1$;
d) $S = -1$.

11. Модель временного ряда предполагает...

a) пренебрежение временными характеристиками ряда;
b) отсутствие последовательности моментов (периодов) времени, в течение которых рассматривается поведение экономического показателя;
c) независимость значений экономического показателя от времени;
d) зависимость значений экономического показателя от времени.

12. Уровень временного ряда может формироваться под воздействием тенденции, сезонных колебаний и...

a) случайных воздействий;
b) циклических колебаний;
c) динамической составляющей;
d) тренда.

13. При построении модели временного ряда проводится...

a) расчет значений компонент для каждого уровня временного ряда;
b) расчет последующих и предыдущих значений уровней временного ряда;
c) расчет каждого уровня временного ряда;
d) расчет средних значений компонент для временного ряда в целом.

14. Стохастическим процессом называется...

a) набор случайных переменных $X(t)$, где t – вещественные числа;
b) набор случайных переменных $X(t)$, где t – иррациональные числа;
c) функциональная связь $X(t)$, где t – вещественные числа;
d) набор неслучайных переменных $X(t)$, где t – вещественные числа.

15. Циклические колебания (с периодичностью более 3 лет) связаны...

a) с сезонностью некоторых видов экономической деятельности (сельское хозяйство, туризм и т. д.);
b) общей динамикой конъюнктуры рынка;
c) трендовыми взаимодействиями между экономическими показателями;
d) воздействиями аномальных факторов.

16. Построена аддитивная модель временного ряда, где Y_t – значение уровня ряда, $Y_t = 10$, T – значение тренда, S – значение сезонной компоненты, E – значение случайной компоненты. Определите вариант правильно найденных значений компонент уровня ряда.

a) $T = 5$, $S = 2$, $E = 0$;
b) $T = 5$, $S = 2$, $E = 1$;
c) $T = 5$, $S = 2$, $E = 3$;
d) $T = 7$, $S = 5$, $E = 2$.

17. Значение коэффициента автокорреляции рассчитывается по аналогии...

a) с линейным коэффициентом корреляции;
b) нелинейным коэффициентом корреляции;
c) линейным коэффициентом регрессии;
d) линейным коэффициентом детерминации.

18. Проверка, является ли временной ряд «белым шумом», осуществляется с помощью...

a) критерия Дарбина-Уотсона;
b) величины лага;
c) коэффициента автокорреляции;
d) Q-статистики Бокса-Пирса.

19. Стационарность характерна для временного ряда...

a) с отрицательной динамикой роста;
b) с положительной динамикой роста;
c) содержащего сезонные колебания;
d) типа «белый шум».

20. Временной ряд называется стационарным, если он является реализацией _______ процесса.
a) неслучайного;
b) нестационарного стохастического;
c) функционального;
d) стационарного стохастического.

21. Под лагом подразумевается число...
a) уровней исходного временного ряда;
b) пар значений, по которым рассчитывается коэффициент автокорреляции;
c) периодов, по которым рассчитывается коэффициент автокорреляции;
d) временных рядов, по которым осуществляется расчет коэффициента автокорреляции.

22. Автокорреляционной функцией временного ряда называется...
a) зависимость приращений значений коэффициентов автокорреляции различных порядков от уровней временного ряда;
b) зависимость коэффициентов автокорреляции первого порядка от уровней временного ряда;
c) последовательность отношений коэффициентов автокорреляции к величинам соответствующих лагов;
d) последовательность значений коэффициентов автокорреляции, рассчитанных для разных порядков.

23. Построена мультипликативная модель временного ряда, где Y_t – значение уровня ряда, $Y_t = 10$, T – значение тренда, S – значение сезонной компоненты, E – значение случайной компоненты. Определите вариант правильно найденных значений компонент уровня ряда.
a) $T = 5$, $S = 2$, $E = 3$;
b) $T = 5$, $S = 2$, $E = 0$;
c) $T = 5$, $S = 2$, $E = 1$;
d) $T = 5$, $S = 2$, $E = -1$.

24. Основной задачей моделирования временных рядов является...
a) исключение уровней из совокупности значений временного ряда;
b) исключение значений каждой из трех компонент из уровней временного ряда;
c) выявление и придание количественного значения каждой из трех компонент;
d) добавление новых уровней к совокупности значений временного ряда.

25. Коррелограммой является...
a) процесс экспериментального нахождения значений автокорреляционной функции;
b) аналитическое выражение для автокорреляционной функции;
c) графическое отображение регрессионной функции;
d) графическое отображение автокорреляционной функции.

26. Если наиболее высоким оказался коэффициент автокорреляции третьего порядка, то исследуемый ряд содержит...
a) случайную величину, влияющую на каждый третий уровень ряда;
b) нелинейную тенденцию полинома третьего порядка;
c) линейный тренд, появляющийся в каждом третьем уровне ряда;
d) сезонные колебания с периодичностью в три момента времени.

27. Экономические временные ряды, представляющие собой данные наблюдений за ряд лет, как правило, являются...
b) функционально зависящими от времени;
c) нестационарными;
d) стационарными;
e) строго возрастающими.

28. Известны значения мультипликативной модели временного ряда: Y_t – значение уровня ряда, $Y_t = 15$, T – значение тренда, $T = 5$, S – значение сезонной компоненты, $S = 3$, определите значение компоненты E (случайные факторы).
a) $E = 1$;
b) $E = -1$;
c) $E = 0$;
d) $E = -3$.

29. Значение коэффициента автокорреляции первого порядка равно 0,9, следовательно, ...
a) линейная связь между последующим и предыдущим уровнями не тесная;
b) линейная связь между последующим и предыдущим уровнями тесная;
c) линейная связь между временными рядами двух экономических показателей тесная;
d) нелинейная связь между последующим и предыдущим уровнями тесная.

30. Модель временного ряда не предполагает...
a) независимость значений экономического показателя от времени;
b) учет временных характеристик;
c) зависимость значений экономического показателя от времени;
d) последовательность периодов времени, в течение которых рассматривается поведение экономического показателя.

31. Максимальный лаг связан с числом уровней временного ряда соотношением не более чем...
a) n/10;
b) n/2;
c) n/4;
d) n/5.

32. При моделировании временных рядов экономических показателей необходимо учитывать...
a) независящий от времени уровень исследуемых показателей;
b) конструктивный характер уровней исследуемых показателей;
c) функциональный характер уровней исследуемых показателей;
d) стохастический характер уровней исследуемых показателей.

33. Может ли ряд содержать только одну из компонент?
a) не может, так как временной ряд не содержит компонент, влияющих на его уровни;
b) не может, так как уровень ряда должен формироваться под воздействием всех трех компонент;
c) может, если он представлен данными, описывающими совокупность различных объектов в определенный момент (период) времени;
d) может, если другие две компоненты не участвуют в формировании уровня ряда.

35. Временной ряд характеризует...
a) данные, описывающие совокупность различных объектов в определенный момент времени;
b) зависимость последовательных моментов (периодов) времени;
c) совокупность последовательных моментов (периодов) времени;
d) данные, описывающие один объект за ряд последовательных моментов (периодов) времени.

36. Укажите справедливые утверждения по поводу критерия Дарбина-Уотсона.
a) равен 0 в случае отсутствия автокорреляции;
b) изменяется в пределах от 0 до 4;
c) позволяет проверить гипотезу о наличии автокорреляции первого порядка;
d) применяется для проверки гипотезы о наличии гетероскедастичности остатков.

37. Тенденция временного ряда характеризует совокупность факторов, ...
a) оказывающих долговременное влияние и формирующих общую динамику изучаемого показателя;
b) оказывающих сезонное воздействие;
c) оказывающих единовременное влияние;
d) не оказывающих влияния на уровень ряда.

Практическая работа 3.6
Системы эконометрических уравнений

Цель – закрепление теоретических основ, методологических принципов и конкретных подходов постановки, решения и анализа задач моделирования систем эконометрических уравнений.

Примеры решения задач

Изучается модель вида

$$\begin{cases} C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + \varepsilon_2, \\ r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + \varepsilon_3, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C_t – расходы на потребление в период t; Y_t – совокупный доход в период t; I_t – инвестиции в период t; r_t – процентная ставка в период t; M_t – денежная масса в период t; G_t – государственные расходы в период t; C_{t-1} – расходы на потребление в период t-1, I_{t-1} инвестиции в период t-1. Первое уравнение – функция потребления, второе уравнение – функция инвестиций, третье уравнение – функция денежного рынка, четвертое уравнение – тождество дохода.

Модель представляет собой систему одновременных уравнений. Проверим каждое ее уравнение на идентификацию.

Модель включает четыре эндогенные переменные (C_t, I_t, Y_t, r_t) и четыре предопределенные переменные (две экзогенные переменные – M_t и G_t и две лаговые переменные – C_{t-1} и I_{t-1}).

Проверим необходимое условие идентификации для каждого из уравнений модели.

Первое уравнение: $Y_t\ C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + \varepsilon_1$. Это уравнение содержит две эндогенные переменные G_t и Y_t и одну предопределенную переменную C_{t-1}. Таким образом, $H = 2$, а $D = 4 - 1 = 3$, т. е. выполняется условие $D + 1 > H$. Уравнение сверхидентифицируемо.

Второе уравнение: $I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + \varepsilon_2$. Оно включает две эндогенные переменные I_t и r_t и одну экзогенную переменную I_{t-1}. Выполняется условие $D + 1 = 3 + 1 > H = 2$. Уравнение сверхидентифицируемо.

Третье уравнение: $r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + \varepsilon_3$. Оно включает две эндогенные переменные Y_t и r_t и одну экзогенную переменную M_t. Выполняется условие $D + 1 = 3 + 1 > H = 2$. Уравнение сверхидентифицируемо.

Четвертое уравнение: $Y_t = C_t + I_t + G_t$. Оно представляет собой тождество, параметры которого известны. Необходимости в идентификации нет.

Проверим для каждого уравнения достаточное условие идентификации. Для этого составим матрицу коэффициентов n и переменных модели (табл. 3.29).

Таблица 3.29

	C_t	I_t	r_t	Y_t	C_{t-1}	I_{t-1}	M_t	G_t
I уравнение	–1	0	0	b_{11}	b_{12}	0	0	0
II уравнение	0	–1	b_{21}	0	0	b_{22}	0	0
III уравнение	0	0	–1	b_{31}	0	0	b_{32}	0
Тождество	1	1	0	–1	0	0	0	1

В соответствии с достаточным условием идентификации ранг матрицы коэффициентов при переменных, не входящих в исследуемое уравнение, должен быть равен числу эндогенных переменных модели без одного.

Первое уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид, приведенный в табл. 3.30.

Таблица 3.30

	I_t	r_t	I_{t-1}	M_t	G_t
II уравнение	–1	b_{21}	b_{22}	0	0
III уравнение	0	–1	0	b_{32}	0
Тождество	1	0	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{22} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{22} b_{32} \neq 0 .$$

Достаточное условие идентификации для данного уравнения выполняется.

Второе уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид, приведенный в табл. 3.31.

Таблица 3.31

	C_t	Y_t	C_{t-1}	M_t	G_t
I уравнение	–1	b_{11}	b_{12}	0	0
III уравнение	0	b_{31}	0	b_{32}	0
Тождество	1	–1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12}b_{32} \neq 0 .$$

Достаточное условие идентификации для данного уравнения выполняется.

Третье уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид, приведенный в табл. 3.32.

Таблица 3.32

	C_t	I_t	C_{t-1}	I_{t-1}	G_t
I уравнение	–1	0	b_{12}	0	0
II уравнение	0	–1	0	b_{22}	0
Тождество	1	1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12}b_{22} \neq 0 .$$

Достаточное условие идентификации для данного уравнения выполняется.

Таким образом, все уравнения модели сверхидентифицируемы. Приведенная форма модели в общем виде будет выглядеть следующим образом:

$$\begin{cases} C_t = A_1 + \delta_{11}C_{t-1} + \delta_{12}I_{t-1} + \delta_{13}M_t + \delta_{14}G_t + u_1, \\ I_t = A_2 + \delta_{21}C_{t-1} + \delta_{22}I_{t-1} + \delta_{23}M_t + \delta_{24}G_t + u_2, \\ r_t = A_3 + \delta_{31}C_{t-1} + \delta_{32}I_{t-1} + \delta_{33}M_t + \delta_{34}G_t + u_3, \\ Y_t = A_4 + \delta_{41}C_{t-1} + \delta_{42}I_{t-1} + \delta_{43}M_t + \delta_{44}G_t + u_1. \end{cases}$$

Тесты

1. Системы эконометрических уравнений классифицируются...

a) по количеству факторов в каждом уравнении системы;
b) количеству уравнений в системе;
c) способу вхождения зависимых и независимых переменных в уравнения регрессии;
d) способу ранжирования факторов в зависимости от силы влияния на моделируемые показатели.

2. Система эконометрических уравнений представляет собой систему...
a) уравнений корреляции;
b) эконометрических показателей;
c) уравнений регрессии;
d) социальных показателей.

3. Изолированное уравнение множественной регрессии может быть использовано для моделирования взаимосвязей экономических показателей, если...
b) система не предполагает использование уравнения множественной регрессии;
c) при изменении одного экономического фактора другие также изменяются;
d) факторы не взаимодействуют друг с другом;
e) изменение переменной влечет за собой изменения во всей системе взаимосвязанных признаков.

4. Структурной формой модели называется система ________ уравнений.
a) взаимозависимых;
b) изолированных;
c) независимых;
d) рекурсивных.

5. В системах рекурсивных уравнений количество переменных в правой части каждого уравнения определяется как
a) сумма количества зависимых переменных последующих уравнений и количества независимых факторов;
b) разность количества зависимых переменных последующих уравнений и количества независимых факторов;
c) сумма количества зависимых переменных предыдущих уравнений и количества независимых факторов;
d) разность количества зависимых переменных предыдущих уравнений и количества независимых факторов.

6. Двухшаговый метод наименьших квадратов предполагает:
a) двукратное использование обычного МНК;
b) однократное использование обычного МНК;
c) трехкратное использование обычного МНК;
d) неиспользование обычного МНК.

7. Выделяют три класса систем эконометрических уравнений:
a) системы взаимозависимых уравнений, системы рекурсивных уравнений и системы возвратных уравнений;
b) системы одновременных уравнений, системы взаимозависимых уравнений и системы рекурсивных уравнений;
c) системы независимых уравнений, системы изолированных уравнений и системы рекурсивных уравнений;
d) системы независимых уравнений, системы взаимозависимых уравнений и системы рекурсивных уравнений.

8. Первопричиной использования систем эконометрических уравнений является то, что...

a) существует доминирующий фактор;
b) изолированное уравнение не отображает истинные влияния факторов на вариацию результативных переменных;
c) случайные факторы не оказывают существенного влияния на моделируемую экономическую систему;
d) отсутствует связь между экономическими показателями.

9. Основной задачей построения систем эконометрических уравнений является описание...

a) структуры связей реальной экономической системы;
b) взаимодействия реальных экономической и политической систем;
c) математических зависимостей;
d) структуры связей реальной политической системы.

10. При оценке параметров приведенной формы модели косвенный метод наименьших квадратов использует алгоритм...

a) обычного метода наименьших квадратов;
b) расчета средней взвешенной величины;
c) метода максимального правдоподобия;
d) метода главных компонент.

11. Система рекурсивных уравнений включает в каждое**...**

a) уравнение в качестве факторов все зависимые переменные с набором собственно факторов;
b) предыдущее уравнение в качестве факторов все зависимые переменные последующих уравнений с набором собственно факторов;
c) последующее уравнение в качестве зависимых переменных собственно факторы предшествующих уравнений;
d) последующее уравнение в качестве факторов все зависимые переменные предшествующих уравнений с набором собственно факторов.

12. Для моделирования сложных экономических систем целесообразно использовать...

a) изолированное уравнение регрессии;
b) систему эконометрических уравнений;
c) стационарный процесс;
d) временной ряд.

13. Приведенная форма модели является результатом преобразования...

a) нелинейных уравнений регрессии;
b) системы рекурсивных уравнений;
c) системы независимых уравнений;
d) структурной формы модели.

14. Модель идентифицируема, если число параметров структурной формы модели...

a) меньше числа параметров приведенной формы модели;
b) равно числу параметров приведенной формы модели;

c) больше числа параметров приведенной формы модели;
d) равно числу уравнений модели.

15. При построении систем независимых уравнений набор факторов в каждом уравнении определяется числом факторов, оказывающих __________ влияние на моделируемый показатель.
a) опосредованное;
b) существенное;
c) как существенное, так и несущественное;
d) краткосрочное.

16. Структурными коэффициентами модели называются коэффициенты __________ __ в структурной форме модели.
a) свободных членов;
b) при экзогенных переменных;
c) при экзогенных и эндогенных переменных;
d) при эндогенных переменных.

17. При изучении взаимодействия спроса и предложения целесообразно использовать...
a) изолированные уравнения;
b) систему эконометрических уравнений;
c) уравнение зависимости предложения от цены;
d) уравнение зависимости спроса от цены.

18. Левая часть системы независимых уравнений представлена вектором...
a) зависимых независимых переменных;
b) зависимых переменных и случайных величин;
c) независимых переменных;
d) зависимых переменных.

19. Система эконометрических уравнений предполагает наличие...
a) одного зависимого и совокупности независимых признаков;
b) нескольких зависимых и одного независимого признаков;
c) нескольких зависимых и нескольких независимых признаков;
d) одного зависимого и нескольких независимых признаков.

20. Приведенная форма модели представляет собой систему...
a) обратных функций эндогенных переменных от экзогенных;
b) случайных функций эндогенных переменных от экзогенных;
c) линейных функций эндогенных переменных от экзогенных;
d) нелинейных функций эндогенных переменных от экзогенных.

21. Косвенный метод наименьших квадратов требует...
a) линеаризации уравнений структурной формы модели;
b) нормализации уравнений структурной формы модели;
c) преобразования структурной формы модели в приведенную;
d) линеаризации уравнений приведенной формы модели.

22. При построении системы эконометрических уравнений необходимо учитывать...
a) структуру связей реальной экономической системы;
b) максимальную величину каждого фактора;
c) значения наблюдений;
d) среднюю величину каждой зависимой переменной.

23. Система независимых уравнений предполагает:
a) совокупность зависимых уравнений регрессии;
b) совокупность независимых временных рядов;
c) одно независимое уравнение регрессии;
d) совокупность независимых уравнений регрессии.

24. При оценке параметров систем одновременных уравнений **не производят**...
a) расчет коэффициентов приведенной формы;
b) преобразование структурной формы модели в приведенную;
c) линеаризацию уравнений системы;
d) идентификацию системы одновременных уравнений.

25. Двухшаговый метод наименьших квадратов применяется для оценки параметров...
a) временных рядов;
b) линеаризованных уравнений регрессии;
c) нелинейных уравнений регрессии;
d) систем эконометрических уравнений.

26. Система взаимозависимых уравнений в ее классическом виде называется также системой ___________ уравнений.
a) независимых;
b) одновременных;
c) изолированных;
d) рекурсивных.

27. Приведенная форма модели получена из _________ формы модели.
a) независимой;
b) рекурсивной;
c) изолированной;
d) структурной.

28. Под идентификационной моделью подразумевается...
a) единственность соответствия между приведенной и структурной формами модели;
b) существование нескольких приведенных моделей для одной структурной формы;
c) адекватность модели;
d) достоверность модели.

4. МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ КОНТРОЛЬНОЙ РАБОТЫ

Перед выполнением контрольного задания студент должен изучить соответствующие разделы курса по пособию и предложенным источникам (см. библиографический список).

Номера вариантов контрольных задач определяются с помощью нижеприведенной таблицы.

Буква	А, Х, О	В, У, Ш	Д, Р, Щ	Е, П	И, Г, Ж	К, Ф, Э	Л, Ч, Ю	Б, М, Я	Н, Т	С, Ц, З
№ вар.	1	2	3	4	5	6	7	8	9	10

Для тем «Виды эконометрических моделей. Введение в регрессионный анализ» и «Множественная линейная регрессия и корреляция» номера находятся по первой букве фамилии студента.

Для тем «Парная линейная регрессия» и «Временные ряды» номера устанавливаются по первой букве имени студента.

Для тем «Нелинейные регрессионные модели» и «Системы эконометрических уравнений» номера находятся по первой букве отчества студента.

При выполнении контрольных работ необходимо строго придерживаться указанных ниже правил. Работы, выполненные без соблюдения этих правил, не зачитываются и возвращаются студенту для переработки.

1. Каждая контрольная работа должна быть выполнена в отдельной тетради в клетку черными или синими чернилами. Необходимо оставлять поля шириной 4–5 см для замечаний рецензента.

2. В заголовке работы на обложке тетради должны быть ясно написаны название учебного заведения, название дисциплины, группа, фамилия, имя и отчество студента, дата отсылки работы и адрес студента. В конце работы следует поставить дату её выполнения и подпись студента.

3. В работу должны быть включены задачи по всем темам дисциплины строго по положенному варианту. Решения задач надо располагать в порядке возрастания их номеров.

4. Перед решением каждой задачи надо полностью выписать её условие. Не следует приступать к выполнению контрольного задания, не решив достаточного количества задач по материалу, соответствующему этому заданию. Опыт показывает, что чаще всего неумение решить ту или иную задачу контрольного задания обусловлено тем, что студент не выполнил это требование.

5. В рецензированной работе студент должен исправить отмеченные рецензентом ошибки и учесть его рекомендации и советы. Рецензии позволяют студенту судить о степени усвоения соответствующего раздела курса; указывают на имеющиеся у него пробелы, на желательное направление работы; помогают сформулировать вопросы для постановки их перед преподавателем. Зачтенные контрольные работы предъявляются студентом при сдаче зачета или экзамена.

6. Контрольные работы должны выполняться самостоятельно. Несамостоятельно выполненная работа не дает возможности преподавателю-рецензенту указать студенту на недостатки в его работе, в усвоении им учебного материала. В результате студент не приобретает необходимых знаний и может оказаться неподготовленным к устному зачету или экзамену.

Контрольная работа 4.1
Виды эконометрических моделей. Введение в регрессионный анализ

По территориям региона приводятся данные за 1990 г. (см. таблицу своего варианта).

1. Постройте поле корреляции и по виду облака рассеяния сделайте предположение о форме связи между переменными x и y.

2. Рассчитайте характеристики случайных величин:

1) средние значения $\bar{x}$, $\bar{y}$;
2) выборочные дисперсии (вариации) $var(x)$, $var(y)$;
3) стандартные (среднеквадратические) отклонения $S(x)$, $S(y)$;
4) выборочную ковариацию (выборочный корреляционный момент) $cov(x,y)$;
5) выборочный коэффициент корреляции r_{xy}.

3. С помощью этих характеристик оцените степень рассеяния случайных величин вокруг средних значений и тесноту связи переменных.

Вариант 1

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
1	81	124
2	77	131
3	85	146
4	79	139
5	93	143
6	100	159
7	72	135
8	90	152
9	71	127
10	89	154
11	82	127
12	111	162

Вариант 2

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
1	74	122
2	81	134
3	90	136
4	79	125
5	89	120
6	87	127
7	77	125
8	93	148
9	70	122
10	93	157
11	87	144
12	121	165

Вариант 3

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., *x*	Среднедневная заработная плата, руб., *y*
1	77	123
2	85	152
3	79	140
4	93	142
5	89	157
6	81	181
7	79	133
8	97	163
9	73	134
10	95	155
11	84	132
12	108	165

Вариант 4

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., *x*	Среднедневная заработная плата, руб., *y*
1	83	137
2	88	142
3	75	128
4	89	140
5	85	133
6	79	153
7	81	142
8	97	154
9	79	132
10	90	150
11	84	132
12	112	166

Вариант 5

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., *x*	Среднедневная заработная плата, руб., *y*
1	79	134
2	91	154
3	77	128
4	87	138
5	84	133
6	76	144
7	84	160
8	94	149
9	79	125
10	98	163
11	81	120
12	115	162

Вариант 6

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
1	92	147
2	78	133
3	79	128
4	88	152
5	87	138
6	75	122
7	81	145
8	96	141
9	80	127
10	102	151
11	83	129
12	94	147

Вариант 7

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
1	75	133
2	78	125
3	81	129
4	93	153
5	86	140
6	77	135
7	83	141
8	94	152
9	88	133
10	99	156
11	80	124
12	112	156

Вариант 8

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
1	69	124
2	83	133
3	92	146
4	97	153
5	88	138
6	93	159
7	74	145
8	79	152
9	105	168
10	99	154
11	85	127
12	94	155

Вариант 9

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
1	78	133
2	94	139
3	85	141
4	73	127
5	91	154
6	88	142
7	73	122
8	82	135
9	99	142
10	113	168
11	69	124
12	83	130

Вариант 10

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
1	97	161
2	73	131
3	79	135
4	99	147
5	86	139
6	91	151
7	85	135
8	77	132
9	89	161
10	95	159
11	72	120
12	115	160

Контрольная работа 4.2
Парная линейная регрессия

По территориям региона приводятся данные за 1990 г. (см. таблицу своего варианта).

1. Постройте линейную регрессионную модель связи переменных, где y интерпретируется как объясняемая переменная, а x – объясняющая, используя оценки наименьших квадратов.

2. Рассчитайте линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.

3. Оцените статистическую значимость параметров регрессии и корреляции на уровне значимости $\alpha = 0{,}05$ с помощью F-критерия Фишера и t-критерия Стьюдента.

4. Выполните прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума x, составляющем 110% от среднего уровня.

5. Оцените точность прогноза, рассчитав 95% доверительные интервалы для среднего и индивидуального значения объясняемой переменной при том же значении x.

6. Найдите с надежностью 0,95 интервальные оценки параметров уравнения регрессии α и β.

7. На одном графике (графике подбора) постройте исходные данные и теоретическую прямую. Сделайте вывод.

Вариант 1

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
1	81	124
2	77	131
3	85	146
4	79	139
5	93	143
6	100	159
7	72	135
8	90	152
9	71	127
10	89	154
11	82	127
12	111	162

Вариант 2

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
1	74	122
2	81	134
3	90	136
4	79	125
5	89	120
6	87	127
7	77	125
8	93	148
9	70	122
10	93	157
11	87	144
12	121	165

Вариант 3

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
1	77	123
2	85	152
3	79	140
4	93	142
5	89	157
6	81	181

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., *x*	Среднедневная заработная плата, руб., *y*
7	79	133
8	97	163
9	73	134
10	95	155
11	84	132
12	108	165

Вариант 4

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., *x*	Среднедневная заработная плата, руб., *y*
1	83	137
2	88	142
3	75	128
4	89	140
5	85	133
6	79	153
7	81	142
8	97	154
9	79	132
10	90	150
11	84	132
12	112	166

Вариант 5

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., *x*	Среднедневная заработная плата, руб., *y*
1	79	134
2	91	154
3	77	128
4	87	138
5	84	133
6	76	144
7	84	160
8	94	149
9	79	125
10	98	163
11	81	120
12	115	162

Вариант 6

Номер региона	Среднедушевой прожиточный инимум одного трудоспособного в день, руб., *x*	Среднедневная заработная плата, руб., *y*
1	92	147
2	78	133
3	79	128
4	88	152

Номер региона	Среднедушевой прожиточный инимум одного трудоспособного в день, руб., *x*	Среднедневная заработная плата, руб., *y*
5	87	138
6	75	122
7	81	145
8	96	141
9	80	127
10	102	151
11	83	129
12	94	147

Вариант 7

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., *x*	Среднедневная заработная плата, руб., *y*
1	75	133
2	78	125
3	81	129
4	93	153
5	86	140
6	77	135
7	83	141
8	94	152
9	88	133
10	99	156
11	80	124
12	112	156

Вариант 8

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., *x*	Среднедневная заработная плата, руб., *y*
1	69	124
2	83	133
3	92	146
4	97	153
5	88	138
6	93	159
7	74	145
8	79	152
9	105	168
10	99	154
11	85	127
12	94	155

Вариант 9

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., *x*	Среднедневная заработная плата, руб., *y*
1	78	133
2	94	139
3	85	141

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
4	73	127
5	91	154
6	88	142
7	73	122
8	82	135
9	99	142
10	113	168
11	69	124
12	83	130

Вариант 10

Номер региона	Среднедушевой прожиточный минимум одного трудоспособного в день, руб., x	Среднедневная заработная плата, руб., y
1	97	161
2	73	131
3	79	135
4	99	147
5	86	139
6	91	151
7	85	135
8	77	132
9	89	161
10	95	159
11	72	120
12	115	160

Контрольная работа 4.3
Нелинейные регрессионные модели

Исследуется зависимость расходов на приобретение некоторого товара (группы товаров) семейными хозяйствами от располагаемого дохода.

В течение года i-я семья, имеющая располагаемый доход x_i, затратила на приобретение этого товара V_i руб. (см. данные своего варианта).

1. Подберите модель зависимости, в которой эластичность потребления рассматриваемого товара по отношению к располагаемому доходу не зависит от размера располагаемого дохода. Постоянство эластичности предполагает оценивание модели, линейной в логарифмах уровней.

2. Постройте график подбора значений регрессии. Рассчитайте среднюю ошибку аппроксимации. Сделайте выводы.

3. Проверьте значимость подобранной модели на уровне $\alpha = 0{,}05$, используя коэффициент детерминации и критерий Фишера.

4. С помощью графического метода оцените соответствие используемых для построения модели статистических данных стандартным предположениям регрессионного анализа.

5. В рамках подобранной модели проверьте гипотезы о том, что:

1) потребление данного товара эластично по отношению к располагаемому доходу. Эластичное потребление соответствует значению эластичности, большему единицы по абсолютной величине ($|\eta| > |\beta| > 1$);
2) потребление данного товара неэластично по отношению к располагаемому доходу ($|\eta| > |\beta| < 1$).

Вариант 1

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	150537,1	6107,689
2	136570,9	4962,273
3	151518,1	6706,055
4	110318,6	4385,618
5	155144,1	5962,619
6	129398,2	4655,843
7	118036	4511,824
8	153232,6	6433,963
9	174761,2	6873,953
10	158744,2	6800,956
11	151702,4	6545,808
12	143872,3	5928,584
13	166110,4	7188,594
14	164493	6408,18
15	114337,7	4649,526
16	136811,3	5138,914
17	135744,2	5269,74
18	120100,7	4088,58
19	169115,2	6510,397
20	156830,3	6362,19

Вариант 2

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	144354,6	3932,599
2	149545,6	3886,132
3	170410,3	3352,526
4	153039,5	3295,72
5	159254,2	3900,406
6	149145,2	3777,833
7	160664,6	4602,151
8	116919,8	3422,525
9	144054,5	3700,39
10	168181,5	3845,074
11	139014,7	3869,305
12	144764	3561,304
13	158904,1	3595,007
14	136085,9	3385,927
15	126789,5	3124,223
16	142027,4	3055,907

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
17	148287,3	3590,945
18	151869,5	3719,061
19	170769,1	3972,955
20	112560,6	2699,635

Вариант 3

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	150537,1	4513,006
2	136570,9	3666,651
3	151518,1	4955,142
4	110318,6	3240,558
5	155144,1	4405,812
6	129398,2	3440,229
7	118036	3333,812
8	153232,6	4754,091
9	174761,2	5079,202
10	158744,2	5025,265
11	151702,4	4836,735
12	143872,3	4380,664
13	166110,4	5311,692
14	164493	4735,04
15	114337,7	3435,561
16	136811,3	3797,173
17	135744,2	3893,84
18	120100,7	3021,075
19	169115,2	4810,569
20	156830,3	4701,058

Вариант 4

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	144354,6	6322,109
2	149545,6	6336,321
3	170410,3	5759,444
4	153039,5	5423,528
5	159254,2	6521,631
6	149145,2	6153,137
7	160664,6	7722,165
8	116919,8	5057,22
9	144054,5	5943,856
10	168181,5	6570,918
11	139014,7	6127,275
12	144764	5731,7
13	158904,1	6005,703
14	136085,9	5316,342
15	126789,5	4768,541

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
16	142027,4	4880,89
17	148287,3	5835,264
18	151869,5	6101,429
19	170769,1	6831,049
20	112560,6	3928,889

Вариант 5

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	150537,1	6492,436
2	136570,9	5378,59
3	151518,1	7119,241
4	110318,6	4960,9
5	155144,1	6300,129
6	129398,2	5101,198
7	118036	5035,106
8	153232,6	6815,03
9	174761,2	7092,135
10	158744,2	7153,027
11	151702,4	6947,431
12	143872,3	6359,383
13	166110,4	7492,453
14	164493	6692,135
15	114337,7	5221,92
16	136811,3	5568,093
17	135744,2	5718,793
18	120100,7	4546,978
19	169115,2	6761,303
20	156830,3	6707,8

Вариант 6

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	144354,6	6776,966
2	149545,6	6744,377
3	170410,3	5972,289
4	153039,5	5746,198
5	159254,2	6854,843
6	149145,2	6552,91
7	160664,6	8102,415
8	116919,8	5654,494
9	144054,5	6374,151
10	168181,5	6831,718
11	139014,7	6617,816
12	144764	6140,599
13	158904,1	6315,333
14	136085,9	5766,466

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
15	126789,5	5246
16	142027,4	5249,091
17	148287,3	6221,558
18	151869,5	6474,361
19	170769,1	7080,518
20	112560,6	4426,413

Вариант 7

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	150537,1	3304,019
2	136570,9	2818,311
3	151518,1	3615,949
4	110318,6	2771,366
5	155144,1	3177,29
6	129398,2	2716,574
7	118036	2756,336
8	153232,6	3449,771
9	174761,2	3451,21
10	158744,2	3582,684
11	151702,4	3527,398
12	143872,3	3280,574
13	166110,4	3701,97
14	164493	3316,259
15	114337,7	2886,033
16	136811,3	2916,069
17	135744,2	3002,036
18	120100,7	2476,208
19	169115,2	3322,795
20	156830,3	3371,934

Вариант 8

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	144354,6	3492,481
2	149545,6	3439,044
3	170410,3	2928,33
4	153039,5	2909,829
5	159254,2	3430,032
6	149145,2	3344,101
7	160664,6	4043,583
8	116919,8	3104,24
9	144054,5	3286,944
10	168181,5	3362,981
11	139014,7	3449,247
12	144764	3161,844
13	158904,1	3162,159

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
14	136085,9	3024,78
15	126789,5	2810,808
16	142027,4	2718,319
17	148287,3	3180,504
18	151869,5	3286,122
19	170769,1	3469,526
20	112560,6	2457,899

Вариант 9

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	150537,1	3453,137
2	136570,9	2778,363
3	151518,1	3793,902
4	110318,6	2403,633
5	155144,1	3381,295
6	129398,2	2592,768
7	118036	2489,58
8	153232,6	3644,066
9	174761,2	3944,787
10	158744,2	3865,559
11	151702,4	3703,694
12	143872,3	3336,731
13	166110,4	4104,461
14	164493	3655,291
15	114337,7	2557,409
16	136811,3	2877,77
17	135744,2	2948,722
18	120100,7	2259,954
19	169115,2	3723,902
20	156830,3	3611,787

Вариант 10

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
1	144354,6	3492,481
2	149545,6	3439,044
3	170410,3	2928,33
4	153039,5	2909,829
5	159254,2	3430,032
6	149145,2	3344,101
7	160664,6	4043,583
8	116919,8	3104,24
9	144054,5	3286,944
10	168181,5	3362,981
11	139014,7	3449,247
12	144764	3161,844

Номер наблюдения, i	Располагаемый доход семейного хозяйства, x (руб.)	Расходы семейного хозяйства на приобретение некоторого товара, V (руб.)
13	158904,1	3162,159
14	136085,9	3024,78
15	126789,5	2810,808
16	142027,4	2718,319
17	148287,3	3180,504
18	151869,5	3286,122
19	170769,1	3469,526
20	112560,6	2457,899

Контрольная работа 4.4
Множественная линейная регрессия и корреляция

По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%) (см. таблицу своего варианта).

1. Постройте линейную модель множественной регрессии. Запишите стандартизованное уравнение множественной регрессии. На основе стандартизованных коэффициентов регрессии и средних коэффициентов эластичности ранжируйте факторы по степени их влияния на результат.

2. Найдите коэффициенты парной, частной и множественной корреляции. Проанализируйте их.

3. Найдите скорректированный коэффициент множественной детерминации. Сравните его с нескорректированным (общим) коэффициентом детерминации.

4. С помощью F-критерия Фишера оцените статистическую надежность уравнения регрессии и коэффициента детерминации $R^2_{yx_1x_2}$.

5. С помощью частных F-критериев Фишера оцените целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1.

6. Составьте уравнение линейной парной регрессии, оставив лишь один значащий фактор.

Вариант 1

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,6	9	11	9	6,3	21
2	6	3,6	12	12	11	6,4	22
3	6	3,9	14	13	11	7	24
4	7	4,1	17	14	12	7,5	25
5	7	3,9	18	15	12	7,9	28
6	7	4,5	19	16	13	8,2	30
7	8	5,3	19	17	13	8	30
8	8	5,3	19	18	13	8,6	31
9	9	5,6	20	19	14	9,5	33
10	10	6,8	21	20	14	9	36

Вариант 2

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,5	10	11	10	6,3	21
2	6	3,6	12	12	11	6,4	22
3	7	3,9	15	13	11	7	23
4	7	4,1	17	14	12	7,5	25
5	7	4,2	18	15	12	7,9	28
6	8	4,5	19	16	13	8,2	30
7	8	5,3	19	17	13	8,4	31
8	9	5,3	20	18	14	8,6	31
9	9	5,6	20	19	14	9,5	35
10	10	6	21	20	15	10	36

Вариант 3

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,7	9	11	11	6,3	22
2	7	3,7	11	12	11	6,4	22
3	7	3,9	11	13	11	7,2	23
4	7	4,1	15	14	12	7,5	25
5	8	4,2	17	15	12	7,9	27
6	8	4,9	19	16	13	8,1	30
7	8	5,3	19	17	13	8,4	31
8	9	5,1	20	18	13	8,6	32
9	10	5,6	20	19	14	9,5	35
10	10	6,1	21	20	15	9,5	36

Вариант 4

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,5	9	11	10	6,3	22
2	7	3,6	10	12	10	6,5	22
3	7	3,9	12	13	11	7,2	24
4	7	4,1	17	14	12	7,5	25
5	8	4,2	18	15	12	7,9	27
6	8	4,5	19	16	13	8,2	30
7	9	5,3	19	17	13	8,4	31
8	9	5,5	20	18	14	8,6	33
9	10	5,6	21	19	14	9,5	35
10	10	6,1	21	20	15	9,6	36

Вариант 5

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,6	9	11	10	6,3	21
2	7	3,6	11	12	11	6,9	23
3	7	3,7	12	13	11	7,2	24
4	8	4,1	16	14	12	7,8	25
5	8	4,3	19	15	13	8,1	27
6	8	4,5	19	16	13	8,2	29
7	9	5,4	20	17	13	8,4	31

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
8	9	5,5	20	18	14	8,8	33
9	10	5,8	21	19	14	9,5	35
10	10	6,1	21	20	14	9,7	34

Вариант 6

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,5	9	11	10	6,3	21
2	7	3,6	10	12	10	6,8	22
3	7	3,8	14	13	11	7,2	24
4	7	4,2	15	14	12	7,9	25
5	8	4,3	18	15	12	8,1	26
6	8	4,7	19	16	13	8,3	29
7	9	5,4	19	17	13	8,4	31
8	9	5,6	20	18	13	8,8	32
9	10	5,9	20	19	14	9,6	35
10	10	6,1	21	20	14	9,7	36

Вариант 7

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,8	11	11	10	6,8	21
2	7	3,8	12	12	11	7,4	23
3	7	3,9	16	13	11	7,8	24
4	7	4,1	17	14	12	7,5	26
5	7	4,6	18	15	12	7,9	28
6	8	4,5	18	16	12	8,1	30
7	8	5,3	19	17	13	8,4	31
8	9	5,5	20	18	13	8,7	32
9	9	6,1	20	19	13	9,5	33
10	10	6,8	21	20	14	9,7	35

Вариант 8

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,8	9	11	11	7,1	22
2	7	4,1	14	12	11	7,5	23
3	7	4,3	16	13	12	7,8	25
4	7	4,1	17	14	12	7,6	27
5	8	4,6	17	15	12	7,9	29
6	8	4,7	18	16	13	8,1	30
7	9	5,3	20	17	13	8,5	32
8	9	5,5	20	18	14	8,7	32
9	11	6,9	21	19	14	9,6	33
10	10	6,8	21	20	15	9,8	36

Вариант 9

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,9	12	11	11	7,1	22
2	7	4,2	13	12	12	7,5	25
3	7	4,3	15	13	13	7,8	26
4	7	4,4	17	14	12	7,9	27
5	8	4,6	18	15	13	8,1	30
6	8	4,8	19	16	13	8,4	31
7	9	5,3	19	17	13	8,6	32
8	9	5,7	20	18	14	8,8	32
9	10	6,9	21	19	14	9,6	34
10	10	6,8	21	20	14	9,9	36

Вариант 10

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,6	12	11	10	7,2	23
2	7	4,1	14	12	11	7,6	25
3	7	4,3	16	13	12	7,8	26
4	7	4,4	17	14	11	7,9	28
5	7	4,5	18	15	12	8,2	30
6	8	4,8	19	16	12	8,4	31
7	8	5,3	20	17	12	8,6	32
8	8	5,6	20	18	13	8,8	32
9	9	6,7	21	19	13	9,2	33
10	10	6,9	22	20	14	9,6	34

Контрольная работа 4.5
Временные ряды

Имеются условные данные по объемам потребления электроэнергии (y_t) жителями региона за 16 кварталов.

Требуется:

1. Построить автокорреляционную функцию и сделать вывод о наличии сезонных колебаний.

2. Построить аддитивную модель временного ряда (для нечетных вариантов) или мультипликативную модель временного ряда (для четных вариантов).

3. Сделать прогноз на два квартала вперед.

Вариант 1

t	y_t	t	y_t
1	5,8	9	7,9
2	4,5	10	5,5
3	5,1	11	6,3
4	9,1	12	10,8
5	7,0	13	9,0
6	5,0	14	6,5
7	6,0	15	7,0
8	10,1	16	11,1

Вариант 2

t	y_t	t	y_t
1	5,5	9	8,0
2	4,6	10	5,6
3	5,0	11	6,4
4	9,2	12	10,9
5	7,1	13	9,1
6	5,1	14	6,4
7	5,9	15	7,2
8	10,0	16	11,0

Вариант 3

t	y_t	t	y_t
1	5,3	9	8,2
2	4,7	10	5,5
3	5,2	11	6,5
4	9,1	12	11,0
5	7,0	13	8,9
6	5,0	14	6,5
7	6,0	15	7,3
8	10,1	16	11,2

Вариант 4

t	y_t	t	y_t
1	5,5	9	8,3
2	4,8	10	5,4
3	5,1	11	6,4
4	9,0	12	10,9
5	7,1	13	9,0
6	4,9	14	6,6
7	6,1	15	7,5
8	10,0	16	11,2

Вариант 5

t	y_t	t	y_t
1	5,6	9	8,2
2	4,7	10	5,6
3	5,2	11	6,4
4	9,1	12	10,8
5	7,0	13	9,1
6	5,1	14	6,7
7	6,0	15	7,5
8	10,2	16	11,3

Вариант 6

t	y_t	t	y_t
1	5,8	9	7,8
2	4,6	10	5,5
3	5,1	11	6,3
4	9,1	12	10,8

t	y_t	t	y_t
5	7,0	13	9,0
6	5,0	14	6,5
7	6,0	15	7,0
8	10,1	16	11,1

Вариант 7

t	y_t	t	y_t
1	5,4	9	8,0
2	4,6	10	5,6
3	5,0	11	6,4
4	9,2	12	10,9
5	7,1	13	9,1
6	5,1	14	6,4
7	5,9	15	7,2
8	10,0	16	11,0

Вариант 8

t	y_t	t	y_t
1	5,3	9	8,2
2	4,8	10	5,5
3	5,2	11	6,5
4	9,1	12	11,0
5	7,0	13	8,9
6	5,0	14	6,5
7	6,0	15	7,3
8	10,1	16	11,2

Вариант 9

t	y_t	t	y_t
1	5,5	9	8,3
2	4,8	10	5,4
3	5,2	11	6,4
4	9,0	12	10,9
5	7,1	13	9,0
6	4,9	14	6,6
7	6,1	15	7,5
8	10,0	16	11,2

Вариант 10

t	y_t	t	y_t
1	5,6	9	8,2
2	4,7	10	5,6
3	5,2	11	6,4
4	9,2	12	10,8
5	7,0	13	9,1
6	5,1	14	6,7
7	6,0	15	7,5
8	10,2	16	11,3

Контрольная работа 4.6
Системы эконометрических уравнений

Даны системы эконометрических уравнений.

Требуется:

1) применив необходимое и достаточное условие идентификации, определить, идентифицируемо ли каждое из уравнений модели;
2) определить метод оценки параметров модели;
3) записать в общем виде приведенную форму модели.

Вариант 1

Модель протекционизма Сальватора (упрощенная версия):

$$\begin{cases} M_t = a_1 + b_{12}N_t + b_{13}S_t + b_{14}E_{t-1} + b_{15}M_{t-1} + \varepsilon_1, \\ N_t = a_2 + b_{21}M_t + b_{23}S_t + b_{26}Y_t + \varepsilon_2, \\ S_t = a_3 + b_{31}M_t + b_{32}N_t + b_{36}X_t + \varepsilon_3. \end{cases}$$

где M – доля импорта в ВВП; N – общее число прошений об освобождении от таможенных пошлин; S – число удовлетворенных прошений об освобождении от таможенных пошлин; E – фиктивная переменная, равная 1, для тех лет, в которые курс доллара на международных валютных рынках был искусственно завышен, и 0 – для всех остальных лет; Y – реальный ВВП; X – реальный объем чистого экспорта; t – текущий период; $t - 1$ – предыдущий период.

Вариант 2

Макроэкономическая модель (упрощенная версия модели Клейна):

$$\begin{cases} C_t = a_1 + b_{12}Y_t + b_{13}T_t + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + b_{24}K_{t-1} + \varepsilon_2, \\ Y_t = C_t + I_t, \end{cases}$$

где C – потребление; I – инвестиции; Y – доход; T – налоги; K – запас капитала; t – текущий период; $t - 1$ – предыдущий период.

Вариант 3

Макроэкономическая модель экономики США (одна из версий):

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + b_{23}r_t + \varepsilon_2, \\ r_t = a_3 + b_{31}Y_t + b_{34}M_t + b_{35}r_{t-1} + \varepsilon_3, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C – потребление; Y – ВВП; I – инвестиции; r – процентная ставка; M – денежная масса; G – государственные расходы; t – текущий период; $t - 1$ – предыдущий период.

Вариант 4

Модель Кейнса (одна из версий):

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}Y_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + \varepsilon_2, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C – потребление; Y – ВВП; I – валовые инвестиции; G – государственные расходы; t – текущий период; $t-1$ – предыдущий период.

Вариант 5

Модель денежного и товарного рынков:

$$\begin{cases} R_t = a_1 + b_{12}Y_t + b_{14}M_t + \varepsilon_1, \\ Y_t = a_2 + b_{21}R_t + b_{23}I_t + b_{25}G_t + \varepsilon_2, \\ I_t = a_3 + b_{31}R_t + \varepsilon_3, \end{cases}$$

где R – процентные ставки; Y – реальный ВВП; M – денежная масса; I – внутренние инвестиции; G – реальные государственные расходы.

Вариант 6

Модифицированная модель Кейнса:

$$\begin{cases} C_t = a_1 + b_{11}Y_t + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + b_{22}Y_{t-1} + \varepsilon_2, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C – потребление; Y – доход; I – инвестиции; G – государственные расходы; t – текущий период; $t-1$ – предыдущий период.

Вариант 7

Макроэкономическая модель:

$$\begin{cases} C_t = a_1 + b_{11}D_t + \varepsilon_1, \\ I_t = a_2 + b_{22}Y_t + b_{23}Y_{t-1} + \varepsilon_2, \\ Y_t = D_t + T_t, \\ D_t = C_t + I_t + G_t, \end{cases}$$

где C – расходы на потребление; Y – чистый национальный продукт; D – чистый национальный доход; I – инвестиции; T – косвенные налоги; G – государственные расходы; t – текущий период; $t-1$ – предыдущий период.

Вариант 8

Гипотетическая модель экономики:

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}J_t + \varepsilon_1, \\ J_t = a_2 + b_{21}Y_{t-1} + \varepsilon_2, \\ T_t = a_3 + b_{31}Y_t + \varepsilon_3, \\ Y_t = C_t + J_t + G_t, \end{cases}$$

где C – совокупное потребление в период t; Y – совокупный доход в период t; J – инвестиции в период t; T – налоги в период t; G – государственные доходы в период t.

Вариант 9

Модель денежного рынка:

$$\begin{cases} R_t = a_1 + b_{11}M_t + b_{12}Y_t + +\varepsilon_1, \\ Y_t = a_2 + b_{21}R_t + b_{22}I_t + \varepsilon_2, \\ I_t = a_3 + b_{33}R_t + \varepsilon_3, \end{cases}$$

где R – процентные ставки; Y – ВВП; M – денежная масса; I – внутренние инвестиции.

Вариант 10

Конъюнктурная модель имеет вид:

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21}r_t + b_{22}I_{t-1} + \varepsilon_2, \\ r_t = a_3 + b_{31}Y_t + b_{32}M_t + \varepsilon_3, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C – расходы на потребление; Y – ВВП; I – инвестиции; r – процентная ставка; M – денежная масса; G – государственные расходы; t – текущий период; $t - 1$ – предыдущий период.

Библиографический список

Основная литература

1. Домбровский, В.В. Эконометрика : учеб. / В.В. Домбровский. – М. : Нов. учеб., 2004. – 342 с.
2. Эконометрика : практикум для студентов экономических специальностей / П.Ф. Зибров [и др.] ; под ред. Ю.К. Черновой. – Тольятти : ТГУ, 2008. – 69 с.
3. Катышев, П.К. Сборник задач к начальному курсу эконометрики / П.К. Катышев, Я.Р. Магнус, А.А. Пересецкий. – М. : Дело, 2002. – 208 с.
4. Кремер, Н.Ш. Эконометрика : учеб. для вузов / Н.Ш. Кремер, Б.А. Путко ; под ред. Н.Ш. Кремера. – М. : ЮНИТИ-ДАНА, 2002. – 311 с.
5. Магнус, Я.Р. Эконометрика. Начальный курс : учеб. / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий. – М. : Дело, 2001. – 400 с.

Дополнительная литература

6. Айвазян, С.А. Прикладная статистика. Основы эконометрики : учеб. для вузов : в 2 т. / С.А. Айвазян, В.С. Мхитарян // Теория вероятностей и прикладная статистика. – М. : ЮНИТИ-ДАНА, 2001. – Т. 1. – 656 с.
7. Айвазян, С.А. Прикладная статистика. Основы эконометрики : учеб. для вузов : в 2 т. / С.А. Айвазян // Основы эконометрики. – М : ЮНИТИ-ДАНА, 2001. – Т. 2. – 432 с.
8. Дорохина, Е.Ю. Сборник задач по эконометрике : учеб. пособие для студентов экономических вузов / Е.Ю. Дорохина, Л.Ф. Преснякова, Н.П. Тихомиров. – М. : Экзамен, 2003. – 224 с.
9. Доугерти, К. Введение в эконометрику : [пер. с англ.] / К. Доугерти. – М. : ИНФРА-М, 1999. – 402 с.
10. Кулинич, Е.И. Эконометрия / Е.И. Кулинич. – М. : Финансы и статистика, 2001. – 304 с.
11. Носко, В.П. Эконометрика. Элементарные методы и введение в регрессионный анализ временных рядов / В.П. Носко. – М. : Институт экономики переходного периода, 2004. – 501 с.
12. Практикум по эконометрике : учеб. пособие / И.И. Елисеева [и др.] ; под ред. И.И. Елисеевой. – М. : Финансы и статистика, 2003. – 192 с.
13. Тихомиров, Н.П. Эконометрика : учеб. / Н.П. Тихомиров, Е.Ю. Дорохина. – М. : Экзамен, 2003. – 512 с.
14. Шалабанов, А.К. Эконометрика : учеб.-метод. пособие / А.К. Шалабанов, Д.А. Роганов. – Казань : ТИСБИ, 2002. – 56 с.
15. Эконометрика : учеб. / И.И. Елисеева [и др.] ; под ред. И.И. Елисеевой. – М. : Финансы и статистика, 2002. – 344 с.

Интернет-ресурсы

1. Ресурсы по статистике и эконометрике [Электронный ресурс] / сост. Сергей Морулис-Якушев и Петр Савельев ; Европейский университет в Санкт-Петербурге, дистанционное обучение. – Электрон. дан. – СПб. : ЕУСПб, 2006. – Режим доступа: http://dist-economics.eu.spb.ru/HTML/predmet/econometrics.htm, свободный. – Загл. с экрана.
2. Орлов, А.И. Эконометрика : учеб. / А.И. Орлов. – М. : Экзамен, 2002 [Электронный ресурс] // Административно-управленческий портал, частный авторский интернет-проект Алексея Катаева. – Электрон. дан. – AUP.Ru, 1999–2011. Режим доступа: http://www.aup.ru/books/m153/, свободный. – Загл. с экрана.

3. Бородич, С.А. / Эконометрика : учеб. пособие / С.А. Бородич. – Минск : Новое знание, 2001. – 408 с. [Электронный ресурс] // Электронная библиотека экономического факультета БГУ. – Электрон. дан. – PDF, полная версия 408 с., 10078 Kb, 11/10/2004, ISBN 985-6516-45-5. – БГУ, 2003. – Режим доступа: hhttp://www.economy.bsu.by/library/%DD%EA%EE%ED%EE%EC%E5%F2%F0%E8%EA%E0.htm, свободный. – Загл. с экрана.
4. Эконометрическая страничка [Электронный ресурс] / Содержит ссылки на разнообразные ресурсы по эконометрике, находящиеся в свободном доступе ; автор Александр Цыплаков. – Электрон. дан. – Интернет-страница, посвященная эконометрике; 12,98 КБ. – Режим доступа: http://www.nsu.ru/ef/tsy/ecmr/index.htm, свободный. – Загл. с экрана.

Глоссарий

№ п/п	Понятие	Содержание
1	Эконометрика	совокупность методов анализа количественных связей между экономическими факторами и показателями на основании реальных статистических данных с использованием аппарата теории вероятностей и математической статистики
2	Обобщенный вид эконометрической модели	$y = f(x_1, x_2, ..., x_n) + \varepsilon$, где y – наблюдаемое значение зависимой переменной (объясняемая переменная); x_1, x_2, ..., x_n – объясняющие переменные; $f(x_1, x_2, ..., x_n)$ – объясненная часть, зависящая от значений объясняющих переменных; ε – случайная составляющая
3	Среднее значение дискретной величины	наиболее простой показатель, характеризующий последовательность $x_1, x_2, ..., x_n$ $(y_1, y_2, .., y_n)$, вычисляется по формуле $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\left(\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\right)$
4	Математическое ожидание дискретной случайной величины	сумма произведений всех значений дискретной величины на их вероятности, оно приближенно равно ее среднему значению: $M(X) = x_1 p_1 + x_2 p_2 + ... + x_n p_n = \sum_{i=1}^{n} x_i p_i$; $\left(M(X) = y_1 p_1 + y_2 p_2 + ... + y_i p_n = \sum_{i=1}^{n} y_i p_i\right)$
5	Выборочная дисперсия (вариации)	характеризуют степень разброса значений $x_1, x_2, ..., x_n$ $(y_1, y_2, ..., y_n)$ вокруг своего среднего значения $\bar{x}$ ($\bar{y}$): $\mathrm{var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$; $\left(\mathrm{var}(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2\right)$
6	Стандартное отклонение (среднеквадратическое отклонение)	характеризует рассеяние дискретной случайной величины вокруг своего среднего значения, более удобно, чем дисперсия, так как измеряется в тех же единицах, что и сама величина: $S(x) = \sqrt{\mathrm{var}(x)}$; $\left(S(x) = \sqrt{\mathrm{var}(x)}\right)$
7	Диаграмма рассеяния (поле корреляции)	удобное графическое средство анализа данных, на нем в прямоугольной системе координат располагаются точки (x_i, y_i), $i = 1, 2, ..., n$
8	Выборочный коэффициент корреляции	характеризует степень выраженности линейной связи между произвольными переменными x и y, принимающими значения x_i, y_i, $i = 1, 2, ..., n$, рассчитывается по формуле $r_{xy} = \frac{\mathrm{cov}(x, y)}{S(x)S(y)} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{S_x S_y}$
9	Выборочная ковариация (выборочный корреляционный момент)	характеризует степень зависимости двух случайных величин и степень их рассеяния: $\mathrm{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \overline{xy} - \bar{x}\cdot\bar{y}$
10	Статистическая зависимость	Зависимость, при которой изменение одной из величин влечет изменение распределения другой
11	Корреляционная зависимость	статистическая зависимость, проявляющаяся в том, что при изменении одной из величин изменяется среднее значение (условное математическое ожидание) другой

№ п/п	Понятие	Содержание
12	Регрессионная модель	модель, где зависимая (объясняемая) переменная у представляется в виде функции $y = M_y(x_1, ..., x_p) + \varepsilon$, где $M_y(x_1, ..., x_p)$ – условное математическое ожидание величины у, полученное при данном наборе значений объясняющих переменных, или так называемая функция регрессии, ε – случайная составляющая
13	Системы одновременных уравнений	модели, описывающиеся системами уравнений. Могут состоять из тождеств и регрессионных уравнений, каждое из которых может, кроме объясняющих переменных, включать в себя также объясняемые переменные из других уравнений системы
14	Эндогенные переменные	взаимосвязанные переменные, описывающие экономический объект и формирующиеся внутри функционирования объекта
15	Экзогенные переменные	переменные, описывающие экономический объект и задаваемые извне
16	Лаговые переменные	взятые в предыдущий момент времени переменные
17	Стандартные предположения регрессионного анализа (предположения о процессе порождения данных)	Предположения, без соблюдения которых построенная регрессионная модель не имеет смысла: 1. В модели не все значения $x_1, x_2, ..., x_n$ совпадают между собой (условие идентифицируемости). 2. Значения функции регрессии $M_y(x_1, ..., x_p)$ рассматриваются как некоторые постоянные, хотя и неизвестные наблюдателю. А значения $y_1, y_2,..., y_n$ носят случайный характер, определяемый случайным характером ε_i. Это условие предполагает отсутствие автокорреляционной зависимости остатков от номера наблюдения. 3. Ошибки $\varepsilon_1, \varepsilon_2,..., \varepsilon_n$ и $y_1, y_2,..., y_n$ – независимые случайные величины, имеющие нормальное распределение. Математическое ожидание ошибок $M\varepsilon_i = 0$. 4. Дисперсия ошибок ε_i (соответственно и величины y_i) постоянна для любого i: $\mathrm{var}(\varepsilon_i) = \delta^2(\varepsilon_i) = \mathrm{const}$; $\mathrm{var}(y_i) = \delta^2(y_i) = \mathrm{const}$
18	Условие равноизменчивости (гомоскедастичности) ошибок ε_i и, соответственно, объясняемой переменной y_i	$\mathrm{var}(\varepsilon_i) = \delta^2(\varepsilon_i) = \mathrm{const}$; $\mathrm{var}(y_i) = \delta^2(y_i) = \mathrm{const}$
19	Условие гетероскедастичности ошибок ε_i и y_i	$\mathrm{var}(\varepsilon_i) = \delta^2(\varepsilon_i) \neq \mathrm{const}$; $\mathrm{var}(y_i) = \delta^2(y_i) \neq \mathrm{const}$
20	Принцип наименьших квадратов	поиск коэффициентов (параметров) уравнения регрессии осуществляется таким образом, чтобы величина остатка регрессии ε_i стремилась к минимуму (в идеале к нулю): $\sum_{i=1}^{n}\varepsilon_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \to \min$. Если $\varepsilon_i = 0$, то все наблюдаемые значения (x_i, y_i) лежат на подобранной линии регрессии. В результате получают теоретическую модель регрессии
21	Метод наименьших квадратов (МНК)	коэффициенты (параметры) уравнения регрессии вычисляются по формулам, обеспечивающим принцип наименьших квадратов ошибок регрессии
22	Модель парной линейной регрессии	$y = (\alpha + \beta x) + \varepsilon$, где y – наблюдаемое значение зависимой переменной (объясняемая переменная), x – объясняющая переменная; $\hat{y} = \alpha + \beta x$ – объясненная часть, зависящая от значения объясняющих переменных, α и β – параметры модели (рассчитываются МНК), ε – случайная составляющая

№ п/п	Понятие	Содержание
23	Несмещенность МНК-оценки параметров уравнения регрессии	статистическая оценка некоторого параметра называется несмещенной, если ее математическое ожидание равно истинному значению этого параметра
24	Состоятельность МНК-оценки параметров уравнения регрессии	при неограниченном возрастании объема выборки значение оценки должно стремиться по вероятности к истинному значению параметра, а дисперсии оценок параметров должны уменьшаться и в пределе стремиться к нулю
25	Эффективность МНК-оценки параметров уравнения регрессии	оценка называется эффективной, если она имеет минимальную дисперсию по сравнению с другими оценками заданного класса
26	Доверительный интервал для прогноза среднего значения y, полученного по функции регрессии $M_{\hat{y}}(x)$	$\hat{y}-t_{1-\alpha;k}S_{\hat{y}} \le M_{\hat{y}}(x) \le \hat{y}+t_{1-\alpha;k}S_{\hat{y}}$, где $S_{\hat{y}}^2 = S_{\varepsilon_i}^2\left(\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$ – оценка дисперсии значения $\hat{y}$, найденного по уравнению регрессии, $t=\frac{\hat{y}-M_{\hat{y}}(x)}{S_{\hat{y}}}$ – статистика, имеющая t – распределение Стьюдента с $k = n-2$ степенями свободы; n – объем выборки
27	Доверительный интервал для прогноза индивидуальных значений y_0 по функции регрессии	$\hat{y}_0-t_{1-\alpha;n-2}S_{\hat{y}_0} \le y_0 \le \hat{y}_0+t_{1-\alpha;n-2}S_{\hat{y}_0}$, где $S_{\hat{y}_0}^2 = S_{\varepsilon_i}^2\left[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]$ – оценка дисперсии значения $\hat{y}_0$, найденного по уравнению регрессии, $t=\frac{\hat{y}_0-y_0}{S_{\hat{y}_0}}$ – статистика, имеющая t – распределение Стьюдента с $k = n-2$ степенями свободы; n – объем выборки
28	Доверительные интервалы для параметров регрессионной модели	при построении доверительного интервала параметра β исходят из того, что статистика $t=\frac{b-\beta}{S}\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ имеет t – распределение Стьюдента с $k = n - 2$ степенями свободы. Интервальная оценка параметра β на уровне значимости α имеет вид: $b-t_{1-\alpha;n-2}\frac{S_{\varepsilon_i}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \le \beta \le b+t_{1-\alpha;n-2}\frac{S_{\varepsilon_i}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$. При построении доверительного интервала для параметра δ^2 исходят из того, что статистика $\frac{nS^2}{\delta^2}$ имеет χ^2 – распределение с $k = n - 2$ степенями свободы. Поэтому интервальная оценка для δ^2 на уровне значимости α имеет вид: $\frac{nS_{\varepsilon_i}^2}{\chi^2_{\frac{\alpha}{2};n-2}} \le \delta^2 \le \frac{nS_{\varepsilon_i}^2}{\chi^2_{\frac{1-\alpha}{2};n-2}}$. Доверительный интервал выбирается таким образом, чтобы вероятность $P\left(\chi^2 \langle \chi^2_{\frac{1-\alpha}{2};n-2}\right)=P\left(\chi^2 > \chi^2_{\frac{1-\alpha}{2};n-2}\right)=\frac{\alpha}{2}$

№ п/п	Понятие	Содержание
29	Средняя ошибка аппроксимации	чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации: $\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left\|\frac{y_i - \hat{y}_i}{y_i}\right\| \cdot 100\% = \sum_{i=1}^{n}\left\|\frac{e_i}{y_i}\right\|$ (не должна превышать 8–10%)
30	Основная идея дисперсионного анализа	согласно основной идее дисперсионного анализа, общая сумма квадратов отклонений переменной y от среднего значения $\overline{y}$ раскладывается на две части – «объясненную моделью» (S_R^2) и «необъясненную» ($S_{\varepsilon_i}^2$). Случайные величины $S_R^2 = \frac{Q_R}{m-1}$ и $S_{\varepsilon_i}^2 = \frac{Q_e}{n-m}$ имеют χ^2 – распределение соответственно с (m-1) и (n-m) степенями свободы
31	F-критерий Фишера	оценка значимости уравнения регрессии в целом производится на основе F-критерия Фишера, которому предшествует дисперсионный анализ. Уравнение регрессии значимо на уровне α, если фактически наблюдаемое значение статистики $F = \frac{Q_R(n-m)}{Q_\ell(m-1)} = \frac{S_R^2}{S_{\varepsilon_i}^2} > F_{\alpha;k_1;k_2}$, где $Q_R = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$ – сумма квадратов, объясненная моделью; $Q_\ell = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}e_i^2$ – остаточная сумма квадратов; m – число оцениваемых параметров уравнения регрессии; n – число наблюдений; $F_{\alpha;k_1;k_2}$ – табличное значение критерия. Если известен коэффициент детерминации R_2, то критерий значимости уравнения регрессии $F = \frac{R^2(n-m)}{(1-R^2)(m-1)} > F_{\alpha;k_1;k_2}$
32	Коэффициент детерминации	характеризует степень выраженности регрессионной связи между переменными. Определяется по формуле $R^2 = \frac{Q_R}{Q} = 1 - \frac{Q_\ell}{Q}$, где $Q_R = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$ - сумма квадратов, объясненная моделью; $Q_\ell = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}e_i^2$ – остаточная сумма квадратов
33	Дисперсионный анализ	оцененная регрессионная модель проверяется на отсутствие автокорреляционной зависимости остатков от номера наблюдения, на независимость случайных ошибок $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$, математическое ожидание которых должно стремиться к нулю ($M\varepsilon_i = 0$), на постоянство или гомоскедастичность дисперсии ошибок [var(ε_i) = $\delta^2(\varepsilon_i)$ = const]
34	Графики стандартизированных остатков	графики зависимости стандартизированных остатков $C_i = \frac{y_i - \hat{y}_i}{S_\varepsilon} = \frac{e_i}{S_{\varepsilon_i}}$ ($i = 1, 2, \ldots, n$) как ординат: – от оцененных значений y (по оси абсцисс); – отдельных объясняющих переменных; – номера наблюдения, если наблюдения производятся в последовательные моменты времени равными интервалами. Позволяют выявлять типичные отклонения от стандартных предположений о модели наблюдений по характеру поведения остатков

№ п/п	Понятие	Содержание
35	Модель парной нелинейной регрессии	$y = f(x) + \varepsilon$, где y — наблюдаемое значение зависимой переменной (объясняемая переменная), x — объясняющая переменная; $\hat{y}$ = f(x) — объясненная часть, зависящая от значения объясняющих переменных, представляет нелинейную функцию, ε — случайная составляющая
36	Линеаризация уравнения нелинейной регрессии	приведение нелинейного уравнения регрессии к линейной форме путем логарифмирования или замены переменной с целью оценки параметров модели
37	Степенная модель парной регрессии с мультипликативными ошибками	модель $y = \alpha \cdot x^{\beta} \cdot \varepsilon$, линеаризуется путем логарифмирования: $\ln y = \ln\alpha + \beta\ln x + \ln\varepsilon$
38	Предельная склонность к потреблению (норма потребления)	величина $C(x)$, которая для заданной величины располагаемого дохода x и затрат Vна некоторый товар определяется формулой $C(x) = \lim\limits_{\Delta x \to 0} \frac{V(x+\Delta x) - V(x)}{\Delta x} = \frac{dV}{dx} = V'(x)$
39	Предельная склонность величины y по отношению к величине x	если имеется связь между какими-то переменными экономическими факторами x и y в виде $Y = f(x)$, то функция $C(x) = \frac{dy}{dx} = f'(x) = \lim\limits_{\Delta x \to 0} \frac{y(x+\Delta x) - y(x)}{\Delta x}$ является предельной склонностью величины y по отношению к величине x
40	Эластичность (функция эластичности)	величина $\eta(x) = \lim\limits_{\Delta x \to 0} \frac{\frac{f(x+\Delta x) - f(x)}{f(x)} \cdot 100\%}{\frac{\Delta x}{x} \cdot 100\%} == \frac{x}{y} \cdot \frac{dy}{dx}$, $\eta(x) = \frac{x}{y} C(x)$
41	Условие постоянства предельной склонности	$C(x) = \beta$ означает линейную связь между уровнями факторов x и y: $y = \alpha + \beta x$
42	Условие постоянства эластичности	$\eta(x) = \beta$ означает линейную связь между логарифмами уровней $\ln y = \alpha' + \beta \ln x$
43	Модель парной регрессии (обратно пропорциональной зависимости)	$y = \alpha + \frac{\beta}{z}$, приводится к линейной заменой переменной $x = \frac{1}{z}$
44	Модель парной регрессии с убывающей эластичностью	$y = \alpha + \beta \ln z$ или $\ln y = \alpha - \frac{\beta}{z}$, сводятся к линейной форме путем перехода от уровней переменных к их логарифмам или обратным величинам
45	Степенная модель парной регрессии с аддитивными возмущениями	$y = \alpha \cdot x^{\beta} + \varepsilon$, то есть имеет аддитивные ошибки. Не преобразуется к линейной модели наблюдений
46	Итерационные методы подбора нелинейных моделей	методы, в процессе реализации которых сначала задаются некоторые «стартовые значения» оцениваемых параметров, а затем производится последовательное приближение значений оценок параметров к реальным величинам, минимизирующим квадраты ошибок $\sum\limits_{i=1}^{n} \varepsilon_i^2 = \sum\limits_{i=1}^{n} (y_i - \hat{y}_i)^2 \to \min$
47	Модели множественной регрессии	регрессионные модели, имеющие более двух переменных

№ п/п	Понятие	Содержание
48	Модель множественной линейной регрессии	$\hat{y}_x = a + b_1x_1 + b_2x_2 + \ldots + b_mx_m$
49	Степенная модель множественной регрессии с мультипликативными возмущениями	$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \ldots \cdot x_n^{b_n} \cdot \varepsilon$, линеаризуется с помощью логарифмирования: $\ln y = \ln a + b_1 \ln x_1 + b_2 \ln x_2 + \ldots + b_n \ln x_n + \ln \varepsilon$
50	Логарифмическая, правая и левая полулогарифмические модели множественной регрессии	$\ln y = a + b_1 \ln x_1 + b_2 \ln x_2 + \ldots + b_n \ln x_n + \varepsilon$, $y = a + b_1 \ln x_1 + b_2 \ln x_2 + \ldots + b_n \ln x_n + \varepsilon$, $\ln y = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n + \varepsilon$. Для линеаризации целесообразно заменить логарифмы уровней: $y' = a + b_1x_1' + b_2x_2' + \ldots + b_nx_n' + \varepsilon$, $y = a + b_1x_1' + b_2x_2' + \ldots + b_nx_n' + \varepsilon$, $y' = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n + \varepsilon$
51	Экспоненциальная модель множественной регрессии	$y = e^{a + b_1x_1 + b_2x_2 + \ldots + b_nx_n + \varepsilon}$ приводится к виду $\ln y = \ln a + b_1 \ln x_1 + b_2 \ln x_2 + \ldots + b_n \ln x_n + \ln \varepsilon$ с помощью логарифмирования
52	Обратная модель множественной регрессии	$\frac{1}{y} = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n + \varepsilon$ линеаризуется заменой переменной $\frac{1}{y} = y'$, $y' = a + b_1x_1 + b_2x_2 + \ldots + b_nx_n + \varepsilon$
53	Полиномиальная модель множественной регрессии	$y = a + b_1x_1 + b_2x_2^2 + b_3x_3^3 + \ldots + b_nx_n^n + \varepsilon$. Применяется замена переменных $x_2^2 = x_2'$; $x_3^3 = x_3'$; ... ; $x_n^n = x_n'$, преобразуется к виду $y = a + bx_1 + b_2x_2' + b_3x_3' + \ldots + b_nx_n' + \varepsilon$
54	Степенная модель множественной регрессии	$y = a + b_1x + b_2x^2 + b_3x^3 + \ldots + b_nx^n + \varepsilon$. Линеаризуют заменой $x^2 = x_2'$; $x^3 = x_3'$; ... ; $x^n = x_n'$, получают: $y = a + b_1x_1 + b_2x_2' + b_3x_3' + \ldots + b_nx_n' + \varepsilon$
55	Интерактивная модель множественной регрессии	$y = a + bx_1 + cx_2 + dx_1x_2 + \ldots + \varepsilon$. После замены переменных $x_1x_2 = x'$, ..., получают линейную модель вида: $y = a + bx_1 + cx_2 + dx' + \ldots + \varepsilon$
56	Степенная модель множественной регрессии с аддитивными возмущениями	$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \ldots \cdot x_n^{b_n} + \varepsilon$. Не может быть приведена к линейной форме. Для оценки ее параметров используются итерационные методы подбора нелинейных моделей
57	Статистическая гипотеза	априорное предположение, например, о значениях параметров регрессионной модели. О проверяемой гипотезе говорят как об исходной «нулевой» гипотезе и обозначают ее H_o
58	Статистический критерий проверки гипотезы Ho	правило решения вопроса об отклонении или неотклонении статистической гипотезы H_o. Определяется заданием: статистической гипотезы H_o; уровня значимости α; статистики критерия (t-статистики, χ^2-статистики, F-статистики); критического множества R
59	Уровень значимости статистического критерия	выбранное при формулировании статистического критерия значение вероятности ошибки α
60	Доверительная вероятность (надежность)	вероятность осуществления статистической гипотезы $\gamma = (1 - \alpha)$

№ п/п	Понятие	Содержание
61	Коэффициент интеркорреляции	коэффициент линейной корреляции между двумя объясняющими переменными
62	Коллинеарные переменные	две переменные явно коллинеарны, т. е. находятся между собой в линейной зависимости, если $r_{x_i x_j} \ge 0{,}7$. Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из множественной регрессии
63	Мультиколлинеарность факторов	когда более чем два фактора в регрессии связаны между собой линейной зависимостью, т. е. имеет место совокупное воздействие факторов друг на друга
64	Методы построения уравнения множественной регрессии	1) метод исключения – отсев факторов из полного его набора; 2) метод включения – дополнительное введение фактора; 3) шаговый регрессионный анализ – исключение ранее введенного фактора
65	Правило отбора факторов в уравнение регрессии	число включаемых факторов обычно в 6–7 раз меньше объема совокупности, по которой строится регрессия
66	Коэффициенты «чистой» регрессии	В линейной множественной регрессии $\hat{y}_x = a + b_1x_1 + b_2x_2 + ... + b_mx_m$ параметры при x называются коэффициентами «чистой» регрессии. Они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизмененном значении других факторов, закрепленных на среднем уровне
67	Уравнение линейной множественной регрессии в стандартизированном масштабе	$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + ... + \beta_m t_{x_m} + \varepsilon$, где t_y, t_{x_1}, ..., t_{x_m} – стандартизированные переменные: $t_y = \frac{y - \overline{y}}{\sigma_y}$, $t_{x_i} = \frac{x_i - \overline{x}_i}{\sigma_{x_i}}$, для которых среднее значение равно нулю: $\overline{t_y} = \overline{t_{x_i}} = 0$, а среднее квадратическое отклонение равно единице: $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$; β_i – стандартизированные коэффициенты регрессии
68	Стандартизованные коэффициенты регрессии	показывают, на сколько единиц изменится в среднем результат, если соответствующий фактор xi изменится на одну единицу при неизменном среднем уровне других факторов. В силу того что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии β_i можно сравнивать между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов «чистой» регрессии, которые несравнимы между собой
69	Частный коэффициент эластичности	$Э_{yx_i} = b_i \cdot \frac{x_i}{\hat{y}_{x_i \cdot x_1, x_2, ... x_{i-1}, x_{i+1}, ..., x_m}}$, где b_i – коэффициент регрессии для фактора x_i в уравнении множественной регрессии, $\hat{y}_{x_i \cdot x_1, x_2, ... x_{i-1}, x_{i+1}, ..., x_m}$ – частное уравнение регрессии
70	Средний коэффициент эластичности	$\overline{Э}_i = b_i \cdot \frac{\overline{x}_i}{\overline{y}_{x_i}}$ показывает, на сколько процентов в среднем изменится результат при изменении соответствующего фактора на 1%. Средние показатели эластичности можно сравнивать друг с другом и соответственно ранжировать факторы по силе их воздействия на результат

№ п/п	Понятие	Содержание
71	Коэффициент множественной корреляции	характеризует тесноту линейной связи рассматриваемого набора факторов с исследуемым признаком, т. е. оценивает тесноту совместного влияния факторов на результат. Определяется по формуле $R_{yx_1x_2\ldots x_m} = \sqrt{1-\frac{S^2_{\varepsilon_i}}{S^2_y}}$, где S^2_y — оценка общей дисперсии результативного признака; $S^2_{\varepsilon_i}$ — оценка остаточной дисперсии. Величина коэффициента множественной корреляции должна быть больше или равна максимальному парному коэффициенту корреляции: $R_{yx_1x_2\ldots x_m} \geq r_{yx_i(\max)}$ $\left(i=\overline{1,m}\right)$
72	Коэффициент множественной детерминации	характеризует степень выраженности регрессионной связи между переменными $R^2_{yx_1x_2\ldots x_m} = 1-\frac{\sum\left(y_i-\hat{y}_{x_1x_2\ldots x_m}\right)^2}{\sum\left(y_i-\bar{y}\right)^2}$
73	Скорректированный коэффициент множественной детерминации	$R^2_{ск} = 1-\frac{\sum(y_i-\hat{y}_i)^2/(n-m)}{\sum(y_i-\bar{y})^2/(n-1)}$, $R^2_{ск} = 1-(1-R^2)\cdot\frac{n-1}{n-m}$, где *m* — число оцениваемых параметров уравнения регрессии; *n* — число наблюдений. Чем больше величина *m*, тем сильнее различия $R_{ск}{}^2$ и R_2
74	Частный коэффициент корреляции	характеризует тесноту линейной связи между результатом и соответствующим фактором при элиминировании (устранении влияния) других факторов, включенных в уравнение регрессии. Коэффициент частной корреляции, измеряющий влияние на y фактора xi, при неизменном уровне других факторов, можно определить по формуле $r_{yx_i\cdot x_1x_2\ldots x_{i-1}x_{i+1}\ldots x_m} = \sqrt{1-\frac{1-R^2_{yx_1x_2\ldots x_i\ldots x_m}}{1-R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_m}}}$, где $R^2_{yx_1x_2\ldots x_i\ldots x_m}$ — множественный коэффициент детерминации всех m факторов с результатом; $R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_m}$ — тот же показатель детерминации, но без введения в модель фактора x_i
75	Частный *F*-критерий	для фактора xi частный *F*-критерий определится как $F_{част,x_l} = \frac{R^2{}_{yx_1\ldots x_i\ldots x_m} - R^2{}_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_m}}{1-R^2{}_{yx_1\ldots x_i\ldots x_{m1}}}\cdot\frac{n-m}{m-1}$, где $R^2_{yx_1\ldots x_i\ldots x_m}$ — коэффициент множественной детерминации для модели с полным набором факторов, $R^2_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_m}$ — тот же показатель, но без включения в модель фактора x_i, *n* — число наблюдений, *m* — число параметров в модели
76	Метод взвешенных наименьших квадратов (обобщенный МНК)	применяется, когда построенная модель множественной регрессии не отвечает требованию гомоскедастичности возмущений. Найденные параметры такой модели не обладают свойствами несмещенности и эффективности, о чем можно судить по гетероскедастичности остатков. Согласно обобщенному (взвешенному) МНК оценки параметров определяются из условия минимума суммы взвешенных квадратов ошибок: $\frac{\sum_{i=1}^{n} e_i^2}{S_{\varepsilon_i}} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{S_{\varepsilon_i}} \to \min$

№ п/п	Понятие	Содержание
77	Фиктивная (искусственная, двоичная) переменная (индикатор)	отражает два противоположных состояния качественного фактора. Например, «фактор действует» – «фактор не действует», «курс валюты фиксированный» – «курс валюты плавающий», «сезон летний» – «сезон зимний» и т. д. Выражается в двоичной форме: $d = \begin{cases} 0, & \text{фактор не действует,} \\ 1, & \text{фактор действует.} \end{cases}$ Значение качественной переменной, для которого принимается $d = 0$, называется базовым или сравнительным
78	Модели дисперсионного анализа (ANOVA-модели)	регрессионные модели, содержащие лишь качественные объясняющие переменные
79	Модели ковариационного анализа (ANCOVA- моделями)	Модели, в которых объясняющие переменные носят как количественный, так и качественный характер
80	Пространственные модели	модели, построенные по данным, характеризующим совокупность различных объектов в определенный момент времени
81	Модель временного ряда	модели, построенные на основе данных, характеризующих один объект за ряд последовательных моментов времени
82	Временной ряд (ряд динамики)	совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени
83	Модель тренда	$y(t) = T(t) + \varepsilon_t$, где $T(t)$ – временной тренд заданного параметрического вида (например, линейный $T(t) = a + bt$), плавно меняющаяся компонента (тенденция), описывающая чистое влияние долговременных факторов – рост населения, экономическое развитие, изменение структуры потребления и т. д., ε_t – случайная (стохастическая) компонента
84	Модель сезонности	$y(t) = S(t) + \varepsilon_t$, где $S(t)$ – периодическая (сезонная) компонента, отражающая повторяемость экономических процессов в течении не очень длительного времени – объем продаж товаров, перевозок пассажиров в различные времена года и т. д., ε_t – случайная (стохастическая) компонента
85	Модель тренда и сезонности	$y(t) = T(t) + S(t) + \varepsilon_t$ (аддитивная), $y(t) = T(t) \cdot S(t) \cdot \varepsilon_t$ (мультипликативная), где $T(t)$ – временной тренд заданного параметрического вида, $S(t)$ – периодическая (сезонная) компонента, ε_t – случайная (стохастическая) компонента
86	Стационарные и нестационарные временные ряды	Некоторые временные ряды не содержат тенденции и циклической компоненты, а каждый следующий их уровень образуется как сумма среднего уровня ряда и некоторой (положительной или отрицательной) случайной компоненты. Такие ряды называются стационарными. Ряды, содержащие в своем составе две или три компоненты, называются нестационарными
87	Стационарный временной ряд	такой случайный процесс, для которого математическое ожиданиие, дисперсия и ковариации между отдельными членами ряда случайно варьируют вокруг постоянного, не зависящего от времени уровня: $M_x(t) = \text{const}$; $D_x(t) = \text{const}$
88	«Белый шум»	чисто случайный процесс, значения которого в различные моменты времени независимы и одинаково распределены

№ п/п	Понятие	Содержание
89	Коэффициент автокорреляции	$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \overline{y}_1)(y_{t-1} - \overline{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \overline{y}_1)^2 \sum_{t=2}^{n}(y_{t-1} - \overline{y}_2)^2}},$$ где $\overline{y}_1 = \frac{1}{n-1}\sum_{t=2}^{n} y_t, \quad \overline{y}_2 = \frac{1}{n-1}\sum_{t=2}^{n} y_{t-1}$. Эту величину называют коэффициентом автокорреляции уровней ряда первого порядка, так как он измеряет зависимость между соседними уровнями ряда t и y_{t-1}. Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями y_t и y_{t-2} и определяется по формуле: $$r_2 = \frac{\sum_{t=3}^{n}(y_t - \overline{y}_3)(y_{t-2} - \overline{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \overline{y}_3)^2 \sum_{t=3}^{n}(y_{t-2} - \overline{y}_4)^2}},$$ где $\overline{y}_3 = \frac{1}{n-2}\sum_{t=3}^{n} y_t, \quad \overline{y}_4 = \frac{1}{n-2}\sum_{t=3}^{n} y_{t-2}.$
90	Лаг	Число периодов, по которым рассчитывается коэффициент автокорреляции. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Считается целесообразным для обеспечения статистической достоверности коэффициентов автокорреляции использовать правило – максимальный лаг должен быть не больше $n/4$
91	Автокорреляционная функция	Последовательность коэффициентов автокорреляции уровней первого, второго и т. д. порядков
92	Коррелограмма	График зависимости значений автокорреляционной функции от величины лага (порядка коэффициента автокорреляции)
93	Аналитическое выравнивание временного ряда	распространенный способ моделирования тенденции временного ряда. Заключается в построении аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда
94	Критерий Дарбина-Уотсона	отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии: $$d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2}$$ Расчет этого критерия – один из наиболее распространенных методов определения автокорреляции в остатках моделей
95	Система независимых уравнений	каждая зависимая переменная у такой системы рассматривается как функция одного и того же набора факторов x: $$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m. \end{cases}$$ Набор факторов x_j в каждом уравнении может варьироваться. Каждое уравнение системы независимых уравнений может рассматриваться самостоятельно. Для нахождения его параметров используется МНК

№ п/п	Понятие	Содержание
96	Система рекурсивных уравнений	в этой системе зависимая переменная у одного уравнения выступает в виде фактора х в другом уравнении. Каждое уравнение этой системы может рассматриваться самостоятельно, и его параметры определяются МНК
97	Система взаимозависимых (совместных, одновременных) уравнений (структурная форма модели)	получила наибольшее распространение в эконометрических исследованиях. В ней одни и те же зависимые переменные в одних уравнениях входят в левую, а в других – в правую часть системы. Каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим. С этой целью используются специальные приемы оценивания
98	Эндогенные переменные	взаимосвязанные переменные, описывающие экономический объект и формирующиеся внутри функционирования объекта. Это зависимые переменные, которые обычно обозначаются через y
99	Экзогенные переменные	переменные, описывающие экономический объект и формирующиеся извне. Это предопределенные переменные, влияющие на эндогенные переменные, но не зависящие от них. Обозначаются через x
100	Структурные коэффициенты	коэффициенты структурной формы модели. Находятся в правой части уравнений при эндогенных (b_{ik}) и экзогенных переменных (a_{ij})
101	Приведенная форма модели	представляет собой систему линейных функций эндогенных переменных от экзогенных: $\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + \ldots + \delta_{1n}x_n + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + \ldots + \delta_{2n}x_n + u_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = \delta_{m1}x_1 + \delta_{m2}x_2 + \ldots + \delta_{mn}x_n + u_m, \end{cases}$ где δ_{ij} – коэффициенты приведенной формы модели, u_i – остаточная величина для приведенной формы. По своему виду приведенная форма модели ничем не отличается от системы независимых уравнений, параметры которой оцениваются традиционным МНК. Применяя МНК, можно оценить δ_{ij}, а затем оценить значения эндогенных переменных через экзогенные
102	Идентификация	единственность соответствия между приведенной и структурной формами модели
103	Идентифицируемая модель	имеет место, когда все структурные коэффициенты модели определяются однозначно только по коэффициентам приведенной формы модели, т. е. если число параметров структурной модели равно числу параметров приведенной формы модели. В этом случае структурные коэффициенты модели оцениваются через параметры приведенной формы модели
104	Неидентифицируемая модель	имеет место, если число приведенных коэффициентов меньше числа структурных коэффициентов и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели
105	Сверхидентифицируемая модель	имеет место, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. В этой модели число структурных коэффициентов меньше числа коэффициентов приведенной формы. Сверхидентифицируемая модель в отличие от неидентифицируемой модели практически решаема, но требует для этого специальных методов исчисления параметров

№ п/п	Понятие	Содержание
106	Условие идентифицируемости структурной модели	модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение
107	Условие идентифицируемости уравнения структурной модели	уравнение будет идентифицируемо, если число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, равно числу эндогенных переменных в данном уравнении без одного
108	Косвенный метод наименьших квадратов (КМНК)	применяется в случае точно идентифицируемой структурной модели. Предполагает три этапа: 1) структурная модель преобразовывается в приведенную форму модели; 2) для каждого уравнения приведенной формы модели обычным МНК оцениваются приведенные коэффициенты δ_{ij}; 3) коэффициенты приведенной формы модели трансформируются в параметры структурной модели
109	Двухшаговый метод наименьших квадратов (ДМНК)	основная идея ДМНК – на основе приведенной формы модели получить для сверхидентифицируемого уравнения теоретические значения эндогенных переменных, содержащихся в правой части уравнения. Далее, подставив их вместо фактических значений, можно применить обычный МНК к структурной форме сверхидентифицируемого уравнения. Если все уравнения системы сверхидентифицируемые, то для оценки структурных коэффициентов каждого уравнения используется ДМНК. Если в системе, кроме сверхидентифицируемых, есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений

Приложение 1

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

k_2 \ k_1	1	2	3	4	5	6	8	12	24	α
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1

Приложение 2

Критические значения t-критерия Стьюдента при уровне значимости 0,10, 0,05, 0,01 (двухсторонний)

Число степеней свободы d.f.	α			Число степеней свободы d.f.	α		
	00,10	0,05	0,01		00,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,5041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Приложение 3

Значения статистик Дарбина-Уотсона $d_L d_U$ при 5%-м уровне значимости

n	*k* = 1		*k* = 2		*k* = 3		*k* = 4		*k* = 5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1,78				
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,85	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,99
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83

СОДЕРЖАНИЕ

Учебное издание

Колачева Наталья Вениаминовна
Палферова Сабина Шехшанатовна

ЭКОНОМЕТРИКА

Сборник учебно-методических материалов

Редактор *Е.Ю. Жданова*
Технический редактор *З.М. Малявина*
Вёрстка: *Л.В. Сызганцева*
Дизайн обложки: *Г.В. Карасева*

Подписано в печать 28.06.2011. Формат 84×108/16.
Печать оперативная. Усл. п. л. 18,06.
Тираж 70 экз. Заказ № 1-72-10.

Тольяттинский государственный университет
445667, г. Тольятти, ул. Белорусская, 14