В. А. ВАЛЕНТИНОВ

ЭКОНОМЕТРИКА

учебник

$$S_{al} = E\sqrt{\frac{1}{\sum(X_i - \bar{X})^2}}$$

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

ДАШКОВ и К°
Издательско-торговая корпорация

Издательско-торговая корпорация «Дашков и К°»

В. А. Валентинов

ЭКОНОМЕТРИКА

Учебник

2-е издание

*Допущено
Министерством образования и науки
Российской Федерации в качестве учебника
для студентов высших учебных заведений,
обучающихся по специальности
"Математические методы в экономике"*

Москва, 2009

УДК 330.115
ББК 65в6
В15

Автор:
В. А. Валентинов — кандидат экономических наук, доцент.

Рецензенты:

З. В. Алферова — доктор экономических наук, профессор кафедры исследования операций МЭСИ;

В. А. Колемаев — заведующий кафедрой прикладной математики Московского государственного университета управления, доктор экономических наук, профессор;

В. Ф. Тулинов — доктор физико-математических наук, профессор.

Валентинов В. А.

В15 **Эконометрика:** Учебник / В. А. Валентинов. — 2-е изд. — М.: Издательско-торговая корпорация «Дашков и К°», 2009. — 448 с.

ISBN 978-5-394-00165-9

В учебнике рассматриваются модели прогнозирования экономических процессов при условии соблюдения и нарушения предпосылок метода наименьших квадратов. Раскрываются свойства экономических объектов и их воспроизведение с помощью математических моделей. Отражены методы определения оценок параметров модели с использованием метода наименьших квадратов, обобщенного метода наименьших квадратов, двухшагового метода наименьших квадратов. Приведены алгоритмы реализации моделей прогнозирования, примечания с целью углубления изучаемых вопросов, а также дискуссии, содержащие авторское видение проблемы. Дан список рекомендованной литературы, тесты, глоссарий.

Для студентов экономических специальностей.

ISBN 978-5-394-00165-9

Оглавление

ПРЕДИСЛОВИЕ

В основу учебника положены материалы лекций, практических занятий и научных разработок, информация по системе качества ISO серии 9000, опыт преподавания данного курса на протяжении нескольких лет с использованием электронного обучающего комплекса Econ3 (авторской разработки), тестирующей программы Study и некоторых пакетов прикладных программ.

Содержание учебника полностью соответствует государственному стандарту по дисциплине “Эконометрика” по специальностям 060400 “Финансы и кредит”, 060500 “Бухгалтерский учет и аудит”, 351400 “Прикладная информатика в экономике”.

При написании учебника автор стремился изложить материал в доступной для студентов форме. Для себя и преподавателей хотелось создать самодостаточную записную рабочую книжку с указанием на возможные ошибки при использовании моделей; на пути решения проблем; на литературу с возможными вариантами решения задач и материалами для написания рефератов, подготовки докладов, сообщений, научных работ.

Основной текст предназначен для всех студентов, в примечаниях и дискуссиях изложен дополнительный, иногда спорный материал, рассчитанный на пытливых студентов и направленный на более углубленное изучение и развитие исследований нерешенных проблем в области эконометрики. Все описанные в учебнике модели реализованы в электронном обучающем комплексе Econ3, издание которого планируется в 2006 г.

В учебном пособии имеется 14 глав по количеству лекций, читаемых по данному курсу. В конце каждой главы представлены вопросы для самоконтроля и ссылки на литературу, которая может быть использована по вопросам главы курса.

С первой по седьмую главы приводится классический регрессионный анализ. С восьмой по одиннадцатую — анализируются временные ряды. Двенадцатая глава посвящена системам одновременных уравнений, тринадцатая — межпредметным связям эконометрики и управление качеством. Этот материал позволяет определить место и роль эконометрики в цикле улучшения процессов деятельности любого предприятия и организации. В четырнадцатой главе изучаются информационные технологии эконометрических исследований, предназначенных для специальности 351400 “Прикладная информатика в экономике”, однако этот материал будет полезен для других экономических специальностей.

В конце учебника расположены список информационных ресурсов, глоссарий, тесты по курсу, перечень основной и дополнительной литературы с небольшими пояснениями.

ВВЕДЕНИЕ

Экономические объекты, производя продукцию или услуги, направляют свои усилия в трех взаимосвязанных направлениях: повышение эффективности производства и управления, повышение конкурентоспособности предприятия или организации и продукции, повышение качества продукции и услуг. Для реализации этой деятельности экономические объекты пользуются принципами эффективного производства и инструментальными средствами. Одним из инструментальных средств управления является эконометрия. Важно знать место и роль эконометрических исследований во всех этапах стратегического планирования жизненного цикла продукции и оперативном управлении каждого этапа. Перед принятием управленческих решений необходимо построить эконометрическую модель прогнозируемых процессов. С использованием этой модели можно рассчитать все последствия принимаемых решений, а затем выбрать оптимальное решение. Основной задачей эконометрики является получение прогноза экономического процесса, который будет использоваться для принятия управленческих решений.

Все модели, которые повышают точность прогноза, очень ценятся и заносятся в золотой фонд эконометрических исследований. Не случайно Энгель получил Нобелевскую премию в 2003 г. за обнаружение закономерностей во временном ряду финансовых индексов, которые позволили повысить точность прогнозирования.

Следует отметить, что проведение эконометрических исследований не является самоцелью, а является одним из инструментов решения проблем выполнения цепочки улучшения процессов в соответствии с циклом Деминга: планируй, делай, анализируй, совершенствуй.

Примечание. Это направление в эконометрической литературе только начинается развиваться и представляется очень перспективным и соответствует международным стандартам системы качества ISO серии 9000. Можно утверждать, что всем существующим математическим методам и моделям можно найти место в цепочке Деминга улучшения процессов. Известна работа [32], в которой автор вводит новое понятие "эконометрика качества", глава 13 посвящена эконометрическим методам управления качеством и сертификации продукции. Однако в этой работе основное внимание уделено контролю качества продукции, а не управлению процессами.

Эконометрика как наука представляет собой систему. Любая система базируется на пяти основаниях: правовом, нормативном, научно-техническом, организационном, информационном.

Правовая основа эконометрики состоит в том, что она признана как наука и занесена в государственный образовательный стандарт России, Украины и других стран.

Нормативная основа эконометрики предполагает наличие стандартов, методик, положений, инструкций. Любой стандарт состоит из словаря, перечня условных обозначений и основного текста. В настоящее время на роль стандарта претендуют работы [1], [6]. Однако единого стандарта по эконометрике нет и многие учебные пособия пользуются своими обозначениями ввиду объективной необходимости.

Примечание. Одной из проблем в обозначении является противоречие в изображении факторов X_i, где i — номер фактора и элемента матрицы X_{ij}, где i — номер строки, j — номер столбца или фактора. В работе [1] вводится универсальное обозначение факторов X_i^j, где i — номер значения фактора или номер строки, j — порядковый номер фактора или номер столбца.

Другой проблемой является определение и обозначение характеристик для генеральной и выборочной совокупностей, обозначение смещенных и несмещенных оценок. Принято обозначать характеристики генеральной совокупности греческими буквами, а характеристики выборочной совокупности — латинскими буквами, принято коэффициентами называть оценки параметров генеральной совокупности. Различий в обозначении смещенных и несмещенных оценок нет, поэтому они указываются по тексту.

Содержание эконометрики открыто и пополняется следующими новыми методами и моделями (не входящими в Государственный стандарт): бутстреп метод, гребневая (ридж-) регрессия, логит- и пробит-модели, многомерный статистический анализ [1].

Научная основа эконометрики. ***Теоретической базой*** эконометрики является теория вероятностей и случайных процессов; математическая статистика, экономическая кибернетика.

В основе всех эконометрических моделей лежит предположение о том, что зависимая переменная является случайной величиной (полученной пространственной выборкой из генеральной совокупности) или реализацией случайного процесса, происходящего во времени. Случайные величины и случайные процессы изучаются в курсе "Теория вероятности и математическая статистика".

Примечание. Детерминированные процессы изучаются в курсе теории чисел и численного анализа.

Средствами описания эконометрических моделей являются математическая символика и линейная алгебра.

Многие учебники по эконометрике содержат главы, посвященные основам теории вероятности и математической статистики, линейной алгебры.

Экономическая кибернетика позволяет системно анализировать экономический объект как систему с входом, обработкой ресурсов, выходом и обратной связью; изучить условия устойчивости экономической системы и саморегулирования. Задачи экономической кибернетики реализуются при изучении систем одновременных уравнений.

Примечание. К сожалению, в эконометрической литературе мало уделено внимания вопросам саморегулирования и устойчивости экономических систем[1].

---

[1] Плодотворные идеи по вопросу устойчивости экономических систем, изложенные в доступной и наглядной форме, имеются в книге: Хейс Д. Причинный анализ в статистических исследованиях. — М.: Финансы и статистика, 1981.

В основном научные школы по эконометрике находятся на Западе, где они существуют давно и имеют разнообразную эконометрическую литературу и специализированный журнал "Econometric". В России научные школы по эконометрике только формируются, однако следует выделить школы С. А. Айвазяна (проведение семинаров по многомерному статистическому анализу) и Я. Р. Магнуса (российская экономическая школа). Об их деятельности см. в Интернете (список информационных ресурсов приведен в прил. 1).

По курсу "Эконометрика" имеется около 20 учебников, предназначенных для разного уровня подготовки читателей. Всю эконометрическую литературу можно условно разделить на группы по признаку принадлежности к следующим направлениям изложения материала: теория, практика, популизация, методика.

К *теоретическим* мы относим те работы, в которых даются обоснование и вывод всех формул, которые используются в эконометрике.

Теоретическими работами являются [1], [6], [23].

К *практическим* мы относим работы, в которых используются результаты теоретических выводов для решения практических задач . Теоретиков очень интересуют результаты работы практиков. Результаты практических популярных исследований дают импульс теоретикам для совершенствования теории. Практической работой является [21].

К *популизаторным* мы относим работы, в которых в доступной для большинства студентов и практических работников форме даются основные положения курса, указываются предпосылки, достоинства и недостатки методов и моделей, сфера их эффективного применения, средства проведения расчетов и методы их анализа. Популизаторными работами являются [10], [30].

К *методическим* мы относим работы, в которых дается методика изложения или преподавания определенных вопросов по курсу.

К *методической* литературе можно отнести практикумы и методические разработки [3], [4], [19].

Имеются работы, которые соединили в себе математическую строгость с доступностью изложения [5], [23].

Для полного понимания вывода всех формул, используемых в эконометрике, необходимы специальные знания, выходящие за пределы базового курса высшей математики. Однако в практической эконометрической деятельности важно знать возможности и ограничения используемых методов и моделей, средства и методы проведения расчетов, правила проведения анализа полученных расчетов, знать где и как можно эффективно использовать эконометрические модели.

Техническая основа эконометрики. Техническую основу эконометрики составляют ЭВМ, средства визуального воспроизведения информации и соответствующее программное обеспечение. Для проведения расчетов используются электронные таблицы Excel, один или несколько специализированных пакетов прикладных программ, обзор которых дается в главе 14, а также электронные обучающие системы по курсу для проведения самостоятельной работы и выполнения практических и лабораторных работ.

Организационная основа эконометрики. В Министерстве образования и науки РФ имеется подразделение УМО (Учебно-методическое объединение), которое координирует преподавание дисциплины "Эконометрика". Имеется острая необходимость в организации издания переводной иностранной литературы, на которую часто имеются ссылки. Наиболее полная библиография по иностранной литературе имеется в работе [6]. Высшая школа испытывает потребность в электронных обучающих системах по курсу "Эконометрика".

Были организованы курсы по подготовке преподавателей эконометрики. Однако имеется потребность в повышении взаимодействия между эконометристами в форме проведения семинаров, съездов, публикаций в специализированном журнале.

Примечание. Назрела острая необходимость написания учебных пособий: 1) большим тиражом для студентов в соответствии с государственным образовательным стандартом, 2) небольшим тиражом для преподавателей с пояснениями всех теоретических и методологических вопросов первого пособия, 3) большим тиражом практикумов с решенными задачами по всему курсу первого пособия. Для разработки этих пособий нужен коллектив теоретиков, практиков и популизаторов эконометрики.

Информационное обеспечение эконометрики. Много полезной информации можно получить по Интернету с использованием ключевых слов "эконометрика" или "econometric". Список информационных ресурсов см. в прил. 1.

Очень полезно в учебном заведении сделать стенд, посвященный эконометрике, где можно регулярно размещать информацию о ведущих эконометристах мира и их творчестве, Нобелевских лауреатах по экономике, научно-практических разработках студентов, текущую информацию об учебном процессе.

Условные обозначения. Обозначим выборочные наблюдения через

$X_1, X_2, ..., X_n$;

$У_1, У_2, ..., У_n$

и введем их арифметические средние

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \overline{У} = \frac{1}{n}\sum_{i=1}^{n} У_i,$$

где $\sum_{i=1}^{n} X_i = \sum{}_{i=1}^{n} X_i = \sum{}_i X = \sum X = X_1 + X_2 + \ldots + X_n$;

X — фактор, объясняемая переменная, влияющая на следствие У;

У — следствие, зависимая переменная.

Греческими буквами обозначают параметры модели для генеральной совокупности.

Например, α_0, α_1 — параметры линейной модели $У = \alpha_0 + \alpha_1 X + \varepsilon$ для генеральной совокупности,

где ε — случайное возмущение или ошибка модели, которая состоит из ошибки уравнения и ошибки измерения.

Латинскими буквами обозначают коэффициенты уравнения регрессии для выборочной совокупности.

Например, a_0, a_1 — коэффициенты уравнения регрессии $У = a_0 + a_1X + e$ для выборочной совокупности, где $e = У - (a_0 + a_1X) = У - У_p$ — отклонение, или остаток (Гаусс называл его убытком, С. А. Айвазян — невязкой. В эконометрической литературе принято e называть остатком), учитывающий влияние всех факторов, не включенных в модель;

У — фактические значения зависимой переменной;

$У_p = a_0 + a_1X$ — расчетные значения У;

X — фактор;

δ — белый шум;

m — число степеней свободы;

n — объем выборки;

T — период периодического колебания;

k — количество всех коэффициентов в модели (включая свободный коэффициент).

Например, уравнение регрессии $У = a_0 + a_1X + e$ имеет два коэффициента: a_0 и a_1, следовательно $k = 2$;

t — индекс времени в моделях временных рядов.

Например, $У_t = a_0 + a_1X_t + e_t$, t — индекс времени;

t_{a1} — фактическое значение критерий Стьюдента для коэффициента a_1;

Например, $t_{a1} = a_1/S_{a1}$,
где S_{a1} _ среднее квадратическое отклонение коэффициента a_1 от своего математического ожидания a_1, или ошибка коэффициента a_1;

$t_{\alpha/2}(\alpha = 0{,}05,\ m = n - 1)$ или $t_{\alpha/2}$ — двухстороннее (критическое) табличное значение критерия Стьюдента;

$t_\alpha(\alpha = 0{,}05,\ m = n - 1)$ или t_α — одностороннее (критическое) табличное значение критерия Стьюдента;
где α — уровень значимости критерия или вероятность ошибки при отклонении верной нулевой гипотезы или вероятность совершить ошибку первого рода;

ошибка первого рода — неправильное отклонение нулевой гипотезы;

$m = n - 1$ — число степеней свободы для критерия Стьюдента.

Например, в уравнении регрессии $У = a_0 + a_1X + e$, $У$ — зависимая переменная, X — фактор.

В системах одновременных уравнений $У$ может выступать как объясняемая переменная.

$$\begin{cases} У_1 = a_0 + a_1X_1 + a_2У_2 + e_1, \\ У_2 = \beta_0 + \beta_1X_2 + \beta_2У_1 + e_2 \end{cases}$$

В первом уравнении: $У_1$ — зависимая переменная, X_1 — фактор, $У_2$ — объясняемая переменная.

Во втором уравнении: $У_2$ — зависимая переменная, X_2 — фактор, $У_1$ — объясняемая переменная.

Введем обозначения уравнения регрессии для трех переменных.

Предположим, мы имеем три связанные между собой переменные, которые обозначим через X_1, X_2, $У$. Переменная $У$ может, например, отражать количество покупок некоторого товара в домашнем хозяйстве семьи, X_1 — цена товара, X_2 — доход семьи. Произведем выборку объемом n из генеральной совокупности, составляющая N семей. Численные значения переменных выборки мы будем записывать в таблице 0.1.

Таблица 0.1

Исходные значения выборочной совокупности

i	X_1	X_2	$У$
1	X_{11}	X_{21}	$У_1$
2	X_{12}	X_{22}	$У_2$
i	X_{1i}	X_{2i}	$У_i$
...	...	...	...
n	X_{1n}	X_{2n}	$У_n$

Здесь X_{di} обозначает величину переменной X_d для i-го домашнего хозяйства. Гипотеза о линейной зависимости $У$ от X_1 и X_2 может быть записана в виде

$$У_i = \alpha_0 + \alpha_1X_{1i} + \alpha_2X_{2i} + \varepsilon_i, \; i = 1, 2, \ldots, n.$$

Примечание. Большинство статистических пакетов предполагает предложенную схему размещения переменных в таблице базы данных: сначала размещают столбцы факторов X_d, последним ставят столбец зависимой переменной У.

Е — ошибка модели.

F — фактическое значение критерия Фишера.

$F_{кр}(\alpha = 0{,}05;\ m_1 = k - 1;\ m_2 = n - k)$ — критическое значение критерия Фишера на уровне значимости α и числе степеней свободы m_1, m_2.

S — среднее квадратическое отклонение для выборочной совокупности.

S^2 — дисперсия для выборочной совокупности.

σ — среднее квадратическое отклонение для генеральной совокупности.

σ^2 — дисперсия для генеральной совокупности.

М — условное обозначение математического ожидания. Например, $M(\varepsilon_i) = 0$ при всех $i = 1, 2, ..., n$.

r — коэффициент корреляции.

r^2 — коэффициент детерминации.

R — множественный коэффициент корреляции.

R^2 — множественный коэффициент детерминации.

Буквы в формулах, выделенные полужирным шрифтом, означают матрицу.

Например, в формуле

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{У}$$

A, X, У — матрицы.

Расчетные формулы. Расчет коэффициентов регрессионного уравнения

$$У = a_0 + a_1 X + e$$

$$a_1 = \frac{\sum (X_i - \overline{X})(У_i - \overline{У})}{\sum (X_i - \overline{X})^2}$$

$$a_0 = \overline{У} - a_1 \overline{X}$$

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{У},$$

где **А** — матрица коэффициентов модели;

Х и **У** — матрицы соответственно факторов и зависимой переменной.

Ошибка модели:

$$E = \sqrt{\frac{\sum (У_i - У_{p_i})^2}{n-k}},$$

где $У_{p_i} = a_0 + a_1 X_i$.

Основное вариационное уравнение

$$\sum (У_i - \bar{У})^2 = \sum (У_i - У_{p_i})^2 + \sum (У_{p_i} - \bar{У})^2,$$

где $\sum (У_i - \bar{У})^2 = C_{общ}$ — вариация общая;

$\sum (У_i - У_{p_i})^2 = C_{ост}$ — вариация остатков;

$\sum (У_{p_i} - \bar{У})^2 = C_{рег} = C_{общ} - C_{ост}$ — вариация регрессии.

$S^2_{общ} = \dfrac{\sum (У_i - \bar{У})^2}{n-1}$ — дисперсия общая.

$S^2_{ост} = \dfrac{\sum (У_i - У_{p_i})^2}{n-k}$ — дисперсия остатков.

$S^2_{рег} = \dfrac{\sum (У_{p_i} - \bar{У})^2}{k-1}$ — дисперсия регрессии.

$R^2 = \dfrac{C_{рег}}{C_{общ}}$ — коэффициент детерминации.

Множественный коэффициент корреляции:

$R = \sqrt{R^2}$.

Критерий Фишера:

$$F = \frac{S^2_{рег}}{S^2_{ост}}.$$

Ошибка коэффициента a_0:

$$S_{a0} = E\sqrt{\frac{\sum X_i^2}{n\sum(X_i - \bar{X})^2}}.$$

Ошибка коэффициента a_1:

$$S_{a1} = E\sqrt{\frac{1}{\sum(X_i - \bar{X})^2}}.$$

Критерий Стьюдента для коэффициента a_1:

$$t_{a1} = \frac{a_1}{S_{a1}}.$$

Частный коэффициент детерминации для фактора X_1:

$$r_{X_1 \cdot X_2, X_3} = \frac{t_1^2 R^2}{t_1^2 + t_2^2 + t_3^2},$$

где t_i — критерий Стьюдента для фактора X_i

Точечный прогноз:

$$У_{пр} = a_0 + a_1 X_{ож},$$

где $X_{ож}$ — ожидаемое значение X.

95% интервальный прогноз для математического ожидания У:

$$У_{пр} \pm t_{\alpha/2}\,(\alpha = 0{,}05,\ m = n - k)\,E\sqrt{\frac{1}{n} + \frac{(X_{ОЖ} - \bar{X})^2}{\sum(X_i - \bar{X})^2}}.$$

Расчет коэффициентов модели методом Эйткена:

$$\mathbf{B} = (\mathbf{X}'_н \mathbf{X}_н)^{-1} \mathbf{X}'_н \mathbf{У}_н = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{У},$$

Парный коэффициент корреляции:

$$r(X_1, X_2) = \frac{\sum(X_{1i} - \bar{X}1)(X_{2i} - \bar{X}2)}{\sqrt{\sum(X_{1i} - \bar{X}1)^2 \sum(X_{2i} - \bar{X}2)^2}}.$$

Частный коэффициент корреляции:

$$r_ч(X_i, X_j) = \frac{-C_{ij}}{\sqrt{C_{ii}C_{jj}}},$$

где C_{ij} — элементы обратной матрицы от матрицы всех парных коэффициентов корреляции.

Критическое значение коэффициента корреляции:

$$r_{кр.} = \frac{t_{\alpha/2}}{\sqrt{{t_{\alpha/2}}^2 + n - 2}},$$

где $t_{\alpha/2} = t_{\alpha/2}(\alpha = 0{,}05;\ m = n - 2)$.

Коэффициент автокорреляции:

$$r(к) = r(У_t, У_{t+к}) = \frac{\sum\limits_{t=1}^{n}(У_t - \bar{У}_1)(У_{t+к} - \bar{У}_2)}{\sqrt{\sum\limits_{t=1}^{n-к}(У_t - \bar{У}_1)^2 \sum\limits_{t=к}^{n}(У_{t+к} - \bar{У}_2)^2}}.$$

Критерий Дарбина — Уотсона (Дарбина — Ватсона):

$$DW = \frac{\sum\limits_{t=2}^{n}(e_t - e_{t-1})^2}{\sum\limits_{t=1}^{n}e_t^2} \approx 2[(1 - r(1)].$$

Виды моделей:

модель распределенных лагов:

$У_t = \alpha_0 + \alpha_1 X_t + \alpha_2 X_{t-1} + \alpha_3 X_{t-2} + \varepsilon_t;$

авторегрессионная модель распределенных лагов:

$У_t = \alpha_0 + \alpha_1 X_t + \alpha_2 X_{t-1} + \alpha_3 У_{t-1} + \varepsilon_t;$

авторегрессионная модель:

$У_t = \alpha_0 + \alpha_1 У_{t-1} + \alpha_2 У_{t-2} + \varepsilon_t;$

модель скользящей средней:

$\varepsilon_t = \delta_t + \rho\delta_{t-1} + \rho^2\delta_{t-2} + \nu_t;$

модель последовательных отклонений:

$\Delta У_t = У_t - У_{t-1}$, $t = 2, \dots n$;

модель периодических составляющих временного ряда:

$У_t = \alpha_0 + \alpha_1 t + \alpha_2 Sin(2\pi t/T_1) + \alpha_3 Cos(2\pi t/T_1) + \alpha_4 Sin(2\pi t/T_2) + \alpha_5 Cos(2\pi t/T_2) + \dots + \varepsilon_t$, где T_1, T_2 периоды для сезонной и длинно периодической составляющей;

модель экспоненциально взвешенного среднего:

$$Z_t = \lambda У_t + \lambda (1-\lambda) У_{t-1} + \lambda (1-\lambda)^2 У_{t-2} + \lambda (1-\lambda)^3 У_{t-3} + \dots = \\ = \lambda У_t + (1-\lambda)[\lambda У_{t-1} + \lambda (1-\lambda) У_{t-2} + \lambda (1-\lambda)^2 У_{t-3} + \dots];$$

линейная модель с автокоррелированными возмущениями:

$У_t = \alpha_0 + \alpha_1 X_t + \rho \varepsilon_{t-1} + \upsilon_t$.

Прогноз по линейной модели с автокоррелированными возмущениями:

$У_{пр(n+1)} = \alpha_0 + \alpha_1 X_{n+1} + \rho \varepsilon_n$.

Метод устранения гетероскедастичности:

$У_i/|e_i| = a_0/|e_i| + a_1 X_I/|e_i| + e_i/|e_i|$.

Глава 1
ОПРЕДЕЛЕНИЕ ЭКОНОМЕТРИКИ

1.1. Термины и определения

Слово “эконометрика” состоит из двух частей “эконо” и “метрика”, которые означают экономические измерения.

Термин “эконометрика” был введен в 1930 г. норвежским статистиком Рагнаром Фришером, который провозгласил в качестве основной задачи эконометрики развитие экономической теории в ее связи со статистикой и математикой.

В настоящее время общепринято следующее определение эконометрики.

Эконометрика — это наука, которая использует методы математической статистики для описания теоретических моделей реальных хозяйственных экономических процессов.

Дискуссия. На Украине принято использовать термин “эконометрия” (по аналогии с понятиями “геометрия”, “биометрия”, “социометрия”), содержание которого не отличается от термина “эконометика”, принятого в России.

Сравним эти два термина. Приставка “метрика” (гр. *metrike* < *metron* мера, размер) — учение о стихотворных размерах, расстояние. “Метрика” (польск. *metryka* < лат. *matricula* список) — свидетельство о рождении, составленное на основании соответствующей записи в метрической книге.

Метрия (гр. *metron* — мера, *metreo* — измеряю) — вторая составная часть сложных слов, соответствующая по значению слову “измерение”, например, геометрия, стереометрия (словарь иностранных слов).

Ради объективности следует признать, что термин “эконометрия” больше соответствует его содержанию, чем “эконометрика”, так как оно возникло после науки биометрии, которая очень тесно связана с эконометрическими исследованиями. Очевидно, слово “эконометрика” произошло от английского *econometric*. В соответствии с ее дословным переводом эконометрику можно рассматривать как науку о взаимном расстоянии между экономическими показателями в определенных координатах. Например, мероприятие можно разместить в координатах прибыль — затраты; все виды управления можно разместить в координатах смелость — радость; модели факторного, компонентного и дискриминатного анализов содержат решения задач расположения объектов в новых координатах таким образом, чтобы получить о них новую информацию.

В настоящее время в СНГ имеется много книг с использованием терминов “эконометрика” и “эконометрия”. Пусть это читателя не смущает. По сути — это два одинаковых термина.

Примечание. Некоторые ученые считают, что термин “эконометрия” впервые введен бухгалтером П. Цьемпой П. в 1910 г.[1].

1.2. Объект, предмет, цели, задачи, методы, структура и область использования эконометрики

Объектом эконометрики являются экономические процессы, происходящие в экономической системе общества.

Предметом эконометрики является количественная оценка взаимосвязи между случайными событиями, признаками, показателями, факторами, пер еменными экономических объектов, проверка теоретические модели реальных экономических процессов для получения прогнозов деятельности экономических систем.

---

[1] *Павло Цьомпа.* Нариси економерії і побудована н національній політекономії теорія бухгалтерського обліку / Пер. з нім. *Я. Гончарук* и др. — Львів: Каменяр, 2001.

Целью эконометрики является оценка точечных и интервальных прогнозов деятельности генеральной совокупности объектов экономической системы на основании расчетов по данным выборочной совокупности или реализации случайного процесса.

Задачами эконометрики являются:

стратегические задачи:

- создание условий эффективного использования эконометрических моделей;
- разработка эконометрических методов, повышающих точность прогноза;
- разработка методов расчета коэффициентов моделей при нарушении предпосылок метода наименьших квадратов;
- определение области эффективного использования эконометрики;

тактические задачи:

- построение математической модели с учетом свойств экономических объектов;
- определение коэффициентов математической модели экономического процесса,
- получение точечного и интервального прогноза экономического процесса;
- расчет оценки качества полученной модели.

Методы эконометрики можно классифицировать по следующим признакам:

по **методам определения** коэффициентов регрессии:

- метод наименьших квадратов;
- двухшаговый метод наименьших квадратов;
- обобщенный метод наименьших квадратов (метод Эйткена);
- метод максимального правдоподобия;
- метод наименьших отклонений и наименьших расстояний;

по **методам построения** модели деятельности экономического объекта:

- методы построения модели сверху вниз и снизу вверх [6, с. 309–312];

по **методам выборки** данных из генеральной совокупности:

- анализ всей генеральной совокупности;
- выборочный метод анализа генеральной совокупности;

по **методам анализа выборочных данных**: пространственных, временных, пространственно- временных данных:

- метод корреляционного анализа;
- метод регрессионного анализа;
- методы с участием лаговых: объясняемых, зависимых переменных и остатков;
- методы с участием фиктивных, инструментальных переменных;
- методы линеаризации нелинейных регрессий;
- методы устранения мультиколлинеарности факторов; гетероскедастичности и автокорреляции остатков;
- методы взвешенной регрессии;
- методы выявления периодических составляющих временных рядов;
- методы адаптивного прогнозирования;
- методы сглаживания временных рядов;
- методы воспроизведения стационарных случайных процессов с помощью белого шума;
- методы анализа системы одновременных уравнений;
- методы связи между качественными переменными.

Большинство перечисленных эконометрических методов будут рассмотрены в данном учебнике. Однако по некоторым методам необходимо привести некоторые замечания.

Метод максимального правдоподобия определения коэффициентов регрессионной модели совпадает с методом наименьших квадратов при условии соблюдения предпосылок этого метода. Но из-за сложности алгоритма метода максимального правдоподобия предпочтение отдается методу наименьших квадратов. Поэтому в данном учебнике метод максимального правдоподобия не рассматривается.

Метод построения модели сверху-вниз и снизу-вверх реализуется в шаговой регрессии построения модели методом включения и исключения.

Корреляционный анализ является наиболее изученным, широко используемым методом анализа между двумя случайными величинами и рассматривается во многих курсах: статистика, математическая статистика и т. д. Поэтому предполагается, что корреляционный анализ хорошо знаком читателям и в учебнике подробно не изучается. Однако следует обратить внимание на шесть типов ложных корреляций (смотрите глоссарий), которые подстерегают исследователей в использовании коэффициента корреляции при нарушении предпосылок этого коэффициента.

Методы анализа всех объектов генеральной совокупности и выборочной совокупности отличаются между собой. Например, формулы расчета среднего значения для генеральной и выборочной совокупности не отличаются между собой. Однако для средней генеральной совокупности вводится понятие ошибки средней, связанной с ошибками округления и измерения значений изучаемого показателя. Для средней выборочной совокупности вводится понятие ошибки средней, связанной с ошибкой разности между средней выборочной и генеральной средней.

Как правило, все объекты генеральной совокупности анализируются методами статистики, а выборочные данные анализируются с помощью методов математической статистики. В эконометрике не производится сплошного обследования всех объектов для изучения связей между показателями экономической системы, а используются методы математической статистики для объектов выборочных данных. Выборочный метод анализа объектов изучения заключается в том, что закономерности, определенные по выборочной совокупности, распространяются на всю генеральную совокупность с определенной вероятностью. При этом численные значения статистических характеристик, рассчитанные по выборочной совокупности, будут отличаться от численных значений соответствующих характеристик, рассчитанных для всех объектов генеральной совокупности. Разность между этими значениями и составляет ошибку статистической характеристики, рассчитанной для объектов выборочной совокупности.

Примечание. Выборочный метод анализа генеральной совокупности подразделяется на два вида. Первый вид — выборка без возврата данных в генеральную совокупность — это обычная выборка (данный метод имеет две разновидности: 1) можно анализировать данные после получения всей выборки, 2) анализируются данные при последовательном получении каждого элемента выборки, такой метод называется последовательным методом Вальда, который оказался очень экономичным и эффективным.

Второй вид выборки — выборка с возвратом в генеральную совокупность, которая называется бутстрепом. В переводе с английского "бутстреп" означает шнурок ботинка. Солдаты английской армии обеспечивались ботинками, у которых шнурок крепился на пятке ботинка и служил для того, чтобы подтягивать его при надевании. Система, состоящая из шнурка и ботинка, обеспечивает выполнение действий с помощью собственных средств. Метод "бутстреп" означает изучение группы данных с помощью множества выборок, содержимое которых возвращается в основную группу, а затем формируется другая выборка. Этот метод является экономичным, особенно при небольшом количестве данных, теоретически обосновывается и периодически вновь изобретается.

Структура эконометрики. Условно курс эконометрики можно разделить на три части: первая часть построение парных и множественных моделей по пространственным данным с использованием классического регрессионного анализа при соблюдении предпосылок метода наименьших квадратов. Эта часть изучается в курсе математической статистики и должна быть знакома студентам. Однако ей уделяется половина всего учебного времени.

Вторая часть эконометрики содержит способы устранения нарушений предпосылок метода наименьших квадратов и некоторые модели, которые изучаются только в курсе эконометрики. Вторая часть содержит следующие модели, устраняющие нарушения предпосылок метода наименьших квадратов: мультиколлинеарность, автокорреляция остатков, гетероскедастичность, обобщенный метод наименьших квадратов Эйткена, а также модели лаговых процессов, решение системы одновременных уравнений, адаптивные методы прогнозирования.

Третью часть эконометрики составляет анализ экономических временных рядов. Анализ экономических временных рядов бурно развивается как в теоретическом, так и в прикладном планах и базируется на теории случайных процессов.

Область использования. Все существующие методы и модели служат инструментальными средствами для совершенствования процессов. В зависимости от степени агрегированности различают макро- и микроэкономические процессы. Совершенствование любого процесса происходит по спирали в такой последовательности (цикл Деминга): планирование, реализация планов в виде мероприятий, анализ результатов выполнения мероприятий, закрепление успехов, затем цикл повторяется [18], [26]. Важно знать место и роль каждого метода и модели в этом цикле улучшения процессов. В зависимости от типа процесса (экономический, биологический, социальный и т. д.) используемые методы и модели приобретают профессиональную направленность.

Отличительная особенность эконометрических методов и моделей от математических методов исследования экономики и математического моделирования экономических систем состоит в том, что они призваны в первую очередь получить прогнозные значения экономического показателя, которые используется для принятия управленческих решений по улучшению экономических процессов. В качестве эффективного инструмента использования эконометрических методов и моделей можно назвать следующие области их использования: бизнес-планирование, маркетинговые исследования, планирование бюджетов государства, регионов и предприятий, математическая обработка данных научных исследований и др.

1.3. Связь эконометрики с родственными науками

Эконометрика тесно связана с биометрией, социометрией, информатикой, имитационным моделированием экономических процессов, управлением качеством.

Использование математической статистики в биологических и социологических исследованиях порождает соответственно науки — биометрию и социометрию.

Каждая из этих наук имеют много общих методов и моделей описания своих объектов исследований. Часто эти науки взаимно обогащают друг друга новыми подходами в решении общих задач. Исторически биометрия зародилась гораздо раньше эконометрики. Поэтому эконометрика включила в себя много биометрических методов и добавила новые разделы, такие как спектральный анализ, анализ временных рядов и решение систем одновременных уравнений.

Например, моделирование оптимальных условий развития растения и предприятия имеет много общего. Благоприятными условиями развития растения являются оптимальные интервальные значения подкормки, влажности, температуры, света. Если значения этих факторов выходят за допустимые интервалы, то растение прекращает рост или погибает.

Благоприятными условиями развития предприятия являются интервальные значения размера инвестиций, процент налогообложения, правовая и законодательная среда, удовлетворение потребностей потребителей и производителей. Если значения этих факторов выходят за допустимые интервалы, то предприятие или разоряется или уходит в теневую экономику.

Информатика позволяет реализовать эконометрические модели на ЭВМ с использованием табличных процессоров и пакетов прикладных программ [28].

Имитационное моделирование экономических процессов может быть использовано на этапе выбора вариантов возможных решений с учетом ограничений на используемые ресурсы.

Цели и задачи эконометрики и управления качеством во многом совпадают между собой. Средства и методы улучшения качеством и эконометрики могут взаимно дополняют друг друга. Глава13 посвящена связи эконометрики и управления качеством.

1.4. История эконометрики

Эконометрический анализ успешно используется в микро- и макроэкономике начиная с 1960 г. Длительное время эконометрические методы входили в состав дисциплины "Экономико-математические методы и модели". Сравнительно недавно эконометрика выделилась в самостоятельную учебную дисциплину и была включена в список обязательного изучения для экономических специальностей.

Одной из первых комплексных эконометрических моделей в США была модель Клейна-Гольдбергера. Она послужила фундаментом, на котором базировались некоторые краткосрочные модели комплексного развития экономики. Эта модель состояла из пятнадцати регрессионных уравнений и пяти тождеств и охватывала сорок макроэкономических показателей. Параметры модели были оценены на базе временных рядов за 20 лет.

На Украине впервые была разработана эконометрическая модель УКР-2 в 1972 г. А. Емельяновым и Ф. Кушнирским. Эта модель увязана с методикой планирования и содержала сто одно регрессионное уравнение, объединенное в семь взаимосвязанных блоков.

В настоящее время на Украине возрождается интерес к эконометрическим исследованиям, центром которых является институт экономического прогнозирования НАН. Сотрудниками этого института рассмотрена система экономико-математических секторных моделей развития экономики Украины, которая охватывает все основные сферы национального хозяйства (реальный сектор, потребление, инвестиции, занятость, внешнюю торговлю, государственные финансы, деньги и кредит). Она предназначена для использования в качестве инструментария имитационного прогнозирования в соответствии с заданными сценариями выбора эффективной экономической политики государства и позволяет получить модельные оценки и прогнозные расчеты в реальном режиме времени. Большинство расчетов производилось с использованием пакета прикладных программ EViews с применением базы данных за период 1985–1998 гг. Через каждые два года этим институтом проводятся международные научно-практические конференции по вопросам эконометрических исследований

В России следует выделить эконометрическую школу С. А. Айвазяна, которая проводит:

• обучение преподавателей эконометрики, издан учебник по эконометрике [1], изданы четыре выпуска серии учебно-методических материалов по эконометрическому моделированию, демонстрирующие возможности методов эконометрики в форме достаточно подробного разбора решения ряда реальных социально-экономических задач (методические материалы можно приобрести в здании МЭСИ);

• еженедельные научные семинары "Многомерный статистический анализ и вероятностное моделирование реальных процессов" (более подробно см. в прил. 1 "Информационные ресурсы").

Для углубленного изучения эконометрических методов существует российская экономическая школа (ее библиотека имеет наиболее полную подборку иностранной литературы по эконометрике), в которой изучаются четыре уровня эконометрики. Некоторые материалы этой школы изложены в учебнике [6]. В российской экономической школе проводятся семинары по преподаванию эконометрики (более подробно см. в прил. 1 "Информационные ресурсы").

В настоящее время разработка эконометрических моделей идет в двух направлениях, которые возглавляют голландская и американская школы.

Основателем голландской школы является Я. Тимберген. Конструкция голландской модели отличается тем, что большинство переменных в ней выражаются в виде относительных приростов, регрессионные уравнения часто содержат лаговые значения эндогенных переменных.

Основателем американской школы является Клейн. Модели американской школы содержат временную и отраслевую детализацию. Известны эконометрические модели, которые содержат более трехсот уравнений. Параметры модели оцениваются методом главных компонент и двухшаговым методом наименьших квадратов.

1.5. Примеры использования эконометрических моделей для решения экономических задач

Приведем решение одной типичной эконометрической задачи для напоминания об основных понятиях регрессионного анализа.

Задача 1

Имеются численные значения двух показателей: количество продавцов и розничного товарооборота по четырем выборочным однородным филиалам одной фирмы.

Таблица 1.1

База данных по четырем филиалам одной фирмы

i	X_i	$У_i$
1	1	4
2	3	6
3	2	7
4	4	10
5	5	?

i — номер филиала фирмы;

X — количество продавцов (чел.);

У — величина розничного товарооборота (тыс. руб.).

Необходимо получить прогноз розничного товарооборота для открывающегося пятого филиала фирмы, в котором ожидаемое количество продавцов равно пяти.

Полученный прогноз необходим для определения ориентировочной величины плана товарооборота для нового филиала фирмы.

Решение задачи.

Экономический анализ задачи показывает, что производительность труда в каждой фирме отличается, причинами этого могут быть шесть основных групп факторов: **персонал, машины, сырье и материалы, технология выполнения работ, среда, время**[1]. Эти группы факторов могут воздействовать на конечный результат любых процессов не только в экономике, но и во всех остальных областях деятельности человека.

Применительно к условию задачи на различие в производительности труда продавцов могут оказывать влияние следующие факторы:

• по фактору **персонал:** стаж работы; уровень квалификации; степень уравновешенности характера; семейное положение и др.;

• по фактору **машины**: уровень механизации труда; количество операций, выполняемых кассовым аппаратом и др.;

• по фактору **сырье и материалы**: наличие сертификата соответствия на реализуемый товар; количество ассортимента товаров и др.;

• по фактору **технология выполнения операций**: обслуживание покупателей продавцом; самообслуживание и т. д.;

• по фактору **среда**:

➤ *физическая среда*: место расположения; уровень: шума, влажности, давления, радиации и т. д.;

---

[1] Предложенные группы факторов взяты из системы качества при построении диаграммы Исикавы.

➤ *абстрактная среда*: уровень налогов, моральный климат в коллективе, традиции и обычаи и т. д.;

➤ по фактору **время:** так как данные анализируются за один день, то влияние фактора времени обнаружить невозможно. Однако если деятельность фирм рассматривать в динамике, то время может оказывать влияние на рассмотренные ранее факторы, которые могут оказывать влияние на результирующий признак.

Нам неизвестна зависимость У от X, поэтому выдвигаем следующие гипотезы:

Предполагаем, что зависимость У от X можно представить в виде классической нормальной линейной модели парной регрессии, при условии, что продавцы оптимально загружены и созданы нормальные условия для их работы. Эти условия обеспечивают линейную зависимость У от X. Приводим спецификацию данной модели:

$$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i,\ i = 1, 2, ..., n,$$

где У — случайная (стохастическая) величина;

X — неслучайная (детерминированная) величина;

относительно возмущения ε_i выдвигаем следующие четыре гипотезы:

а) математическое ожидание возмущения ε_i равно нулю для каждого фиксированного значения i

$$M(\varepsilon_i) = 0 \qquad \text{при всех } i = 1, 2, ..., n;$$

б) ε_i независимы

$$M(\varepsilon_i\, \varepsilon_j) = 0 \qquad \text{при } i \neq j;\ i, j = 1, 2, ..., n;$$

в) ε_i имеют одинаковую дисперсию σ_ε^2

$$M(\varepsilon_i\, \varepsilon_j) = \sigma_\varepsilon^2 \qquad \text{при } i = j;\ i, j = 1, 2, ..., n;$$

г) ε_i — нормально распределенные случайные величины со средним $\bar{\varepsilon} = 0$ и дисперсией σ_ε^2;

α_0, α_1 — неизвестные параметры;

σ_ε^2 — неизвестная дисперсия возмущения ε;

M — знак математического ожидания;

n — объем выборки.

Напомним некоторые основные понятия регрессионного анализа с помощью графического представления данных.

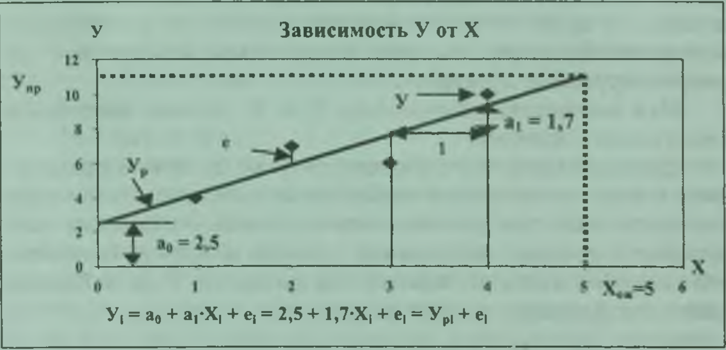


Рис. 1.1. Графический анализ данных.

Проведем аппроксимационную линию так, чтобы разницы между фактическими и расчетными значениями розничного товарооборота (остатки) были минимальными. Если говорить более строго, то необходимо так провести линию, чтобы сумма квадратов остатков была минимальной.

Для оценки параметров α_0, α_1 и возмущения ε, будем использовать коэффициенты a_0, a_1 и остатки е следующего линейного парного уравнения регрессии, вычисленные по выборочной совокупности объемом $n = 5$:

$$У_i = a_0 + a_1 X_i + e_i = У_{pi} + e_i,$$

где $У_i$ — фактические значения розничного товарооборота;

i — порядковый номер измерения;

a_0 — свободный коэффициент линейного уравнения регрессии равный значению $У_p$ при $X = 0$. В нашем примере если в фирме не будет продавцов, то товарооборот составит a_0, и если a_0 не будет равно нулю, то возникнет противоречие с содержательной частью задачи. Для устранения этого противоречия необходимо на значение a_0 наложить ограничение: $a_0 = 0$[1];

a_1 — коэффициент пропорциональности между У и X, равный приросту $У_р$ при изменении X на единицу. В нашем примере коэффициент a_1 означает среднею производительность труда продавцов;

$У_{pi} = a_0 + a_1X_i$ — расчетные значения розничного товарооборота;

$e_i = У_i - У_{pi}$ — остатки модели, учитывающие влияние факторов, не включенные в модель.

Визуальный анализ графика зависимости У от X позволяет приблизительно вычислить коэффициенты парной регрессии, при этом не станем вводить ограничения на значение $a_0 = 0$.

$a_0 = 2{,}5$
$a_1 = 1{,}7$
$У_р = 2{,}5 + 1{,}7X$

Если в уравнение регрессии подставить ожидаемое значение $X_{ож} = 5$, то можно получить прогнозное значение $У_{пр}$.

$У_{пр} = 2{,}5 + 1{,}7X_{ож} = 2{,}5 + 1{,}7\times5 = 11.$

Вывод. Для выборочной совокупности фирм зависимость розничного товарооборота от количества продавцов можно представить в виде парного линейного уравнения регрессии:

$У_i = 2{,}5 + 1{,}7X_i + e_i.$

Средняя производительность труда продавцов составляет 1,7 тыс. руб.

Если в новой фирме ожидается 5 продавцов, то среднее прогнозное значение товарооборота будет равно 11 тыс. руб.

---

[1] Если расчеты коэффициентов a_0 и a_1 проводить в Excel с помощью функции “Линейн”, то имеется возможность задать значение $a_0 = 0$ и расчет коэффициента a_1 выполнится оптимизируя остатки модели с учетом ограничения на значение a_0.

Однако коэффициенты a_0, a_1, определенные по выборке равной четырем фирмам, будут отличаться от истинных параметров α_0, α_1, которые можно определить по всей генеральной совокупности.

Приводим некоторые примеры типичных эконометрических задач, которые решаются в экономике.

Задача 2

Цена товара в определенное время зависит прежде всего от объемов его поставок в данный период времени и от цены конкурирующих товаров. Насколько изменится цена данного товара в результате изменения объемов поставок или цен конкурирующих товаров?

Задача 3

Величина спроса на определенный товар в определенное время зависит прежде всего от цены этого товара, от цены товаров, производимых конкурентами, а также от реальных доходов потребителей в данный период.

Определить, как зависит спрос на данный товар от величин названных факторов?

Задача 4

Средние издержки производства на предприятии в определенное время зависит в основном от цен и используемого количества производственных ресурсов.

Насколько сильна функциональная зависимость средних издержек производства от названных факторов?

Задача 5

Допустим, что объем сбыта продукции предприятия (У) положительно зависит от затрат на рекламу (X_1) и индекса чистоты производимой продукции (X_2), который должен быть для соответствующего предприятия важным экологическим индикатором.

Необходимо определить функциональную зависимость величины (У) от факторов (X_1) и (X_2).

Задача 6

Предприятия отрасли, работающие примерно в одинаковых условиях, имеют различные результаты хозяйственной деятельности.

Как функционально зависит рентабельность предприятия от таких определяющих факторов, как его удельный вес на рынке товаров, расходы на маркетинг, научные исследования и развитие, инвестиционные расходы?

Задача 7

Различия между избирательными округами в удельном весе голосов, отданных за определенную партию, зависит прежде всего от половозрастного состава населения, удельного веса избирателей-верующих, уровня образования населения, а также от привлекательности программ и кандидатов конкурирующих партий.

Какую форму имеет функциональная зависимость? Является ли зависимость постоянной во времени?

Задача 8

Как вероятность того, что семья купит в следующем квартале товары длительного пользования (видеомагнитофон, автомобиль и пр.), функционально зависит от таких основных факторов, как величина бюджета семьи, доход на одного человека, срок использования, имеющиеся в семье аналогичные товары длительного пользования, уровень образования, вероисповедания, общая конъюнктура рынка?

Задача 9

Имеется макроэкономическая мультипликаторная модель Кейнса. Это статическая модель, построенная на предположении о том, что народное хозяйство является системой закрытого типа без государственного регулирования экономики. Она состоит из двух уравнений, в частности из функции потребления и тождества, определяющего формирование национального дохода.

$У_{1t} = a_1 + a_2 У_{2t} + e_t,$

$$У_{2t} = У_{1t} + У_{3t},$$

где $У_{1t}$ — личное потребление в постоянных ценах за период времени t;

$У_{2t}$ — национальный доход в постоянных ценах за период времени t;

$У_{3t}$ — частные инвестиции плюс государственные расходы плюс баланс внешней торговли в постоянных ценах за период времени t;

e — случайная составляющая.

Задача 10

Имеется макроэкономическая модель Людеке. Это динамическая модель, построенная на предположении о том, что народное хозяйство является системой открытого типа с государственным регулированием экономики. Она состоит из четырех уравнений.

$$У_{1t} = a_1 + a_2У_{4t} + a_3У_{1(t-1)} + e_{1t},$$
$$У_{2t} = b_1 + b_2У_{4t} + b_3X_{1(t-1)} + e_{2t},$$
$$У_{3t} = c_1 + c_2У_{4t} + c_3У_{3(t-1)} + e_{3t},$$
$$У_{4t} = У_{1t} + У_{2t} + У_{3t} + X_{2t},$$

где $У_{1t}$ — личное потребление в постоянных ценах за период t;

$У_{2t}$ — частные чистые инвестиции и основной капитал (без резерва инвестиций) за период t;

$У_{3t}$ — импорт за период t;

$У_{4t}$ — национальный доход за период t;

$X_{1(t-1)}$ — запаздывающие на один период доходы населения, полученные от предпринимательской деятельности, дивиденды и нераспределенная прибыль предприятий до налогообложения за период t-1;

X_{2t} — государственные расходы плюс государственные чистые инвестиции и основной капитал плюс изменения в товарных запасах плюс экспорт минус косвенные налоги плюс субсидии за период t;

e_{1t}, e_{2t}, e_{3t} — случайные составляющие за период t.

Необходимо определить коэффициенты модели и прогнозные значения $У_1$, $У_2$, $У_3$, $У_4$ при известных значениях X_1 и X_2.

1.6. Элементы математической статистики

Напомним основные правила работы с массивами чисел и основные формулы математической статистики, которые потребуются нам в дальнейшем изложении курса.

Операции суммирования

Пусть величина X задается последовательностью данных X_1, X_2, ..., X_n, каждое из которых можно записать как X_i, $i = 1, ..., n$.

Сумма этих чисел записывается следующим образом:

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + ... + X_n.$$

Если из текста понятно, какие начальные и конечные суммируемые члены, то можно использовать сокращенное обозначение:

$$\sum_{i=1}^{n} X_i = \sum X_i = \sum {}_i X = \sum X.$$

Обозначим:

$\bar{X} = \frac{1}{n} \sum X_i$ — среднее значение величины X.

Правила суммирования (a, b — константы):

1. $\sum a = na.$

2. $\sum aX_i = a \sum X_i = an\bar{X}.$

3. $\sum (a + bX_i) = na + bn\bar{X}.$

4. $\sum (X_i + У_i) = \sum X_i + \sum У_i = n\,(\bar{X} + \bar{У}).$

5. $\sum(X_i - \bar{X}) = 0.$

6. $\frac{1}{n}\sum(X - \bar{X})^2 = \overline{X^2} - (X)^2.$

7. $\frac{1}{n}\sum(X_i - \bar{X})(У_i - \bar{У}) = \overline{XУ} - \overline{X}\overline{У}.$

Математическое ожидание дискретной случайной величины — это взвешенное среднее всех ее возможных значений, причем в качестве весового коэффициента берется вероятность соответствующего исхода.

Предположим, что X может принимать n конкретных значений

(X_1, X_2, ..., X_n) и что вероятность получения X_i равна p_i, тогда

$M(X) = X_1p_1 + X_2p_2 + ... + X_np_n = \Sigma X_ip_i$.

Если $p_i = n_i/n$, $i = 1, 2, ..., k$, где $\Sigma n_i = n$, то

$$M(X) = X_1p_1 + X_2p_2 + \ldots + X_kp_k =$$
$$= (X_1n_1 + X_2n_2 + \ldots + X_kn_k)/n = \left(\sum X_in_i\right)/n = \bar{X}.$$

Если $n_i = 1$ для каждого $i = 1, 2, ..., n$, то

$$M(X) = (X_1 + X_2 + \ldots + X_n)/n = \left(\sum X_i\right)/n = \bar{X}.$$

Свойства математического ожидания:

(a, b — константы; X, У — случайные величины; p — вероятность случайной величины)

1) $M(a) = a$;

2) $M(aX) = aM(X)$;

3) $M(a+bX) = a + bM(X)$;

4) $M(X + У) = M(X) + M(У)$;

5) $M(XУ) = M(X)\,M(У)$, при условии, что X и У не связаны между собой;

6) математическое ожидание функции f(X) определяется выражением:

$M(f(X)) = \Sigma f(X_i)p_i$.

Например, если $f(X) = X^2$, то $M(f(X)) = M(X^2) = \Sigma X_i^2\,p_i$.

Свойства дисперсии:

- **Дисперсия для генеральной совокупности**

$$\sigma^2 = D(X) = M(X_i - \overline{X})^2;$$

a, b — константы, X — случайная величина

1) $D(a) = 0$;
2) $D(aX) = a^2D(X)$;
3) $D(a+bX) = b^2D(X)$.

- **Дисперсия для выборочной совокупности**

$$Var(X) = \frac{1}{n}\sum(X_i - \overline{X})^2 = \overline{X^2} - \left(\overline{X}\right)^2.$$

Выборочная дисперсия var(X) является смещенной оценкой генеральной дисперсии σ^2, при этом

$$M[Var(X)] = \frac{n-1}{n}\sigma^2.$$

В качестве несмещенной оценки генеральной дисперсии используется величина

$$S^2 = \frac{n}{n-1}Var(X) = \frac{\sum(X_i - \overline{X})^2}{n-1}.$$

1) $Var(a) = 0$;
2) $Var(aX) = a^2Var(X)$;
3) $Var(a+bX) = b^2Var(X)$.

Свойства весовых коэффициентов $\omega_i = \dfrac{X_i - \overline{X}}{\sum(X_i - \overline{X})^2}$

1) $\sum\omega_i = \dfrac{\sum(X_i - \overline{X})}{\sum(X_i - \overline{X})^2} = 0;$

2) $\sum\omega_i^2 = \dfrac{\sum(X_i - \overline{X})^2}{[\sum(X_i - \overline{X})^2]^2} = \dfrac{1}{\sum(X_i - \overline{X})^2};$

3) $\sum[\omega_i(X_i - \overline{X})] = \sum\dfrac{(X_i - \overline{X})(X_i - \overline{X})}{\sum(X_i - \overline{X})^2} = \dfrac{\sum(X_i - \overline{X})^2}{\sum(X_i - \overline{X})^2} =$

$$= \frac{\sum (X_i - \bar{X}) X_i}{\sum (X_i - \bar{X})^2} = 1.$$

Основные понятия регрессионного анализа

Линейная регрессионная модель. Зависимость следствия У от причины X моделируется с помощью линейного регрессионного уравнения

$$У_i = a_0 + a_1 X_i + e_i,$$

где $У_i$ — фактическое значение товарооборота;

X_i — фактическое значение затрат на рекламу;

i — порядковый номер измерения.

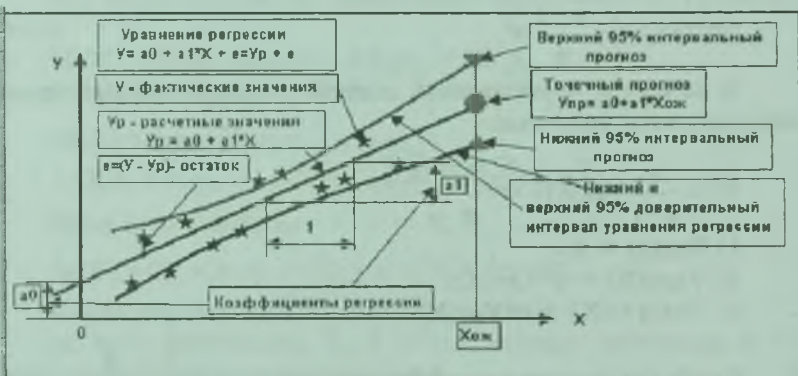


Рис. 1.2. Структура регрессионного уравнения
$У_i = a_0 + a_1 X_i + e_i = У_{pi} + e_i$

$У_{pi} = a_0 + a_1 X_i$ — расчетные значения товарооборота отражают существующую связь между У и X;

$e_i = (У_i - У_{pi})$ — остатки модели, отражают влияние неучтенных факторов;

a_1 — коэффициент модели, определенный методом наименьших квадратов, численно равный приросту значения У при изменении X на единицу.

$$a_1 = \frac{\sum(X_i - \bar{X})(У_i - \bar{У})}{\sum(X_i - \bar{X})^2},$$

где $\sum(X_i - \bar{X})(У_i - \bar{У})$ — ковариация или сумма произведений отклонений значений X и У от своих средних значений;

$\bar{X}$ — среднее значение X_i;

$\bar{У}$ — среднее значение $У_i$;

$\sum\left(X_i - \bar{X}\right)^2$ — вариация переменной X или сумма квадратов отклонений значений X от своего среднего значения;

$a_0 = \bar{У} - a_1\bar{X}$ — коэффициент модели, определенный методом наименьших квадратов, численно равный значению $У_р$ при значении X равном нулю.

Вопросы для самоконтроля

1. Дайте определение эконометрики.
2. Укажите объект, предмет, цели, задачи, методы, модели, теоретическую базу и структуру эконометрики.
3. Изложите историю эконометрики.
4. Выясните связь эконометрики с родственными науками.
5. Приведите примеры использования эконометрических методов для решения экономических задач.
6. Укажите этапы цикла Деминга улучшения процессов.
7. Укажите основные задачи выборочного метода анализа генеральной совокупности.
8. Приведите спецификацию линейной модели парной регрессии.
9. Приведите физический и экономический смысл коэффициентов a_0 и a_1 линейной регрессии.
10. Нарисуйте график зависимости У от X для любых исходных данных, проведите линию регрессии и укажите: фактические и теоретические значения, значения коэффициентов a_0 и a_1, остатки модели.

Глава 2
МОДЕЛИРОВАНИЕ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

2.1. Классификация моделей. Этапы моделирования

Изучить взаимосвязь между всеми переменными экономической системы практически невозможно, поэтому приходится упрощать объект исследования, выделять самые важные для цели исследования факторы и логические связи между переменными, иными словами, мы должны заниматься моделированием экономических процессов.

Модель — условный образ объекта исследования. Экономический объект — это сложная саморазвивающаяся динамическая система с множеством взаимосвязей между всеми ее переменными как внутри системы, так и вне ее с различными видами деятельности. Если изучать взаимосвязи между переменными, то получается множество связей, которые не ясно как использовать. Поэтому моделирование экономической системы можно проводить в следующем порядке: необходимо определить назначение экономического объекта и сформулировать цель его деятельности, для реализации цели необходимо определить задачи, которые необходимо решить, для решения задач необходимо выбрать ресурсы и эффективно их использовать. Решение каждой задачи предполагает изучение всех переменных и их связей, которые влияют на конечный результат. На каждом из этих этапов перед исследователем ставятся определенные задачи, которые надо решить.

Слово "условный" в определении модели означает наложение ограничений на переменные и условия деятельности предприятия. Например, при решении задачи используются не все переменные экономической системы, а только часть из них, причем те, которые существенно влияют на конечный результат при предположении, что для предприятия созданы нормальные условия его развития. Предполагаемая зависимость будет справедливой, если численные значения переменных не будут выходить из допустимого интервала.

Слово "образ" в определении модели означает не зеркальное отражение всех связей между переменными экономического объекта, а только самые важные, которые определяют его деятельность.

Проанализируем все виды условностей, которые предполагает модель.

Известны два вида условий создания модели: необходимые и достаточные.

Необходимые условия модели могут включать следующие положения.

Объект исследования должен представлять систему. Система — это совокупность взаимосвязанных элементов. Взаимосвязь элементов может быть трех видов: несвязанной — хаос, или полностью состоящей из случайных связей; слабосвязанной с наличием детерминированной и случайной составляющей; сильносвязанной, состоящей из детерминированной составляющей.

Система с хаотичной связью не способна создать продукт, поэтому она используется другими системами как источник ресурсов. Однако для того, чтобы воспользоваться ресурсами в хаотичной системе, должны быть обеспечены минимальные необходимые условия ее существования.

Система со слабой и средней связью элементов является типичной для большинства экономических объектов и по умолчанию является объектом изучения эконометрики.

Системы с сильной связью между элементами являются высокоорганизованными и жестко детерминированными системами, однако даже в таких системах может присутствовать небольшая случайная составляющая. Детерминированные процессы не являются объектом изучения эконометрики.

Модель должна отражать объективно существующие закономерности между переменными системы. Исследователь на основании изучения деятельности объекта как в теоретическом, так и в практическом планах выдвигает предположения о наличии определенной зависимости между переменными, при этом он должен указать, при каких условиях эта закономерность будет присутствовать. Модель конструируется субъектом так, чтобы отобразить взаимосвязи переменных, существенные для цели исследования. Например, известно, что с увеличением числа продавцов будет возрастать товарооборот однородных товаров, но при условии, что в магазине будет очередь. Если в магазине не будет очереди, то линейная зависимость между количеством продавцов и товарооборотом будет нарушена. Следовательно, необходимым условием модели должно быть указание на вид зависимости между переменными с указанием на область допустимых их значений и свойств, при которых эта зависимость будет проявляться.

Необходимо указать: метод и расчетные формулы, с помощью которых будут вычисляться оценки параметров модели и способ проверки достоверности модели.

Достаточным условием создания модели должно быть указание на область эффективного ее использования в цикле улучшения процессов, так как построение модели является не самоцелью, а средством улучшения процессов.

Примечание. Постараемся ответить на вопрос: почему между переменными экономического объекта возникают взаимосвязи? Если предприятие имеет необходимые ресурсы и технологии, но все процессы происходят хаотично, то оно не способно получить продукцию и взаимосвязь между переменными будет отсутствовать. Для выпуска продукции необходимо уменьшить хаос за счет выделения четкой структуры управления; определения прав, ответственности и взаимодействия между всеми исполнителями процессов[1]; выявления и соблюдения

---

[1] Более подробно о равностороннем треугольнике права, обязанности, взаимодействие можно ознакомиться в работе [25, с. 31–33].

принципов эффективного выполнения всех процессов с соблюдением интересов внутренних и внешних потребителей[1], а также заинтересованной среды[2]; уменьшения случайной составляющей во всех процессах. Все эти мероприятия приводят к тому, что степень хаоса уменьшается, а степень управляемости увеличивается, при этом увеличивается взаимосвязь между всеми показателями деятельности предприятия.

Можно утверждать, что при наличии полного хаоса в деятельности предприятия связь между переменными отсутствует. Чем больше порядка, организованности выполнения всех процессов, тем больше будет взаимосвязь между переменными.

Взаимосвязь между переменными на предприятии является объективной реальностью, но эта реальность является вещью в себе — она есть, но ее еще надо увидеть. Для управления процессами нужно выявить наличие этих связей (что не совсем просто), определить степень этой связи, определить форму и количественные значения этой зависимости, выяснить ограничения при которых выявленная форма будет корректно отражать существующею связь между переменными[3]. При этом мы заранее не можем знать точно о форме зависимости между переменными, поэтому исследователь выдвигает предположения или гипотезы об этой связи на основе каких-то теоретических знаний и опыта выполнения процессов, а затем проверяет их на фактических данных. Если гипотеза подтверждается, то выявленная форма связи используется в управлении. Если гипотеза отвергается, то выдвигается другая гипотеза до тех пор, пока она не подтвердится. В этом состоит суть научного метода изучения действительности.

При этом предполагается, что существующего уровня знания исследователя достаточно для понимания происходя-

---

[1] Более подробно восемь принципов эффективного выполнения процессов изложены в гл. 13.

[2] Любой статистический метод потерпит неудачу, если менеджмент не подготовит заинтересованную среду.

[3] Увидеть связь между переменными и наметить цель улучшения процессов свойственно лидеру (руководителю предприятия).

щих процессов. Если существующего уровня знаний человечества недостаточно для понимания некоторых процессов, происходящих в окружающем мире, то используется метод черного ящика, при котором фиксируются факты наличия значений переменных на входе и выходе системы. Связь между этими переменными устанавливается чисто формально с использованием различных методов.

Классификация моделей. Все модели можно разделить по средствам моделирования на материальные и абстрактные.

Материальные модели воспроизводят характеристики объекта с помощью материальных средств. Например, используя объект магазина, можно искать оптимальный вариант размещения оборудования, перемещения покупателей, путей движения товаров.

Абстрактные модели воспроизводят характеристики объекта с помощью умозаключений, которые являются плодом человеческого мышления.

По способу моделирования абстрактные модели подразделяются на графические, математические, словесно-описательные.

Графические модели — это представление взаимосвязи переменных в виде графиков, диаграмм. Визуальный анализ графических моделей является одним из самых мощных средств обнаружения вида тенденции и прогнозной оценки динамики экономического процесса, степени взаимосвязи переменных.

Графические модели позволяют провести качественный анализ взаимосвязи переменных, а также выдвинуть гипотезу о форме зависимости между переменными, которую необходимо проверить с помощью математических методов. Представление результатов моделирования в графическом виде значительно облегчает восприятие информации и значительно уменьшает текст доклада или сообщения, позволяет принять более обоснованное управленческое решение.

Математическая модель — это совокупность уравнений, отображающих зависимости между экономическими показателями или переменными.

Математические модели позволяют количественно измерить степень взаимосвязи переменных и получить расчетные значения признаков объектов.

Словесно-описательные модели — это совокупность умозаключений, которые качественно характеризуют логическую взаимозависимость экономических показателей экономической системы.

Экономические процессы происходят в пространстве и во времени, в правовой и законодательной среде общества в сочетании с потребностями членов общества при соблюдении законов сохранения ресурсов: денежных, трудовых, сырья и материалов, духовных, интеллектуальных, моральных и богатства, которое можно взять с данной территории.

Приведенная словесно-описательная модель представляет пример системного рассмотрения всех процессов, происходящих на предприятии, организации или в государстве.

Все три вида абстрактных моделей взаимно дополняют друг друга и широко используются в эконометрическом анализе.

Все экономические процессы происходят в пространстве и во времени. Следовательно, эконометрические модели могут быть пространственными и временными, а также пространственно-временными.

Пространственные модели описывают связь между переменными объектов, взятых за определенный момент времени.

Временные модели описывают динамику переменной объекта за определенные фиксированные моменты времени.

Пространственно-временные модели описывают динамику переменной нескольких объектов за определенные фиксированные моменты времени.

Этапы эконометрического моделирования. При изучении любой науки необходимо в ней выделить стержень, на котором будет находиться весь теоретический материал. В физике этим стержнем могут быть законы сохранения массы, количества движения, энергии, импульса; в теории относительности — закон независимости скорости света от его источника. В эконометрике таким стержнем нам представляются этапы эконометрического моделирования. Рассмотрение этих этапов позволяет системно осветить все основные положения эконометрики.

Этап 1. ***Постановочный.*** Формирование цели исследования. Выявление экономических проблем и выделение из них наиболее существенной. Цель исследования соответствует решению тех проблем, которые возникают у предприятия, находящегося на определенном этапе своего развития.

Этап 2. ***Априорный.*** Анализ сущности изучаемого объекта. Определяют главные и второстепенные факторы, влияющие на проблему.

Этап 3. ***Информационный.*** Сбор данных и статистической информации по всем переменным причинно-следственных связей зависимости проблемы от предполагаемых причин. Визуальный анализ графиков всех переменных и причинно-следственных связей для определения тенденций и вида зависимостей между переменными.

Этап 4. ***Спецификация*** математической модели. Определение вида математической функции, которая описывает влияние объясняемых переменных на зависимую переменную.

Этап 5. ***Идентификация модели.*** Статистический анализ модели и оценка параметров регрессионной модели.

Этап 6. ***Определение качества модели*** по обучающей выборке. Определение качества спецификации модели по обучающей выборке с использованием следующих характеристик:

E — стандартная ошибка модели;

F — критерий Фишера для проверки достоверности модели;

S_{ai} — стандартное отклонение коэффициента a_i;

t_{ai} — критерий Стьюдента для проверки достоверности коэффициента a_i.

Проверка достоверности модели и коэффициентов регрессионной модели.

Получение точечного и интервального прогноза.

Этап 7. ***Верификация модели.*** Определение адекватности модели по контрольной выборке.

Принятие решения об улучшении качества спецификации модели.

Этап 8. ***Выводы и предложения.***

Выводы и предложения должны содержать оценку качества математической модели и возможные варианты решения экономической проблемы. Выбор варианта решения проблемы предполагает использование методов оптимизации целевой функции, при соблюдении ограничений на ресурсы.

2.2. Основные свойства экономической системы, которые учитываются в моделях

Можно выделить следующие свойства экономической системы, которые можно воспроизвести эконометрической моделью:

- Все экономические процессы происходят в пространстве и во времени. Свойства времени двигаться только вперед используется во всех моделях временных рядов.
- Экономическая система — самонастраиваемая система, которая может находиться в состоянии динамического равновесия. Это свойство экономической системы используется в решении систем одновременных уравнений.
- Экономическая система обладает инерционными свойствами. Движущей силой общества являются потребности членов общества, которые нельзя быстро изменить. Также нельзя быстро изменить форму собственности, способ производства, культуру производства, производительные силы и производственные отношения. Инерционные свойства экономической системы являются методологической предпосылкой прогнозирования.
- Текущее состояние экономической системы испытывает влияния прошлых, настоящих и будущих значений пе-

ременных этой системы. Это свойство используется в авторегрессионных и автокорреляционных моделях, моделях адаптивных ожиданий и частичной корректировки, адаптивных методах прогнозирования.

- Для всех явлений в природе между причиной и следствием существует временно́й лаг, или временная задержка. Для обнаружения временно́й задержки необходимо, чтобы ее величина была больше временного шага изучения связи между причиной и следствием. Это свойство используется в моделях сосредоточенного и распределенного лага.
- Все временные экономические процессы происходят циклически. Это свойство используется в моделях сезонных и длиннопериодических волн, существующих во временных рядах.
- Последние значения временно́го ряда оказывают большее влияние на прогнозное значение, чем первые значения временно́го ряда. Это свойство реализуется с помощью взвешенной регрессии.
- Прошлые значения показателя временно́го ряда оказывают влияние на его текущее значение, но не зависит от него. Это свойство значений временно́го ряда используется в системах одновременных уравнений для получения экзогенной переменной с помощью лаговой эндогенной переменной.
- На некоторых участках временных рядов экономического показателя могут наблюдаться закономерности изменения их дисперсии. Дисперсии участков временно́го ряда можно рассматривать как зависимую переменную, численные значения которой можно прогнозировать методами регрессионного анализа. Зная амплитуду колебаний и их длительность, можно уточнить доверительные прогнозные интервалы изучаемого показателя. Использование дисперсии временного ряда как зависимой переменной является новым направлением в эконометрике (Интернет-поисковая система Rambler, ключевые слова Нобель, экономика, Engle R. E.).

2.3. Классификация переменных в эконометрических исследованиях

Микроэкономические данные количественно описывают процессы на отдельных предприятиях народного хозяйства. Микроэкономический анализ ограничивается изучением явлений и их взаимосвязей на заводах, предприятиях и учреждениях.

Макроэкономические ряды составляются на основе микроэкономических данных и характеризуют народное хозяйство или отдельные его отрасли.

Переменная — показатель экономической системы, численные значения которого изменяются.

Переменные можно классифицировать по различным признакам. По признаку принадлежности ко входу или выходу экономической системы переменные подразделяются на входные и выходные.

Входные переменные экономической системы — это ресурсы и условия существования предприятия. Например, количество работников, объемы сырья и материалов.

Выходные переменные — это результирующие показатели деятельности предприятия. Например, объем валовой продукции, прибыль.

Производные переменные — это переменные, которые получаются вследствие определенных отношений выходных показателей к входным. Например, производительность труда, фондоотдача.

Как правило, между производными переменными и их составляющими переменными существует ложная корреляция Пирсона (о типах ложной корреляции см. в глоссарии).

Как правило, входные переменные являются причинами или факторами, а выходные переменные следствиями или зависимыми переменными.

По признаку принадлежности к причине или следствию в причинно-следственных отношениях переменные подразделяются на факторы и зависимые переменные.

Фактор — причина, которая влияет на зависимую переменную.

Зависимая переменная — следствие, которое испытывает влияние со стороны факторов. Как правило, зависимыми переменными являются результаты деятельности процессов. В эконометрике недавно появился еще один вид зависимой переменной — дисперсия экономического показателя, зависящая от времени.

Примечание. Зависимую переменную не принято называть фактором.

По характеру влияния на зависимую переменную можно выделить главные и второстепенные факторы.

Главный фактор — это такая причина, без которой не будет существовать зависимая переменная.

Второстепенный фактор — это такая причина, которая влияет на зависимую переменную, но без которой зависимая переменная сможет существовать.

Например, по отношению к товарообороту ассортимент товара, спрос являются главными факторами, а реклама и культура обслуживания являются второстепенными факторами.

При построении модели обычно пользуются следующим правилом: в модель нужно включить все главные факторы и несколько второстепенных[1].

Примечание. В Японии в магазинах стремятся очень хорошо благоустроить бесплатные туалеты. Замечено, что если посетитель побывает в туалете, то в качестве благодарности что-то купит в магазине. Фактор наличия туалета в магазине является второстепенным, но приносящим ощутимую прибыль.

По признаку значений переменные подразделяются на числовые и качественные.

Числовая переменная — переменная, которая имеет дискретные или непрерывные численные значения.

---

[1] Предложенное правило было сформулировано в ответ на вопросы студентов, выполняющие дипломные работы, о том "какие факторы включать в модель?"

Примечание. По умолчанию предполагается, что численной переменной является характеристика деятельности предприятия, используемая в статистической отчетности. В эконометрических исследованиях расширяется область численных переменных и распространяется на характеристики выборочных совокупностей или временных рядов. В частности, в качестве числовой переменной может выступать дисперсия временного ряда. К этой дисперсии могут быть применимы методы регрессионного анализа, авторегрессии и автокорреляции.

Качественная переменная — переменная, значения которой принадлежат к определенному классу. Например, предприятие может быть приватизированным или неприватизированным, убыточным или прибыльным, расположенным далеко или близко от источников сырья.

По признаку преобразования переменные могут быть исходные и преобразованные.

Исходные переменные подразделяются на первичные и агрегированные.

Первичные переменные — это переменные, полученные в процессе деятельности экономической системы в реальном времени. Например, данные по каждой покупке, фиксируемые в кассовом аппарате.

Агрегированные переменные — это переменные, полученные суммированием первичных данных за определенный промежуток времени. Например, данные о сумме товарооборота, зафиксированного кассовым аппаратом за один день. При этом полностью сглаживается варьирование товарооборота в течение рабочего дня.

Преобразованные переменные — переменные, которые подвергаются определенным преобразованиям. Преобразования можно производить с помощью простых функций ($1/X$, $\ln X$ и т. д.), ортогональных преобразований, преобразований факторного анализа, делением одной переменной на другую. Очень часто при преобразованиях теряется экономический смысл новой переменной.

Дискуссия. Часто модели, построенные по преобразованным данным. действительно позволяют получить хорошие оценки параметров модели, но иногда отражают виртуальную действительность и возникают проблемы перехода от полученной модели к реальной действительности. Однако имеются удачные преобразования, которые как-то позволяют интерпретировать новые переменные, но всегда возникают проблемы перехода от точечных и интервальных прогнозов, полученных по преобразованным переменным, к прогнозам для фактических значений переменных.

По отношению к возможности получения численных значений переменные группируются на доступные и скрытные, или латентные

Доступные переменные — переменные, которые можно получить из каких либо источников.

Скрытные, или ***латентные переменные*** — переменные, численные значения которых нельзя получить. Например, реальные доходы населения являются скрытной переменной. Обычно, в модель вместо латентного фактора включают доступный фактор (инструментальная переменная), который сильно связан с латентной переменной. Например, реальные доходы населения (латентная переменная) тесно связаны с затратами на предметы роскоши (доступная инструментальная переменная).

Инструментальная переменная — переменная, которая может заменить в модели исходную переменную и обладает двумя свойствами: во-первых, она тесно связана с исходной переменной, во-вторых, она не связана с остатками модели.

Фиктивная переменная — переменная, которая количественным образом описывает качественный признак.

По времени действия переменные группируются на текущие и лаговые.

Текущие переменные — переменные, которые измерены в текущий момент времени.

Лаговые переменные — переменные, численные значения которых измерены в предшествующие моменты времени по отношению к текущим значениям зависимой переменной. Лаговая переменная обладает удивительным свойством — она может влиять, но не может быть зависимой, так как прошлое может влиять на будущее, но прошлое не зависит от текущего времени.

По отношению к месту нахождения в экономической системе переменные разделяются на внутренние (эндогенные) и внешние (экзогенные).

Эндогенные переменные — внутренние переменные, которые принадлежат к экономической системе. При этом эндогенные переменные могут влиять на другие эндогенные переменные и могут от них зависеть. Например, если в магазине имеется очередь, то увеличение продавцов приведет к увеличению товарооборота, увеличение товарооборота приведет к увеличению продавцов. Процесс увеличения продавцов и товарооборота будет продолжаться до тех пор, пока будет сохраняться очередь. Товарооборот и количество продавцов являются эндогенными переменными.

Экзогенные переменные — внешние переменные, которые влияют на переменные экономической системы, но от них не зависят. Например, суточное вращение Земли влияет на показатели экономической системы, но вращение Земли не зависит от экономической системы.

Лаговая эндогенная переменная является экзогенной[1].

В зависимости от расположения переменной во времени могут быть статические и динамические переменные.

Статические переменные — это переменные, которые изучаются в определенный фиксированный момент времени. Статическими переменными можно считать данные пространственной выборки.

Динамические переменные — это переменные, которые изучаются в течение определенного времени. Динамическими переменными можно считать данные временных рядов. Динамические переменные могут иметь средние ди-

---

[1] Это утверждение справедливо, так как прошлые значения эндогенной переменной (лаговой переменной) могут влиять на текущее значение зависимой переменной, но не зависят от текущего значения зависимой переменной, что характерно для экзогенной переменной.

намические равновесные значения. В экономической системе все эндогенные переменные связаны между собой. Если одна переменная изменилась, то по экономической системе побегут волны до установления среднединамических равновесных значений переменных, при условии что система имеет отрицательные обратные связи и колебания в системе носят затухающий характер[1]. Однако при положительной обратной связи амплитуда колебаний в системе увеличивается, система не может самонастроиться и при определенных условиях может разрушиться, например, при гиперинфляции.

Среднее динамическое равновесное значение переменной — значение эндогенной переменной, которое устанавливается стабильным в экономической системе.

Примечание. Среднее динамическое равновесное значение переменной подразделяют на статическое (установившееся за определенный интервал времени) и динамическое (изменение статических значений во времени на протяжении нескольких интервалов времени). Исследованию динамических и среднединамических равновесных значений посвящена работа [24, с. 387–397], где различают равновесные значения экономических показателей трех порядков, дан механизм возникновения периодических волн как отклонения от равновесного состояния экономики.

Синхронизирующий фактор — причина, которая оказывает влияние на все или на часть факторов, однонаправленно влияющих на зависимую переменную (праздники, катастрофы, собрание сотрудников перед выполнением работы).

Эконометрические исследования накладывают следующие ограничения на значения переменных:

• численные значения переменной должны изменяться. Это условие обусловлено тем, что нельзя определить степень влияния переменной, значения которой не изменяются. Степень синхронности изменения переменных влияет на величину критерия их связи между собой;

---

[1] Примером таких процессов является взаимовлияние спроса, цены и потребления.

- количество значений в переменной или объем выборки должен быть в 2–3 раза больше, чем количество факторов в модели;
- все переменные в модели должны иметь изменяющиеся численные значения. Если в модели используется качественная переменная, то необходимо каждому классу дать численное значение;
- численные значения переменных по возможности не должны содержать ошибок измерений;
- в значениях переменной не должно быть пропусков. Если какие-то значения в переменной отсутствуют, то следует заполнить их средними значениями или интерполяционными значениями.

Эконометрические исследования должны учитывать следующие свойства значений переменных:

- сильно выделяющееся значение зависимой переменной (выброс) может быть результатом влияния сильно изменившегося одного влияющего фактора или результатом однонаправленного воздействия большинства объясняющих переменных. Вероятность однонаправленного воздействия большинства объясняющих переменных очень мала. Поэтому причиной этого явления могло послужить изменение такого синхронизирующего фактора, который одновременно повлиял на эти объясняющие переменные. В экономических исследованиях такими синхронизирующими факторами могут быть: праздники; культурно-массовые мероприятия; стихийные бедствия; предвыборные кампании; сезонность; события, которые вызывают ожидание перемен, и другие явления, которые одновременно воздействуют на все население или только на ее часть;
- если переменные изменяются синхронно в текущие моменты времени или с временной задержкой, то эти переменные связаны между собой. Если переменные изменяются хаотично или случайным образом, то эти переменные не связаны между собой;

Мы привели перечень всех переменных, которые используются в эконометрике.

Вопросы для самоконтроля

1. Дайте определение модели.
2. Приведите классификацию моделей.
3. Приведите виды абстрактных моделей.
4. Приведите цикл Деминга улучшения процессов и этапы эконометрического моделирования.
5. Приведите основные свойства экономической системы как объекта моделирования.
6. Приведите классификацию переменных в эконометрических исследованиях.
7. Приведите основные свойства значений переменных.
8. Приведите особенности влияния синхронизирующего фактора на результирующий признак.
9. Приведите причины появления выброса у зависимой переменной.
10. Приведите условия связанности переменных.

Глава 3

ОБЩИЙ ВИД МНОЖЕСТВЕННОЙ РЕГРЕССИИ. ОЦЕНКА ПАРАМЕТРОВ МОДЕЛИ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

3.1. Выявление проблем и их причин, существующих на предприятии

Продолжаем изучение этапов эконометрического моделирования. Многие этапы эконометрического моделирования можно выполнять с помощью средств системы качества, которые собраны в гл. 14.

Этап 1. Постановочный. Выявление экономических проблем и выделение из них наиболее существенной. Формирование цели исследования.

Цель исследования соответствует тем проблемам, которые существуют у предприятия, находящегося в определенном состоянии своего развития. Известны 6 видов состояний предприятия, изображенных на рис. 3.1.

У — объем реализованной продукции в денежном выражении, X — время;

$У_{мин}$ — минимальный объем реализации продукции, при которой можно рассчитаться с затратами и получить минимальную прибыль для нормального существования предприятия;

$У_{max}$ — максимальный объем реализации продукции, при котором предприятие явно превосходит своих конкурентов, экономика стабильна и есть резерв для дальнейшего роста.

Рис. 3.1. Точечный график возможных объемов реализации продукции в определенные интервалы времени:

Выделенные состояния имеют следующие названия:
Состояние 1 — почти разорение;
Состояние 2 — чрезвычайное положение;
Состояние 3 — нормальная деятельность;
Состояние 4 — изобилие;
Состояние 5 — могущество;
Состояние 6 — опасность.

Примечание. Более подробное изучение состояний предприятий изложено в курсе "Менеджмент". Имеются перечни мероприятий, которые необходимо провести предприятию, чтобы учучшить свое состояние. Эти мероприятия могут относиться и к человеку, так как вид состояния предприятий может соответствовать человеку.

Перед предприятием, которое находится в определенном состоянии, ставятся определенные цели.

Цель должна обладать следующими свойствами:

• быть тесно согласована с общей целью деятельности организации;

• направлена на повышение удовлетворенности потребителей и совершенствование эффективности и результативности процессов;

• ставиться так, чтобы было возможным измерение показателей достигнутого прогресса;

• быть понятной, масштабной, соответствовать стоящим перед организацией задачам;

• быть разъяснена всем, кому предстоит работать в этом направлении, и согласована с ними;

• регулярно пересматриваться с тем, чтобы она отражала изменения ожиданий потребителей [22, с. 7–8].

Если цель не обладает всеми этими свойствами, то достичь ее бывает очень трудно. “Если капитан корабля не ставит перед собой никаких целей места прихода, то ни один ветер не будет попутным”. Это утверждение справедливо и для любой организации.

Примечание. В качестве цели деятельности предприятия может быть достижение уровня передового предприятия отрасли. Текущее состояние предприятия рассматривается как плацдарм для дальнейшего роста. Прошлое предприятия важно с точки зрения сложившихся темпов роста и степени эффективности используемых ресурсов, а также направлений развития благоприятных сложившихся условий его деятельности. При этом статистическая отчетность деятельности предприятия должна отражать не только темпы роста, но и степень достижения уровня передового предприятия отрасли или уровня достижения долгосрочной цели. Для передового предприятия важно сохранить лидирующее положение в отрасли, а для этого надо “бежать из всех сил”, иначе конкуренты догонят его. При этом необходимо установить долгосрочные цели, определить возможности ресурсов и направления их эффективного использования для достижения цели в заданное время в сочетании с благоприятными сложившимися условиями. Для решения подобных задач необходимы модели, учитывающие ожидаемое значение зависимой переменной и будущие значения объясняемых переменных.

Постановка и решение подобных задач в настоящее время очень актуальны, так как 12 августа 2004 г. правительство РФ публично объявило и целях своей деятельности, сроках их выполнения, численных значениях ожидаемых основных показателей и ответственности министерств за их выполнение. При этом указанные цели обладают всеми свойствами в соответствии с системой качества.

Частным случаем решения подобных задач может быть следующее направление. Допустим имеется пространственная выборка двух переменных X и У, в которых точка ($X_{мах}$ и $У_{мах}$) характеризует самое передовое предприятие. Постулируется линейная зависимость между У и X: $У = a_0 + a_1X + e$. При этом в расчетных формулах для коэффициентов a_0 и a_1 вместо $\bar{X}$ и $\bar{У}$ можно поставить $X_{мах}$ и $У_{мах}$.

$$a_0 = \bar{У} - a_1 X_{мах}, \quad a_1 = \frac{\sum (X_i - X_{мах})(У_i - У_{мах})}{\sum (X_i - X_{мах})^2}.$$

Тогда уравнение регрессии пройдет точно через точку ($X_{мах}$ и $У_{мах}$).

В качестве точки ($X_{мах}$ и $У_{мах}$) можно выбрать планируемые значения: $У_{пл}$ и $X_{пл}$. В дальнейшем можно ввести экономическую интерпретацию полученной модели. Суть ее будет состоять в том, что отклонения всех переменных будет определяться не от средних значений, а от значений, характеризующих или передовое предприятие, или значение долгосрочной планируемой цели.

Для выделенной цели подбираются задачи, для решения которых составляются мероприятия. В дальнейшем проводится анализ выполненных мероприятий в соответствии с поставленной целью, и формируются управленческие решения по дальнейшему совершенствованию процессов

Этап 2. Априорный. Анализ сущности изучаемого объекта. Определяют главные и второстепенные факторы, влияющие на проблему. После определения главной проблемы, которую следует решать, необходимо провести анализ этой проблемы; выявить, какая переменная отражает в численном выражении значение этой проблемы; определить, какие причины X (факторы) влияют на зависимую переменную У.

Определение главных и второстепенных факторов, оказывающих влияние на зависимый признак У. Методом мозговой атаки определяются основные факторы, способные оказывать влияние на проблему. Любые проблемы зависят в разной степени от следующих основных причин (факторов):

• условия окружающей среды (физическая среда: температура, влажность, давление и степень загрязнения воздуха, освещенность рабочего места, уровень шума и радиации и т. д.; абстрактная среда: правовая и законодательная среда деятельности предприятия, стиль руководства, моральный климат и традиции, существующие на предприятии и т. д.);

• оборудование; материалы; измерения; методы; люди; время.

Этап 3. Информационный. После определения предполагаемых факторов, влияющих на зависимую переменную У, необходимо проверить эти связи на фактических данных. Для этой цели надо собрать данные по всем переменным причинно-следственных связей зависимости проблемы от предполагаемых причин. Произвести визуальный анализ графиков для всех переменных и причинно-следственных связей для определения тенденций и вида зависимостей между переменными. Рассчитать матрицу парных коэффициентов корреляции и проверить наличие линейной связи между переменными.

В результате выполнения третьего этапа исследователь должен выявить причины (факторы), которые влияют на проблему. Визуальный анализ графиков должен определить регулярности, зависимости между проблемой и причинами, ее вызывающими. При этом выдвигаются гипотезы о форме тенденции этой связи, которые должны проверяться с помощью их моделирования. На следующем этапе необходимо определить вид математической функции, которая воспроизведет визуально обнаруженную тенденцию.

Визуальный анализ графиков показателей является самым мощным средством определения закономерностей между переменными. Не случайно в Японии все инструменты менеджмента качества имеют визуальное представление. Однако очень сложно формализовать процесс визуального анализа. Часто закономерность, обнаруженную визуальным анализом, очень трудно выявить математически, но всегда можно воспроизвести с помощью математической модели. Приводим попытку формализации визуального анализа получения прогноза по исходным данным, представленным графически на рис. 3.4.

Рис. 3.4. График динамики розничного товарооборота и первых разниц:

i — порядковый номер месяца;

$У_{факт}$ — розничный товарооборот магазина (тыс. руб.);

$У_i - У_{i-1}$ — первые разности.

Численные значения временного ряда имеют следующие характеристики:

характеристики подъемов:

- количество серий подъемов 7 шт.
- минимальная длительность серий подъемов 1 мес.
- максимальная длительность серий подъемов 4 мес.
- средняя длительность серии подъемов 2,5 мес.
- минимальная величина размера первого подъема 2 тыс. руб.
- максимальная величина размера первого подъема 7 тыс. руб.
- средняя величина размера первого подъема 4,5 тыс. руб.
- минимальная величина серий подъемов 1 тыс. руб.

- максимальная величина серий подъемов 8 тыс. руб.
- средняя величина серий подъемов 4,5 тыс. руб.

характеристики спадов:

- количество спадов 6 шт.
- минимальная длительность спадов 1 мес.
- максимальная длительность спадов 1 мес.
- средняя длительность спадов 1 мес.
- минимальная величина размера первого спада −1 тыс. руб.
- максимальная величина размера первого спада −4 тыс. руб.
- средняя величина размера первого спада −2,5 тыс. руб.
- минимальная величина серий спадов −1 тыс. руб.
- максимальная величина серий спадов −4 тыс. руб.
- средняя величина серий спадов −2,5 тыс. руб.

характеристика одинаковых значений:

- количество серий одинаковых значений 0 шт.
- минимальная длительность одинаковых значений 0 мес.
- максимальная длительность одинаковых значений 0 мес.
- средняя длительность одинаковых значений 0 мес.

характеристики серий подъемов, спадов и одинаковых значений:

- количество серий чередований подъемов и спадов 6 шт.
- количество серий чередований подъемов и одинаковых значений 0 шт.
- количество серий чередований спадов и одинаковых значений 0 шт.

характеристики прогнозного периода

- ожидается подъем, спад или одинаковый уровень по сравнению с последним значением временного ряда спад
- **численное значение временного ряда на прогнозный период** 56,5 тыс. руб.

Дискуссия. Очень перспективным в изучении тенденций является использование чисел Фибоначчи — представляющих последовательность 0, 1, 1, 2, 3, 5, 8, ..., в которых каждый последующий член равен сумме двух предыдущих. При этом каждый последующий член ряда зависит от всех предыдущих и в большей степени испытывает влияние от последнего значения этого ряда. Этими свойствами обладают все экономические временные ряды.

Числа Фибоначчи реально существуют в природе, и их изучению посвящены монографии. При этом интересна следующая цепочка логических рассуждений: если в природе существуют числа Фибоначчи, а человек является частью природы, то характеристики человека должны подчиняться этим числам, а так как экономические показатели зависят от деятельности человека, то они тоже должны испытывать влияние этих чисел. Таким образом, в последовательности значений вре́менных рядов экономических показателей должны проявляться закономерности чисел Фибоначчи. Такие исследования уже проводились, и был обнаружен идеальный цикл, состоящий из пяти подъемов и трех спадов. Согласно этому циклу, если во временном ряду цен на товар наблюдается пять подъемов и затем два спада, то очень велика вероятность того, что следующее значение будет спадом. Анализ рис. 3.4 показывает, что в наших данных такого идеального цикла нет.

Если проводить аналогию экономики и природы, то можно поставить соответствие растения экономике. Например, ель по своей кроне напоминает структуру управления, темпы роста камыша и бамбука напоминает темпы роста развивающихся стран, размещение веток и листьев на дереве напоминает структуру организации и количество сотрудников в отделах и т. д. [37].

3.2. Спецификация модели

После обнаружения причин (факторов), влияющих на проблему, и выдвижения гипотез, как о наличии регулярностей в зависимой переменной, так и форме влияния факторов, необходимо составить их абстрактную модель в виде математической функции. Четвертый этап посвящен спецификации модели.

Этап 4. Спецификация математической модели.

Спецификация математической модели — определение такой математической функции, которая должна воспроизводить определенное количество закономерностей зависимой переменной.

Определение вида математической функции, которая описывает влияние факторов на зависимую переменную.

Общий вид множественного регрессионного уравнения для выборочных данных можно представить следующим математическим выражением, которое имеет следующую структуру:

$$У_i = f(X_{1i}, X_{2i}, X_{3i}, ...) + e_i = У_{pi} + e_i,$$

где $У_i$ — случайная зависимая переменная;

$У_{pi}$ — расчетные значения зависимой переменной;

$f(X_{1i}, X_{2i}, X_{3i},...)$ — математическая функция, отражающая детерминированную составляющую $У_i$, которая диктуется законодательной, правовой средой общества в сочетании с потребностями членов общества;

X_{1i}, X_{2i}, X_{3i} — причины, объясняющие переменные (факторы), оказывающие существенное влияние на $У_i$;

$e_i = У_i - У_{pi}$- случайная составляющая (остатки) $У_i$, которая учитывает влияние факторов, не вошедшие в модель;

i — порядковый номер измерения $i = 1, ..., n$;

n — объем выборки.

В зависимости от количества переменных в модели различают парные и множественные модели.

Приводим общий вид линейного парного регрессионного уравнения

$$У_i = a_0 + a_1 X_i + e_i = У_{pi} + e_i,$$

где a_0 — свободный коэффициент, равный $У_p$ при $X = 0$;

a_1 — коэффициент пропорциональности зависимости $У_р$ от X, численно равный приросту $У_р$ при изменении X на 1.

Однако линейная модель парной регрессии не является единственной. Существует несколько видов моделей, которые используются в эконометрике. Для их определения необходимо выявить механизм генерации всех видов эконометрических моделей, которые уже известны и которых еще нет.

Генерация эконометрических моделей. В настоящее время в научной и учебной литературе можно встретить большое количество различных эконометрических моделей, которые необходимо классифицировать, и предсказать появление новых пока неизвестных моделей. Для этой цели необходимо выявить механизм генерации этих моделей.

Располагая механизмом генерации эконометрических моделей можно существенно улучшить известные модели и расширить область их использования в экономических исследованиях, а также может появиться возможность системного изложения видов моделей в учебном процессе.

Генерацию эконометрических моделей можно производить с использованием белого шума δ_t и регрессионного уравнения следующего вида:

$$У_t = f(X_t) + \varepsilon_t,$$

где У — зависимая переменная;

X — объясняемая переменная, количество объясняемых переменных может быть разным;

$f(X_t)$ — общий вид математической функции;

t — текущее время, время может быть прошлым и будущим:

(t − м) — прошлое время, сдвинутое на м дат назад от текущего времени t;

(t + м) — будущее время, сдвинутое на м дат вперед от текущего времени t;

$X_{t-м}$ — лаговая переменная порядка м;

$X_{t+м}$ — будущее значение объясняемой переменной порядка м (пока не имеет специального названия);

$У_{t-м}$ — лаговая зависимая переменная порядка м;

$У_{t+м}$ — будущее значение зависимой переменной порядка м (пока не имеет специального названия);

ε — случайное возмущение или ошибка модели, включающая ошибку уравнения и ошибку измерения;

$\varepsilon_{t-м}$ — лаговая переменная возмущения порядка м;

$\varepsilon_{t+м}$ — будущее значение возмущения порядка м (пока не имеет специального названия);

δ_t — белый шум;

$\delta_{t-м}$ — лаговые значения белого шума;

$\delta_{t+м}$ — будущее значение белого шума.

Элементы линейного регрессионного уравнения показаны на рис. 3.5.

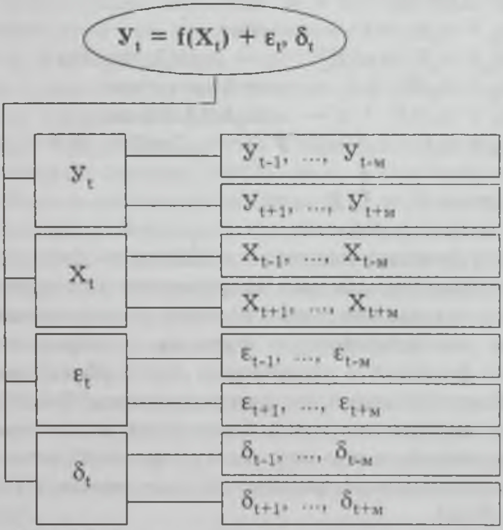


Рис. 3.5. Структура эконометрических моделей

Над переменными можно производить различные преобразования, например, по следующим функциям: Z^2, $Z^{1/2}$,

$1/Z$, $\ln Z$, $\sin Z$, $\cos Z$, $\Delta Z_t = Z_t - Z_{t-1}$, где Z — условное обозначение любой переменной: $У_t$, X_t, ε_t, δ_t, t.

Предлагаем следующий механизм генерации эконометрических моделей — можно комбинировать элементы регрессионного уравнения и получать различные виды моделей. Некоторые из этих моделей уже известны и хорошо изучены, другие мало изучены, однако могут появиться абсурдные или неизвестные модели, которые представляют особый интерес.

Генерацию эконометрических моделей начнем с определения вида математической функции $f(X)$.

В эконометрике часто используются следующие виды математических функций, отражающие определенный вид тенденции зависимости У от X:

$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$ — линейная;

$У_i = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \varepsilon_i$ — параболическая;

$У_i = \alpha_0 + \alpha_1 / X_i + \varepsilon_i$ — гиперболическая;

$У_i = \alpha_0 + \alpha_1 \ln X_i + \varepsilon_i$ — логарифмическая;

$У_t = \alpha_0 + \alpha_1 t + \alpha_2 \sin(2\pi t/T) + \alpha_3 \cos(2\pi t/T) + \varepsilon_t$ — периодическая,

где t — время ($t = 1, 2, \ldots, n$);

T — период колебания.

Все эти функции являются линейными относительно коэффициентов модели, так как коэффициенты находятся в первой степени и сами не стоят в степени к переменной. Можно составлять комбинированную функцию, которая состоит из нескольких функций с различными преобразованиями переменных. Известен пакет прикладных программ "TableCurve2D", который позволяют получить более двух тысяч комбинированных функций, и все они являются линейными относительно коэффициентов, расчет которых можно повторить с помощью Excel.

В эконометрике имеются функции, которые учитывают определенные свойства экономической системы:

$У_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \varepsilon_i$ — многофакторная линейная аддитивная модель, где X_1, X_2, X_3 — факторы, оказывающие влияние на У.

$У_i = \alpha_0 X_1^{\alpha_1} X_2^{\alpha_2} \varepsilon_i$ — многофакторная мультипликативная модель Кобба-Дугласа.

Аддитивность — сложение.

Мультипликативность — умножение.

Построим серии моделей с учетом лаговых переменных.

$X_{t-м}$, $У_{t-м}$, $\varepsilon_{t-м}$, $S_{\varepsilon\ t-м}^2$

$У_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{1(t-1)} + \alpha_3 X_{1(t-2)} + \varepsilon_t$ — модель распределенного лага второго порядка РЛ(2) — зависимость последующих значений зависимой переменной от двух предыдущих значений объясняемой переменной.

$У_{pt} = \alpha_0 + \alpha_1 У_{t-1} + \alpha_2 У_{t-2} + \varepsilon_t$ — авторегрессионная модель второго порядка АР(2), учитывающая влияние на У их же двух предыдущих значений.

Скомбинируем эти две модели и получим:

$У_{pt} = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{1(t-1)} + \alpha_3 X_{1(t-2)} + \alpha_4 У_{t-1} + \alpha_5 У_{t-2} + \varepsilon_t$ — модель распределенного лага второго порядка и авторегрессии второго порядка РЛАР(2, 2) .

$\varepsilon_t = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-1} + \nu_t$ — модель авторегрессии возмущения второго порядка АР(2), где ν_t — случайная составляющая. Модель авторегрессии возмущения первого порядка принято называть автокорреляцией возмущения.

$\varepsilon_t = \rho \varepsilon_{t-1} + \delta_t$ — модель автокорреляции возмущения АК или авторегрессии возмущения первого порядка АР(1), где ρ — параметр автокорреляции первого порядка.

Если в линейную модель

$У_t = \alpha_0 + \alpha_1 X_t + \varepsilon_t$

вставить модель авторегрессии остатков второго порядка, то получится следующая математическая запись

$У_t = \alpha_0 + \alpha_1 X_t + \alpha_2 \varepsilon_{t-1} + \alpha_3 \varepsilon_{t-2} + \nu_t$ — линейная модель с учетом АР(2) возмущения (авторегрессия возмущения второго порядка).

Если во временном ряду наблюдаются регулярности изменений остатков, то в качестве переменной для изучения может выступать дисперсия остатков $S_{\varepsilon\ t}^2$, численные значения которой можно получить следующим образом:

• приведем временной ряд к стационарному виду, например с помощью линейной функции

$У_t = a_0 + a_1 t + e_t$;

• вычислим остатки

$e_t = У_t - (a_0 + a_1 t)$;

• разобьем временной ряд на равные участки, в которых квадраты остатков будут примерно одинаковыми, вычислим для каждого участка дисперсию остатков $S_e{}^2{}_i$, где i — номер участка.

Построим модель зависимости дисперсии остатков от своих прошлых значений

$S_e{}^2{}_i = a_0 + a_1 S_e{}^2{}_{i-1} + a_2 S_e{}^2{}_{i-2} + v_i$ — модель авторегрессии дисперсии остатков второго порядка АР(2) или авторегрессия гетероскедастичности (при условии неравенства дисперсий остатков между собой, что является признаком наличия гетероскедастичности).

Рассмотрим модели с участием белого шума.

Авторегрессию возмущения первого порядка можно представить с помощью белого шума δ_t

$\varepsilon_t = \rho\varepsilon_{t-1} + \delta_t = \delta_t + \rho\delta_{t-1} + \rho^2\delta_{t-2} + \rho^3\delta_{t-3} + \ldots + \rho^\infty\delta_{t-\infty}$,

которую называют моделью скользящего среднего СС бесконечного порядка [1, с. 262].

Важно отметить, что с помощью белого шума δ_t можно воспроизвести стационарные возмущения ε_t.

В эконометрике могут использоваться комбинации этих моделей.

$У_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{1(t-1)} + \alpha_3 У_{t-1} + \alpha_4 \varepsilon_{t-1} + \delta_t$ — модель распределенного лага, авторегрессии и автокорреляции возмущения РЛАРАК(1,1,1).

Для моделирования временных рядов можно использовать в качестве фактора белый шум. Например, в модель распределенного лага, авторегрессии вместо автокорреляции возмущения поставим авторегрессию белого шума d_t.

$$У_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{1(t-1)} + \alpha_3 У_{t-1} + \delta_t + \alpha_4 \delta_{t-1} + \alpha_5 \delta_{t-2} + \nu_t$$

Если монотонную тенденцию исключить разностями определенного порядка и использовать их как зависимую переменную, которую воспроизведем ее авторегрессией с интегрированной (суммирование) авторегрессией белого шума, то получится модель Бокса-Дженкинса.

Модель Бокса-Дженкинса авторегрессии первого порядка, интегрирован-ной первого порядка, скользящей средней первого порядка АРИСС(1, 1, 1) имеет следующий вид:

$$У_t - У_{t-1} = \Delta У_t = \alpha_1 \Delta У_{t-1} + \delta_t + \alpha_2 \delta_{t-1}.$$

$У_i/|e_i| = a_0/|e_i| + a_1(X_i/|e_i| + e_t/|e_t|$ — модель взвешенной регрессии, устраняющая гетероскедастичность остатков.

Если экономический процесс описывается линейной моделью

$$У_t = a_0 + a_1 X_t + e_t,$$

в которой X зависит от У и эту связь можно выразить уравнением

$$X_t = в_0 + в_1 У_t + v_t,$$

то получается система одновременных уравнений.

$$\left.\begin{array}{l} У_t = a_0 + a_1 X_t + e_t \\ X_t = в_0 + в_1 У_t + v_t \end{array}\right\}$$

Класс моделей систем одновременных уравнений хорошо изучен, имеются эффективные алгоритмы определения коэффициентов модели, а также методы определения прогнозных значений эндогенных переменных.

Примечание. Если в качестве факторов использовать будущие значения объясняемой переменной то получается класс моделей сосредоточенного и распределенного будущего (пока этот класс моделей не получил специального названия).

Например, модель сосредоточенного будущего порядка м имеет следующий вид —

$$У_t = a_0 + a_1 X_t + a_2 X_{t+м} + e_t.$$

Обсудим модель сосредоточенного будущего.

Во-первых, нет принципиальных трудностей в составлении базы данных с переменными $У_t$, X_t, $X_{t+м}$, с помощью которой методом наименьших квадратов (МНК) можно определить коэффициенты модели.

Во-вторых, порядок м сосредоточенного будущего не должен быть слишком большим, так как при прогнозировании придется прогнозировать объясняемую переменную на большую глубину, с увеличением прогнозного доверительного интервала.

В-третьих, представляет особый интерес влияние будущих выбросов объясняемой переменной на выбросы зависимой переменной, а также влияние будущего выброса на все значения зависимой переменной.

В-четвертых, известно, что адаптивные и авторегрессионные модели обладают недостатком, который заключается в том, что расчетные значения этих моделей плохо воспроизводят выбросы. Включение в модель будущих значений факторов способно учитывать будущие значения выбросов.

Модель распределенного будущего второго порядка имеет следующий вид:

$$У_t = a_0 + a_1X_t + a_2X_{t+1} + a_3X_{t+2} + e_t,$$

в которой численные значения X_t не должны изменяться с равным интервалом. Для определения коэффициентов этой модели используется МНК.

Если в качестве факторов использовать лаговые переменные зависимой переменной, то получается класс авторегрессионных моделей.

Например, модель авторегрессии порядка м имеет следующий вид:

$$У_t = a_0 + a_1X_t + a_2У_{t-м} + e_t.$$

Если в качестве факторов использовать будущие значения зависимой переменной, то получается класс моделей авторегрессии сосредоточенного и распределенного будущего (пока этот класс моделей не получил специального названия).

Например, модель авторегрессии сосредоточенного будущего порядка м имеет следующий вид:

$$У_t = a_0 + a_1X_t + a_2У_{t+м} + e_t.$$

Модель авторегрессии распределенного будущего второго порядка имеет следующий вид:

$$У_t = a_0 + a_1X_t + a_2У_{t+1} + a_3У_{t+2} + e_t,$$

в которой измерения У не должны равняться между собой. Для определения коэффициентов этой модели в первом приближении можно использовать МНК.

Если в качестве факторов использовать будущие значения остатков, то получается класс моделей сосредоточенного и распределенного будущего остатков (пока этот класс моделей не получил специального названия).

Например, модель сосредоточенного будущего остатков порядка м имеет следующий вид:

$$У_t = a_0 + a_1X_t + a_2e_{t+м} + v_t.$$

Модель распределенного будущего остатков второго порядка имеет следующий вид:

$$У_t = a_0 + a_1X_t + a_2e_{t+1} + a_3e_{t+2} + v_t.$$

Для определения коэффициентов этой модели в первом приближении можно использовать МНК.

Выбор вида функции производится на основе экономической теории изучаемого процесса при соблюдении основных принципов построения модели: простота и ясность экономической интерпретации.

Параметры модели оцениваются методом наименьших квадратов или обобщенным методом наименьших квадратов.

Если выбранная функция не соответствует экономическому процессу, то ее заменяют другой [11].

3.3. Идентификация модели

После определения вида модели необходимо оценить параметры этой модели.

Этап 5. Идентификация модели. Статистический анализ модели и оценка параметров регрессионной модели.

Рассмотрим простейшую модель экономического процесса, состоящую из одного уравнения, которое содержит только две переменные. Обозначив переменные буквами У и X, мы постулируем между ними зависимость

$$У = f(X) + \varepsilon.$$

Содержательные соображения, на основе которых было принято введенное соотношение, должны подсказать и его конкретную функциональную форму или указать на дополнительные условия, которым должна обладать кривизна функции. Поскольку одним и тем же условиям могут удовлетворять различные функции, то необходимо обратиться к статистическому анализу и осуществить выбор среди возможных альтернативных вариантов. Простейшим соотношением между двумя переменными является линейная связь между ними:

$$У = \alpha_0 + \alpha_1 X + \varepsilon,$$

где α_0, α_1 — неизвестные параметры,

ε — случайное возмущение, которое отражает влияние тех факторов, которые не вошли в модель, случайности человеческих реакций, ошибок наблюдений или измерений.

Спецификация взаимосвязи между переменными должна сопровождаться рядом гипотез о свойствах распределения вероятностей для случайного возмущения. Этими гипотезами являются следующие утверждения:

- средние значения возмущений для каждого фиксированного значения X равны нулю;
- дисперсии возмущений являются постоянными и не зависящими от X;
- различные значения е не зависят друг от друга и не зависят от X.

Если гипотеза о линейной зависимости справедлива и если предположения относительно характера случайного возмущения удовлетворяются, то возникшая при этом ситуация может быть представлена на рис. 3.7.

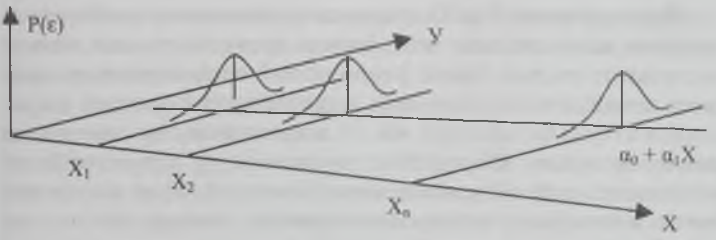


Рис. 3.7. График зависимости У от X:
P(ε) — плотность распределения случайного возмущения

Если бы нам удалось собрать все пары значений X_i $У_i$, относящиеся к генеральной совокупности экономического процесса, то можно построить диаграмму рассеивания, изображенную на рис. 3.8.

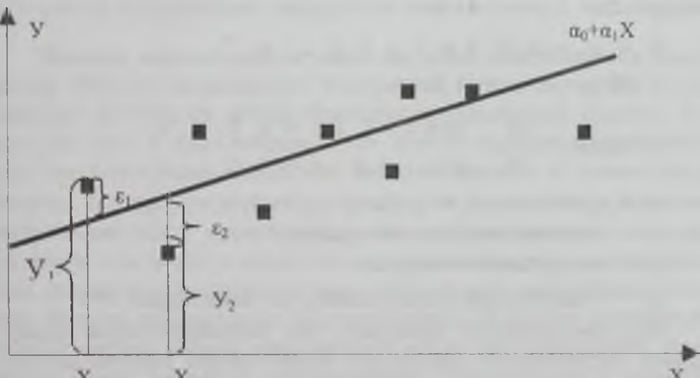


Рис. 3.8. Диаграмма рассеивания У и X для генеральной совокупности

Если значения У и X отражают деятельность предприятия за определенный интервал времени, а количество предприятий, составляющих генеральную совокупность, является очень большим, то мы не сможем собрать данные по всей генеральной совокупности, следовательно нам не известны значения α_0 и α_1.

Если значения У и X отражают деятельность одного предприятия за несколько интервалов времени, то мы можем располагать только одной реализацией экономического процесса предприятия. Повторить экономический процесс в прошлые интервалы времени мы не в состоянии, так как время движется только вперед. Если величины α_0 и α_1 отражают зависимость для всех возможных значений У, но на протяжении нескольких интервалов времени реализуется только их небольшая часть, следовательно, нам не известны значения α_0 и α_1.

На практике линия $\alpha_0 + \alpha_1 X$ отсутствует, ее следует определить, используя следующие гипотезы.

Оценки параметров модели, определенные методом наименьших квадратов, будут обладать свойствами: несмещенности, состоятельности, эффективности, если выполняются следующие предпосылки:

$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$, $i = 1, 2, ..., n$.

а) $M(\varepsilon_i) = 0$ — при всех i,

б) $M(\varepsilon_i \varepsilon_j) = \begin{cases} 0 & \text{при } i \neq j,\ i, j = 1, 2, ..., n \\ \sigma_\varepsilon^2 & \text{при } i = j;\ i, j = 1, 2, ..., n \end{cases}$

$M(X_i \varepsilon_j) = 0$ при $i = j$, $i \neq j$; $i, j = 1, 2, ..., n$,
где M — математическое ожидание,

ε_i — возмущения модели,

σ_ε^2 — дисперсия возмущений, не зависящая от i.

Эта предпосылка означает, что автокорреляция возмущений отсутствует, дисперсии возмущений гомоскедастичны (однородны).

с) Факторы не должны быть связанными между собой, или между объясняемыми переменными отсутствует мультиколлинеарность.

д) Возмущения должны иметь совместное нормальное распределение.

е) У — случайная переменная, не должна содержать ошибочных измерений.

ж) X — детерминированная или случайная переменная.

з) Объем выборки n должен быть больше числа коэффициентов в модели.

и) Значения объясняемых переменных не должны зависеть от возмущений модели.

Параметры α_0, α_1 и σ_{ε}^2 неизвестны. Мы хотим на основании наших выборочных наблюдений над У и X не только статистически оценить эти параметры, но и проверить по отношению к ним некоторые гипотезы. Служит ли линейная зависимость адекватным отражением эмпирических данных? Оправдана ли в свете этой выборки гипотеза о постоянстве дисперсии возмущающей составляющей при всех значениях X?

Известно несколько методов оценки параметров модели. Сравнительный анализ этих методов см. в книге [6, с. 33–34). Широкое распространение оценки параметров модели получил метод наименьших квадратов.

Оценка параметров модели методом наименьших квадратов. Методу наименьших квадратов, предложенному Гауссом для воспроизведения траектории движения планет, более 200 лет. В дальнейшем этот метод стал использоваться для биологических объектов, экономических и социальных систем. Метод наименьших квадратов продолжает активно использоваться в эконометрических исследованиях благодаря тому, что в его основе лежит предположение о нормальном законе распределения анализируемых случайных величин. Многие переменные биологических объектов и экономических систем имеют нормальный закон распределения. Этот метод имеет хорошее теоретическое обоснование и практическую апробацию для различных объектов. Основные положения метода наименьших квадратов рассмотрим на конкретном примере.

Пример

Имеются численные значения двух показателей: количество продавцов и розничного товарооборота по четырем выборочным однородным филиалам одной фирмы.

Таблица 3.2

База данных по четырем филиалам одной фирмы

№	X	У
1	1	4
2	3	6
3	2	7
4	4	10
5	5	?

где № — номер филиала фирмы;

X — количество продавцов (чел.);

У — величина розничного товарооборота (тыс. руб.).

Необходимо оценить параметры α_0, α_1 линейной модели

$$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$$

с помощью вычисления коэффициентов a_0 и a_1 парной линейной регрессии по выборочным данным У и X:

$$У_i = a_0 + a_1 X_i + e_i.$$

Решение. Необходимо определить такие значения коэффициентов a_0 и a_1, при которых сумма квадратов остатков регрессионной модели

$$У_i = a_0 + a_1 X_i + e_i$$

будут минимальными.

Представим уравнение $У_i = a_0 + a_1 X_i + e_i$ в развернутом виде:

$$4 = 1a_0 + 1a_1 + e_1;$$
$$6 = 1a_0 + 3a_1 + e_2;$$
$$7 = 1a_0 + 2a_1 + e_3;$$
$$10 = 1a_0 + 4a_1 + e_4.$$

Представим полученную систему уравнений в матричном виде:

$$\mathbf{У} = \mathbf{XA} + \mathbf{e},$$

где $\mathbf{У} = \begin{pmatrix} 4 \\ 6 \\ 7 \\ 10 \end{pmatrix}$, $\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 2 \\ 1 & 4 \end{pmatrix}$, $\mathbf{A} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$, $\mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{pmatrix}$.

Метод наименьших квадратов заключается в том, что он позволяет определить такие значения коэффициентов a_0 и a_1, при которых сумма квадратов остатков будет минимальной:

$$\mathbf{e'e} \Rightarrow \min, \text{ где } \mathbf{e} = \mathbf{У} - \mathbf{XA}.$$

$$\mathbf{e'e} = (\mathbf{У} - \mathbf{XA})'(\mathbf{У} - \mathbf{XA}) = (\mathbf{У'} - \mathbf{A'X'})(\mathbf{У} - \mathbf{XA}) =$$
$$= \mathbf{У'У} - \mathbf{A'X'У} - \mathbf{У'XA} + \mathbf{A'X'XA} =$$
$$= \mathbf{У'У} - 2\mathbf{A'X'У} + \mathbf{A'X'XA} \tag{3.1}$$

Так как **У'ХА** величина скалярная, то она не меняется при транспонировании, т. е.

$$\mathbf{У'XA} = (\mathbf{У'XA})' = \mathbf{A'X'У}.$$

Примечание. Свойства операций транспонирования.
А, **В**, **С** — матрицы;
$(\mathbf{A})' = \mathbf{A}$;
$(\mathbf{A}+\mathbf{B})' = \mathbf{A'} + \mathbf{B'}$;
$(\mathbf{AB})' = \mathbf{B'A'}$;
$(\mathbf{ABC})' = \mathbf{C'B'A'}$.

Возьмем первую частную производную от **e'e** по переменной **А**, приравняем ее к нулю и определим оценки параметров модели, при которых **e'e** будет иметь наименьшее значение .

$$\frac{\partial(\mathbf{e'e})}{\partial \mathbf{A}} = \frac{\partial}{\partial \mathbf{A}}(\mathbf{У'У} - 2\mathbf{A'X'У} + \mathbf{A'X'XA}) = -2\mathbf{X'У} + 2\mathbf{X'XA} = 0.$$

Примечание. Для вектора частных производных справедливы следующие формулы:

$$\frac{\partial(\mathbf{B'C})}{\partial \mathbf{B}} = \mathbf{C};$$

$$\frac{\partial(\mathbf{B'DB})}{\partial \mathbf{B}} = 2\mathbf{DB},$$

где **В**, **С** — вектор-столбцы, **D** — симметрическая матрица.

Доказательство см. [5, с. 84].

Полагая, что **А** и **Х'У** векторы-столбцы, найдем

$$\frac{\partial(-2\mathbf{A'X'У})}{\partial \mathbf{A}} = -2\mathbf{X'У}.$$

Полагая, что **X'X** — симметрическая матрица, **A** — вектор-столбец, найдем

$$\frac{\partial(\mathbf{A'X'XA})}{\partial \mathbf{A}} = 2\mathbf{X'XA}.$$

После несложных преобразований можно получить несколько выражений.

X'XA = X'У — система нормальных уравнений, представленная в матричном виде.

$\mathbf{A} = (\mathbf{X'X})^{-1}\mathbf{X'У}$ — расчетная формула коэффициентов линейной модели или оценка параметров модели для генеральной совокупности. Данная формула позволяет рассчитывать коэффициенты многофакторной линейной модели для любого количества факторов.

$(\mathbf{X'X})^{-1}$ — обратная матрица, полученная от матрицы **X'X**, является матрицей ошибок оценок параметров модели.

X' — транспонированная матрица, полученная от матрицы **X**.

Расчетную формулу коэффициентов модели следует запомнить, так как она будет применяться на протяжении всего курса эконометрики.

Можно расчетные формулы коэффициентов уравнения регрессии представить в развернутом виде.

Необходимо оценить параметры α_0, α_1 линейной модели

$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$

с помощью вычисления коэффициентов a_0 и a_1 парной линейной регрессии:

$У_i = a_0 + a_1 X_i + e_i = У_{pi} + e_i$, при условии, что $\Sigma e_i^2 \to \min$, где $У_{pi} = a_0 + a_1 X_i$; $e_i = У_i - У_{pi}$;

$$\Sigma e_i^2 = \Sigma(У_i - У_{pi})^2 = \Sigma(У_i^2 - 2У_i У_{pi} + У_{pi}^2) =$$
$$= \Sigma[У_i^2 - 2У_i(a_0 + a_1 X_i) + (a_0 + a_1 X_i)^2] =$$
$$= \Sigma У_i^2 - 2a_0 \Sigma У_i - 2a_1 \Sigma X_i У_i + n a_0^2 + 2a_0 a_1 \Sigma X_i + a_1^2 \Sigma X_i^2$$

Возьмем от выражения Σe_i^2 первую производную по a_0, затем по a_1, приравняем их к нулю и получим систему нормальных уравнений:

$$\begin{cases} \dfrac{\partial}{\partial a_0}\sum e_i^2 = -2\sum У_i + 2na_0 + 2a_1\sum X_i = 0 \\ \dfrac{\partial}{\partial a_1}\sum e_i^2 = -2\sum X_i У_i + 2a_0\sum X_i + 2a_1\sum X_i^2 = 0 \end{cases}$$

После несложных преобразований получаем систему нормальных уравнений в развернутом виде:

$$\begin{cases} na_0 + a_1\sum X_i = \sum У_i, \\ a_0\sum X_i + a_1\sum X_i^2 = \sum X_i У_i. \end{cases}$$

Систему нормальных уравнений в развернутом виде можно получить другим способом, если выполнить матричные операции системы нормальных уравнений, записанных в матричном виде:

$X'XA = X'У$.

Решим систему нормальных уравнений, представленной в развернутом виде:

Из первого уравнения определим a_0

$$a_0 = \frac{\sum У_i}{n} - a_1\frac{\sum X_i}{n} = \bar{У} - a_1\bar{X}.$$

Найденное значение a_0 подставим во второе уравнение и произведем преобразования:

$$(\bar{У} - a_1\bar{X})\sum X_i + a_1\sum X_i^2 = \sum(X_i У_i);$$

$$a_1(\sum X_i^2 - \bar{X}\sum X_i) = \sum(X_i У_i) - \bar{У}\sum X_i;$$

$$a_1 = \frac{\sum(X_i У_i) - \bar{У}\sum X_i}{\sum X_i^2 - \bar{X}\sum X_i} = \frac{\sum(X_i - \bar{X})(У_i - \bar{У})}{\sum(X_i - \bar{X})^2} =$$

$$= \frac{\sum(X_i - \bar{X})У_i}{\sum(X_i - \bar{X})^2} = \sum(\omega_i У_i).$$

Так как $\sum(X_i - \bar{X})(У_i - \bar{У}) =$

$$= \sum[(У_i - \bar{У})X_i - (У_i - \bar{У})\bar{X}] = \sum(У_i - \bar{У})X_i - \sum(У_i - \bar{У})\bar{X} =$$

$$= \sum(У_i - \bar{У})X_i - \bar{X}\sum(У_i - \bar{У}) = \sum(У_i - \bar{У})X_i - \bar{X} \times 0 =$$

$$= \sum(У_i - \bar{У})X_i = \sum(X_i У_i) - \bar{У}\sum X_i;$$

$$\sum(X_i - \bar{X})(У_i - \bar{У}) = \sum[(X_i - \bar{X})У_i - (X_i - \bar{X})\bar{У}] =$$

$$= \sum[(X_i - \bar{X})У_i - \sum(X_i - \bar{X})\bar{У}] = \sum[(X_i - \bar{X})У_i -$$

$$- У_i\sum(X_i - \bar{X})] = \sum(X_i - \bar{X})У_i - У_i \times 0 = \sum(X_i - \bar{X})У_i;$$

$$\sum(X_i - \bar{X})^2 = \sum(X_i - \bar{X})(X_i - \bar{X}) =$$

$$= \sum(X_i - \bar{X})X_i - \bar{X}\left[\sum(X_i - \bar{X}) = 0\right] = \sum X_i^2 - \bar{X}\sum X_i;$$

$$\omega_i = \frac{X_i - \bar{X}}{\sum(X_i - \bar{X})^2}.$$

Представим формулы расчета коэффициентов a_1 a_0 в развернутом виде:

$$a_1 = \frac{\sum(X_i - \bar{X})(У_i - \bar{У})}{\sum(X_i - \bar{X})^2},$$

$$a_0 = \bar{У} - a_1\bar{X},$$

где $\bar{X}, \bar{У}$, — средние значения переменных X и У.

Знак коэффициента a_1 зависит от числителя $\sum(X_i - \bar{X}) \times \times (У_i - \bar{У})$. Если величины X_i и $У_i$ изменяются синхронно и в одной фазе, то числитель будет иметь большое положительное значение. Если X_i и $У_i$ будут изменяться синхронно и в противофазе, то числитель будет иметь большое отрицательное значение. Если X_i и $У_i$ будут изменяются случайным образом, то числитель будет равен нулю и $a_1 = 0$, а переменные X и У не будут иметь линейной связи.

Примечание. Большинство учебных пособий по регрессионному анализу предлагают проводить расчеты коэффициента a_1 по формуле

$$a_1 = \frac{\sum(X_i У_i) - \bar{У}\sum X_i}{\sum X_i^2 - \bar{X}\sum X_i},$$

позволяющей проще выполнять расчеты на калькуляторе. Если числа X и У имеют много разрядов и объем выборки достаточно большой, то числитель и знаменатель могут иметь большие ошибки округлений, так как значения числителя и знаменателя являются разностью между накопленными суммами и если эти суммы превышают разрядность ЭВМ, то ошибки округления будут неизбежны. Много лет назад при выполнении расчетов парного коэффициента корреляции по аналогичной упрощенной формуле на ЭВМ "Проминь2", у нас один раз получился коэффициент корреляции больше 1, именно из-за ошибки округления больших накопленных сумм. С тех пор нигде и никогда не пользуюсь этими упрощенными формулами.

Идентифицируемость модели. Идентифицируемость модели — возможность вычисления коэффициентов регрессионной модели методом наименьших квадратов.

Метод наименьших квадратов применим к линейной аддитивной функции:

$$У_p = a_0 + a_1X_1 + a_2X_2 + \ldots$$

Известны некоторые регрессионные модели, коэффициенты которых непосредственно нельзя вычислить методом наименьших квадратов, но их можно привести к линейному аддитивному виду, как правило, логарифмированием переменных. Но есть такие модели, которые нельзя привести к аддитивному виду и для них используется приближенные итерактивные методы вычисления коэффициентов модели.

Например, в степенной функции $У_{p_i} = a_0(X_i)^{a1}$ коэффициент a_1 является степенью фактора X, следовательно, она не является линейной относительно коэффициентов. Степенная функция приводится к линейному виду (с помощью процесса линеаризации) путем логарифмирования левой и правой частей уравнения.

В эконометрике понятие идентифицируемости модели распространяется на возможность вычисления коэффициентов структурной системы одновременных уравнений по коэффициентам приведенной системы одновременных уравнений. Для этого случая вводятся понятия строгой идентифицируемости, недостаточной идентифицируемости и сверхидентифицируемости, которые будут изучаться позже.

Коэффициенты регрессионной модели нельзя вычислить методом наименьших квадратов, если факторы тесно связаны между собой.

Расчет коэффициентов модели теряет смысл, если объем выборки равен или меньше количества коэффициентов в модели.

3.4. Свойства оценок параметров модели

Оценки параметров модели, определенные методом наименьших квадратов, будут несмещенными, состоятельными и эффективными, если соблюдаются предпосылки этого метода.

Несмещенность — свойство оценок параметров модели, которое заключается в том, что математическое ожидание коэффициентов модели должно равняться их истинному значению при любом объеме выборки. Условие несмещенности часто записывают в виде $M_\beta(b) = \beta$ или $M(b) = \beta$ [6, с. 533].

Несмещенность означает, что при расчете оценки мы не получим систематической ошибки.

Состоятельность — свойство оценок параметров модели, которое заключается в том, что с ростом объема выборки численное значение коэффициента модели должно стремиться к соответствующему параметру генеральной совокупности.

Состоятельность оценки гарантирует приближение оценки к истинному значению параметра при увеличении объема выборки.

Эффективность — свойство оценок параметров модели, которое заключается в том, что для выборок равного объема они должны иметь минимальную дисперсию.

Эффективность оценки является наилучшей в смысле минимума среднеквадратического отклонения.

Несмещенность и эффективность — это свойства, которые не зависят от объема выборки. Состоятельность является асимптотическим свойством оценки при стремлении n к бесконечности [6, с. 533–534].

При первом чтении доказательства свойств оценок параметров модели можно пропустить.

Проверим полученные оценки параметров линейной модели на несмещенность по следующим шагам:

шаг 1 — уточнить постановку задачи;

шаг 2 — в формулу расчета коэффициента a_1 (или a_0) необходимо ввести параметр α_1 (или α_0);

шаг 3 — для полученного выражения взять математическое ожидание;

шаг 4 — выполнить анализ полученного результата.

Проверим на несмещенность формулу расчета коэффициента a_1, полученной методом наименьших квадратов.

Шаг 1. Уточним постановку задачи. Мы располагаем следующими объектами для анализа:

1) Модель зависимости У от Х для генеральной совокупности

$$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i,$$

при соблюдении следующих ограничений:

а) $M(\varepsilon_i) = 0$ — при всех i;

б) $M(\varepsilon_i \varepsilon_j) = \begin{cases} 0 & \text{при } i \neq j,\ i, j = 1, 2, ..., n \\ \sigma_\varepsilon^2 & \text{при } i = j,\ i, j = 1, 2, ..., n; \end{cases}$

$M(X_i \varepsilon_j) = 0$ при $i = j$; $i \neq j$; $i, j = 1, 2, ..., n$,

где М — математическое ожидание;

ε_i — возмущения модели;

σ_ε^2 — дисперсия возмущений, не зависящая от i.

2) Имеется модель зависимости У от Х для выборочной совокупности

$$У_i = a_0 + a_1 X_i + e_i$$

3) Коэффициент a_1 вычисляется по следующей формуле

$$a_1 = \frac{\sum(X_i - \overline{X})(У_i - \overline{У})}{\sum(X_i - \overline{X})^2} = \frac{\sum(X_i - \overline{X})У_i}{\sum(X_i - \overline{X})^2} = \sum(\omega_i У_i)$$

4) Имеется гипотеза. На интуитивном уровне мы предполагаем, что коэффициент α_1, вычисленный по выборочной

совокупности будет отличаться от истинного параметра α_1, вычисленного для данных всей генеральной совокупности. Интересно знать величину этой разницы и от каких переменных она зависит.

Шаг 2. В формулу расчета коэффициента a_1 необходимо ввести параметр α_1.

В формуле расчета коэффициента a_1

$$a_1 = \frac{\sum(X_i - \bar{X})(У_i - \bar{У})}{\sum(X_i - \bar{X})^2} = \frac{\sum(X_i - \bar{X})У_i}{\sum(X_i - \bar{X})^2} = \sum(\omega_i У_i)$$

вместо $У_i$ (фактических значений выборочной совокупности) подставим значения, вычисленные по истинной модели

$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$

$$\begin{aligned} a_1 &= \sum(\omega_i У_i) = \sum\omega_i(\alpha_0 + \alpha_1 X_i + \varepsilon_i) = \\ &= \alpha_0\sum\omega_i + \alpha_1\sum\omega_i X_i + \sum\omega_i\varepsilon_i = \\ &= \alpha_0\left(\sum\omega_i = 0\right) + \alpha_i\left[\left(\sum\omega_i X_i\right) = 1\right] + \sum\omega_i\varepsilon_i = 0 + \alpha_1 + \sum\omega_i\varepsilon_i \end{aligned}$$

$$a_1 = \alpha_1 + \sum(\omega_i\varepsilon_i) = \alpha_1 + \frac{\sum(X_i - \bar{X})\varepsilon_i}{\sum(X_i - \bar{X})^2},$$

где в преобразованиях использовались следующие правила:

$\sum\omega_i\alpha_0 = \alpha_0\sum\omega_i$ (α_0 — константа и ее можно выносить за знак суммы);

$\sum\omega_i = 0,\quad \sum(\omega_i X_i) = 1,$ где $\omega_i = \dfrac{X_i - \bar{X}}{\sum(X_i - \bar{X})^2}$ (доказательство этих соотношений уже приводилось).

Вывод. Выборочный коэффициент a_1

$$a_1 = \alpha_1 + \sum(\omega_i\varepsilon_i) = \alpha_1 + \frac{\sum(X_i - \bar{X})\varepsilon_i}{\sum(X_i - \bar{X})^2}$$

состоит из двух частей: своего истинного значения α_1 и случайной составляющей, которая называется ошибкой коэффициента, зависящей от взаимодействия остатков ε_i с объясняемой переменной X_i

Шаг 3. Вычислим математическое ожидание коэффициента a_1.

$$M(a_1) = M(\alpha_1) + M\frac{\sum(X_i - \bar{X})\varepsilon_i}{\sum(X_i - \bar{X})^2} =$$

$$= M(\alpha_1) + M\frac{\sum(\varepsilon_i X_i - \varepsilon_i \bar{X})}{\sum(X_i - \bar{X})^2} = \alpha_1,$$

где $M\sum(\varepsilon_i X_i - \varepsilon_i \bar{X}) = \sum[M(\varepsilon_i X_i) - M(\varepsilon_i \bar{X})] =$

$$= \sum[M(\varepsilon_i)M(X_i) - \bar{X}M(\varepsilon_i)] = \sum[0 \times M(X_i) - \bar{X} \times 0)] = 0;$$

$M(\varepsilon_i) = 0$ согласно предпосылке метода наименьших квадратов;

$M(\varepsilon_i X_i) = M(\varepsilon_i)M(X_i)$, при условии, что ε_i и X_i не связаны между собой, а так как $M(\varepsilon_i) = 0$, то $M(\varepsilon_i)M(X_i) = 0$, или согласно предпосылке метода наименьших квадратов $M(\varepsilon_i X_i) = 0$.

Если ε_i и X_i будут связаны между собой, то $M(\varepsilon_i X_i) \neq$ $\neq M(\varepsilon_i)M(X_i)$ и нужны дополнительные исследования.

Примечание. При определении математического ожидания частного

$$M\frac{\sum(\varepsilon_i X_i - \varepsilon_i \bar{X})}{\sum(X_i - \bar{X})^2}$$

мы предполагали, что математическое ожидание знаменателя не равно нулю, так сумма квадратов числа всегда будет положительным числом, а математическое ожидание положительного числа будет тоже положительным и отличным от нуля, поэтому было корректно определять математическое ожидание числителя. В общем случае математическое ожидание частного не равно математическому ожиданию числителя, деленному на математическое ожидание знаменателя в случаях возникновения неопределенности типа 0/0, если математическое ожидание числителя и знаменателя равно нулю.

Шаг 4. Математическое ожидание коэффициента a_1 равно истинному значению α_1, следовательно коэффициент a_1, определенный методом наименьших квадратов является несмещенным.

Анализ полученного результата. Ответим на вопрос, что необходимо предпринять исследователю, чтобы уменьшить ошибку коэффициента a_1?

Ошибка коэффициента a_1 равна

$$\frac{\sum(\varepsilon_i X_i - \varepsilon_i \overline{X})}{\sum(X_i - \overline{X})^2}.$$

Для уменьшения ошибки необходимо увеличивать знаменатель с помощью увеличения размаха значений объясняемого фактора X, а также уменьшать числитель с помощью подбора таких факторов, которые не связаны с возмущениями при каждом фиксированном значении i. Так как

$$M\sum(\varepsilon_i X_i) = \sum M(\varepsilon_i X_i) = M(\varepsilon_1 X_1) + M(\varepsilon_2 X_2) + \ldots + M(\varepsilon_n X_n).$$

Пример 1. При увеличении значений X_i амплитуда возмущения возрастает, но математическое ожидания для каждого значения i равно нулю (гетероскедастичность). Как эта зависимость окажет влияние на величину

$$\sum M(\varepsilon_i X_i) = ?$$

С увеличением X_i возрастает амплитуда возмущения, а не математическое ожидание, следовательно X_i и e_i не связаны между собой математическими ожиданиями, тогда

$$\begin{aligned}\sum M(\varepsilon_i X_i) &= M(\varepsilon_1 X_1) + M(\varepsilon_2 X_2) + \ldots + M(\varepsilon_n X_n) = \\ &= M(\varepsilon_1)M(X_1) + M(\varepsilon_2)M(X_2) + \ldots + M(\varepsilon_n)M(X_n) = \\ &= 0 \times M(X_1) + 0 \times M(X_2) + \ldots + 0 \times M(X_n) = 0.\end{aligned}$$

Следовательно при наличии гетероскедастичности коэффициент a_1 будет несмещенным.

Пример 2. В системах одновременных уравнений зависимая случайная переменная может выступать как случайная объясняемая переменная $У_i$, при этом для каждого фиксированного значения $У_i$ математические ожидания возмущения не равно нулю и, следовательно, имеется зависимость $У_i$ от возмущения e_i, тогда

$$\sum M(\varepsilon_i У_i) = M(\varepsilon_1 У_1) + M(\varepsilon_2 У_2) + \ldots + M(\varepsilon_n У_n) \neq 0$$

и коэффициент a_1 будет смещенным. Для устранения смещения необходимо вместо случайной объясняемой переменной поставить ее теоретические значения. Для каждого фиксированного теоретического значения зависимой переменной ($У_{тi}$) математические ожидания возмущений равно нулю, тогда будет справедливо соотношение

$\sum M(\varepsilon_i У_{тi}) = M(\varepsilon_1 У_{т1}) + M(\varepsilon_2 У_{т2}) + ... + M(\varepsilon_n У_{тn}) = 0$ и коэффициент a_1 будет несмещенным. Описанная процедура называется двухшаговым методом наименьших квадратов. Более подробно об этом методе изложено при изучении систем одновременных уравнений.

Проверим на несмещенность формулу расчета коэффициента a_0, полученной методом наименьших квадратов.

Шаг 1. Уточним постановку задачи. Мы располагаем следующими объектами для анализа:

1) Имеется модель зависимости У от X для генеральной совокупности

$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i,$

при соблюдении следующих ограничений:

а) $M(\varepsilon_i) = 0$ — при всех i,

б) $M(\varepsilon_i \varepsilon_j) = \begin{cases} 0 & \text{при } i \neq j,\ i, j = 1, 2, ..., n \\ \sigma_\varepsilon^2 & \text{при } i = j,\ i, j = 1, 2, ..., n \end{cases}$

$M(X_i \varepsilon_j) = 0$ при $i = j$; $i \neq j$; $i, j = 1, 2, ..., n$,

где M — математическое ожидание,

ε_i — возмущения модели,

σ_ε^2 — дисперсия возмущений, не зависящая от i.

2) Имеется модель зависимости У от X для выборочной совокупности

$У_i = a_0 + a_1 X_i + e_i.$

3) Коэффициент a_0 вычисляется по следующей формуле:

$a_0 = \overline{У} - a_1 \overline{X}.$

4) Имеется гипотеза. На интуитивном уровне мы предполагаем, что коэффициент a_0, вычисленный по выборочной

совокупности будет отличаться от истинного параметра α_0, вычисленного для данных всей генеральной совокупности. Интересно знать величину этой разницы и от каких переменных она зависит.

Шаг 2. В формулу расчета коэффициента a_0 необходимо ввести параметр α_0.

В уравнении вычисления a_0 имеется коэффициент a_1, который надо заменить его формулой.

Подставим a_1

$$a_1 = \frac{\sum(X_i - \bar{X})У_i}{\sum(X_i - \bar{X})^2}$$

в уравнение определения a_0

$$a_0 = \bar{У} - a_1\bar{X}.$$

$$a_0 = \bar{У} - \frac{\bar{X}\sum(X_i - \bar{X})У_i}{\sum(X_i - \bar{X})^2} = \sum\left(\frac{1}{n} - \omega_i\bar{X}\right)У_i.$$

В уравнении определения коэффициента a_0 имеется фактическое значение зависимой переменной $У_i$. Подставим вместо $У_i$ значения, полученные по истинному уравнению

$$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i.$$

При этом фактические значения $У_i$ и значения, вычисленные по истинному уравнению будут равны между собой.

$$a_0 = \Sigma\left(\frac{1}{n} - \omega_i\bar{X}\right)(\alpha_0 + \alpha_1 X_i + \varepsilon_i) =$$

$$\alpha_0 - \alpha_0\bar{X}\left(\sum\omega_i = 0\right) + \alpha_1\bar{X} - \alpha_1\bar{X}\left(\sum\omega_i X_i = 1\right) + \sum\left(\frac{1}{n} - \bar{X}\omega_i\right)\varepsilon_i =$$

$$= \alpha_0 - 0 + \alpha_1\bar{X} - \alpha_1\bar{X} + \sum\left(\frac{1}{n} - \bar{X}\omega_i\right)\varepsilon_i = \alpha_0 + \sum\left(\frac{1}{n} - \bar{X}\omega_i\right)\varepsilon_i.$$

$$a_0 = \alpha_0 + \sum\left(\frac{1}{n} - \bar{X}\omega_i\right)\varepsilon_i = \alpha_0 + \sum\left(\frac{1}{n} - \bar{X}\frac{X_i - \bar{X}}{\sum(X_i - \bar{X})^2}\right)\varepsilon_i.$$

Шаг 3. Вычислим математическое ожидание коэффициента a_0.

$$M(a_0) = M\left(a_0 + \sum\left(\frac{1}{n} - \overline{X}\frac{X_i - \overline{X}}{\sum(X_i - \overline{X})^2}\right)\varepsilon_i\right) =$$

$$= M(a_0) + M\sum\left(\frac{\varepsilon_i}{n}\right) - M\sum\left(\frac{\overline{X}(X_i - \overline{X})}{\sum(X_i - \overline{X})^2}\varepsilon_i\right) =$$

$$= \alpha_0 + \sum\left(\frac{M(\varepsilon_i) = 0}{n}\right) - \sum\left(\frac{\overline{X}M(X_i\varepsilon_i) - \overline{X}^2 M(\varepsilon_i)}{M\sum(X_i - \overline{X})^2}\right) =$$

$$= \alpha_0 + 0 - \sum\left(\frac{\overline{X}M(X_i)[M(\varepsilon_i) = 0] - \overline{X}^2[M(\varepsilon_i) = 0]}{M\sum(X_i - \overline{X})^2}\right) =$$

$$= \alpha_0 + 0 - 0 = \alpha_0$$

$M(a_0) = \alpha_0$

Анализ условий несмещенности коэффициента a_0 полностью совпадает с уже проводимым анализом коэффициента a_1.

Шаг 4. Так как математическое ожидание коэффициента a_0 равно истинному значению α_0, то коэффициент a_0, определенный методом наименьших квадратов, является несмещенным.

Выводы, полученные для коэффициента a_1 полностью распространяются на коэффициент a_0.

Вычисление дисперсии коэффициентов a_1 и a_0.

Вычислим дисперсию коэффициента a_1.

Для оценки параметра α_1 имеем:

$a_1 = \alpha_1 + \sum(\omega_i\varepsilon_i)$.

Дисперсию коэффициента a_1 относительно параметра α_1 можно вычислить по формуле

$$Var(a_1) = M(a_1 - \alpha_1)^2 = M\left[\sum(\omega_i\varepsilon_i)^2\right] = M(\omega_1^2\varepsilon_1^2 + \ldots + \omega_n^2\varepsilon_n^2 +$$
$$+ 2\omega_1\omega_2\varepsilon_1\varepsilon_2 + \ldots + 2\omega_{n-1}\omega_n\varepsilon_{n-2}\varepsilon_n) = \sigma_\varepsilon^2\sum\omega_i^2 = \sigma_\varepsilon^2 / \sum(X_i - \overline{X})^2,$$

где $M(\varepsilon_i^2)=\sigma_\varepsilon^2$; $M(\varepsilon_i\varepsilon_j)=0$, при $i \neq j$, $\sum\omega_i^2 = 1/\sum(X_i-\bar{X})^2$.

Окончательно получим дисперсию коэффициента a_1

$Var(a_1)=\sigma_\varepsilon^2/\sum(X_i-\bar{X})^2$.

Вычислим дисперсию коэффициента a_0.

Для оценки параметра α_0 имеем:

$$a_0=\alpha_0+\sum\left(\frac{1}{n}-\bar{X}\omega_i\right)\varepsilon_i.$$

Дисперсию коэффициента a_0 относительно своего параметра α_0 вычислим по следующей формуле

$$Var(a_0)=M(a_0-\alpha_0)^2=M\left[\sum(1/n-\bar{X}\omega_i)\varepsilon_i\right]^2=$$

$$=M\left[\varepsilon_i^2\sum(1/n-\bar{X}\omega_i)^2\right]=\sigma_\varepsilon^2\sum(1/n-\bar{X}\omega_i)^2=$$

$$=\sigma_\varepsilon^2\left[1/n+\bar{X}^2\sum\omega_i^2-(2\bar{X}/n)\sum\omega_i\right]=$$

$$=\sigma_\varepsilon^2\left[1/n+\bar{X}^2\sum\omega_i^2-0\right]=\sigma_\varepsilon^2\left[1/n+\bar{X}^2/\sum(X_i-\bar{X})^2\right]=$$

$$=\sigma_\varepsilon^2\sum X_i^2/\left[n\sum(X_i-\bar{X})^2\right],$$

где $\omega_i=(X_i-\bar{X})/\sum(X_i-\bar{X})^2$; $\sum\omega_i^2=1/\sum(X_i-\bar{X})^2$; $\sum\omega_i=0$.

Для получения преобразования $1/n+\bar{X}^2/\sum(X_i-\bar{X})^2=$ $=\sum X_i^2/\left[n\sum(X_i-\bar{X})^2\right]$ было использовано следующее соотношение:

$$\sum(X_i-\bar{X})^2=\sum\left(X_i^2-2X_i\bar{X}+\bar{X}^2\right)=\sum X_i^2-2\bar{X}\sum X_i+n\bar{X}^2=$$

$$\sum X_i^2-2\left[\left(\sum X_i\right)/n\right]\sum X_i+n\left[\left(\sum X_i\right)/n\right]\left[\left(\sum X_i\right)/n\right]=\sum X_i^2-n(\bar{X})^2.$$

Окончательно получим дисперсию коэффициента a_0:

$$Var(a_0)=\left(\sigma_\varepsilon^2\sum X_i^2\right)/\left[n\sum(X_i-\bar{X})^2\right].$$

Дисперсии коэффициентов a_0 и a_1 содержат дисперсию случаной составляющей σ_ε^2, которую необходимо оценить с помощью выборочных данных.

Оценка дисперсии случайной составляющей σ_ε^2

Правдоподобной выглядит попытка оценить дисперсию возмущающей составляющей σ_ε^2 путем возведения в квадрат и последующего усреднения отклонений наблюдаемых значений от линии, полученной методом наименьших квадратов.

По выборочным данным определим зависимость У от Х с помощью уравнения регрессии

$$У_i = a_0 + a_1 X_i + e_i. \tag{3.1}$$

Коэффициенты a_0 и a_1 определим методом наименьших квадратов, при этом справедливо соотношение (см. первое уравнение нормальных систем уравнений)

$$\overline{У} = a_0 + a_1\overline{X}. \tag{3.2}$$

Вычтем из (3.1) уравнение (3.2), получим уравнение регрессии, с преобразованными переменными

$$У_i - \overline{У} = a_1\left(X_i - \overline{X}\right) + e_i,$$

$$e_i = \left(У_i - \overline{У}\right) - a_1\left(X_i - \overline{X}\right). \tag{3.3}$$

Известно, что зависимость У от Х для данных всей генеральной совокупности определяется по уравнению регрессии

$$У_{pi} = \alpha_0 + \alpha_1 X_i.$$

Применим это уравнение к выборочным данным и получим равенство

$$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i. \tag{3.4}$$

Применим его к выборочным данным, усредним по всем n значениям нашей выборки и получим выражение

$$\overline{У} = \alpha_0 + \alpha_1\overline{X} + \overline{\varepsilon}. \tag{3.5}$$

При этом среднее значение $\overline{\varepsilon} \neq 0$, так как параметры α_0, α_1 вычислялись по данным генеральной совокупности, а не по выборочным данным.

Из уравнения (3.4) вычтем уравнение (3.5), получим выражение

$$У_i - \overline{У} = \alpha_1\left(X_i - \overline{X}\right) + \left(\varepsilon_i - \overline{\varepsilon}\right).$$

Подставим его в уравнение (3.3) и получим выражение

$$e_i = \alpha_1\left(X_i - \bar{X}\right) + \left(\varepsilon_i - \bar{\varepsilon}\right) - a_1\left(X_i - \bar{X}\right) = -\left(a_1 - \alpha_1\right)\left(X_i - \bar{X}\right) + \left(\varepsilon_i - \bar{\varepsilon}\right).$$

Возведем в квадрат e_i и просуммируем по всем n значениям выборочной совокупности.

$$\sum e_i^2 = \left(a_1 - \alpha_1\right)^2 \sum\left(X_i - \bar{X}\right)^2 + \sum\left(\varepsilon_i - \bar{\varepsilon}\right)^2 -$$
$$-2\left(a_1 - \alpha_1\right)\sum\left(X_i - \bar{X}\right)\left(\varepsilon_i - \bar{\varepsilon}\right).$$

Возьмем математическое ожидание для левой и правой части этого уравнения.

$$M\left(\sum e_i^2\right) = M\left[\left(a_1 - \alpha_1\right)^2 \sum\left(X_i - \bar{X}\right)^2\right] + M\left[\sum\left(\varepsilon_i - \bar{\varepsilon}\right)^2\right] -$$
$$-M\left[2\left(a_1 - \alpha_1\right)\sum\left(X_i - \bar{X}\right)\left(\varepsilon_i - \bar{\varepsilon}\right)\right].$$

Правая часть уравнения состоит из трех составляющих, математические ожидания которых вычислим по отдельности, а затем соберем их вместе.

Определим математическое ожидание первой составляющей дисперсии остатков выборочной совокупности:

$$M\left[\left(a_1 - \alpha_1\right)^2 \sum\left(X_i - \bar{X}\right)^2\right] = M\left(a_1 - \alpha_1\right)^2 M\left[\sum\left(X_i - \bar{X}\right)^2\right] =$$
$$= M\left[\left(a_1 - \alpha_1\right)^2\right]\sum\left(X_i - \bar{X}\right)^2 = \sigma_\varepsilon^2 .$$

Так как $\operatorname{Var}(a_1) = M\left(a_1 - \alpha_1\right)^2 = \sigma_\varepsilon^2 / \sum\left(X_i - \bar{X}\right)^2$.

Определим математическое ожидание второй составляющей дисперсии остатков выборочной совокупности:

$$M\sum\left(\varepsilon_i - \bar{\varepsilon}\right)^2 = M\left[\sum\left(\varepsilon_i\right)^2 - \left(\sum \varepsilon_i\right)^2 / n\right] =$$
$$= M\left[\sum\left(\varepsilon_i\right)^2\right] - M\left[\left(\sum \varepsilon_i\right)^2 / n\right] = n\sigma_\varepsilon^2 - \sigma_\varepsilon^2 = (n-1)\sigma_\varepsilon^2.$$

Полученные преобразования выполнялись по следующим шагам:

Шаг 1. Необходимо разложить выражение $\sum\left(\varepsilon_i - \bar{\varepsilon}\right)^2$ на составляющие части.

$$\sum\left(\varepsilon_i - \bar{\varepsilon}\right)^2 = \sum\left(\varepsilon_i\right)^2 - \left(\sum \varepsilon_i\right)^2 / n.$$

Докажем справедливость этого разложения на привычных данных

$$\sum\left(X_i-\bar{X}\right)^2=\sum\left(X_i^2-2X_i\bar{X}+\bar{X}^2\right)=$$
$$=\sum X_i^2-2\bar{X}\sum X_i+n\sum X_i\sum X_i/n^2=$$
$$=\sum X_i^2-2\sum X_i\sum X_i/n+\sum X_i\sum X_i/n=$$
$$=\sum X_i^2-\sum X_i\sum X_i/n=\sum X_i^2-\left(\sum X_i\right)^2/n.$$

Вывод завершен.

Шаг 2. Возьмем математическое ожидание от полученного разложения

$$M[\sum\left(\varepsilon_i-\bar{\varepsilon}\right)^2]=M[\sum\left(\varepsilon_i\right)^2-\left(\sum\varepsilon_i\right)^2/n].$$

Используя свойство математического ожидания M(X+У) = = M(X)+M(У) продолжим преобразования

$$M\sum\left(\varepsilon_i-\bar{\varepsilon}\right)^2=M[\sum\left(\varepsilon_i\right)^2-\left(\sum\varepsilon_i\right)^2/n]=$$
$$=M[\sum\left(\varepsilon_i\right)^2]-M[\left(\sum\varepsilon_i\right)^2/n].$$

В полученном выражении имеется две составляющие: $M[\sum(\varepsilon_i)^2]$ и $M[\left(\sum\varepsilon_i\right)^2/n]$, которые преобразуем по отдельности

Шаг 3. Преобразоване $M\left[\sum(\varepsilon_i)^2\right]$

$$M\left[\sum(\varepsilon_i)^2\right]=M\left(\varepsilon_1^2+\ldots+\varepsilon_n^2\right)=M\varepsilon_1^2+\ldots+M\varepsilon_n^2=\sigma_1^2+\ldots+\sigma_n^2=n\sigma_\varepsilon^2.$$

Так как $\sigma_1^2=\ldots=\sigma_\varepsilon^2$ согласно предпосылкам метода наименьших квадратов.

Шаг 4. Преобразование $M\left[\left(\sum\varepsilon_i\right)^2/n\right]$.

$$M[\left(\sum\varepsilon_i\right)^2/n]=M[\left(\varepsilon_1+\ldots+\varepsilon_n\right)^2/n]=$$
$$=M[\left(\varepsilon_1^2+\ldots+\varepsilon_n^2+2\varepsilon_1\varepsilon_2+\ldots+2\varepsilon_{n-1}\varepsilon_n\right)/n]=$$
$$=[M\left(\varepsilon_1^2\right)+\ldots+M\left(\varepsilon_n^2\right)+2M\left(\varepsilon_1\varepsilon_2\right)+\ldots+2M\left(\varepsilon_{n-1}\varepsilon_n\right)]/n=$$
$$=\left(\sigma_1^2+\ldots+\sigma_n^2+0+\ldots+0\right)/n=n\sigma_\varepsilon^2/n=\sigma_\varepsilon^2.$$

Так как $\sigma_1^2 = \ldots = \sigma_n^2$ и $M(\varepsilon_1\varepsilon_2) = \ldots = M(\varepsilon_{n-1}\varepsilon_n) = 0$, согласно предпосылкам метода наименьших квадратов.

Шаг 5. Соединим вместе преобразования $M[\sum(\varepsilon_i)^2]$, $M[(\sum\varepsilon_i)^2/n]$ и получим конечный результат

$$M\sum(\varepsilon_i - \bar{\varepsilon})^2 = M\left[\sum(\varepsilon_i)^2 - \left(\sum\varepsilon_i\right)^2/n\right] =$$
$$= M\left[\sum(\varepsilon_i)^2\right] - M\left[\left(\sum\varepsilon_i\right)^2/n\right] = n\sigma_\varepsilon^2 - \sigma_\varepsilon^2 = (n-1)\,\sigma_\varepsilon^2.$$

Определим математическое ожидание третьей составляющей дисперсии остатков выборочной совокупности:

$$M\left[2(a_1 - \alpha_1)\sum(X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})\right] =$$
$$= 2M\left\{\sum(X_i - \bar{X})\varepsilon_i\left[\sum(X_i - \bar{X})\varepsilon_i - \varepsilon\sum(X_i - \bar{X})\right]/\sum(X_i - \bar{X})^2\right\} =$$
$$= 2M\left\{\sum(X_i - \bar{X})\varepsilon_i\sum(X_i - \bar{X})\varepsilon_i/\sum(X_i - \bar{X})^2\right\} =$$
$$= 2M\left\{\left[\sum(X_i - \bar{X})\varepsilon_i\right]^2/\sum(X_i - \bar{X})^2\right\} =$$
$$= 2\sigma_\varepsilon^2\sum(X_i - \bar{X})^2/\sum(X_i - \bar{X})^2 = 2\sigma_\varepsilon^2.$$

Преобразования выполнялись по следующим шагам.

Шаг 1. Определим значение $(a_1 - \alpha_1)$

Так как

$$a_1 = \alpha_1 + \sum(\omega_i\varepsilon_i) = \alpha_1 + \frac{\sum(X_i - \bar{X})\varepsilon_i}{\sum(X_i - \bar{X})^2},$$

то $a_1 - \alpha_1 = \sum(X_i - \bar{X})\varepsilon_i/\sum(X_i - \bar{X})^2$, подставим найденное выражение в исходную формулу и продолжим преобразования.

Шаг 2.

$$M\left[2(a_1 - \alpha_1)\sum(X_i - \bar{X})(\varepsilon_i - \bar{\varepsilon})\right] =$$
$$= 2M\left\{\sum(X_i - \bar{X})\varepsilon_i\left[\sum(X_i - \bar{X})\varepsilon_i - \bar{\varepsilon}\sum(X_i - \bar{X})\right]/\sum(X_i - \bar{X})^2\right\}.$$

В полученном выражении $\bar{\varepsilon}\sum(X_i - \bar{X}) = \bar{\varepsilon}\times 0 = 0$, так как $\sum(X_i - \bar{X}) = 0$ согласно свойствам среднего значения. Упростим выражение и перейдем к шагу 3.

Шаг 3.

$$M\left[2(a_1-\alpha_1)\sum\left(X_i-\bar{X}\right)\left(\varepsilon_i-\bar{\varepsilon}\right)\right]=$$

$$=2M\left\{\sum\left(X_i-\bar{X}\right)\varepsilon_i\left[\sum\left(X_i-\bar{X}\right)\varepsilon_i-\bar{\varepsilon}\sum\left(X_i-\bar{X}\right)\right]/\sum\left(X_i-\bar{X}\right)^2\right\}=$$

$$=2M\left\{\sum\left(X_i-\bar{X}\right)\varepsilon_i\sum\left(X_i-\bar{X}\right)\varepsilon_i/\sum\left(X_i-\bar{X}\right)^2\right\}.$$

Возведем в квадрат повторяющийся множитель и перейдем к шагу 4.

Шаг 4.

$$M\left[2(a_1-\alpha_1)\sum\left(X_i-\bar{X}\right)\left(\varepsilon_i-\bar{\varepsilon}\right)\right]=2M\{\sum\left(X_i-\bar{X}\right)\varepsilon_i[\sum\left(X_i-\bar{X}\right)\varepsilon_i-$$

$$-\bar{\varepsilon}\sum\left(X_i-\bar{X}\right)]/\sum\left(X_i-\bar{X}\right)^2\}=2M\{\sum\left(X_i-\bar{X}\right)\varepsilon_i\sum\left(X_i-\bar{X}\right)\varepsilon_i:$$

$$:\sum\left(X_i-\bar{X}\right)^2\}=2M\left\{\left[\sum\left(X_i-\bar{X}\right)\varepsilon_i\right]^2/\sum\left(X_i-\bar{X}\right)^2\right\}.$$

Шаг 5. Возьмем математические ожидания для числителя и знаменателя.

В знаменателе имеется выражение $\sum\left(X_i-\bar{X}\right)^2$, которое состоит из фиксированных значений, поэтому его математическое ожидание равно этому же значению.

$$M\left[\sum\left(X_i-\bar{X}\right)^2\right]=\sum\left(X_i-\bar{X}\right)^2.$$

В числителе имеется случайное возмущение ε_i, поэтому математическое ожидание найти будет гораздо сложнее.

$$M\left[\sum\left(X_i-\bar{X}\right)\varepsilon_i\right]^2=$$

$$=M\left\{\left[\sum\left(X_i-\bar{X}\right)\varepsilon_i\right]^2+\ldots+\left[\left(X_n-\bar{X}\right)\varepsilon_n\right]^2+\right.$$

$$\left.+2\left(X_1-\bar{X}\right)\left(X_2-\bar{X}\right)\varepsilon_1\varepsilon_2+\ldots+2\left(X_{n-1}-\bar{X}\right)\left(X_n-\bar{X}\right)\varepsilon_{n-1}\varepsilon_n\right\}=$$

$$=M\left[\left(X_1-\bar{X}\right)^2\varepsilon_1^2+\ldots+\left(X_n-\bar{X}\right)^2\varepsilon_n^2+2\left(X_1-\bar{X}\right)\left(X_2-\bar{X}\right)\varepsilon_1\varepsilon_2+\ldots\right.$$

$$\left.+2\left(X_{n-1}-\bar{X}\right)\left(X_n-\bar{X}\right)\varepsilon_{n-1}\varepsilon_n\right]=\left(X_1-\bar{X}\right)^2M(\varepsilon_1^2)+\ldots+$$

$$+\left(X_n-\bar{X}\right)^2M(\varepsilon_n^2)+2\left(X_1-\bar{X}\right)\left(X_2-\bar{X}\right)M(\varepsilon_1\varepsilon_2)+\ldots+$$

$$+2\left(X_{n-1}-\bar{X}\right)\left(X_n-\bar{X}\right)M\left(\varepsilon_{n-1}\varepsilon_n\right)=\sigma_\varepsilon^2\sum\left(X_i-\bar{X}\right)^2.$$

Так как $M(\varepsilon_i^{\,2}) = \sigma_1^2 = \ldots = M\sum(\varepsilon_n^{\,2}) = \sigma_n^2 = \sigma_\varepsilon^2$ и $M(\varepsilon_1\varepsilon_2) = \ldots = = M(\varepsilon_{n-1}\varepsilon_n) = 0$ согласно предпосылкам метода наименьших квадратов.

Шаг 6. Подставим математическое ожидание в исходное выражение и получим

$$
\begin{aligned}
&M\left\{2(a_1-\alpha_1)\sum(X_i-\bar{X})(\varepsilon_i-\bar{\varepsilon})\right\}= \\
&=2M\left\{\sum(X_i-\bar{X})\varepsilon_i\left[\sum(X_i-\bar{X})\varepsilon_i-\bar{\varepsilon}\sum(X_i-\bar{X})\right]/\sum(X_i-\bar{X})^2\right\}= \\
&=2M\left\{\left[\sum(X_i-\bar{X})\varepsilon_i\sum(X_i-\bar{X})\varepsilon_i\right]/\sum(X_i-\bar{X})^2\right\}= \\
&=2M\left\{\left[\sum(X_i-\bar{X})\varepsilon_i\right]^2/\sum(X_i-\bar{X})^2\right\}= \\
&=2\sigma_\varepsilon^2\sum(X_i-\bar{X})^2/\sum(X_i-\bar{X})^2=2\sigma_\varepsilon^2.
\end{aligned}
$$

Соберем вместе все составляющие математического ожидания дисперсии остатков выборочной совокупности.

$$M\left(\sum\varepsilon_i^2\right)=\sigma_\varepsilon^2+(n-1)\sigma_\varepsilon^2-2\sigma_\varepsilon^2=(n-2)\sigma_\varepsilon^2$$

Если мы определим дисперсию случайного возмущения с помощью формулы

$$S^2=\sum\varepsilon_i^2/(n-2),$$

то S^2 будет несмещенной оценкой истинного значения σ_ε^2.

Проверка эффективности.

Далее установим, что оценки, полученные методом наименьших квадратов являются эффективными и представляют собой наилучшие линейные несмещенные оценки, т. е. что в классе всех линейных несмещенных операторов оценивания оценки наименьших квадратов обладают наименьшей дисперсией.

Шаг 1. Постановка задачи. Известно, что коэффициент a_1, определенный по методу наименьших квадратов, вычисляется по формуле

$$a_1=\sum\omega_i У_i,$$

где $\omega_i = \left(X_i - \bar{X}\right) / \sum\left(X_i - \bar{X}\right)^2$.

Введем произвольную линейную оценку параметра α_1 как

$$a_{н1} = \sum c_i У_i,$$

где $c_i = \omega_i + d_i$;

d_i — произвольные константы.

Необходимо:

• для корректности выводов потребовать от нового коэффициента $a_{н1}$, чтобы он был несмещенной оценкой параметра α_1;

• найти дисперсии коэффициентов $a_{н1}$ и a_1;

• сравнить дисперсии коэффициентов $a_{н1}$ и a_1;

• проверить вывод: если дисперсия коэффициента $a_{н1}$ будет больше дисперсии коэффициента a_1, то коэффициент a_1 будет эффективным, т. е. в классе всех линейных несмещенных операторов оценивания оценки наименьших квадратов обладают наименьшей дисперсией.

Шаг 2. Выполнение условий несмещенности коэффициента $a_{н1}$.

Для того, чтобы $a_{н1}$ была несмещенной оценкой параметра α_1, константы d_i должны удовлетворять некоторым свойствам.

Подставим в выражение для $a_{н1}$

$$a_{н1} = \sum c_i У_i,$$

вместо $У_i$ выражение

$$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$$

и произведем несложные преобразования

$$a_{н1} = \sum c_i У_i = \sum c_i\,(\alpha_0 + \alpha_1 X_i + \varepsilon_i) = \alpha_0 \sum c_i + \alpha_1 \sum c_i X_i +$$
$$+ \sum c_i \varepsilon_i = \alpha_0 \sum c_i + \alpha_1 \sum c_i X_i + 0 = \alpha_0 \sum c_i + \alpha_1 \sum c_i X_i,$$

где $\sum c_i \varepsilon_i = \sum \varepsilon_i \sum c_i = 0 \times \sum c_i = 0$, так как $\sum \alpha_i = 0$.

Для того, чтобы коэффициент $a_{н1}$ был несмещенной оценкой параметра α_1 необходимо, чтобы математическое ожидание $a_{н1}$ равнялось параметру α_1. Следовательно, должно выполняться равенство

$$M(a_{н1}) = \alpha_0\sum c_i + \alpha_1\sum c_iX_i + M\sum c_i\varepsilon_i = \alpha_0\sum c_i + \alpha_1\sum c_iX_i +$$
$$+ \sum M(c_i\varepsilon_i) = \alpha_0\sum c_i + \alpha_1\sum c_iX_i = \alpha_1,$$

где c_i и ε_i не связаны между собой.

Поэтому

$$\sum M(c_i\varepsilon_i) = \sum M(c_i)M(\varepsilon_i).$$

Так как $M(\varepsilon_i) = 0$ согласно предпосылкам метода наименьших квадратов, то

$$\sum M(c_i\varepsilon_i) = \sum M(c_i)M(\varepsilon_i) = \sum M(c_i)\times 0 = 0.$$

Полученное выражение

$$M(a_{н1}) = \alpha_0\sum c_i + \alpha_1\sum c_iX_i = \alpha_1$$

будет справедливым при всех a_0, a_1 только в том случае, если выполняются условия

$$\sum c_i = 0, \sum c_iX_i = 1.$$

Учитывая следующие свойства весовых коэффициентов c_i и ω_i:

$$c_i = \omega_i + d_i;$$
$$\sum \omega_i = 0;$$
$$\sum \omega_i X_i = \sum \omega_i \left(X_i - \bar{X}\right) = 1;$$
$$\sum c_i = \sum \omega_i + \sum d_i = 0;$$
$$\sum c_iX_i = 1$$

получим ограничения на значения d_i

$\sum d_i = 0$, так как только при $\sum d_i = 0$ будет соблюдаться условие $\sum c_i = 0$,

где $\sum c_i = \sum \omega_i + \sum d_i = 0 + \sum d_i = 0$;

$\sum d_iX_i = 0$, так как только при $\sum d_iX_i = 0$ будет выполняться условие

$$\sum c_iX_i = 1,$$

где $\sum c_iX_i = \sum \omega_iX_i + \sum d_iX_i = 1 + 0 = 1$, при соблюдении условия $\sum \omega_iX_i = 1$.

Вывод. Коэффициент $a_{н1}$ будет несмещенной оценкой параметра $б_1$, если константы d_i будут обладать следующими свойствами

$\sum d_i = 0,$

$\sum d_i X_i = 0.$

Шаг 2. Определим ошибку коэффициента $a_{н1}$.

Определим коэффициент $a_{н1}$

$$a_{н1} = \sum c_i У_i = \sum c_i (\alpha_0 + \alpha_1 X_i + \varepsilon_i) =$$
$$= \alpha_0 \sum c_i + \alpha_1 \sum c_i X_i + \sum c_i \varepsilon_i = \alpha_1 + \sum c_i \varepsilon_i,$$

где $\sum c_i = 0$, $\sum c_i X_i = 1$ — согласно условий несмещенности коэффициента $a_{н1}$.

Вывод. Ошибка коэффициента $a_{н1}$ равна

$a_{н1} - \alpha_1 = \sum c_i \varepsilon_i.$

Шаг 3. Определим математическое ожидание дисперсии коэффициента $a_{н1}$. Дисперсия произвольной линейной несмещенной оценки $a_{н1}$будет равна

$\mathrm{Var}(a_{н1}) = M(a_{н1} - \alpha_1)^2 = M[(\sum c_i \varepsilon_i)^2] = \sigma_e^2 \sum c_i^2.$

В полученном выражении

$$\sum c_i^2 = \sum(\omega_i + d_i)^2 = \sum \omega_i^2 + \sum d_i^2 + 2\sum d_i \omega_i =$$
$$= \sum \omega_i^2 + \sum d_i^2 + 0 = \sum \omega_i^2 + \sum d_i^2,$$

где $\sum d_i \omega_i = \left[\sum d_i \left(X_i - \bar{X}\right)\right] / \sum \left(X_i - \bar{X}\right)^2 =$

$$= \left(\sum d_i X_i - \sum d_i \bar{X}\right) / \sum \left(X_i - \bar{X}\right)^2 = \left(\sum d_i X_i - \bar{X} \sum d_i\right) / \sum \left(X_i - \bar{X}\right)^2 =$$
$$= (0 - 0) / \sum \left(X_i - \bar{X}\right)^2 = 0,$$

так как по условиям несмещенности $a_{н1}$ должны выполняться условия:

$\sum d_i X_i = 0,$

$\sum d_i = 0.$

Продолжим получение дисперсии $a_{н1}$ с учетом полученных ограничений на $\sum c_i^2 = \sum \omega_i^2 + \sum d_i^2$.

$$\mathrm{Var}(a_{н1}) = M(a_{н1} - \alpha_1)^2 = M[(\sum c_i\varepsilon_i)^2] = \sigma_e^2 \sum c_i^2 =$$
$$= \sigma_e^2 \sum \omega_i^2 + \sigma_e^2 \sum d_i^2 = \mathrm{Var}(a_1) + \sigma_e^2 \sum d_i^2,$$

так как $\sigma_e^2 \sum \omega_i^2 = \mathrm{Var}(a_1)$.

Окончательно получаем

$$\mathrm{Var}(a_{н1}) = \mathrm{Var}(a_1) + \sigma_e^2 \sum d_i^2.$$

Вывод. Дисперсия коэффициента $a_{н1}$ равна дисперсии коэффициента a_1 плюс случайная составляющая, равная $\sigma_e^2 \sum d_i^2$.

Шаг 4. Анализ дисперсии $a_{н1}$.

В полученном выражении дисперсии $a_{н1}$

$$\mathrm{Var}(a_{н1}) = \mathrm{Var}(a_1) + \sigma_e^2 \sum d_i^2$$

$\sum d_i^2$ будет обязательно положительной и превратится в нуль, если все d_i будут равны нулю. Следовательно, дисперсия коэффициента $a_{н1}$ будет больше или равна дисперсии коэффициента a_1.

Вывод. Оценки наименьших квадратов обладают наименьшей дисперсией в классе всех линейно несмещенных оценок. Аналогичный результат получается при рассмотрении $\mathrm{Var}(a_0)$.

Проверка состоятельности.

Докажем состоятельность оценок параметров модели, определенных методом наименьших квадратов.

Выпишем дисперсии коэффициентов a_0 и a_1.

$$\mathrm{Var}(a_0) = \sigma_\varepsilon^2 \sum X_i^2 / \left[n \sum \left(X_i - \overline{X}\right)^2 \right],$$

$$\mathrm{Var}(a_1) = \sigma_\varepsilon^2 / \sum \left(X_i - \overline{X}\right)^2,$$

Преобразуем полученные выражения таким образом, чтобы они содержали в знаменателе объем выборки, а все остальные переменные при увеличении выборки стремились к постоянной величине.

$$\mathrm{Var}(a_0) = \sigma_\varepsilon^2 \sum X_i^2 / \left[n \sum \left(X_i - \overline{X}\right)^2 \right] = \sigma_\varepsilon^2 \overline{X}^2 / \left[n / n \sum \left(X_i - \overline{X}\right)^2 \right] =$$

$$= \sigma_\varepsilon^2 \overline{X}^2 / n \overline{\left(X_i - \overline{X}\right)^2},$$

где $\overline{X}^2$ — среднее значение квадратов переменной X;

$\overline{(X_i - \overline{X})^2}$ — среднее значений квадратов отклонений $X_i - \overline{X}$.

При увеличении выборки такие переменные, как σ_ε^2, $\overline{X}^2$, $\overline{(X_i - \overline{X})^2}$ будут стремиться к своему конечному математическому ожиданию, а так как объем выборки n стоит в знаменателе, то с его увеличением дисперсия коэффициента a_0 будет стремиться к нулю и тем ближе выборочный коэффициент a_0 будет находиться около своего математического ожидания α_0.

Следовательно, условия состоятельности для коэффициента a_0 выполняются.

Произведем аналогичные преобразования для дисперсии коэффициента a_1

$$\mathrm{Var}(a_1) = \sigma_\varepsilon^2 / \sum (X_i - \overline{X})^2 = \sigma_\varepsilon^2 / \left[n / n \sum (X_i - \overline{X})^2 \right] =$$
$$= \sigma_\varepsilon^2 / \left[n \overline{(X_i - \overline{X})^2} \right],$$

где $\overline{(X_i - \overline{X})^2}$ — среднее значений квадратов отклонений $X_i - \overline{X}$.

При увеличении объема выборки n такие переменные, как σ_ε^2, $\overline{(X_i - \overline{X})^2}$ будут стремиться к своему конечному математическому ожиданию, а так как объем выборки n стоит в знаменателе, то с его увеличением дисперсия коэффициента a_1 будет стремиться к нулю и тем ближе выборочный коэффициент a_1 будет находиться около своего математического ожидания α_1.

Следовательно, условия состоятельности для коэффициента a_1 выполняются.

Общий вывод. Если соблюдаются предпосылки метода наименьших квадратов, то коэффициенты a_0 a_1, вычисленные с помощью этого метода, будут несмещенными, состоятельными и эффективными[1].

Примечание. Мы намеренно приводим подробный вывод всех формул по двум причинам: первая — научить студентов проводить доказательства; вторая — при выводе формул необходимо учитывать много нюансов, которые в учебной литературе не всегда расшифровываются. При первом чтении можно ограничиться только выводами, но постепенно надо изучать обоснованность всех формул. В последних публикациях по эконометрике приводятся более компактные выводы [23].

Вопросы для самоконтроля

1. Приведите шесть возможных состояний предприятий.

2. Приведите восемь принципов эффективной работы предприятия.

3. Приведите основные свойства цели.

4. Приведите алгоритм выбора наиболее значимой проблемы.

5. Приведите алгоритм анализа причин, влияющих на зависимый признак.

6. Приведите вид множественной модели для выборочных данных.

7. Приведите алгоритм генерации эконометрических моделей.

8. Опишите предпосылки метода наименьших квадратов.

9. Выведите формулы расчета коэффициентов a_0 и a_1 методом наименьших квадратов в матричной и скалярной формах.

10. Что означает идентифицируемость модели?

11. Дайте определение состоятельности, эффективности и несмещенности оценок параметров регрессионной модели.

---

[1] При выводе доказательств несмещенности, состоятельности и эффективности оценок параметров модели использовалась литература: 1) *Доугерти К.* Введение в эконометрику. — М.: ИНФРА-М, 1997; 2) *Джонстон Д.* Эконометрические методы. — М.: Статистика, 1980.

Глава 4
ХАРАКТЕРИСТИКИ РЕГРЕССИОННОЙ МОДЕЛИ

Приступаем к изучению характеристик регрессионной модели в ее обычном классическом изложении, не разбивая выборочные данные на обучающую и контрольную совокупности.

Однако необходимо отметить, что ошибка модели и все статистические критерии, вычисленные по выборочной обучающей совокупности, являются внутренними критериями и не могут быть критериями достоверности модели. Зависимая переменная состоит из двух составляющих: случайной и детерминированной. При ограниченной выборке для каждого фиксированного значения X, как правило, бывает одно значение У, т. е. отсутствуют повторности. Это приводит к тому, что нельзя определить случайную составляющую зависимой переменной равной средней дисперсии воспроизводимости. Если увеличивать количество коэффициентов модели, то расчетные значения модели сначала пройдут через детерминированную составляющую, затем будут проходить и через случайную составляющую. При этом ошибка модели может сначала уменьшаться и при числе свободы, стремящемся к нулю, ошибка модели будет стремиться к бесконечности. Как определить, при каком количестве коэффициентов в модели расчетные значения У пройдут точно через детерминированную составляющую при использовании только обучающей выборки, никто не знает, так как используются внутренние критерии. По теореме Геделя нельзя судить о достоверности модели по внутреннему критерию.

Однако у этой проблемы есть решение. Разделим всю совокупность данных на обучающую и контрольную выборки. Если по обучающей выборке рассчитать коэффициенты модели, а по контрольной выборке — все характеристики модели и проверить достоверность модели, то полученные выводы будут основанием утверждения о достоверности модели. Например, для обучающей выборки мы использовали простую и сложную функции, которые позволили получить ошибку модели больше у простой, чем у сложной. Расчетные значения у сложной функции будут воспроизводить детерминированную и случайную составляющие. С помощью этих функций рассчитаем ошибку моделей на контрольной выборке и получим ошибку модели больше у сложной, чем у простой.

Вывод. Не стремитесь в своих исследованиях усложнения вида математической функции, если в этом нет необходимости.

4.1. Основные характеристики регрессионной модели

Исследователь на основании выборочных данных изучает регрессионную модель

$У_i = a_0 + a_1X_i + e_i = У_{pi} + e_i$

Основными характеристиками регрессионной модели являются:

- E — ошибка модели;
- E% — процентная ошибка модели;
- R^2 — коэффициент детерминации;
- S_{a0} — ошибка коэффициента a_0;
- S_{a1} — ошибка коэффициента a_1;
- F — критерий Фишера;
- $t_{\alpha/2}$ — двухсторонний критерий Стьюдента;
- $r^2(X_1.X_2,X_3)$ — частный коэффициент детерминации для фактора X_1 при условии, что факторы X_2 и X_3 не изменяются.

$$E = \sqrt{\frac{\sum (У_i - У_{pi})^2}{n - k}},$$

где E — ошибка модели, выраженная в единицах измерения У;

n — объем выборки;

k — количество всех коэффициентов модели.

$$E\% = \frac{E \cdot 100\%}{\overline{У}},$$

где E% — ошибка модели, выраженная в процентах.

При проведении предварительного анализа принято считать, что если ошибка модели будет меньше 5%, то модель является достоверной без указания вероятности этого утверждения. При этом надо всегда указывать, что количество коэффициентов в модели значительно меньше объема выборки. Для более обоснованной проверки достоверности модели используется критерий Фишера.

Примечание. Ошибка модели и критерий Фишера очень тесно связаны между собой, но эта связь меняется в зависимости от соотношения количества коэффициентов в модели и объема выборки. Если объем выборки больше количества коэффициентов в модели, то с усложнением модели ошибка модели уменьшается, а критерий Фишера возрастает. При этом становится справедливым утверждение о том, что если ошибка модели будет меньше 5%, то модель является достоверной. Если количество коэффициентов в модели приближается к объему выборки, то ошибка модели и критерий Фишера стремятся к нулю. При этом по ошибке модели, вычисленной в процентах, модель является достоверной, а по критерию Фишера — недостоверной.

Для расчета остальных характеристик нам потребуется произвести дисперсионный анализ регрессионной модели.

Дисперсия — свойство переменной X, которая вычисляется делением вариации на ее число степеней свободы по формуле

$$S^2 = \frac{\sum (X_i - \overline{X})^2}{n - 1},$$

где n − 1 — число степеней свободы вариации;

n — объем выборки;

$\sum(X_i - \bar{X})^2$ — вариация признака X.

Дисперсионный анализ регрессионной модели — разложение вариации зависимой переменной на вариации уравнения регрессии и остатков.

Примечание. Следует избегать следующей формулировки: дисперсионный анализ — разложение общей дисперсии на дисперсию регрессии и дисперсию остатков. При этом возникает неопределенность, что понимать под дисперсией. Если под дисперсией понимать вариацию, то определение верно, если под дисперсией понимать вариацию, деленную на число степеней свободы, то утверждение неверно, так как число степеней у разных вариаций не равны между собой.

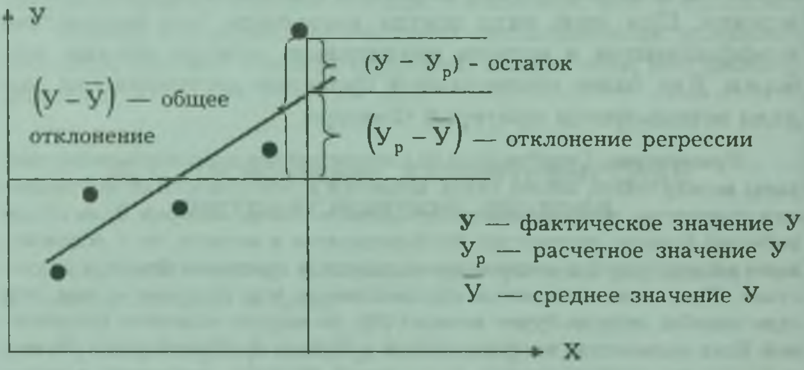


Рис. 4.1. Графическое представление разложения отклонения от среднего на составляющие элементы

Для каждой точки X_i справедливо следующее соотношение

$$(У - \bar{У}) = (У - У_р) + (У_р - \bar{У})$$

Если левую и правую часть этого соотношения возвести в квадрат и просуммировать по всем измерениям, то будет справедливо соотношение

$$\sum(У_i - \bar{У})^2 = \sum(У_i - У_{p_i})^2 + 2\sum(У_i - У_{p_i})(У_{p_i} - \bar{У}) + \\ + \sum(У_{p_i} - \bar{У})^2,$$

где $2\sum(У_i - У_{p_i})(У_{p_i} - \bar{У}) = 0.$

Получаем основное вариационное уравнение:

$$\sum\left(У_i - \bar{У}\right)^2 = \sum\left(У_i - У_{p_i}\right)^2 + \sum\left(У_{p_i} - \bar{У}\right)^2,$$

которое читается следующим образом — вариация общая равна вариации регрессии плюс вариация остатков,

где $\sum\left(У_i - \bar{У}\right)^2 = C_{общ}$ — вариация общая (общая сумма квадратов) ($SS_{общ}$);

$\sum\left(У_i - У_{p_i}\right)^2 = C_{ост}$ — вариация остатков (сумма квадратов остатков) ($SS_{ост}$);

$\sum\left(У_{p_i} - \bar{У}\right)^2 = C_{рег}$ — вариация регрессии (сумма квадратов регрессии) ($SS_{рег}$).

Составляющие элементы основного вариационного уравнения имеют соответствующие значения степеней свободы, которые связаны тождеством:

(n — 1) = (k — 1) + (n — k),

где (n − 1) — число степеней свободы ($df_{общ}$) для вариации общей;

(n − k) — число степеней свободы ($df_{ост}$) для вариации остатков,

(k − 1) — число степеней свободы ($df_{рег}$) для вариации регрессии.

Дисперсии регрессионной модели вычисляются делением соответствующей вариации на число степеней свободы по следующим формулам:

$$S^2_{общ} = \frac{\sum(У_i - \bar{У})^2}{n-1}$$ — дисперсия общая ($MS_{общ}$);

$$S^2_{ост} = \frac{\sum(У_i - У_{p_i})^2}{n-k}$$ — дисперсия остатков ($MS_{ост}$);

$$S^2_{рег} = \frac{\sum(У_{р_i} - \overline{У})^2}{k-1}$$ — дисперсия регрессии ($MS_{рег}$).

В скобках приведены обозначения, которыми пользуются некоторые пакеты прикладных программ, в частности программа "Регрессия", входящая в состав "Анализ данных" табличного процессора Excel.

Примечание. Следует предупредить, что дисперсия общая не равна сумме дисперсии остатков и дисперсии регрессии.

Продолжаем изучение характеристик модели.

Коэффициент детерминации равен доли объясненной вариации от общего варьирования зависимой переменной:

$$R^2 = \frac{C_{рег}}{C_{общ}}.$$

Приведенный (нормированный) коэффициент детерминации с учетом степеней свободы:

$$R^2_{пр} = \frac{S^2_{рег}}{S^2_{общ}} = 1 - \frac{(n-1)(1-R^2)}{n-k},$$

Множественный коэффициент корреляции:

$$R = \sqrt{R^2}.$$

Статистические критерии проверки значимости модели и ее коэффициентов

Нулевая гипотеза — это предположение о том, что две совокупности, рассматриваемые с точки зрения одного или нескольких признаков, одинаковы. При этом предполагается, что действительное различие равно нулю, а найденное из эксперимента отличие от нуля носит случайный характер.

Нулевые гипотезы проверяются с помощью статистических критериев.

Поскольку статистические критерии могут установить только отличие, но не одинаковость совокупностей относительно рассматриваемых признаков, то нуль-гипотеза, как правило, выдвигается для проверки, нет ли оснований для ее отбрасывая и принятия альтернативной гипотезы.

Проверка достоверности модели производится с помощью статистического критерия Фишера.

Шаг 1. Выдвигается нулевая гипотеза

H_0: $У_{рас} = \overline{У}$ или между У и X нет связи.

Шаг 2. Вычисляется фактическое значение критерия Фишера

$$F = \frac{S^2_{рег}}{S^2_{ост}}.$$

Шаг 3. Определяется критическое значение критерия Фишера на уровне значимости $\alpha = 0{,}05$, числе степеней свободы m_1 и m_2. Критическое значение критерия Фишера можно найти по таблицам, которые есть в каждом учебнике по эконометрике, или вычислить в Excel по функции FРАСПОБР.

Критическое значение критерия Фишера имеет следующие параметры:

$F_{кр}(\alpha = 0{,}05;\ m_1 = k - 1;\ m_2 = n - k)$,

где a — уровень значимости критерия;

m_1 — число степеней свободы для большей дисперсии регрессии;

m_2 — число степеней свободы для меньшей дисперсии остатков;

n — объем выборки;

k — количество всех коэффициентов модели.

Шаг 4. Сравниваются фактические значения критерия Фишера с его критическим значением.

Если $F > F_{кр}$ $(\alpha = 0{,}05;\ m_1 = k - 1;\ m_2 = n - k)$, то нулевая гипотеза отвергается с вероятностью 1- a и считается, что модель является достоверной.

Если $F < F_{кр}$ $(\alpha = 0{,}05;\ m_1 = k - 1;\ m_2 = n - k)$, то нулевая гипотеза принимается и считается, что достоверность модели не доказана, при этом не указывается вероятность этого утверждения.

Примечание. Предложенный критерий Фишера проверки достоверности модели имеет существенный недостаток, который заключается в том, что он является внутренним критерием. При этом чем меньше ошибка модели, тем достовернее она становится. Это справедливо до того момента, когда модель описывает существующую тенденцию, при дальнейшем уменьшении ошибки модели уравнение регрессии будет проходить через случайные составляющие, что приведет к фактическому увеличению ошибки прогноза. Для устранения этого недостатка необходимо использовать внешние критерии, которые реализованы в методе ретроспективного прогноза (более подробно см. в [1]). Метод ретроспективного прогноза реализован в пакете программ "Стат Эксперт."

Проверка достоверности коэффициентов модели производится с помощью статистического критерия Стьюдента.

Шаг 1. Выдвигается нулевая гипотеза

H_0: $\alpha_1 = 0$.

Шаг 2. Вычисляются ошибки коэффициентов модели по формулам:

$$S_{a0} = E\sqrt{\frac{\sum X_i^2}{n\sum(X_i - \bar{X})^2}}$$ — ошибка коэффициента a_0;

$$S_{a1} = E\sqrt{\frac{1}{\sum(X_i - \bar{X})^2}}$$ — ошибка коэффициента a_1.

Примечание. Ошибка коэффициента a_1 имеет обозначение среднего квадратического отклонения выборочных коэффициентов α_1 от своего математического ожидания a_1. Вывод формул ошибок коэффициентов a_0 и a_1 см. в [1].

Шаг 3. Вычисляется фактическое значение критерия Стьюдента

$$t = \frac{a_1}{S_{a1}}.$$

Шаг 4. Определяется критическое значение критерия Стьюдента на уровне значимости $\alpha = 0,05$. Критическое значение Стьюдента можно найти по таблицам или вычислить в Excel по функции СТЬЮДРАСПОБР.

$t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k)$,

где α — уровень значимости критерия;

m — число степеней свободы.

Шаг 5. Сравниваются фактические значения критерия Стьюдента с его критическим значением $ta_{/2}$.

Если $|t| > t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k)$, то нулевая гипотеза отвергается с вероятностью $1 - \alpha$ и считается, что коэффициент a_1 достоверно отличается от нуля.

Если $|t| < t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k)$, то нулевая гипотеза принимается и считается, что достоверность a_1 статистически не доказана.

Определение частного коэффициента детерминации

Частный коэффициент детерминации равен доли влияния фактора, включенного в модель, при условии что все остальные факторы не изменяются. Например, частный коэффициент детерминации для фактора X_1 (при наличии в модели факторов: X_1, X_2, X_3) имеет следующее обозначение $r^2(X_1.X_2,X_3)$, которое означает степень влияния фактора X_1, при условии что факторы X_2 и X_3 не изменяются.

Критерии Стьюдента можно использовать для определения частного критерия детерминации для каждого фактора, включенного в модель (при отсутствии мультиколлинарности).

Допустим имеется многофакторная модель

$У = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + e$.

Составляется пропорция.

R^2 — $(t^2_{a1} + t^2_{a2} + t^2_{a3})$,

$r^2(X_1.X_2,X_3)$ — t^2_{a1}.

Решая данную пропорцию, можно найти частный коэффициент детерминации для первого фактора (аналогично рассчитываются частные коэффициенты детерминации для любого фактора)

$$r^2_{X_1 \cdot X_2, X_3} = \frac{t^2_{a1} R^2}{t^2_{a1} + t^2_{a2} + t^2_{a3}},$$

где $r^2(X_1 \cdot X_2, X_3)$ — частный коэффициент детерминации для фактора X_1;

$t_{a1} = a_1/S_{a1}$ — критерий Стьюдента для коэффициента a_1 первого фактора X_1;

$t_{a2} = a_2/S_{a2}$ — критерий Стьюдента для коэффициента a_2 второго фактора X_2;

$t_{a3} = a_3/S_{a3}$ — критерий Стьюдента для коэффициента a_3 третьего фактора X_3.

R^2 — множественный коэффициент детерминации для всей модели.

Примечание. Приводим вывод частного коэффициента детерминации.

Анализ многофакторной регрессии должен содержать информацию о степени влияния каждого фактора, включенного в модель.

Если соблюдается предпосылка регрессионного анализа в том, что объясняемые факторы не должны быть связанные между собою, то степень влияния каждого фактора можно вычислить по таким формулам, которые используют различные варианты подходов к решению данной проблемы.

Постановка задачи.

Предположим, что деятельность экономической системы можно предоставить в виде математической модели

$$У_i = a_0 X_{0_i} + a_1 X_{1_i} + a_2 X_{2_i} + e_i,$$

где i — порядковый номер измерения;

$У_i$ — зависимая переменная;

X_0 — фиктивная переменная;

X_{ji} — ортогональные (независимые) объясняющие переменные;

J — номер фактора;

e_i — случайная составляющая модели;

a_j — коэффициенты модели, определенные методом наименьших квадратов.

Необходимо. Определить долю влияния каждого фактора, включенного в модель.

Предусматривается, что между факторами отсутствует мультиколлинеарность.

Решение задачи.

Вариант 1.

Используя выражение (3.1)

e'e = У'У – 2А'Х'У +А'Х'ХА,

можно получить основное уравнение дисперсионного анализа регрессионной модели, которое представим в матричной форме[1] :

$$\mathbf{У'У = A'X'У + e'e,} \tag{4.1}$$

где

$$У = \begin{bmatrix} У_1 \\ У_2 \\ .... \\ У_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ & ... & \\ 1 & X_{1n} & X_{2n} \end{bmatrix} \quad A = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ ... \\ e_n \end{bmatrix};$$

$\mathbf{У'У} = \sum У_i^2$ — общая сумма квадратов $У_i$, не скорректированная на среднее $\bar{У}$;

$e'e = \sum e_i^2$ — сумма квадратов остатков;

$A'X'У = a_0 \sum X_{0i} У_i + a_1 \sum X_{1i} У_i + a_2 \sum X_{2i} У_i$. — скалярное разложение матричного выражения.

Представим уравнения (4.1) в скалярной форме:

$$\sum У_i^2 = a_0 \sum X_{0i} У_i + a_1 \sum X_{1i} У_i + a_2 \sum X_{2i} У_i + \sum e_i^2. \tag{4.2}$$

Предлагается использовать формулу (4.2) для определения доли влияния каждого фактора, включенного в модель.

Общий коэффициент детерминации, равный доле суммы квадратов $У_i$, обусловленный влиянием всех факторов, которые включены в модель, можно определить по формуле

$$R_1^2 = \frac{a_0 \sum X_{0i} У_i + a_1 \sum X_{1i} У_i + a_2 \sum X_{2i} У_i}{\sum У_i^2}.$$

---

[1] *Джонстон Д.* Эконометрические методы. — М.: Статистика, 1980. С. 130.

Частный коэффициент детерминации фактора X_j, равный доли суммы квадратов $У_i$, обусловленный фактором X_j, можно вычислить по формуле

$$R_{xj}^2 = \frac{a_j \sum X_{j_i} У_i}{\sum У_i^2}. \tag{4.3}$$

Должно соблюдаться равенство

$$\sum_{j=1}^{m} R_{xj}^2 = R_1^2, \tag{4.4}$$

где j — номер фактора;

m — общее количество факторов, которые включены в модель;

$a_j > 0$ — коэффициенты регрессионной модели для всех факторов X_j, включенных в модель;

$R_{x_j}{}^2$ — частный коэффициент детерминации для фактора X_j;

$R_1{}^2$ — общий коэффициент детерминации.

К сожалению, формула (4.3) не имеет свойства массовости на случай, если коэффициенты a_j будут иметь отрицательные значения, тогда частные коэффициенты детерминации теряют математическое и экономическое содержание.

Соотношение (4.4) выполняется при любых значениях a_j.

Вариант 2.

Степень влияния фактора, который включен в модель, можно вычислить по формуле

$$R_{xj}{}^2 = C_{xj}{}^2 / C_{общ}{}^2,$$

где $R_{xj}{}^2$ — коэффициент детерминации, учитывающий влияние фактора X_j;

$C_{xj}{}^2$ — сумма квадратов зависимой переменной, обусловленная фактором X_j;

$C_{общ}{}^2$ — общая сумма квадратов зависимой переменной;

$C_{общ}{}^2 = (C_{x0}{}^2 + C_{x1}{}^2 + C_{x2}{}^2) + C_{ост}{}^2$;

$C_{ост}{}^2$ — остаточная сумма квадратов зависимой переменной.

Предположим, что известны следующие величины по результатом расчета классического регрессивного анализа:

$C_{рег}{}^2 = C_{x0}{}^2 + C_{x1}{}^2 + C_{x2}{}^2$ — сумма квадратов зависимой переменной, обусловленная факторами, включенными в модель;

$C_{ост}{}^2$ — остаточная сумма квадратов;

t_{aj} — критерий Стьюдента, рассчитанный для коэффициента a_j фактора X_j.

Используя соотношения

$$t_{aj}^2 = F_{ч\ xj}, \tag{4.5}$$

где $F_{ч\ xj}$ — частный коэффициент Фишера для фактора X_j,

$$F_{ч\ xj} = C_{xj}^2/[C_{ост}^2/(n - k)], \tag{4.6}$$

можно получить формулу расчета объясненной вариации за счет влияния фактора X_j.

$$C_{xj}^2 = F_{ч\ xj}S_{ост}^2 = t_{aj}^2C_{ост}^2/(n - k),$$

$$C_{общ}^2 = C_{рег}^2 + C_{ост}^2.$$

Доля влияния фактора, который включен в модель, вычисляется по следующей формуле (по условию все факторы не мультиколлинеарны):

$$R_{xj}^2 = C_{xj}^2 / C_{общ}^2 = \frac{t_{aj}^2 C_{ост}^2 /(n-k)}{C_{рег}^2 + C_{ост}^2}, \tag{4.7}$$

где t_{aj}^2 — критерий Стьюдента, определенный для коэффициента a_j фактора X_j;

$C_{ост}^2$ — остаточная сумма квадратов;

$C_{рег}^2$ — сумма квадратов, обусловленная регрессией;

n — объем выборки;

k — количество коэффициентов в регрессионной модели, включая свободный коэффициент.

Расчетная формула (4.7) имеет только одно ограничение — факторы, включенные в модель, должны быть ортогональными, при этом сумма частных коэффициентов детерминации совпадает с общим коэффициентом детерминации.

Если факторы мультиколлинеарны, то сумма частных коэффициентов детерминации становится меньше общего коэффициента детерминации. По-видимому, в состав частных коэффициентов детерминации необходимо включить взаимовлияние объясняемых факторов.

Вариант 3.

Идея предложенного метода расчета степени влияния факторов, включенных в модель, состоит в том, что степень влияния факторов должна быть пропорциональна соответствующему значению квадрата критерия Стьюдента. Обоснование использования квадрата критерия Стьюдента можно получить в результате анализа формул (4.5) и (4.6).

Расчетная формула степени влияния фактора, включенного в модель, выводится с помощью решения пропорции — сумма всех квадратов критериев Стьюдента обеспечивает получение общего коэффициента детерминации, а отдельный квадрат критерию Стьюдента обеспечивает обусловленное значение частного коэффициента детерминации

$$R_{xj}^2 = \frac{t_{aj}^2 R^2}{t_{a0}^2 + t_{a1}^2 + t_{a2}^2}, \qquad (4.8)$$

где t_{aj} — критерий Стьюдента для коэффициента a_j;
R^2 — общий коэффициент детерминации;
R_{xj}^2 — частный коэффициент детерминации для фактора X_j.

Расчетная формула (4.8) не содержит ограничений на знаки и численные значения коэффициентов регрессионной модели.

Вывод частного коэффициента детерминации основан на следующих предпосылках: при отсутствии мультиколлинеарности справедливо соотношение $t^2 = F$ }[1]; составляющие критерия Фишера можно разложить на частные вариации, обусловленные влиянием факторов, включенных в модель. Расчет частных критериев Фишера имеется в алгоритмах метода Эфроимсона шаговой регрессии[2].

4.2. Методологические основы прогнозирования

Методологической основой прогнозирования являются инерционные свойства экономической системы. В настоящем присутствуют элементы прошлого и зачатки будущего.

Прогнозирование будет соответствовать действительности, или корректным, если в прогнозном периоде не изменится внутренняя и внешняя среда функционирования экономической системы.

Внутренняя среда экономической системы включает в себя:

- количество ресурсов трудовых, денежных, интеллектуальных и т. д.;
- стиль руководства;
- форму собственности;

[1] *Грубер Й.* Эконометрика. Т. 1. Введение в эконометрию. — Киев: Астарта, 1996. С. 153.

[2] *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ. — М.: Статистика, 1973.

- производственные отношения и производительные силы;
- потребности членов общества.

Внешняя среда экономической системы включает в себя:

- правовую и законодательную среду общества;
- политику государства;
- природные катастрофы, которые происходят на планете.

В настоящее время принято рассчитывать прогнозы в двух вариантах: оптимистический и пессимистический, соответственно при благоприятной и неблагоприятной внутренней и внешней среде.

4.3. Точечный и интервальный прогноз

Постановка задачи

Необходимо распространить нашу модель, содержащую две переменные, на другие значения независимой переменной и определить прогноз среднего значения У, соответствующего некоторому данному значению X, допустим X_0, которое может находиться как между выборочными наблюдениями от X_1 до X_n (прогноз называют интерполяцией), так и вне соответствующего интервала (прогноз называют экстраполяцией). Прогноз может быть точечным и интервальным. Определим точечный прогноз как некоторую линейную функцию от $У_i$ ($i = 1, ..., n$). Допустим

$$У_{пр0} = \sum c_i У_i,$$

где веса c_i должны быть выбраны так, чтобы сделать $У_{пр0}$ наилучшим линейным несмещенным прогнозом.

В соответствии с введенными предположениями мы имеем модель зависимости У от X для генеральной совокупности

$$У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$$

при соблюдении следующих ограничений:

а) $M(\varepsilon_i) = 0$ — при всех i,

б) $M(\varepsilon_i \varepsilon_j) = \begin{cases} 0 & \text{при } i \neq j,\ i, j = 1, 2, ..., n \\ \sigma_\varepsilon^2 & \text{при } i = j,\ i, j = 1, 2, ..., n \end{cases}$

$M(X_i\varepsilon_j) = 0$ при $i = j$; $i \neq j$; $i, j = 1, 2, ..., n$

где M — математическое ожидание;

ε_i, ε_j — возмущения модели;

σ_ε^2 — дисперсия возмущений, не зависящая от i.

Условное математическое ожидание $У_0$ при $X = X_0$ будет равно

$$M(У_0 \mid X_0) = \alpha_0 + \alpha_1 X_0.$$

В выражение $У_{пр0} = \sum c_i У_i$ подставим $У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$ и получим

$$У_{пр0} = \sum c_i У_i = \alpha_0 \sum c_i + \alpha_1 \sum c_i X_i + \sum c_i \varepsilon_i.$$

Условное математическое ожидание $У_{пр0}$ при $X = X_0$ будет равно

$$M(У_{пр0} \mid X_0) = \alpha_0 \sum c_i + \alpha_1 \sum c_i X_i$$

Следовательно, $У_{пр0}$ будет несмещенным линейным прогнозом для

$M(У_0 \mid X_0)$, если $\sum c_i = 1$, $\sum c_i X_i = X_0$.

Дисперсия $У_{пр0}$ будет равна

$$M[(У_{пр0} - M(У_{пр0} \mid X_0)^2] = M[(\sum c_i \varepsilon_i)^2] = \sigma_e^2 \sum c_i^2.$$

Решим оптимизационную задачу: необходимо определить такие значения весовых коэффициентов c_i, при которых функция $\sum c_i^2$ будет иметь условный экстремум при соблюдении следующих ограничений:

$$\sum c_i = 1, \sum c_i X_i = X_0.$$

Для решения поставленной задачи воспользуемся методом множителей Лагранжа[1].

Составим функцию Лагранжа:

$$f(c_j, \lambda, \mu) = \sum c_i^2 - 2\lambda(\sum c_i - 1) - 2\mu(\sum c_i X_i - X_0),$$

где λ, μ — множители Лагранжа.

[1] *Салманов О. Н.* Математическая экономика с применением Mathcad и Excel. — СПб.: БХВ-Петербург, 2003. С. 276–278.

Найдем частные производные функции Лагранжа, приравняем их нулю и получим систему уравнений:

$$\begin{cases} \partial f/\partial c_i = 2c_i - 2\lambda - 2\mu X_i = 0, \quad i = 1, \ldots, n; \\ \partial f/\partial \lambda = -2\left(\sum c_i - 1\right) = 0; \\ \partial f/\partial \mu = -2\left(\sum c_i X_i - X_0\right) = 0; \end{cases}$$

Просуммируем первое уравнение по всем i и воспользуемся тем, что во втором уравнении $\sum c_i = 1$. Тогда

$$\sum c_i = n\lambda + \mu \sum X_i = 1,$$

откуда

$$\lambda = 1/n - \mu \bar{X}.$$

Полученное значение λ вновь подставим в первое уравнение и найдем

$$c_i = \lambda - \mu X_i = (1/n - \mu \bar{X}) - \mu X_i = 1/n + \mu(X_i - \bar{X}).$$

Умножим это уравнение на X_i, вновь произведем суммирование по i и, пользуясь третьим уравнением $\sum c_i X_i = X_0$, получим

$$\sum c_i X_i = \bar{X} + \mu \sum [(X_i - \bar{X}) X_i] = X_0,$$

что дает нам

$$\mu = (X_0 - \bar{X}) / [\sum (X_i - \bar{X}) X_i] = (X_0 - \bar{X}) / \sum (X_i - \bar{X})^2,$$

так как

$$X_i = (X_i - \bar{X}) + \bar{X} \text{ и } \sum (X_i - \bar{X}) = 0.$$

Подставим значение μ в выражение

$$c_i = 1/n + \mu(X_i - \bar{X}),$$

получим окончательное значение для весовых коэффициентов

$$c_i = 1/n + \mu(X_i - \bar{X}) = 1/n + (X_0 - \bar{X})(X_i - \bar{X}) / \sum (X_i - \bar{X})^2.$$

Подстановка найденного значения для весовых коэффициентов c_i в

$$У_{пр0} = \sum c_i У_i$$

позволяет получить наилучшую несмещенную линейную оценку, так как

$$У_{пр0} = \sum c_i У_i = \sum [1/n + (X_0 - \bar{X})(X_i - \bar{X}) / \sum (X_i - \bar{X})^2] У_i =$$

$$= \sum [1/n + \bar{X}(X_i - \bar{X}) / \sum (X_i - \bar{X})^2 +$$

$$+ X_0 (X_i - \bar{X}) / \sum (X_i - \bar{X})^2] У_i =$$

$$= \bar{У} - \bar{X} \sum [(X_i - \bar{X}) У_i / \sum (X_i - \bar{X})^2] +$$

$$+ X_0 \sum [(X_i - \bar{X}) У_i / \sum (X_i - \bar{X})^2] = (\bar{У} - a_1 \bar{X}) + a_1 X_0 = a_0 + a_1 X_0,$$

где $a_1 = \sum [(X_i - \bar{X}) У_i] / \sum (X_i - \bar{X})^2,$

$$a_0 = \bar{У} - a_1 \bar{X}.$$

Таким образом, наилучшей несмещенной линейной оценкой

$M(У_0 \mid X_0) = \alpha_0 + \alpha_1 X_0$ будет $a_0 + a_1 X_0$, где a_0 и a_1 коэффициенты, определен-ные методом наименьших квадратов.

Вычислим дисперсию $У_{пр0}$ по формуле $M\{[У_{пр0} - M(У_0|X_0)]^2\}$ с учетом того, что

$$var(a_0) = \sigma_\varepsilon^2 \sum X_i^2 / [n \sum (X_i - \bar{X})^2];$$

$$var(a_1) = \sigma_\varepsilon^2 / \sum (X_i - \bar{X})^2;$$

$$cov(a_0 a_1) = -\sigma_\varepsilon^2 \bar{X} / \sum (X_i - \bar{X})^2.$$

$$M\{[У_{пр0} - M(У_0 \mid X_0)]^2\} = M\{[(a_0 - \alpha_0) + (a_1 - \alpha_1) X_0]^2\} =$$

$$= M[(a_0 - \alpha_0)^2 + (a_1 - \alpha_1)^2 X_0^2 + 2X_0 (a_0 - \alpha_0)(a_1 - \alpha_1)] =$$

$$= M(a_0 - \alpha_0)^2 + X_0^2 M(a_1 - \alpha_1)^2 + 2X_0 M[(a_0 - \alpha_0)(a_1 - \alpha_1)] =$$

$$= var(a_0) + X_0^2 \, var(a_1) + 2X_0 \, cov(a_0 a_1) =$$

$$= \sigma_\varepsilon^2 \sum X_i^2 / n \sum (X_i - \bar{X})^2 + \sigma_\varepsilon^2 X_0^2 / \sum (X_i - \bar{X})^2 -$$

$$- 2X_0 \sigma_\varepsilon^2 \bar{X} / \sum (X_i - \bar{X})^2 = \sigma_\varepsilon^2 [\sum X_i^2 / n \sum (X_i - \bar{X})^2 +$$

$$+ X_0^2 / \sum (X_i - \bar{X})^2 - 2X_0 \bar{X} / \sum (X_i - \bar{X})^2] =$$

$$= \sigma_\varepsilon^2 [\sum X_i^2 / n\sum(X_i - \bar{X})^2 + nX_0^2 / n\sum(X_i - \bar{X})^2 -$$
$$- 2nX_0\bar{X} / n\sum(X_i - \bar{X})^2] =$$
$$= \sigma_\varepsilon^2 (\sum X_i^2 + nX_0^2 - 2nX_0\bar{X}) / [n\sum(X_i - \bar{X})^2].$$

Введем в числитель выражение $У_p = 2n(\bar{X})^2 - 2\bar{X}\sum X_i$, так как

$$2n(\bar{X})^2 = 2\bar{X}\sum X_i.$$

Продолжим преобразования

$$M\{[У_{пр0} - M(У_0 | X_0)]^2\} =$$
$$= \sigma_\varepsilon^2[\sum X_i^2 + nX_0^2 - 2nX_0\bar{X} + 2n(\bar{X})^2 - 2\bar{X}\sum X_i]/[n\sum(X_i - \bar{X})^2] =$$
$$= \sigma_\varepsilon^2[\sum X_i^2 - 2\bar{X}\sum X_i + n(\bar{X})^2 + nX_0^2 - 2nX_0\bar{X} + n(\bar{X})^2]/[n\sum(X_i - \bar{X})^2] =$$
$$= \sigma_\varepsilon^2 [(\sum(X_i - \bar{X})^2 + n(X_0 - \bar{X})^2]/[n\sum(X_i - \bar{X})^2] =$$
$$= \sigma_\varepsilon^2 [1/n + (X_0 - \bar{X})^2 / \sum(X_i - \bar{X})^2]. \quad (4.9)$$

Для оценки σ_ε^2 воспользуемся несмещенной величиной ошибки модели

$$S_e^2 = E^2 = \frac{\sum(У_i - У_{pi})^2}{n-k},$$

где n — объем выборки;

k — количество коэффициентов в модели.

Поскольку $У_{пр0}$ — линейная функция от a_0 и a_1, имеющих двумерное нормальное распределение, распределение $У_{пр0}$ также удовлетворяет нормальному закону распределения со средним $\alpha_0 + \alpha_1 X_0$ и дисперсией, заданной формулой (4.9). Известно, что величина $(n\text{-}2)S_e^2/\sigma_e^2$ удовлетворяет независимому распределению по критерию χ^2 с n − 2 степенями свободы. Поэтому

$$t = \frac{У_{пр0} - M(У_0 | X_0)}{S_e\sqrt{1/n + (X_0 - \bar{X})^2 / \sum(X_i - \bar{X})^2}}$$

будет иметь t-распределение с n — 2 степенями свободы. Так как среднее квадратическое отклонение остатков равно ошибке модели $S_e = E$, то $100(1 - \alpha)\%$-ный доверительный интервал для $M(У_0|X_0)$ задается формулой

$$(a_0 + a_1 X_0) \pm t_{\alpha/2}(m = n-2)E\sqrt{\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum (X_i - \overline{X})^2}}. \qquad (4.10)$$

Определим доверительный прогнозный интервал не для среднего значения, а для некоторого значения $У_0$, которое связано с X_0.

Если сохраняется прежнее линейное соотношение, то

$У_0 = \alpha_0 + \alpha_1 X_0 + \varepsilon_0$;

$У_{пр0} = a_0 + a_1 X_0$.

Определим новую переменную z:

$z = У_0 - У_{пр0} = \varepsilon_0 - (a_0 - \alpha_0) - (a_1 - \alpha_1)X_0$,

для которой получаем $M(z) = 0$.

Возведем z в квадрат и вычислим математическое ожидание, воспользовавшись независимостью ε_0 от $\varepsilon_1, ..., \varepsilon_п$, на основании которой были получены a_0 и a_1.

$Var(z) = M(У_0 - У_{пр0})^2 = M(e_0^2) + M\{[(a_0 - a_0) + (a_1 - a_1)X_0]^2\}$

Учитывая выражение (4.9),

$$M[(a_0 - \alpha_0) + (a_1 - \alpha_1)X_0)^2] = \sigma_\varepsilon^2\{1/n + (X_0 - \overline{X})^2 / \sum (X_i - \overline{X})^2\},$$

найдем

$$M(У_0 - У_{пр0})^2 = M(\varepsilon_0^2) + M\{[(a_0 - \alpha_0) + (a_1 - \alpha_1)X_0]^2\} =$$

$$= \sigma_\varepsilon^2[1 + 1/n + (X_0 - \overline{X})^2 / \sum (X_i - \overline{X})^2]$$

Для оценки $\sigma_\varepsilon{}^2$ воспользуемся несмещенной величиной ошибки модели

$$S_e^2 = E^2 = \frac{\sum (У_i - У_{pi})^2}{n-k},$$

где n — объем выборки;

k — количество коэффициентов в модели[1].

---

[1] *Джонстон Д.* Эконометрические методы. — М.: Статистика, 1980. С. 46–51.

Величина

$$t = \frac{У_0 - У_{про}}{S_e \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum (X_i - \overline{X})^2}}}$$

имеет t-распределение с n − 2 степенями свободы, а 100(1 − α)%-ный доверительный интервал для индивидуального значения $У_0$ задается формулой

$$(a_0 + a_1 X_0) \pm t_{\alpha/2}(m = n - 2) E \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum (X_i - \overline{X})^2}} \qquad (4.11)$$

Определение точечного прогноза

Точечный прогноз — это среднее значение прогнозной переменной, который вычисляется по формуле

$У_{пр} = a_0 + a_1 X_{ож}$,

где $У_{пр}$ — прогнозное значение зависимой переменной на ожидаемый период;

$X_{ож}$ — численное значение объясняемой переменной на ожидаемый период.

Интервальный прогноз — это интервал, в котором с определенной вероятностью может находиться фактическое значение прогнозной величины.

Существуют два вида интервальных прогнозов: для индивидуальных значений и для математических ожиданий или среднего значения.

На основании формулы (4.11) можно утверждать, что с вероятность 1 − α индивидуальное значение прогноза будет находиться в интервале от $У_{пр} - W_и$ до $У_{пр} + W_и$,

где $W_и = t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k) E \sqrt{1 + \frac{1}{n} + \frac{(\overline{X} - X_{ож})^2}{\sum (X_i - \overline{X})^2}}$;

$W_и$ — ошибка прогноза для индивидуальных значений зависимой переменной;

$t_{\alpha/2}$ — двухсторонний критерий Стьюдента на уровне значимости α и числе степеней свободы $m = n - 2$;

n — объем выборки;

k — количество всех коэффициентов в уравнении регрессии;

E — ошибка модели;

X_i — текущее значение объясняемой переменной;

$\bar{X}$ — среднее значение объясняемой переменной;

$X_{ож}$ — ожидаемое значение объясняемой переменной.

Ошибка прогноза для индивидуальных значений обладает следующими свойствами:

Свойство 1. Чем дальше находится $X_{ож}$ от $\bar{X}$, тем больше прогнозный доверительный интервал.

Свойство 2. Чем больше ошибка модели E, тем больше прогнозный доверительный интервал.

Свойство 3. Чем меньше a, тем больше t, тем больше прогнозный доверительный интервал.

Свойство 4. Чем больше n, тем меньше прогнозный доверительный интервал,

Свойство 5. Чем больше вариация X, тем меньше прогнозный доверительный интервал.

Вывод. Для того, чтобы уменьшить прогнозный доверительный интервал для индивидуальных значений зависимой переменной необходимо улучшить спецификацию модели, увеличить объем выборки, увеличить дисперсию объясняемой переменной.

На основании формулы (4.10) можно, что с вероятность $1 - \alpha$ прогнозное значение математического ожидания или среднее значение зависимой переменной будет находиться в интервале от $У_{пр} - W_м$ до $У_{пр} + W_м$,

где $W_м = t_{\alpha/2}(\alpha = 0{,}05,\ m = n - k)E\sqrt{\dfrac{1}{n} + \dfrac{(X_{ож} - \bar{X})^2}{\sum(X_i - \bar{X})^2}}$;

$W_м$ — ошибка прогноза для математического ожидания или среднего значения зависимой переменной;

$t_{\alpha/2}$ — двухсторонний критерий Стьюдента на уровне значимости α и числе степеней свободы $m = n - 2$;

n — объем выборки;

k — количество всех коэффициентов в уравнении регрессии;

E — ошибка модели;

X_i — текущее значение объясняемой переменной;

$\bar{X}$ — среднее значение объясняемой переменной;

$X_{ож}$ — ожидаемое значение объясняемой переменной.

4.4. Доверительный интервал функции регрессии

Существуют два вида интервальных прогнозов уравнения регрессии для: индивидуальных значений и математических ожиданий.

С вероятность $1 - \alpha$ можно утверждать, что уравнение регрессии для индивидуальных значений будет находиться в интервале от $У_р - W_и$ до $У_р + W_и$,

где $W_и = t_{\alpha/2}(\alpha = 0{,}05,\ m = n - k)\, E \sqrt{1 + \frac{1}{n} + \frac{(\bar{X} - X_i)^2}{\sum (X_i - \bar{X})^2}}$;

$W_и$ — ошибка прогноза для индивидуальных значений зависимой переменной,

$У_{pi} = a_0 + a_1 X_i$ — расчетные значения зависимой переменной.

С вероятность $1 - \alpha$ можно утверждать, что уравнение регрессии для математических ожиданий зависимой переменной будет находиться в интервале от $У_р - W_м$ до $У_р + W_м$,

где $W_м = t_{\alpha/2}(\alpha = 0{,}05,\ m = n - k) E \sqrt{\frac{1}{n} + \frac{(\bar{X} - X_i)^2}{\sum (X_i - \bar{X})^2}}$;

$W_м$ — ошибка прогноза для математического ожидания зависимой переменной.

Построение изолиний

В множественном эконометрическом анализе часто возникает проблема в определении таких значений объясняемых переменных, при которых зависимая переменная будет иметь постоянное желаемое значение — это и означает построить изолинию.

Изолиния — это такая линия на многомерном графике, на которой численные значения зависимой переменной имеют постоянные значения.

Построение изолинии позволяет определить такие значения комбинаций объясняемых переменных, при которых значения зависимой переменной не изменяется.

4.5. Эконометрический анализ регрессионной модели

Эконометрический анализ проводится в такой последовательности:

1. Приводится условие задачи и база данных всех переменных, которые будут участвовать в построении модели.
2. Строится график зависимости между переменными.
3. Приводятся все характеристики модели.
4. Проверяется достоверность модели и ее коэффициентов.
5. Приводится точечный и интервальный прогноз на ожидаемый период.
6. Приводится графическое представление всех результатов расчетов с указанием фактических и расчетных значений зависимой переменной, 95% доверительных интервалов для уравнения регрессии, точечный прогноз и 95% прогнозный доверительный интервал для зависимой переменной.
7. Делаются выводы и предложения по результатам эконометрического анализа.

Приводим пример эконометрического анализа линейной регрессионной модели, оформленной в виде лабораторной работы, в которой используются основные формулы регрессионного анализа, реализованные в Excel.

Задача

Имеются выборочные данные зависимости прироста объема валовой продукции предприятия от количества рационализаторских предложений, реализованных на однородных предприятиях за один и тот же интервал времени (месяц) (см. в табл. 4.1).

Таблица 4.1

Исходные данные зависимости У от X

i	X_i	$У_i$
1	1	14
2	2	21
3	3	20
4	4	29
5	5	36
6	6	34
7	7	33
8	8	40
9	9	41
10	10	52
11	11	50
12	12	60
Ожидаем	13	?

где $У_i$ — значения прироста валовой продукции производства за месяц (тыс. руб.);

X_i — количество рационализаторских предложений, реализованных в течении месяца (шт.);

i — порядковый номер измерения;

n = 12 — объем выборки.

Необходимо.

1. Вычислить выборочные коэффициенты и характеристики линейной модели

$$У_i = a_0 + a_1 X_i + e_i,$$

где $У_i$ — значения прироста валовой продукции производства за месяц (тыс. руб.);

X_i — количество реализованных рационализаторских предложений в течение месяца (шт.);

i — порядковый номер измерения;

e_i — остатки модели, которые учитывают влияние всех факторов, которые не вошли в модель.

2. Вычислить точечный и интервальный прогноз У при ожидаемом количестве рационализаторских предложений $X_{ож} = 13$.

3. Произвести эконометрический анализ линейной модели.

4. Решение задачи представить в соответствии с этапами эконометрического моделирования.

Решение задачи.

Приводим графическое представление решения задачи, представленной на рис. 4.2.

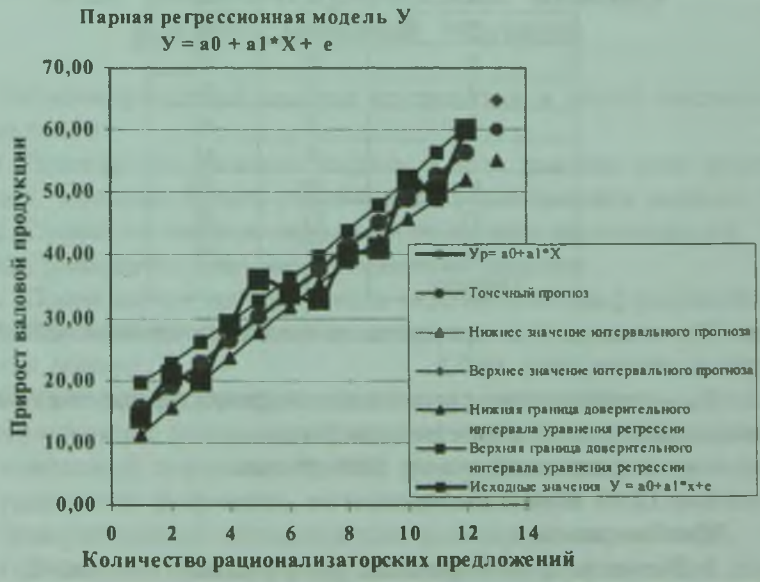


Рис. 4.2. Парная регрессионная модель прироста валовой продукции

На рис. 4.2 показаны результаты моделирования с получением точечных и интервальных значений уравнения регрессии и прогноза.

Визуальный анализ регулярностей зависимости У от X показывает, что она имеет четко выраженную линейную тенденцию с однородными остатками.

Примечание. Возникает естественный вопрос, зачем в начале задачи дается ее решение? В соответствии с соционикой люди подразделяются на рационалистов и ирационалистов. Рационалисты будут выполнять задание, если им четко ясен весь путь решения. Для ирационалиста важны конечная цель и возможность импровизации в поиске пути решения задачи. Поэтому для рационалистов и ирационалистов важен конечный результат, к которому надо прийти. Решение задачи займет много времени и только в конце будет известен результат. Поэтому решение выносится в начало задачи. Этот дидактический прием используется нами при решении всех эконометрических задач.

Представляем решение задачи в соответствии с этапами эконометрического моделирования.

Этап 1. Постановка задачи.

Цель моделирования.

Целью моделирования является решение приоритетной проблемы, связанной с повышением валовой продукции предприятия по сравнению с лучшими предприятиями данного профиля.

Выбор переменных.

Величина валовой продукции зависит от шести основных факторов: сырье и материалы (качество сырья и материалов); машины (техническое оборудование предприятия); методы (технология производства, технологические карты процессов); люди (количество рабочих, уровень квалификации, образование, возраст, деловые характеристики); физическая и абстрактная среда (температура, влажность, степень загрязнения воздуха, освещенность рабочего места; традиции, психологический климат, существующие на предприятии); правовая среда существования предприятия, информационное обеспечение предприятия; время (в данной задаче время не оказывает влияния на У, так как оно не изменяется и равно одному месяцу, в течение которого производились измерения). Влияние всех факторов на зависимый признак представлено на рис. 4.3, который называется "рыбий скелет", или диаграмма Исикавы. С помощью диаграммы Исикавы можно учесть все факторы, которые могут влиять на зависимую переменную в любой области исследований.

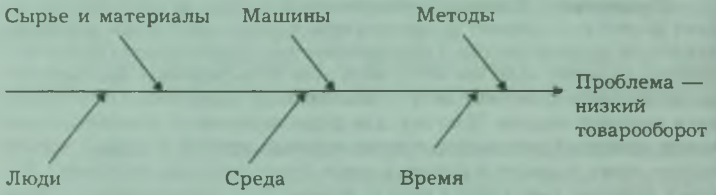


Рис. 4.3. Причинно-следственная диаграмма (Исикавы)

Выделение входных и выходных переменных

Предприятие имеет множество характеристик, но в соответствии с возникшей проблемой, связанной с малым уровнем валовой продукции по сравнению с конкурентами, возникла потребность изучить зависимость и прогноз прироста валовой продукции от количества реализованных рационализаторских предложений. Очевидно, что прирост валовой продукции может происходить по многим причинам (временная тенденция, улучшение работы предприятия при внедрении системы качества во всех подразделения предприятия). Рационализаторские предложения направлены на повышение качества выпускаемой продукции за счет улучшения процессов. Премию работникам можно выдавать только за совершенствование процессов. Совершенствование процессов можно осуществлять на основе критериев Европейской модели идеального предприятия и международных стандартов ИСО серии 9000 систем менеджмента качества.

Входной переменной будет количество поданных рационализаторских предложений, выходным показателем — прирост валовой продукции.

Сбор статистических данных.

Была произведена пространственная группировка по двум признакам: У — величина прироста валовой продукции и X — количество поданных рационализаторских предложений. Исходные данные представлены в табл. 4.2. График зависимости У от X изображен на рис. 4.4.

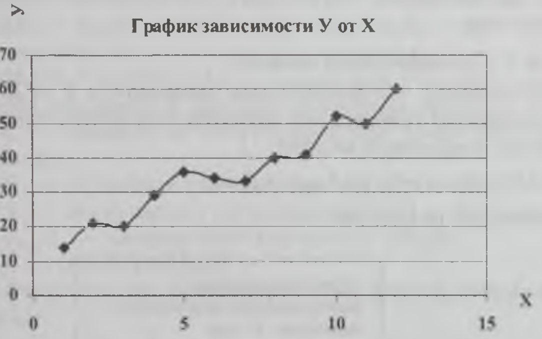


Рис. 4.4. График зависимости У от X

Анализ рис. 4.4 показывает, что с ростом числа поданных рационализаторских предложений увеличивается прирост валовой продукции предприятия. Однако эта зависимость является стохастической.

Этап 2. Предмодельный анализ.

Выдвижение гипотез.

В теоретическом плане при увеличении числа рационализаторских предложений должна наблюдаться тенденция увеличения прироста валовой продукции предприятия. Однако прирост валовой продукции может осуществляться от других мер совершенствования процессов не только производства продукции, но и от совершенствования процессов управления производством. Поэтому зависимость У от X является стохастической.

Предполагаем, что прирост валовой продукции линейно зависит от количества поданных рационализаторских предложений.

Формулировка допущений.

Предполагаем, что прирост валовой продукции в большей степени зависит от количества поданных рационализаторских предложений, чем от других мер совершенствования процессов.

Этап 3. Спецификация модели.

Предположим, что фактическая зависимость У от X для всех предприятий генеральной совокупности можно представить в виде следующей модели:

$$У = M(У \mid X) + \varepsilon = У_p + \varepsilon = \alpha_0 + \alpha_1 X + \varepsilon,$$

представленной на рис. 4.5.

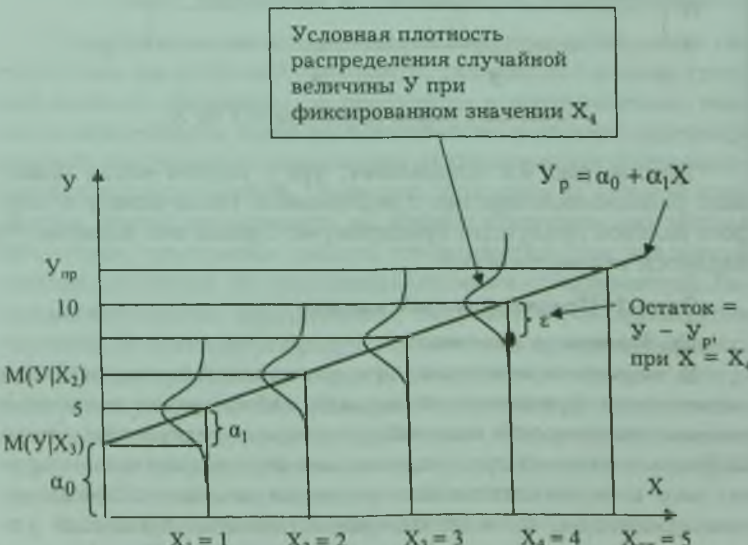


Рис. 4.5. Зависимость У от X для данных генеральной совокупности

Сформулируем в сжатой форме спецификацию модели.

Классическая нормальная линейная модель парной регрессии.

Предполагаем, что зависимость У от X можно представить в виде классической нормальной линейной модели парной регрессии. Приводим спецификацию данной модели

$$У_i = f(X_i) + \varepsilon_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i \quad i = 1, 2, ..., n,$$

где $f(X_i) = M_x(У) = M(У|X = X_i) = \alpha_0 + \alpha_1 X_i$ — функция зависимости У от X;

$M_x(У) = M(У|X = X_i)$ — две записи условного математического ожидания У при фиксированном значении X;

ε_i — неизвестное возмущение, при фиксированном значении X_i;

$У_i$ — случайная (стохастическая) переменная;

X_i — неслучайная (детерминированная) переменная;

α_0, α_1 — неизвестные параметры модели;

n — объем выборки.

Относительно возмущения e_i выдвигаем следующие гипотезы:

$$M(\varepsilon_i) = 0 \text{ при всех } i = 1, 2, ..., n;$$

$$M(\varepsilon_i \varepsilon_j) = \begin{cases} 0 & \text{при } i \neq j; \text{ при } i, j = 1, 2, ..., n; \\ \sigma^2 & \text{при } i = j; \text{ при } i, j = 1, 2, ..., n; \end{cases}$$

ε — нормально распределенная случайная величина со средним $\bar{\varepsilon} = 0$ и дисперсией равной σ^2;

σ^2 — неизвестная дисперсия возмущения ε.

Первая гипотеза означает, что расчетные значения $У_р$ должны пройти через условные математические ожидания.

Вторая гипотеза означает, что ковариация остатков любого порядка равна нулю.

Третья гипотеза означает, что дисперсии остатков являются однородными (гомоскедастичными) и равны константе.

Этап 4. Идентифицируемость модели.

(Можно ли в принципе однозначно восстановить значения неизвестных параметров по выбранной модели).

Классическая нормальная линейная модель парной регрессии соответствует объекту исследования и может быть применена к изучению зависимости У от Х.

Этап 5. Идентификация модели.

(Предложить и реализовать процедуру оценивания параметров модели).

Для оценки параметров и характеристик истинной зависимости У от Х воспользуемся регрессионным уравнением, которое называется выборочной регрессионной моделью зависимости У от Х. Коэффициенты и характеристики которой определим по выборочной совокупности объемом n:

$У_i = a_0 + a_1X_i + e_i = У_{pi} + e_i$,

где $У_i$ — исходные значения зависимой переменной;

X_i — фактор, влияющий на У;

i — порядковый номер измерения;

$У_{pi} = a_0 + a_1X_i$ — расчетные значения зависимой переменной, отражают существующую связь между У и Х;

$e_i = (У_i - У_{pi})$ — остатки модели, отражают влияние неучтенных факторов;

a_1- коэффициент модели, численно равный приросту значения У при изменении Х на единицу, вычисляется с помощью метода наименьших квадратов по следующей формуле

$$a_1 = \frac{\sum(X_i - \overline{X})(У_i - \overline{У})}{\sum(X_i - \overline{X})^2};$$

$\sum(X_i - \overline{X})(У_i - \overline{У})$ — ковариация или сумма произведений отклонений значений Х и У от своих средних значений, отражает синхронность изменений У и Х;

$\overline{X}$ — среднее значение X_i;

$\overline{У}$ — среднее значение $У_i$;

$\sum(X_i - \overline{X})^2$ — вариация переменной Х или сумма квадратов отклонений значений Х от своего среднего значения;

a_0 — коэффициент модели, численно равный значению $У_p$ при значении Х равном нулю и определяется методом наименьших квадратов по следующей формуле

$a_0 = \overline{У} - a_1\overline{X}$.

Модуль 1. Расчет коэффициентов a_1, a_0.

Таблица 4.2

Расчет коэффициентов a_1, a_0

i	X_i	$У_i$	$X_i - \overline{X}$	$У_i - \overline{У}$	$(X_i - \overline{X})(У_i - \overline{У})$	$(X_i - \overline{X})^2$
1	2	3	4	5	6	7
1	1	14	-5,5	-21,83	120,08	30,25
2	2	21	-4,5	-14,83	66,75	20,25
3	3	20	-3,5	-15,83	55,42	12,25
4	4	29	-2,5	-6,83	17,08	6,25
5	5	36	-1,5	0,17	-0,25	2,25
6	6	34	-0,5	-1,83	0,92	0,25
7	7	33	0,5	-2,83	-1,42	0,25
8	8	40	1,5	4,17	6,25	2,25
9	9	41	2,5	5,17	12,92	6,25
10	10	52	3,5	16,17	56,58	12,25
11	11	50	4,5	14,17	63,75	20,25
12	12	60	5,5	24,17	132,92	30,25
Сумма					531	143
Среднее	6,5	35,83				

Расчет коэффициента a_1

$$a_1 = \frac{\sum (X_i - \overline{X})(У_i - \overline{У})}{\sum (X_i - \overline{X})^2} = \boxed{3{,}7133} \qquad \boxed{}$$

где a_1 — коэффициент модели $У_i = a_0 + a_1X_i + e_i$.

$\sum (X_i - \overline{X})(У_i - \overline{У}) =$ $\boxed{531}$ — ковариация переменных X и У

$\sum (X_i - \overline{X})^2 =$ $\boxed{143}$ — вариация переменной X.

Расчет коэффициента a_0

$a_0 = \overline{У} - a_1\overline{X}$	11,697		
где a_0 — коэффициент модели $У_i = a_0 + a_1X_i + e_i$.			
	$\overline{X}$ =	6,5	— среднее значение X,
	a_1 =	3,7133	— коэффициент модели $У_i = a_0 + a_1X_i + e_i$.
	$\overline{У}$ =	35,833	— среднее значение У,

Верификация модели

(Проверка достоверности коэффициентов модели и самой модели на обучающей совокупности. Проверка прогноза модели для контрольной совокупности.) Верификацию модели будем проводить по обучающей выборки.

Ошибка модели

E — стандартная ошибка модели или среднее квадратическое отклонение остатков

$$E = \sqrt{\frac{\sum e_i^{\ 2}}{n-k}}$$

где n — объем выборки;

k — количество коэффициентов в модели;

(n − k) — число степеней свободы остатков модели.

Стандартная ошибка модели зависит от единиц измерения У и не сопоставима с стандартными ошибками других моделей. Стандартная ошибка модели, выраженная в процентах к среднему значению У, лишена этого недостатка. E% — стандартная ошибка модели, выраженная в процентах к среднему значению У.

$$E\% = \frac{E}{\overline{У}}100.$$

В экономических исследованиях принято считать, что если E% не превышает 15%, то модель хорошая; если E%

меньше 5%, то модель очень хорошая. Если E% больше 20%, то модель считается плохой. E% имеет тесную обратную связь с критерием Фишера.

Модуль 2. Вычисление ошибки модели E.

Таблица 4.3

Расчет ошибки модели

i	X_i	$У_i$	$У_{pi}$	$e_i = У_i - У_{pi}$	$(У_i - У_{pi})^2$
1	2	3	4	5	6
1	1	14	15,41	-1,41	1,989
2	2	21	19,12	1,88	3,521
3	3	20	22,84	-2,84	8,048
4	4	29	26,55	2,45	6,002
5	5	36	30,26	5,74	32,909
6	6	34	33,98	0,02	0,001
7	7	33	37,69	-4,69	21,996
8	8	40	41,40	-1,40	1,969
9	9	41	45,12	-4,12	16,946
10	10	52	48,83	3,17	10,050
11	11	50	52,54	-2,54	6,467
12	12	60	56,26	3,74	14,014
Сумма					123,91

Расчет ошибки модели E

$$E = \sqrt{\frac{\sum\left(У_i - У_{pt}\right)^2}{n-k}} = \boxed{3,5201} \qquad \boxed{}$$

где E — стандартная ошибка модели;

$У_{pi} = a_0 + a_1X_i$ — расчетные значения У;

$a_0 = \boxed{11,697}$ — коэффициент a_0 модели $У_i = a_0 + a_1X_i + e_i$;

$a_1 = \boxed{3,7133}$ — коэффициент a_1 модели $У_i = a_0 + a_1X_i + e_i$;

$\sum(У_i - У_{pt})^2 = \boxed{123,91}$ — вариация остатков;

n	=	12	— объем выборки;
k	=	2	— количество коэффициентов в модели.

Проверка достоверности коэффициента a_1.

Выдвигается нулевая гипотеза H_0: $\alpha_1 = 0$.

Проверка нулевой гипотезы H_0.

Если $|t_{a1}| \geq t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k)$, то нулевая гипотеза H_0 отвергается с вероятность $1 - \alpha = 1 - 0{,}05 = 0{,}95$ и утверждается, что коэффициент a_1 достоверно отличается от нуля с вероятностью 0,95.

Если $|t_{a1}| < t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k)$, то нулевая гипотеза H_0 принимается и утверждается, что коэффициент a_1 не отличается от нуля.

Где t_{a1} — критерий Стьюдента для коэффициента a_1 показывает во сколько раз коэффициент a_1 больше его ошибки. Критерий Стьюдента для коэффициентов a_0 и a_1 вычисляется по соответствующим формулам:

$$t_{a1} = \frac{a_1}{S_{a1}}; \quad t_{a0} = \frac{a_0}{S_{a0}},$$

где S_{a0} — стандартное отклонение коэффициента a_0, равное среднеквадратическому отклонению коэффициента a_0 от своего математического ожидания α_0

$$S_{a0} = E\sqrt{\frac{\sum X_i^2}{n\sum(X_i - \bar{X})^2}};$$

S_{a1} — стандартное отклонение коэффициента a_1, равное среднеквадратическому отклонению коэффициента a_1 от своего математического ожидания α_1;

$$S_{a1} = E\sqrt{\frac{1}{\sum(X_i - \bar{X})^2}};$$

$t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k)$ вычисляется по функции “СТЬЮДРАСПОБР”.

Модуль 3. Расчет ошибок коэффициентов a_1 и a_0.

Таблица 4.4

Расчет ошибок коэффициентов a_1 и a_0

i	X_i	X_i^2	$(X_i - \bar{X})^2$
1	2	3	4
1	1	1	30,25
2	2	4	20,25
3	3	9	12,25
4	4	16	6,25
5	5	25	2,25
6	6	36	0,25
7	7	49	0,25
8	8	64	2,25
9	9	81	6,25
10	10	100	12,25
11	11	121	20,25
12	12	144	30,25
Сумма		650	143
Среднее	6,5		

$$S_{a1} = E\sqrt{\frac{1}{\sum(X_i - \bar{X})^2}} = \boxed{0,2944} \qquad \boxed{}$$

где S_{a1} — ошибка коэффициента a_1 или среднее квадратическое отклонение коэффициента a_1 от параметра α_1, рассчитанного для всех объектов генеральной совокупности.

$E = \boxed{3,5201}$ — ошибка модели.

$\sum(X_i - \bar{X})^2 = \boxed{143}$ — вариация переменной X.

$$S_{a0} = E\sqrt{\frac{\sum(X_i)^2}{n\sum(X_i - \bar{X})^2}} = \boxed{2,1665} \qquad \boxed{}$$

где S_{a0} — ошибка коэффициента a_0 или среднее квадратическое отклонение коэффициента a_0 от параметра α_0, рассчитанного для всех объектов генеральной совокупности.

E	=	3,5201	— ошибка модели.
$\sum(X_i - \bar{X})^2$	=	143	— вариация переменной X.
$\sum(X_i)^2$	=	650	— сумма квадратов переменной X.
n	=	12	— объем выборки.

Модуль 4. Вычисление критерия Стьюдента для коэффициента a_1 и a_0.

$$t_{a1} = \frac{a_1}{Sa_1} = \boxed{12,615} \qquad \boxed{}$$

где t_{a1} — критерий Стьюдента для коэффициента a_1.

$a_1 = \boxed{3,7133}$ — коэффициент модели $У_i = a_0 + a_1X_i + e_i$.

$S_{a1} = \boxed{0,2944}$ — среднее квадратическое отклонение коэффициента a_1.

$$t_{a0} = \frac{a_0}{Sa_0} = \boxed{5,3991} \qquad \boxed{}$$

где t_{a0} — критерий Стьюдента для коэффициента a_0.

$a_0 = \boxed{11,697}$ — коэффициент модели $У_i = a_0 + a_1X_i + e_i$.

$S_{a0} = \boxed{2,1665}$ — среднее квадратическое отклонение коэффициента a_0.

Модуль 5. Определим табличное значение критерия Стьюдента на уровне значимости a = 0,05 с помощью статистической функции "СТЬЮДРАСПОБР".

$t_{\alpha/2}(\alpha = 0,05;\ m = n - k) =$	2,2281	— табличное значение критерия Стьюдента.
где $n =$	12	— объем выборки;

k =	2	— количество коэффициентов в модели.

Модуль 6. Проверим достоверность коэффициента a_1.

Проверим нулевую гипотезу H_0: "$a_1 = 0$".
Так как $|t_{a1}| = 12{,}615 > t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k) = 2{,}2281$, то нулевая гипотеза H_0: " $a_1 = 0$ " отвергается с вероятностью $1 - \alpha = 0{,}95$.

Определение доли объясненной вариации.

R^2 — коэффициент детерминации равен доли общего варьирования, которое объясняется за счет уравнения регрессии.

$$R^2 = \frac{\sum(У_{pi} - \bar{У})^2}{\sum(У_i - \bar{У})^2}.$$

Модуль 7. Вычислим долю объясненной вариации переменной У.

Таблица 4.5

Расчет коэффициента детерминации и критерия Фишера

i	$У_i$	$Ур_i$	$(У_i - \bar{У})^2$	$(У_{pi} - \bar{У})^2$	$(У_i - У_{pi})^2$
1	14	15,41	476,69	417,10	1,989
2	21	19,12	220,03	279,22	3,521
3	20	22,84	250,69	168,91	8,048
4	29	26,55	46,69	86,18	6,002
5	36	30,26	0,03	31,02	32,909
6	34	33,98	3,36	3,45	0,001
7	33	37,69	8,03	3,45	21,996
8	40	41,40	17,36	31,02	1,969
9	41	45,12	26,69	86,18	16,946
10	52	48,83	261,36	168,91	10,050
11	50	52,54	200,69	279,22	6,467
12	60	56,26	584,03	417,10	14,014
Сумма			2095,7	1971,8	123,91
Среднее	35,833	35,833			

$$R^2 = \frac{\sum\left(У_{pi} - \overline{У}\right)^2}{\sum\left(У_i - \overline{У}\right)^2} = \quad 0{,}9409$$

Где R^2 — коэффициент детерминации.

$\sum(У_{pi} - \overline{У})^2 \quad = \quad 1971{,}8$ — вариация регрессии.

$\sum(У_i - \overline{У})^2 \quad = \quad 2095{,}7$ — вариация переменной У.

Проверка достоверности модели.

Выдвигается нулевая гипотеза H_0: Модель недостоверна или $У_{pi} = \overline{У}$ или

$$S^2_{рег} = S^2_{ост}.$$

Проверка нулевой гипотезы H_0.

Если $F \geq F_{кр}$ ($\alpha = 0{,}05$; $m_1 = k - 1$; $m_2 = n - k$), то нулевая гипотеза H_0 отвергается с вероятностью совершить ошибку, равной $\alpha = 0{,}05$, или нулевая гипотеза H_0 отвергается с вероятностью $1 - \alpha = 1 - 0{,}05 = 0{,}95$ и утверждается, что модель является достоверной с вероятностью 0,95.

Если $F < F_{кр}$ ($\alpha = 0{,}05$; $m_1 = k - 1$; $m_2 = n - k$), то нулевая гипотеза H_0 принимается и утверждается, что модель является недостоверной. Где F-критерий Фишера означает во сколько раз дисперсия регрессии больше дисперсии остатков. Чем точнее модель, тем больше критерий Фишера и меньше процент стандартной ошибки модели Е%.

$$F = \frac{S^2_{рег}}{S^2_{ост}},$$

где $S_{рег}{}^2$ — дисперсия, обусловленная регрессией, или сумма квадратов отклонений расчетных значений У от своего среднего значения, деленная на число степеней свободы,

$$S^2_{рег} = \frac{\sum(У_{pi} - \overline{У})^2}{k - 1};$$

$S_{ост}{}^2$ — дисперсия остатков или сумма квадратов отклонения фактических от расчетных значений переменной У, деленная на число степеней свободы,

$$S^2_{ост} = \frac{\sum(У_i - У_{pi})^2}{n-k};$$

$F_{кр}(\alpha = 0{,}05;\ m_1 = k - 1;\ m_2 = n - k)$ — табличное значение критерия Фишера, определенное на уровне значимости $\alpha = 0{,}05$, числе степеней свободы для большей дисперсии m_1 и числе степеней свободы для меньшей дисперсии m_2. $F_{кр}$ определяется по функции "FРАСПОБР".

Модуль 8. Вычисление критерия Фишера.

$$F = \frac{S^2_{рег}}{S^2_{ост}} = \frac{(n-k)\sum\left(У_{pi} - \bar{У}\right)^2}{(k-1)\sum(У_i - У_{pi})^2} = \boxed{159{,}13} \qquad \boxed{}$$

где F-критерий Фишера.

$\sum(У_{pi} - \bar{У})^2 = \boxed{1971{,}8}$ — вариация регрессии.

$\sum(У_i - У_{pi})^2 = \boxed{123{,}91}$ — вариация остатков.

$n = \boxed{12}$ — объем выборки.

$k = \boxed{2}$ — количество коэффициентов в модели.

$S^2_{рег}$ — дисперсия регрессии.

$S^2_{ост}$ — дисперсия остатков.

Модуль 9. Расчет табличного критерия Фишера.

Вычислим табличное значение критерия Фишера на уровне значимости $\alpha = 0{,}05$ и числе степеней свободы для большей дисперсии m_1 и числе степеней свободы для меньшей дисперсии m_2 с помощью статистической функции "FРАСПОБР"

$F_{кр}(\alpha = 0{,}05;\ m_1 = k - 1;\ m_2 = n - k) = 4{,}9646$

$n = 12$ — объем выборки.
$k = 2$ — количество коэффициентов в уравнении.

Модуль 10. Проверка достоверности модели.

Проверим нулевую гипотезу H_0: "модель недостоверна", или H_0: "$У_р = \bar{У}$", или H_0: "$S_{рег}{}^2 = S_{ост}{}^2$",

где $S_{рег}{}^2$ — дисперсия регрессии;

$S_{ост}{}^2$ — дисперсия остатков.

Так как $F = 159{,}13 > F_{кр}(\alpha = 0{,}05;\ m_1 = k - 1;\ m_2 = n - k) = 4{,}964$, то нулевая гипотеза H_0:"модель недостоверна" отвергается с вероятностью $1 - \alpha = 0{,}95$.

Получение точечного и интервального прогноза.

Точечный прогноз — это прогнозное среднее значение зависимой переменной У.

$У_{пр} = a_0 + a_1 X_{ож}$,

где $X_{ож}$ — ожидаемое значение регрессора X, выходящее за пределы значений X обучающей выборочной совокупности;

$У_{пр}$ — точечное прогнозное значение переменной У — является оценкой математического ожидания или индивидуального прогнозного значения У.

Интервальный прогноз — это интервал, в котором с заданной вероятностью будет находиться фактическое значение зависимой переменной У.

95%-ный прогнозный доверительный интервал нахождения прогнозного значения математического ожидания У определяется по формуле:

$$У_м = У_{пр} \pm t_{\alpha/2} W_м,$$

где $У_м$ — математическое ожидание У для значения $X_{ож}$,

$W_м$ — стандартная ошибка прогноза математического ожидания У, равная среднеквадратическому отклонению между точечным прогнозом и математическим ожиданием У;

$$W_м = E\sqrt{\frac{1}{n} + \frac{(X_{ож} - \bar{X})^2}{\sum(X_i - \bar{X})^2}};$$

$t_{\alpha/2}$ ($\alpha = 0{,}05$; $m = n - k$) — табличное значение двухстороннего критерия Стьюдента.

95%-ный прогнозный доверительный интервал нахождения прогнозного индивидуального значения У определяется по формуле:

$$У_и = У_{пр} \pm t_{\alpha/2} W_и,$$

где $У_и$ — индивидуальное значение У для значения $X_{ож}$,

$W_и$ — стандартная ошибка прогноза индивидуального значения У, равная среднеквадратическому отклонению между точечным прогнозом и индивидуальным значением У.

$$W_и = E\sqrt{1 + \frac{1}{n} + \frac{(X_{ож} - \bar{X})^2}{\sum(X_i - \bar{X})^2}}.$$

Получение доверительных интервалов уравнения регрессии.

95%-ный доверительные интервалы математических ожиданий $У_i$ для значений X_i вычисляются по формулам:

$У_{min(м)i} = У_{pi} - t_{\alpha/2} W_{мi}$ — нижний 95%-ный доверительный интервал для математического ожидания У;

$У_{max(м)i} = У_{pi} + t_{\alpha/2} W_{мi}$ — верхний 95%-ный доверительный интервал для математического ожидания У,

где $W_{мi}$ — стандартная ошибка расчетного значения $У_{pi}$ для значения X_i по отношению к математическому ожиданию значения $У_i$;

$$W_{мi} = E\sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum(X_i - \bar{X})^2}}.$$

Экономическая интерпретация доверительных интервалов для математических ожиданий У.

С вероятностью 0,95 можно утверждать, что математические значения У или средние значения У, рассчитанные для всей генеральной совокупности для каждого фиксированного значения X_i, или уравнение регрессии, рассчитанное для всех данных генеральной совокупности, будут находиться в интервале от $У_{min(м)i}$ до $У_{max(м)i}$.

95%-ный доверительные интервалы индивидуальных значений $У_i$ для значений X_i вычисляются по формулам:

$У_{min(и)i} = У_{pi} - t_{\alpha/2} W_{иi}$ — нижний 95%-ный доверительный интервал для индивидуального значения У;

$У_{max(и)i} = У_{pi} + t_{\alpha/2} W_{иi}$ — верхний 95%-ный доверительный интервал для индивидуального значения У,

где $W_{иi}$ — стандартная ошибка расчетного значения $У_{pi}$ для значения X_i по отношению к индивидуальному значению $У_i$,

$$W_{иi} = E\sqrt{1 + \frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum (X_i - \overline{X})^2}}.$$

Экономическая интерпретация доверительных интервалов для индивидуальных значений У.

С вероятностью 0,95 можно утверждать, что индивидуальное значения $У_i$ или значения $У_i$, которые относятся к выборочным объектам, взятые из генеральной совокупности, для каждого фиксированного значения X_i будут находиться в интервале от $У_{min(и)i}$ до $У_{max(и)i}$.

Вычислим точечный и интервальный прогноз У при ожидаемом значении $X_{ож} = 13$.

Получим точечное прогнозное значение $У_{пр}$ при ожидаемом значении $X_{ож}$.

$У_{пр} = a_0 + a_1 X_{ож}$ = $\boxed{59,97}$ $\boxed{}$

где a_0 = $\boxed{11,697}$ — коэффициент a_0 модели $У_i = a_0 + a_1 X_i + e_i$,

a_1 = $\boxed{3,7133}$ — коэффициент a_1 модели $У_i = a_0 + a_1 X_i + e_i$,

$X_{ож}$ = $\boxed{13}$ — ожидаемое значение X.

Модуль 11. Рассчитаем 95%-ный доверительный интервальный прогноз нахождения прогнозного математического ожидания (среднего значения)

Упрг по всем предприятиям отрасли (для всей генеральной совокупности) при ожидаемом значении $X_{ож}$, который вычисляется по формуле

$$У_м = У_{пр} \pm t_{\alpha/2} E \sqrt{\frac{1}{n} + \frac{\left(X_{ож} - \overline{X}\right)^2}{\sum (X_i - \overline{X})^2}}$$

$У_м = \boxed{9,97} \pm \boxed{4,8272}$ – $(1 - \alpha = 0,05) \cdot 100 = 95\%$-ный доверительный интервал прогноза математического ожидания У.

где $У_м$ — прогнозное значение У для генеральной совокупности или математическое ожидание У;

$У_{пр} = a_0 + a_1 X_{ож} = \boxed{59,97}$ — прогнозное значение У для выборочной совокупности;

$X_{ож} = \boxed{13}$ — ожидаемое значение X;

$t_{\alpha/2(\alpha = 0,05;\ m = n - k)} = \boxed{2,2281}$ — табличное значение критерия Стьюдента;

$n = \boxed{12}$ — объем выборки;

$k = \boxed{2}$ — количество коэффициентов в модели;

$E = \boxed{3,5201}$ — ошибка модели;

$\sum \left(X_i - \overline{X}\right)^2 = \boxed{143}$ — вариация переменной X;

$\overline{X} = \boxed{6,5}$ — среднее значение X;

$t_{\alpha/2} E \sqrt{\frac{1}{n} + \frac{\left(X_{ож} - \overline{X}\right)^2}{\sum (X_i - X)^2}} = \boxed{4,8272}$ — ошибка прогноза. $\boxed{}$

Модуль 12. Вывод по определению интервального прогноза.

> С вероятностью 95% можно утверждать, что среднее значение объема производства для генеральной совокупности ($У_м$) при ожидаемом значении $X_{ож} = 13$ будет находиться в интервале от 55,142 до 64,797.

Модуль 13. Рассчитаем 95-процентный доверительный интервал нахождения математического ожидания (среднего значения) $У_{мi}$ по всем предприятиям отрасли (для всей генеральной совокупности) при каждом значении X, который вычисляется по формуле

$$У_{мi} = У_{pi} \pm t_{\alpha/2} E \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}},$$

где $У_{мi}$ — математическое ожидание У для i-го измерения;

$У_{pi} = a_0 + a_1 X_i$ — расчетное значение У для i-го измерения,

X_i — значения X для i-го измерения;

$t_{\alpha/2(\alpha = 0{,}05;\ m = n - k)} = \boxed{2{,}2281}$ — табличное значение критерия Стьюдента;

$n = \boxed{12}$ — объем выборки;

$k = \boxed{2}$ — количество коэффициентов в модели;

$E = \boxed{3{,}5201}$ — ошибка модели;

$\sum (X_i - \bar{X})^2 = \boxed{143}$ — вариация переменной X;

$\bar{X} = \boxed{6{,}5}$ — среднее значение X.

Таблица 4.6

Расчет 95%-ного доверительного интервала математического ожидания У

i	X_i	$У_{pi}$	$У_{ммin}(i)$	$У_{ммax}(i)$
1	2	3	4	5
1	1	15,41	11,15	19,67
2	2	19,12	15,40	22,84
3	3	22,84	19,61	26,06
4	4	26,55	23,75	29,35
5	5	30,26	27,79	32,73
6	6	33,98	31,69	36,26
7	7	37,69	35,40	39,98
8	8	41,40	38,93	43,87
9	9	45,12	42,32	47,91
10	10	48,83	45,61	52,05
11	11	52,54	48,82	56,26
12	12	56,26	52,00	60,52
Сумма	78	430,00	392,49	467,51
Среднее	6,5	35,83		

где $У_{ммin(i)}$ — 95%-ный нижний доверительный интервал математических ожиданий У;

$У_{ммax(i)}$ — 95%-ный верхний доверительный интервал математических ожиданий У.

Модуль 14. Произведем эконометрический анализ линейной модели. Линейная модель экономического процесса имеет следующий вид

$$У_i = a_0 + a_1X_i + e_i$$

$У_i = 11,697 + 3,7133X_i + e_i$

$S_{a0} = 2,1665$ $S_{a1} = 0,2944$

$t_{a0} = 5,3991$ $t_{a1} = 12,615$ $t_{\alpha/2} = 2,2281$

$E = 3,5201$

$R^2 = 0,9409$

$F = 159,13$ $F_{кр} = 4,9646,$

где a_0 — свободный коэффициент модели $У_i = a_0 + a_1X_i + e_i$,

a_1 — коэффициент модели $У_i = a_0 + a_1X_i + e_i$,

Sa_0 — ошибка коэффициента a_0,

Sa_1 — ошибка коэффициента a_1,

t_{a0} — критерий Стьюдента для коэффициента a_0,

t_{a1} — критерий Стьюдента для коэффициента a_1,

E — ошибка модели,

R^2 — коэффициент детерминации или доля объясненной вариации У,

F — критерий Фишера,

$t_{\alpha/2}$ — табличное значение критерия Стьюдента,

$F_{кр}$ — табличное значение критерия Фишера.

Проверка достоверности модели.

Проверим нулевую гипотезу H_0: "модель недостоверна", или H_0: "$У_р = \overline{У}$", или H_0: "$S_{рег}^2 = S_{ост}^2$",

где $S_{рег}^2$ — дисперсия регрессии;

$S_{ост}^2$ — дисперсия остаточная.

Так как $F = 159{,}13 > F_{кр(\alpha = 0{,}05;\ m_1 = k-1;\ m_2 = n-k)} = 4{,}9646$, то нулевая гипотеза H_0: "модель недостоверна" отвергается с вероятностью $1 - \alpha = 0{,}95$.

Вывод: модель зависимости Y от X, представленная в виде регрессионного уравнения

$У_i\ a_0 + a_1X + e_i = 11{,}697 + 3{,}7133X_i + e_i$,

построенного по выборочной совокупности равной 12 месяцам, является достоверной с вероятностью 0,95.

Проверка достоверности коэффициента a_1.

Проверим нулевую гипотезу H_0: "$a_1 = 0$".

Так как $|t_{a1}| = 12{,}615 > t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k) = 2{,}228$, то нулевая гипотеза H_0: "$\alpha_1 = 0$" отвергается с вероятностью $1 - \alpha = 0{,}95$.

Вывод.

Коэффициент $a_1 = 3{,}7133$ достоверно отличается от нуля.

Модуль 15. Общий вывод.

Линейная модель динамики объема производства может быть представлена в виде регрессионного уравнения

$У_i = 11{,}697 + 3{,}7133\ X_i + e_i$

является достоверной с вероятностью 0,95.

Коэффициент $a_1 = 3{,}7133$ достоверно отличается от нуля.

При $X_{ож} = 13$ точечный прогноз или наиболее вероятное прогнозное значение прироста объема производства будет равно $У_{пр} = 59{,}97$.

С вероятностью 0,95 можно утверждать, что математическое ожидание (среднее значение) прироста объема производства или интервальный прогноз будет находиться в интервале от 55,142 до 64,797

Графическое представление результатов расчета представлено на рис. 4.6.

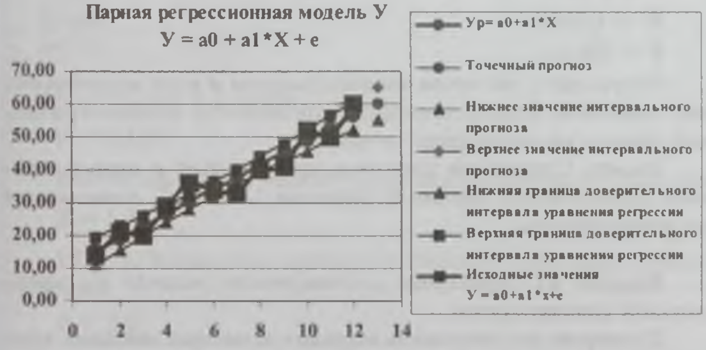


Рис. 4.6. График полученных расчетов

Модуль 16. Проверка расчетов.

Произведем проверку всех расчетов с использованием статистической функции "Линейн" в среде Excel 97, которая выполняется в такой последовательности:

• ввести в рабочую ячейку функцию "Линейн" с указанием диапазона ячеек для зависимой переменной У и объясняемой переменной X, ввести условие вывода всех характеристик функции "Линейн";

• выделить два столбца и пять строчек, F2, Cirl+Shifi+Enter.

Представляем результат выполнения статистической функции "Линейн".

3,713287	11,69697
0,294366	2,166475
0,940873	3,520105
159,1262	10
1971,755	123,9114

$У = 11{,}697 + 3{,}7133X$

$S_{a0} = 2{,}1665;$ $S_{a1} = 0{,}2944;$

$t_{a0} = 5{,}3991;$ $t_{a1} = 12{,}615;$

$E = 3{,}5201;$

$R^2 = 0{,}9409;$

$F = 159{,}13.$

Результаты расчетов коэффициентов и всех характеристик линейной модели позволяют произвести эконометрический анализ регрессионной модели.

Вывод. Сравнение расчетов, полученных с использованием электронных таблиц и функции "Линейн", показывают их идентичность.

Модуль 17. Проверка достоверности модели на контрольной совокупности.

Проверим достоверность модели с помощью внешних критериев в такой последовательности.

(Все расчеты выполнены в Excel и изображены на рис. 4.7 и рис. 4.8)

Шаг 1. Разделим всю выборочную совокупность на две части — обучающую, которая составит 2/3 объема выборки и контрольную, которая составит 1/3 объема выборки.

Таблица 4.7

Обучающая выборка

i	X_i	$У_i$
1	1	14
2	2	21
3	3	20
4	4	29
5	5	36
6	6	34
7	7	33
8	8	40

Таблица 4.8

Контрольная выборка

i	X_i	$У_i$
1	9	41
2	10	52
3	11	50
4	12	60

Шаг 2. Вычислим коэффициенты, ошибку линейной модели для обучающей выборки с помощью функции "Линейн".

3,464286	12,78571
0,541241	2,733135
0,872254	3,507645
40,96807	6
504,0536	73,82143

$У_p = 12{,}78 + 3{,}464X$

$E = 3{,}507$

Шаг 3. Вычислим ошибку модели на контрольной выборке.

$E = 5{,}58$

Шаг 4. Вычислим критерий Фишера для контрольной выборки.

$F = 2{,}2;$

$F_{кр}(0{,}05;\ m_1 = к - 1 = 2 - 1;\ m_2 = n - к = 4 - 2) = 18{,}51.$

Шаг 5. Проверим достоверность модели по контрольной выборке.

Обучающая выборка

i	Xi	$У_{обуч}$	$У_{рОбуч}$	
1	1	14	16,25	
2	2	21	19,71429	
3	3	20	23,17857	
4	4	29	26,64286	
5	5	36	30,10714	
6	6	34	33,57143	
7	7	33	37,03571	
8	8	40	40,5	
F =	40,96807		$F_{кр}(0{,}05;\ к-1=2-1;\ n-к=8-2)=$	5,987374152
E =	3,507645			

Контрольная выборка

i	X_i	$У_{кон}$	$У_{рКон}$	$(У_{кон}-У_{рКон})^2$	$(У_{рКон}-\overline{У})^2$
1	9	41	43,96429	8,786989796	46,04591837
2	10	52	47,42857	20,89795918	11,03188776
3	11	50	50,89286	0,797193878	0,020408163
4	12	60	54,35714	31,84183673	13,01147959
$\sum$		203		62,32397959	70,10969388
$\overline{У}$		50,75			
$S^2_{ост}=(\sum(У_{кон}-У_{рКон})^2)/(n-к)=$				31,1619898	
$S^2_{рег}=(\sum(У_{рКон}-\overline{У})^2)/(к-1)=$				70,10969388	
$F=S_{рег}{}^2/S_{ост}{}^2=$	2,2498	$F_{кр}$ (0,05; к-1 = 2-1; n-к = 4-2) =			18,51276465
$E=\sqrt{S^2_{ост}}=$	5,5822				

Рис. 4.7. Протокол проведения расчетов

Анализ графика 4.8 показывает, что остатки модели на обучающей и контрольной совокупности имеют одинаковые значения. Для предварительного анализа этого достаточно, чтобы говорить о хорошем воспроизведении исходных данных контрольной совокупности с помощью модели, построенной по обучающей совокупности.

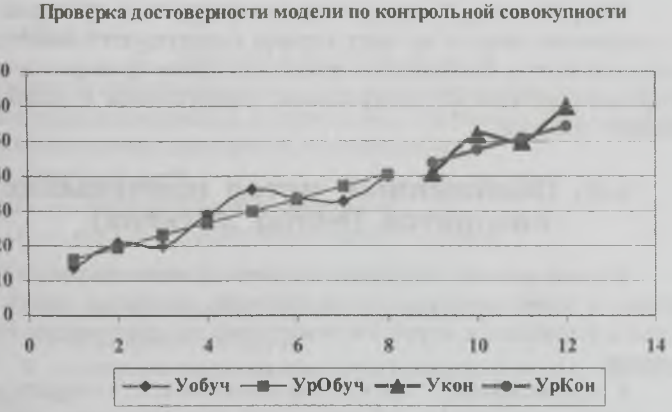


Рис. 4.8. График проверки достоверности модели на контрольной совокупности

Стандартный статистический анализ показывает следующее:

• ошибка модели для контрольной совокупности больше, чем по обучающей выборки. Это различие объясняется разностью числа степеней свободы. Если вычислить средние ошибок модели, то они будут равны как для обучающей, так и для контрольной совокупностей;

• по критерию Фишера модель на обучающей совокупности является достоверной, так как $F = 40{,}9 > F_{кр}(0{,}05;\ m_1 = k - 1 = 2 - 1;\ m_2 = n - k = 8 - 2) = 5{,}98$. На контрольной совокупности модель является недостоверной, так как $F = 2{,}249 < F_{кр}(0{,}05;\ m_1 = k - 1 = 2 - 1;\ m_2 = n - k = 4 - 2) = 18{,}51$. Причиной недостоверности модели на контрольной совокупности является малый объем ее выборки. Стандартный статистический анализ показывает, что модель плохо воспроизводит тенденцию на контрольной совокупности, хотя визуальный анализ показывает, что модель хорошо воспроизводит тенденцию на контрольной совокупности.

Общий вывод. Стандартная статистическая проверка достоверности модели на контрольной совокупности имеет малую мощность. Необходимы новые методики проверки достоверности модели по контрольной совокупности в условиях малых выборок.

4.6. Обобщенный метод наименьших квадратов (метод Эйткена)

Использование обобщенного метода наименьших квадратов в экономических исследованиях является одним из самых сложных в курсе эконометрики по следующим причинам:

- использование сложного математического аппарата для получения расчетов коэффициентов и характеристик модели;
- нет четкой интерпретации новой модели после преобразований данных;
- нет методики перехода от модели с преобразованными данными к исходным данным.

С помощью обобщенного метода наименьших квадратов Эйткена можно найти коэффициенты линейной модели преобразованных данных, для которых остатки станут гомоскедастичными при отсутствии автокорреляция остатков. Однако имеются трудности в экономической интерпретации новых преобразованных данных и нет методики построения доверительных интервалов уравнения регрессии для исходных данных при наличии гетероскедастичности остатков.

Если использовать коэффициенты, определенные обобщенным методом наименьших квадратов, для исходных непреобразованных данных, то уравнение регрессии будет проходить ближе к тем измерениям, для которых остатки модели были наименьшими. Это удивительное свойство коэффициентов метода Эйткена можно использовать для учета информативности данных.

Допустим для модели $\mathbf{У} = \mathbf{X}\alpha + \varepsilon$ нарушается предпосылка метода наименьших квадратов

$$M(\varepsilon_i \varepsilon_j) = \begin{cases} 0 & \text{при } i \neq j,\ i, j = 1, 2, \dots, n \\ \sigma_\varepsilon^2 & \text{при } i = j,\ i, j = 1, 2, \dots, n \end{cases}$$

и вводится предположение о том, что возмущения могут быть автокоррелированными и гетероскедастичными. Это предположение можно записать в матричном виде

$$\mathbf{M}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathbf{V} \neq \sigma_\varepsilon^2 \mathbf{I},$$

где M — математическое ожидание;

$\boldsymbol{\varepsilon}$ — вектор-столбец возмущения модели;

σ_ε^2 — дисперсия возмущения;

$\boldsymbol{\alpha}$ — вектор-столбец параметров модели;

$\mathbf{I}$ — единичная матрица;

$\mathbf{V}$ — ковариационная матрица (предполагается, что она известна).

Если оценить параметры $\boldsymbol{\alpha}$ с помощью коэффициентов модели по выборочным данным методом наименьших квадратов по формуле

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{У},$$

то полученные коэффициенты будут несмещенными, состоятельными, но неэффективными.

Если исходные значения **X** и **У** соответствующим образом преобразовать и по преобразованным значениям вычислить коэффициенты модели методом наименьших квадратов, то по отношению к преобразованным данным полученные коэффициенты будут обладать свойствами несмещенности, состоятельности и эффективности. Предложенный метод называется обобщенным методом наименьших квадратов, или методом Эйткена.

После преобразования исходных данных получается фактически новая модель с новыми переменными, которые часто не имеют экономического смысла. При определенных условиях удается получить практическую применимость данного метода.

Если матрица **V** является симметрической и положительно определенной, то верно соотношение:

$$\mathbf{V} = \mathbf{P}\mathbf{P}',$$

где матрица **P** является невырожденной.

Матрицу **P** можно вычислить по формуле[1]

$$\mathbf{P} = [(\mathbf{h}\mathbf{D})^{-1}]',$$

$$\mathbf{D} = \mathbf{I}\frac{1}{\sqrt{\mathbf{h}'\mathbf{V}\mathbf{h}}} = \mathbf{I}\frac{1}{\sqrt{\mathbf{k}}},$$

где **I** — единичная матрица;

h — собственный (характеристический) вектор матрицы **V**;

k — собственное (характеристическое) значение матрицы **V**.

Собственные значения **k** и соответствующие им собственные вектора **h** матрицы **V** можно определить с помощью статистических или математических пакетов, или из уравнения[2]:

$$\mathbf{V}\mathbf{h} = \mathbf{k}\mathbf{h}.$$

Умножим уравнение модели $\mathbf{У} = \mathbf{X}\mathbf{A} + \mathbf{e}$ слева на матрицу $\mathbf{P}^{-1}$, получим

$$\mathbf{У}_{н} = \mathbf{X}_{н}\mathbf{B} + \mathbf{e}_{н},$$

где $\mathbf{У}_{н} = \mathbf{P}^{-1}\mathbf{У}$; $\mathbf{X}_{н} = \mathbf{P}^{-1}\mathbf{X}$ и $\mathbf{e}_{н} = \mathbf{P}^{-1}\mathbf{e}$.

Коэффициенты **B** можно вычислить с помощью обобщенного метода наименьших квадратов (метод Эйткена) по формуле

$$\mathbf{B} = (\mathbf{X}_{н}'\mathbf{X}_{н})^{-1}\mathbf{X}_{н}'\mathbf{У}_{н} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{У}.$$

Коэффициенты **B** будут несмещенными, состоятельными и эффективными по отношению к преобразованным данным $\mathbf{У}_{н}$ и $\mathbf{X}_{н}$. К преобразованным данным можно применить весь арсенал эконометрического анализа с проверкой статис-

---

[1] *Джонстон Д.* Эконометрические методы. — М.: Статистика, 1980. С. 112–113.

[2] Алгоритмы (с контрольными примерами) определения **k** и **h**, которые можно реализовать в Excel имеются в книгах: *Иберла К.* Факторный анализ. — М.: Статистика, 1980; *Харман Г.* Современный факторный анализ. — М.: Мир, 1972.

тических гипотез с получением точечного и интервального прогноза (для преобразованных данных). Однако возникают проблемы получения значений точечного и интервального прогноза для фактических данных.

Известны формулы получения характеристик модели по исходным данным с использованием матрицы **V**, но они очень громоздкие и не отличаются от характеристик модели, полученные по преобразованным данным. Поэтому важно знать правила получения матрицы преобразования $\mathbf{P}^{-1}$.

Если условие $\mathbf{M}(\varepsilon\varepsilon') = \mathbf{V} = \sigma_\varepsilon^2 \mathbf{I}$ выполняется, то коэффициенты **A** и **B** будут равны между собой, которые вычисляются по соответствующим формулам:

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{У}; \;\; \mathbf{B} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{У}.$$

Вопросы для самоконтроля

1. Приведите формулу ошибки модели.
2. Выведите основное вариационное уравнение.
3. Укажите разницу между коэффициентом детерминации и приведенным коэффициентом детерминации.
4. Дайте определение нулевой гипотезы.
5. Дайте определение уровня значимости.
6. Приведите шаги проверки достоверности модели с помощью критерия Фишера.
7. Приведите шаги проверки достоверности коэффициентов модели.
8. Приведите формулу расчета частного критерия детерминации.
9. Приведите методологические основы прогнозирования.
10. Приведите формулы точечного и интервального прогноза для среднего значения.
11. Укажите способы уменьшения ошибки прогноза.
12. Приведите этапы эконометрического анализа регрессионной модели.

Глава 5
ПРОВЕДЕНИЕ РАСЧЕТОВ ХАРАКТЕРИСТИК МОДЕЛИ С ПОМОЩЬЮ ЭВМ

В эконометрике для проведения расчетов часто используют табличный процессор Excel 97, 2000 и несколько специализированных пакетов прикладных программ.

Необходимо выяснить, каким образом можно выполнить расчеты с использованием ЭВМ: как ввести данные, как получить расчеты, что означают выводимые характеристики, и как их можно использовать в эконометрическом анализе. Содержание данной главы можно использовать при проведении лабораторной работы как первое знакомство с функциями Excel и статистическими пакетами прикладных программ, позволющих выполнить прогноз.

5.1. Расчет характеристик модели с помощью табличного процессора Excel

5.1.1. Возможности табличного процессора Excel

Выясним возможности табличного процессора Excel 97 для проведения эконометрического анализа.

Excel располагает почти всеми инструментальными средствами для проведения эконометрического анализа.

Предполагаем, что читатель знаком с основными элементами проведения расчетов данных, представленных в виде таблицы, которые изучались в курсе информатики. Более подробно остановимся на использовании некоторых функций и программ, которые часто применяются в эконометрическом анализе: графических средств Excel, функции “Линейн”, пакета программ “Анализ данных”, программы “Поиск решения”, функции “Предсказ”, ППП Stadia.

Технология проведения расчетов состоит из четырех этапов: ввод данных, проверка данных, вычисление прогноза, анализ полученных расчетов.

Ввод данных осуществляется в электронную таблицу при выполнении следующих правил:

- факторы вводятся последовательно по колонкам, каждая колонка обозначается соответственно X_1, X_2 и т. д., последняя колонка должна содержать значения зависимой переменной, обозначаемой как У;
- количество значений в каждой колонке должно быть одинаковым и не должны содержать пропусков и нулевых значений;
- внизу таблицы даются пояснения к введенным обозначениям переменных.

Проверка данных осуществляется с целью выяснения возможности проведения прогноза. Прогноз будет иметь практическую ценность, если данные обладают следующими свойствами: *сопоставимостью, полнотой, однородностью, устойчивостью, репрезентативностью.*

Сопоставимость данных предполагает использование одинаковых: единиц измерения уровней временного ряда; шага наблюдения, интервала изучаемого времени; методики сбора данных.

Полнота данных предполагает включение в модель факторов, составляющих полную группу причин, влияющих на зависимую переменную.

Однородность данных бывает двух видов: по форме и по структуре. Однородность по форме предполагает отсутствие в данных резких изменений тенденции. Однородность по структуре означает отсутствие в данных ошибочных и резко выделяющихся значений.

Устойчивость временного ряда означает наличие четко выраженной определенной тенденции в значениях уровней ряда, которое легко можно определить с помощью визуального анализа.

Репрезентативность (представительность по количеству измерений, способу выборки, размеру обучающей выборки для получения прогноза) данных временного ряда будет означать, что количество измерений уровней временного ряда должно быть в три раза больше количества факторов, включенных в модель; количество обучающих уровней временного ряда, используемых для прогноза, должно быть не менее 10; глубина прогноза должна быть в три раза меньше количества обучающих уровней временного ряда.

Вычисление прогноза можно проводить с помощью различных инструментальных средств: графических средств Excel, функции “Линейн”, пакета программ “Анализ данных”, программы “Поиск решения”, функции “Предсказ”, ППП Stadia. Результаты расчетов должны совпадать между собой, если имеются расхождения в расчетах, то ищите ошибку в процессе ввода и обработке данных.

Анализ полученных расчетов состоит из двух этапов: проверка правильности расчетов и выяснение возможности использования расчетов для принятия управленческих решений.

Правильность расчетов проверяется несколькими методами: выполнение контрольного примера, сравнение результатов расчета, выполненных разными инструментальными средствами.

Возможность использования расчетов для принятия управленческих решений проверяется в такой последовательности: проверяется соблюдение требований к исходным данным; проверяется достоверность модели, по которой проводился прогноз; сравнивается рассчитанное значение прогноза с результатом визуального анализа (так как визуальный анализ данных является одним из самых мощных методов определения прогноза среди известных методов прогнозирования).

Варианты получения прогнозов с использованием различных инструментальных средств можно выполнять на различных листах табличного процессора Excel.

Варианты индивидуальных заданий отражают благоприятное влияние внедрения системы качества на динамику финансовых показателей любых предприятий, в частности, и на динамику показателей предприятий потребительской кооперации.

5.1.2. Прогнозирование с помощью графических средств Excel

Задача 1.

Данные динамики выпуска продукции предприятия за 10 лет представлены в табл. 5.1.

Таблица 5.1

Данные выпуска валовой продукции предприятия (данные условные)

X	У
1	12
2	15
3	13
4	16
5	17
6	18
7	16
8	19
9	20
10	21

X — порядковый номер года с 1993 по 2002 г.;

У — объем валовой продукции предприятия (млн руб.).

Необходимо:

1. Получить прогноз выпуска валовой продукции предприятия на 2003 г. с использованием следующих функций: линейная, логарифмическая, полиномиальная второй степени, степенная, экспоненциальная.

2. С использованием коэффициента детерминации R^2 определить лучшую модель, по которой произвести прогноз на 2003 г.

Решение.

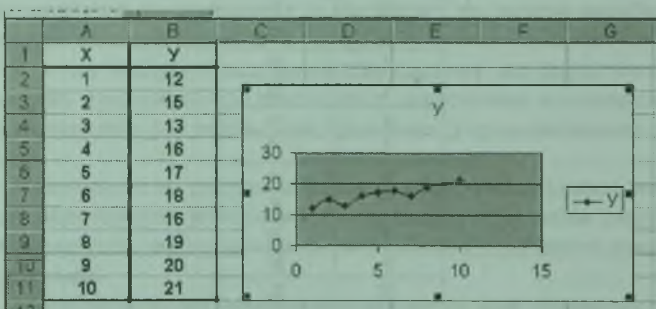
1.1. Загрузите Excel 97.

1.2. Введите данные в таблицу. На первом листе Excel по колонкам введите значения X и У, взятые из табл. 5.1. Размещение таблицы показано на рис. 5.1.

1.3. Постройте график зависимости У от X.

Этап 1. Выделите всю таблицу с названиями колонок. Курсор мышки поставьте на ячейку A1, нажмите левую клавишу мышки и протяните курсор до ячейки B11, отпустите кнопку мышки.

Этап 2. Постройте диаграмму "Точечная". На строке инструментов щелкните пиктограмму "Добавить диаграмму", выберите на вкладке "Стандартные" тип диаграммы "Точечная", выберите вид: "Точечная диаграмма со значениями, соединенными сглаживающими линиями"; нажмите кнопку "Готово". Результат действий представлен на рис. 5.1.



	A	B
1	X	У
2	1	12
3	2	15
4	3	13
5	4	16
6	5	17
7	6	18
8	7	16
9	8	19
10	9	20
11	10	21

Рис. 5.1. Таблица данных и график зависимости У от X

1.4. Ознакомьтесь с меню "Линия тренда".

Этап 1. Вызовите на рабочее поле электронной таблицы окно "Линия тренда". Поставьте курсор мышки на область построенной диаграммы, щелкните левой клавишей мышки, в строке главного меню Excel вместо кнопки "Данные" появится кнопка “Диаграмма”, щелкните кнопкой “Диаграмма”, выберите в ниспадающем меню пункт “Добавить линию тренда”, на экране появится окно “Линия тренда”, изображенное на рис. 5.2.

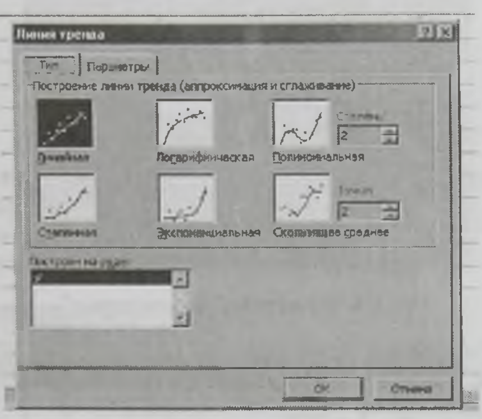


Рис. 5.2. Окно “Линия тренда”

Этап 2. Ознакомьтесь с содержанием надписей на окне “Линия тренда”. Щелкните вопрос на окне “Линия тренда” и перенесите вопрос на необходимую надпись, щелкните левой клавишей мышки и прочтите содержание надписей.

1.5. Получите прогноз по линейной функции на одну дату вперед. Выберите тип тренда “Линейная”, щелкните кнопкой вкладки “Параметры”, установите параметры, показанные на рис. 5.3, нажмите кнопку “ОК”. На экране появится изображение представленное на рис. 5.4.

1.6. Красиво оформите диаграмму. Один из вариантов самого простого оформления диаграммы изображен на рис. 5.4. При наличии времени можно воспользоваться всеми средствами красочного оформления диаграммы, представленного. На рис. 5.5 показан еще один пример оформления диаграммы.

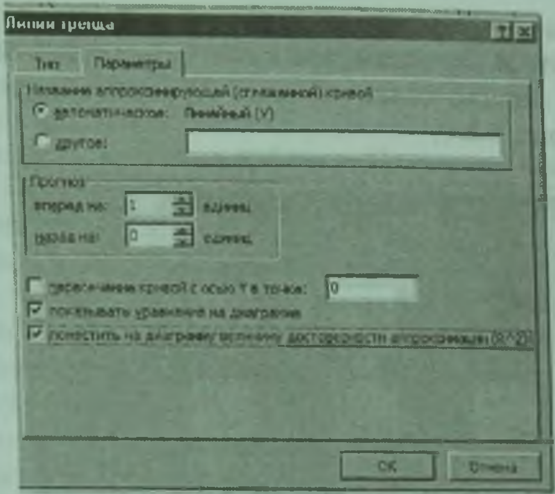


Рис. 5.3. Параметры "Линия тренда"

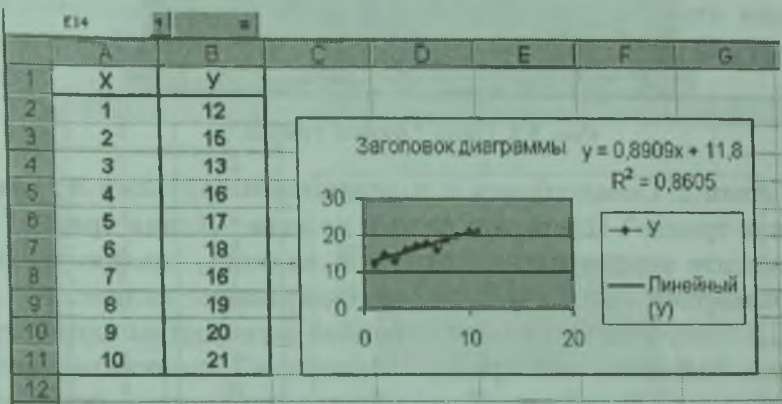


Рис. 5.4. Результат прогнозирования по линейной функции

1.7. Произведите расчеты теоретических значений $У_p$ по линейной функции.

$У_p$ вычисляется по формуле

$У_{pi} = a_o + a_1 X_i$ при $i = 1, 2, \dots, 10$,

где значения a_0 и a_1 можно взять из диаграммы.

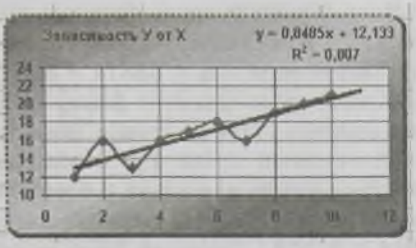
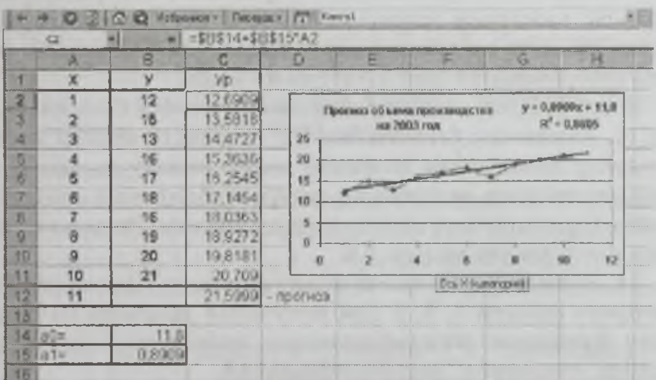


Рис. 5.5. Пример оформления диаграммы

Дополните таблицу расчетными значениями $У_{pi}$ для каждого значения X_i. При Х = 11 получите прогнозное значение У на 2003 г. Результаты расчетов и содержимое расчетной формулы изображены на рис. 5.6.



	A	B	C	D
1	X	У	Ур	
2	1	12	12.6909	
3	2	16	13.5818	
4	3	13	14.4727	
5	4	16	15.3636	
6	5	17	16.2545	
7	6	18	17.1454	
8	7	16	18.0363	
9	8	19	18.9272	
10	9	20	19.8181	
11	10	21	20.709	
12	11		21.5999	- прогноз
13				
14	a0=	11.8		
15	a1=	0.8909		
16				

Рис. 5.6. Расчет теоретических значений $У_p$ и прогнозного значения У на 2003 г.

Пояснения к проведению расчетов по формуле $У_p = a_0 + a_1X$

Этап 1. В ячейку B14 и B15 введите соответственно значения коэффициентов a_0 и a_1, взятых из диаграммы рис. 5.6. Результаты действия изображены на рис. 5.6.

Этап 2. В ячейку С2 введите формулу = В14 + В15*А2, адреса ячеек: В14 и В15 сделайте абсолютными, т. е. формула должна иметь следующий вид = $В$14 + $В$15*А2 (для того, чтобы при копировании полученной формулы адреса коэффициентов не изменялись). Эти действия можно выполнить непосредственной вставкой знака $ в необходимое место или автоматизировать расстановку знака $ с использованием клавиши F4 в таком порядке: активизируйте ячейку С2, введите формулу = В14 + В15*А2, клавишу Enter не нажимайте (если уже нажали, то вызовите содержимое ячейки С2 для редактирования), поставьте курсор мышки в начале или в конце адреса В14 (перед буквой В или после цифры 4), нажмите на клавиатуре клавишу F4. В результате этого действия формула примет следующий вид = $В$14 + В15*А2. Аналогичным образом сделайте абсолютным адрес В15, завершите ввод формулы нажатием клавиши Enter. В результате формула должна принять следующий вид = $В$14 + $В$15*А2. Эта формула изображена на рис. 5.6 в строке ввода.

Этап 3. Произведите копирование ячейки С2 в диапазон С3:С11. Активизируйте ячейку С2, поставьте курсор мышки на маркер заполнения, нажмите левую кнопку мышки и перетащите маркер заполнения через заполняемые ячейки до ячейки C_{11}, отпустите левую кнопку мышки. Результат расчетов представлен на рис. 5.6.

1.8. Выполните самостоятельную работу. Самостоятельно повторите пункты 3, 4, 5 для получения прогноза по следующим функциям: логарифмическая, полиномиальная второй степени, степенная, экспоненциальная.

1.9. Выберите лучшую тенденцию. По критерию R^2 выберите функцию, которая лучше всех описывает тенденцию зависимости У от X. Сделайте выводы о прогнозе по лучшей функции.

Примечания. 1. R^2 — коэффициент детерминации равен доле объясненной вариации за счет выбранной тенденции или упрощенно — R^2 равен доле исходных данных, которые подчиняются выбранной тенденции.

2. Очень грубо коэффициент детерминации можно вычислить по формуле $R_{пр}^2 = в_1/в_2$, где $в_1$ — количество первых разниц ($У_i - У_{i-1}$), у которых направление изменений совпадает с выбранной тенденцией, $в_2$ — общее количество первых разниц ($У_i - У_{i-1}$). Например, в соответствии с рис. 5.6 $R_{пр}^2 = в_1/в_2 = 7/9 = 0{,}777$. Точное значение коэффициента детерминации равно $R^2 = 0{,}86$. Сравнение двух коэффициентов детерминации: $R_{пр}^2 = 0{,}777$ и $R^2 = 0{,}86$ показывает наличие не очень большой ошибки, но для предварительного анализа вполне достаточно.

3. Очень грубо ошибку модели можно вычислить по формуле $E_{пр} = (e_{man} - e_{mix})/6$, где e_{min} и e_{max} соответственно минимальные и максимальные значения остатков модели.

5.1.3. Прогнозирование с помощью расчетных формул

Задача 2

Имеются данные зависимости У от X. Вычислить коэффициенты линейной модели методом наименьших квадратов остатков с помощью расчетных формул. Получить прогнозные значения У по линейной модели.

Решение.

Сущность метода наименьших квадратов. Необходимо найти такие значения коэффициентов a_0 и a_1 линейной модели $У_i = a_0 + a_1X_i + e_i = У_{pi} + e_i$, при которых сумма квадратов (остатков) отклонений $\sum e_i^2 = \sum(У_i - У_{pi})^2$ будет минимальной, где $У_{pi} = a_0 + a_1X_i$.

Чтобы найти минимум функции $\sum e_i^2$ необходимо вычислить частные производные по каждому коэффициенту a_0 и a_1, приравнять их нулю и решить полученную систему нормальных уравнений.

$$\partial\sum e_i^2/\partial a_0 = -2\sum У_1 + 2n + 2a_1\sum X_i = 0;$$

$$\partial\sum e_i^2/\partial a_0 = -2\sum У_iX_i + 2a_0\sum X_i + 2a_1\sum X_i^2 = 0.$$

После несложных преобразований можно получить систему нормальных уравнений

$$\begin{cases} na_0 + a_1 \sum X_i = \sum У_i; \\ a_0 \sum X_i + a_1 \sum X_i^2 = \sum У_i X_i. \end{cases}$$

Решение системы уравнений можно выполнить разными способами, один из которых предполагает использовать следующие формулы, представленные соответственно в развернутом или в матричном виде:

$$a_1 = \frac{\sum (X_i - \bar{X})(У_i - \bar{У})}{\sum \left(X_i - \bar{X}\right)^2};$$

$$a_0 = \bar{У} - a_1 \bar{X};$$

$$\mathbf{A = (X'X)^{-1}X'У}.$$

$У_{пр} = a_0 + a_1 X_{ож}$ — прогнозное значение У при ожидаемом значении $X_{ож}$.

Результаты расчетов коэффициентов a_0 и a_1, выполненные по развернутым формулам, изображены на рис. 5.7 и рис. 5.8.

H16 = =E17+E15*H15

	A	B	C	D	E	F	G	H
1	i	Xi	Уi	Xi-Xср	Уi-Уср	(Xi-Xср)*(Уi-Уср)	(Xi-Xср)^2	Урi
2	1	1	12	-4,5	-4,7	21,15	20,25	12,69
3	2	2	15	-3,5	-1,7	5,95	12,25	13,58
4	3	3	13	-2,5	-3,7	9,25	6,25	14,47
5	4	4	16	-1,5	-0,7	1,05	2,25	15,36
6	5	5	17	-0,5	0,3	-0,15	0,25	16,25
7	6	6	18	0,5	1,3	0,65	0,25	17,15
8	7	7	16	1,5	-0,7	-1,05	2,25	18,04
9	8	8	19	2,5	2,3	5,75	6,25	18,93
10	9	9	20	3,5	3,3	11,55	12,25	19,82
11	10	10	21	4,5	4,3	19,35	20,25	20,71
12	Сумма					73,5	82,5	
13	Среднее	5,5	16,7					
14								
15		$a1 = \frac{\sum (Xi - Xср)*(Уi - Уср)}{\sum (Xi - Xср)^2}$			0,890909		Xож =	11,00
16							Упр =	21,60
17		a0 = Уср - a1*Xср =			11,8000			

Рис. 5.7. Расчет коэффициентов a_0, a_1 и прогноза на $X_{ож} = 11$

Расчеты коэффициентов a_0 и a_1 с помощью матричных операций выполним при решении задачи 3.

	A	B	C	D	E	F	G	H
1	1	Xi	Уi	Xi-Xср	Уi-Уср	(Xi-Xср)*(Уi-Уср)	(Xi-Xср)^2	Уpi
2	1	1	12	=B2-$B$13	=C2-$C$13	=D2*E2	=D2^2	=$E$17+$E$15*B2
3	2	2	15	=B3-$B$13	=C3-$C$13	=D3*E3	=D3^2	=$E$17+$E$15*B3
4	3	3	13	=B4-$B$13	=C4-$C$13	=D4*E4	=D4^2	=$E$17+$E$15*B4
5	4	4	16	=B5-$B$13	=C5-$C$13	=D5*E5	=D5^2	=$E$17+$E$15*B5
6	5	5	17	=B6-$B$13	=C6-$C$13	=D6*E6	=D6^2	=$E$17+$E$15*B6
7	6	6	18	=B7-$B$13	=C7-$C$13	=D7*E7	=D7^2	=$E$17+$E$15*B7
8	7	7	16	=B8-$B$13	=C8-$C$13	=D8*E8	=D8^2	=$E$17+$E$15*B8
9	8	8	19	=B9-$B$13	=C9-$C$13	=D9*E9	=D9^2	=$E$17+$E$15*B9
10	9	9	20	=B10-$B$13	=C10-$C$13	=D10*E10	=D10^2	=$E$17+$E$15*B10
11	10	10	21	=B11-$B$13	=C11-$C$13	=D11*E11	=D11^2	=$E$17+$E$15*B11
12	Сумм					=СУММ(F2:F11)	=СУММ(G2:G11)	
13	Сред	=СРЗНАЧ(B2:B11)	=СР					
14								
15		$a_1 = \frac{\Sigma(Xi - Xcp)*(Yi - Ycp)}{\Sigma(Xi - Xcp)^2}$			=F12/G12		Хож =	11
16							Упр =	=E17+E15*H15
17		a0 = Уср - a1*Xср =			=C13-E15*B			

Рис. 5.8. Расчетные формулы к рис. 5.7

Вывод. С помощью расчетных формул были получены значения коэффициентов a_0 и a_1, а также расчетные значения У и прогноз для $X_{ож} = 11$.

5.1.4. Прогнозирование с помощью матричных операций

Задача 3.

Имеются данные зависимости У от X. Вычислить коэффициенты линейной модели методом наименьших квадратов остатков с помощью матричных операций. Получить прогнозные значения У по линейной модели.

Решение.

В предыдущей задаче мы производили расчет коэффициентов модели с использованием расчетных формул, представленных в скалярном виде. Существует тождественный способ вычисления коэффициентов модели с использованием расчетных формул, представленных в матричном виде.

Использование матричной формы записи формул и проведения расчетов имеет несколько преимуществ и недостатков.

Преимущества заключаются в том, что запись формул приобретает очень компактный вид; вид формул, представленных в матричном виде, не зависит от количества факторов, включенных в модель, и является очень удобным при расчетах характеристик многофакторных моделей.

Недостатком использования в расчетах матричных формул является необходимость хорошего знания матричной алгебры и наличия программных средств проведения матричных операций. Математическое обеспечение Excel-97 позволяет выполнять все необходимые матричные операции, кроме вычисления собственных значений и собственных векторов матрицы.

Приводим перечень используемых матричных операций и место их нахождения.

Транспонирование — Вставка функции, Категория: Ссылки и массивы, Функции: ТРАНСП.

Вычисление обратной матрицы — Вставка функции, Категория: Математические, Функции: МОБР.

Умножение матриц — Вставка функции, Категория: Математические, Функции: МУМНОЖ.

Выполнение матричных функций имеют следующие особенности:

➤ для результирующей матрицы нужно выделить необходимое количество ячеек;

➤ для распространения действий функции на массив:

• выделите первую ячейку с расчетами и все ячейки, на которые будет распространено действие функции;

• нажмите и отпустите клавишу "F2";

• последовательно нажмите, не отпуская, клавиши "Ctrl" "Shift" "Enter", отпустите сразу все три клавиши, после этого на экране появится содержимое всей матрицы.

3.1. Введите данные в таблицу. Скопируйте весь лист, в котором решалась задача 1.

3.2. Выполните расчет коэффициентов модели с помощью матричных операций.

Расчет коэффициентов модели a_0, a_1 в матричном виде производится по формуле:

$$\mathbf{A} = (\mathbf{X'X})^{-1}\mathbf{X'У},$$

где **А** — вектор-столбец коэффициентов модели;

Х — матрица исходных данных, включающая вектор-столбец переменной для свободного коэффициента, векторы-столбцы объясняемых факторов;

Х' — транспонированная матрица;

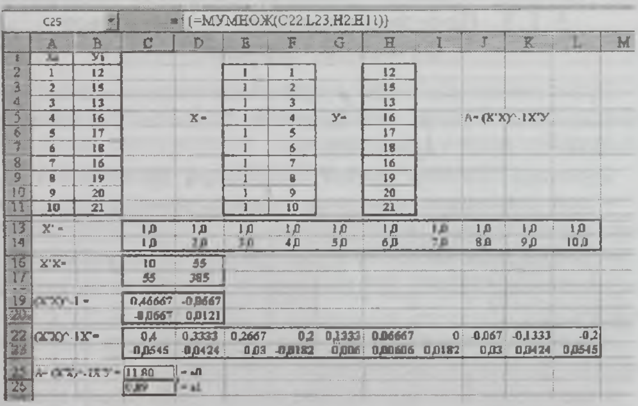
$(\mathbf{X'X})^{-1}$ — обратная матрица от произведения двух матриц;

У — вектор-столбец зависимой переменной.

Вычислим коэффициенты модели a_0, a_1 в матричном виде по формуле

$$\mathbf{A} = (\mathbf{X'X})^{-1}\mathbf{X'У}$$

в последовательности, изображенной на рис. 5.9 и 5.10.



C25 = {=МУМНОЖ(C22:L23;H2:H11)}

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Xi	Yi											
2	1	12			1	1		12					
3	2	15			1	2		15					
4	3	13			1	3		13					
5	4	16		X =	1	4	У =	16		A = (X'X)^-1X'У			
6	5	17			1	5		17					
7	6	18			1	6		18					
8	7	16			1	7		16					
9	8	19			1	8		19					
10	9	20			1	9		20					
11	10	21			1	10		21					
13	X' =		1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	
14			1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0	9,0	10,0	
16	X'X=		10	55									
17			55	385									
19	(X'X)^-1 =		0,46667	-0,0667									
20			-0,0667	0,0121									
22	(X'X)^-1X'=		0,4	0,3333	0,2667	0,2	0,1333	0,06667	0	-0,067	-0,1333	-0,2	
23			-0,0545	-0,0424	0,03	-0,0182	0,006	0,00606	0,0182	0,03	0,0424	0,0545	
25	A= (X'X)^-1X'У =		11,80	= a0									
26			0,89	= a1									

Рис. 5.9. Последовательность расчета коэффициентов модели

После расчета коэффициентов модели вычисление $У_р$ и $У_{пр}$ выполняется аналогично предыдущей задаче.

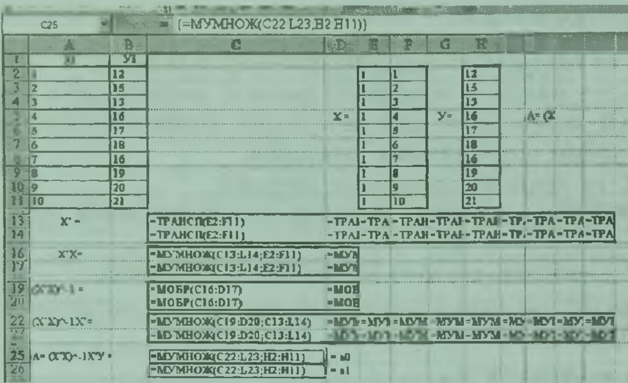


Рис. 5.10. Матричные формулы вычисления коэффициентов модели

5.1.5. Прогнозирование с помощью функции "Линейн"

Практически все эконометрические модели можно идентифицировать с помощью функции "Линейн", поэтому эту функцию следует хорошо знать.

Функция "Линейн" имеет следующие достоинства:

- при изменении исходных данных происходит автоматический перерасчет этой функции;
- результаты расчетов содержат необходимую информацию, чтобы проверить достоверности модели и ее коэффициентов;
- можно получить результаты расчетов для любого количества факторов, включенных в модель;
- имеется возможность присвоить свободному коэффициенту уравнения регрессии значение, равное нулю.

Функция "Линейн" имеет следующие недостатки:

- процедура расчетов плохо запоминается;
- не выдаются точечный интервальный прогноз;
- не выдаются доверительные интервалы уравнения регрессии;

• не выдаются критические значения критерия Фишера и Стьюдента.

Задача 4.

Имеются данные зависимости У от Х. Необходимо получить прогноз значения У по линейной модели с использованием функции "Линейн".

Решение.

4.1. Введите данные в таблицу. Скопируйте весь лист, в котором решалась задача 1.

4.2. Выполните расчеты основных показателей линейной модели по функции "Линейн". Поставьте курсор на свободное место электронной таблицы (свободными должны быть два столбца и пять строчек); введите знак =; щелкните по пиктограмме "Вставка функции"; выберите категорию "Статистические"; выберите функцию "Линейн"; нажмите клавишу "ОК", на экране появится окно параметров функции "Линейн", заполните любым способом содержимое параметров (см. рис. 5.11), нажмите кнопку "ОК". На экране появится результат расчетов, но содержимое только одной ячейки. Для вывода содержимого всех ячеек расчетной таблицы необходимо выполнить описанную ниже процедуру.

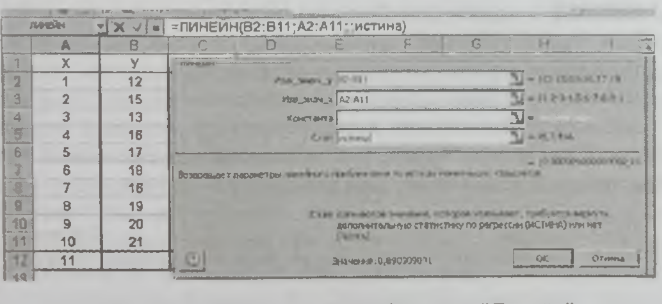
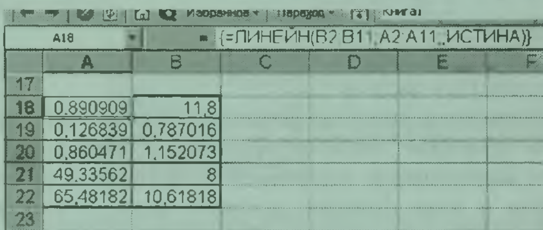


Рис. 5.11. Меню параметров функции "Линейн"

4.3. Выведите содержимое всех ячеек таблицы расчетов по функции "Линейн". Выделите первую ячейку с расчетами и дополнительно одну колонку справа и еще пять строчек вниз, нажмите и отпустите клавишу “F2”, последовательно нажмите, не отпуская, клавиши “Ctrl” “Shift” “Enter”, отпустите сразу все три клавиши. На экране появится протокол расчетов основных характеристик линейной модели, изображенной на рис. 5.12.



	A	B	C	D	E	F
17						
18	0,890909	11,8				
19	0,126839	0,787016				
20	0,860471	1,152073				
21	49,33562	8				
22	65,48182	10,61818				
23						

Рис. 5.12. Протокол расчетов по функции “Линейн”

11,8 — коэффициент a_0 уравнения $У = a_0 + a_1X$;
0,890909 — коэффициент a_1 уравнения $У = a_0 + a_1X$;
0,787016 — ошибка коэффициента a_0;
0,126839 — ошибка коэффициента a_1;
1,152073 — ошибка модели;
0,860471 — коэффициент детерминации R^2;
49,33562 — критерий Фишера;
8 — число степеней свободы для остатков;
65,48182 — сумма квадратов регрессии;
10,61818 — сумма квадратов остатков.

Запомните положение характеристики линейной модели в таблице расчетов. Числа, расположенные в расчетной таблице, имеют содержание, представленное в табл. 5.2.

Таблица 5.2

Условные обозначения результатов расчета по функции “Линейн”

a_1	a_0
S_{a1}	S_{a0}
R^2	E
F	n-k
$C_{рег}$	$C_{ост}$

a_1, a_0 — коэффициенты линейной модели;

$У = a_0 + a_1X_1 + e$;

S_{a1}, S_{a0} — среднеквадратические отклонения коэффициентов a_1, a_0 от своих математических ожиданий;

E — ошибка модели или среднеквадратическое отклонение остатков;

F — критерий Фишера;

R^2 — коэффициент детерминации;

$(n - k)$ — число степеней свободы для остатков модели;

n — объем выборки;

k — количество коэффициентов в модели, включая свободный коэффициент;

$C_{рег}$ — сумма квадратов отклонений У, обусловленных регрессией;

$C_{ост}$ — сумма квадратов отклонений остатков.

Выводимые характеристики модели связаны между собой следующими соотношениями:

$$E = \frac{C_{ост}}{n-k};\ F = \frac{C_{рег}(n-k)}{C_{ост}(k-1)};\ R^2 = \frac{C_{рег}}{C_{рег} + C_{ост}};\ t_{a0} = \frac{a_0}{S_{a0}};\ t_{a1} = \frac{a_1}{S_{a1}},$$

где t_{a1}, t_{a0} — критерии Стьюдента для коэффициентов a_1 и a_0 соответственно.

Расчеты, полученные с помощью функции “Линейн”, не содержат точечных и интервальных прогнозов, нет доверительных интервалов уравнения регрессии, которые следует выполнять самостоятельно с использованием уже изученных формул.

4.4. Прочтите в справочнике описание функции “Линейн”. Нажмите на кнопку “?” главного меню, в ниспадающем меню выберите “Вызов справки”, нажмите вкладку “Содержание”, последовательно открывайте следующие рубрики: “Создание формул и проверка книг”, “Использование функций”, “Статистические функции”, “Линейн”. Прочитайте подробное описание функции “Линейн”. Обратите особое внимание на решение контрольного примера 3 “Множественная линейная регрессия”.

4.5. Вычислить $У_p = a_0 + a_1X$. При вычислении $У_p$ для всех значений X необходимо воспользоваться значениями a_0 и a_1, взятыми из протокола расчетов функции “Линейн”, значения X взять из таблицы данных. Последовательность выполнения расчетов подробно описана в задаче 1, п. 1.7. Результаты расчетов поместите рядом с таблицей данных.

4.6. Эконометрический анализ

Регрессия имеет следующий вид:

$У = 11{,}8 + 0{,}890909X_1 + e;$

$S_{a0} = 0{,}787;\ S_{a1} = 0{,}126;$

$t_{a0} = 14{,}9;\ t_{a1} = 7{,}06;$

$R^2 = 0{,}860471;$

$E = 1{,}152073;$

$F = 49{,}33562.$

По таблицам или в Excel определим критические значения критерия Стьюдента и Фишера.

$t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k = 10 - 4 = 6) = 2{,}44.$

$F_{кр}(\alpha = 0{,}05;\ m_1 = k - 1 = 2 - 1 = 1;\ m_2 = n - k =$
$= 10 - 2 = 8) = 5{,}31.$

Так как фактическое значение критерия Фишера, равное 49,33, больше критического значения, равного 5,31, то модель является достоверной.

Так как фактические значения критерия Стьюдента для фактора X (7,06) больше его критического значения 2,44, то фактор X оказывает достоверное влияние на величину У. Имеется высокая доля объясненной вариации переменной У, равная 0,86.

Вывод. Полученная модель является достоверной, по которой можно проводить прогнозирование.

4.7. Вычислите прогнозное значение У на следующий год. Прогнозное значение У на следующий год равно

$$У = a_0 + a_1X = 11{,}8 + 0{,}89 \times 11,$$

где $a_0 = 11{,}8$, $a_1 = 0{,}89$, $X = 11$.

При X = 11 получается прогноз на следующий год.

5.1.6. Прогнозирование с помощью пакета программ "Анализ данных"

В состав Microsoft Excel входит набор средств анализа данных (называемый пакет анализа), предназначенный для решения сложных статистических и инженерных задач. Для проведения анализа данных с помощью этих инструментов следует указать входные данные и выбрать параметры; анализ будет проведен с помощью подходящей статистической или инженерной макрофункции, а результат будет помещен в выходной диапазон. Другие инструменты позволяют представить результаты анализа в графическом виде.

Чтобы вывести список доступных инструментов анализа, выберите команду "Анализ данных" в меню Сервис. Если команда "Анализ данных" отсутствует в меню Сервис, то необходимо запустить программу установки Microsoft Excel. После установки пакета анализа его необходимо выбрать и активизировать с помощью команды "Настройки".

Для успешного применения процедур анализа необходимы начальные знания в области статистических и инженерных расчетов, для которых эти инструменты были разработаны.

Пакет программ "Анализ данных" включает в себя следующие программы:

- однофакторный дисперсионный анализ;
- двухфакторный дисперсионный анализ с повторениями;
- двухфакторный дисперсионный анализ без повторений;
- корреляция;
- ковариация;
- описательная статистика;
- экспоненциальное сглаживание;
- двухвыборочный F-тест для дисперсии;
- анализ Фурье;
- гистограмма;
- скользящее среднее;
- генерация случайных чисел;
- ранг и персентиль;

- регрессия;
- выборка;
- парный двухвыборочный t тест для средних;
- двухвыборочнiй t тест с одинаковыми дисперсиями;
- двухвыборочнiй t тест с разными дисперсиями;
- двухвыборочный z тест для средних.

Многие из этих программ используются при статистическом анализе данных. Для проведения расчетов регрессионного анализа можно использовать программу "регрессия".

Достоинства программы "Регрессия":

- простота обращения к программе и ввода диапазонов переменных;
- выводятся значения коэффициентов, ошибки коэффициентов, критерий Стьюдента для коэффициентов, ошибка модели, критерий Фишера и его уровень значимости;
- выводятся фактические и расчетные значения У, значения остатков;
- строятся графики исходных, расчетных данных и значений остатков.

Недостатки программы "Регрессия":

- не рассчитываются доверительные интервалы уравнения регрессии;
- не считаются точечный и интервальный прогнозы;
- при изменении исходных данных не происходит перерасчет характеристик модели;

Задача 5.

Имеются данные зависимости У от X. Необходимо получить прогноз значения У по линейной модели с использованием программы "регрессия", которая входит в состав пакета программ "Анализ данных".

Решение.

Для прогнозирования можно использовать программу "Регрессия".

5.1. Прочтите в справочной системе Excel описание программы "Регрессия".

Найдите описание программы "регрессия" по такому пути: "?", "Вызов справки", "Содержание", "Статистический анализ", "Регрессия". К сожалению, в справке имеется только описание меню программы, но нет описания расчетных формул этой программы.

5.2. Введите данные в таблицу. Скопируйте весь лист, в котором решалась задача 1.

5.3. Выполните программу "регрессия". Щелкните кнопкой "Сервис", выберите пункт "Анализ данных", выберите программу "регрессия", нажмите кнопку "ОК", на экране появится окно программы "Регрессия". Заполните параметры программы в соответствии с рис. 5.13, нажмите кнопку "ОК". На экране появится протокол расчетов, изображенный на рис. 5.14.

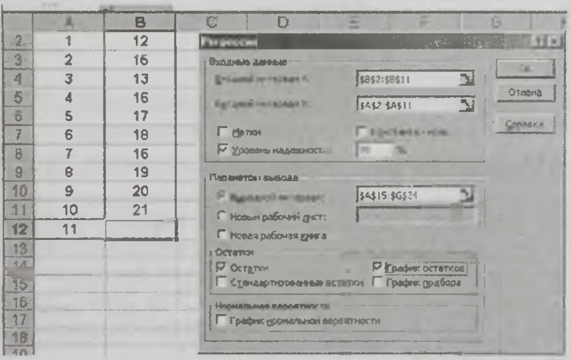
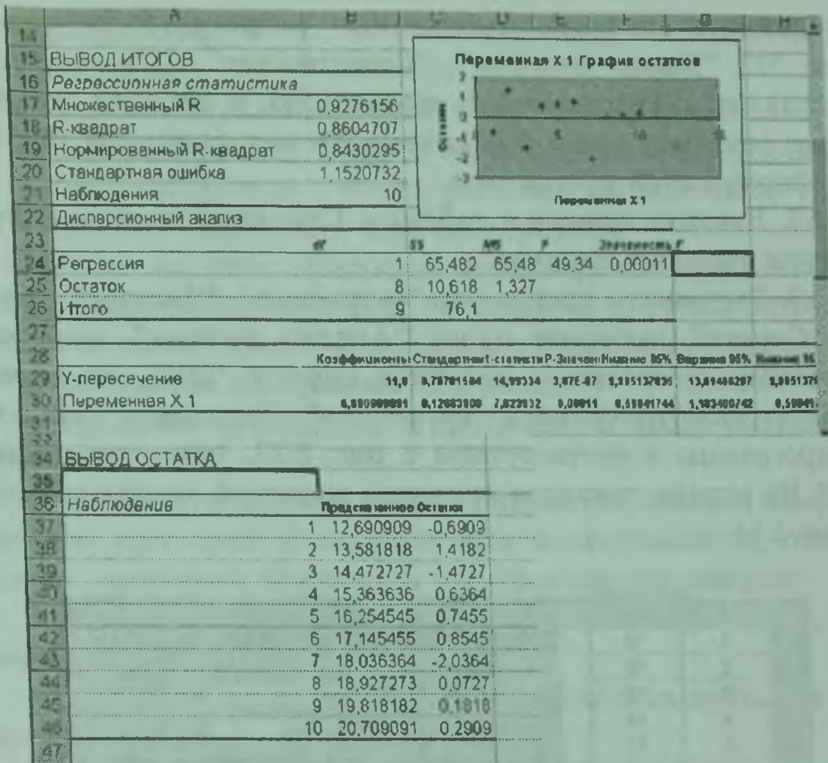


Рис. 5.13. Параметры программы "Регрессия"

5.4. Проведите анализ расчетов. Найдите в протоколе расчетов "Регрессия" значения коэффициентов a_0 и a_1 и сравните их значения с протоколом расчетов функции "Линейн".

Найдите в протоколе расчетов "регрессия" "Предсказанное значение У" и сравните их со значениями $У_р$, представленными на рис. 5.4. Найдите в протоколе расчетов "регрессия"



ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,9276156
R-квадрат	0,8604707
Нормированный R-квадрат	0,8430295
Стандартная ошибка	1,1520732
Наблюдения	10

Дисперсионный анализ

	df	SS	MS	F	Значимость F
Регрессия	1	65,482	65,48	49,34	0,00011
Остаток	8	10,618	1,327		
Итого	9	76,1			

	Коэффициенты	Стандартная	t-статисти	P-Значени	Нижние 95%	Верхние 95%	Нижние 95
Y-пересечение	11,8	[illegible]	14,99334	3,87E-07	[illegible]	[illegible]	[illegible]
Переменная X 1	[illegible]	[illegible]	7,023032	0,00011	[illegible]	[illegible]	[illegible]

ВЫВОД ОСТАТКА

Наблюдение	Предсказанное	Остатки
1	12,690909	-0,6909
2	13,581818	1,4182
3	14,472727	-1,4727
4	15,363636	0,6364
5	16,254545	0,7455
6	17,145455	0,8545
7	18,036364	-2,0364
8	18,927273	0,0727
9	19,818182	0,1818
10	20,709091	0,2909

Рис. 5.14. Вывод результатов расчетов по программе "Регрессия"

значение критерия Фишера и сделайте вывод о достоверности линейной модели. Пояснение: если значимость критерия Фишера будет меньше 0,05, то модель является достоверной.

5.5. Получите прогнозное значение У на 2003 г. при X = 11. Значения a_0 и a_1 можно взять из протокола расчетов программы "регрессия".

Приводим пояснения к рис. 5.14.

- стандартная ошибка — ошибка модели E;
- df — число степеней свободы;
- SS — значение вариации или сумма квадратов отклонений;
- MS — дисперсия, равная вариации деленной на соответствующее число степеней свободы;
- F — критерий Фишера;

• значимость F — вероятность совершить ошибку при отклонении нулевой гипотезы: модель является недостоверной. Достоверность модели проверяется на уровне значимости, равной 0,05. Для данного примера значимость F равна 0,07, что больше 0,05, следовательно, модель является недостоверной;

• "У — пересечение" — означает свободный коэффициент a_0;

• стандартная ошибка — ошибка коэффициента;

• t-статистика — критерий Стьюдента для коэффициента;

• "Р — значение" — уровень значимости критерия Стьюдента;

• "Нижние 95%", "Верхние 95%" — нижние и верхние 95% доверительные интервалы нахождения математических ожиданий коэффициентов модели;

• Предсказанное У — расчетные значения У.

Расчеты содержат графики, которые в пояснениях не нуждаются.

5.1.7. Прогнозирование с использованием программы "Поиск решения"

Определение коэффициентов любого уравнения регрессии можно представить как оптимизационный процесс по минимизации целевой функции методом линейного программирования, предложенный Канторовичем. Программа "Поиск решения" позволяет итерационным методом определить коэффициенты любого уравнения регрессии, если в качестве целевой функции выбрать сумму квадратов остатков. Под остатком понимается разность между фактическим и расчетным значением зависимой переменной.

Задача 6.

Имеются данные зависимости У от X. Вычислить коэффициенты линейной модели методом наименьших квадратов остатков с помощью программы "Поиск решения". Проверить полученные значения коэффициентов на оптимальность. Получить прогнозные значения У по линейной модели.

Решение.

Сущность метода наименьших квадратов. Необходимо найти такие значения коэффициентов a_0 и a_1 линейной модели $У = a_0 + a_1X + e = У_p + e$, при которых сумма квадратов (остатков) отклонений $e = У - У_p$ будет минимальной, где $У_p = a_0 + a_1X$. Расчет коэффициентов a_0 и a_1 можно выполнить разными способами, один из которых предполагает использовать итерактивную программу "Поиск решения". Итерактивными называют алгоритмы, которые осуществляют поиск решения с помощью последовательных приближений или итераций.

6.1. Введите данные в таблицу. Скопируйте весь лист, в котором решалась задача 1.

6.2. Назначьте ячейки для коэффициентов a_0 и a_1. Назначьте ячейки B14 и B15 для коэффициентов соответственно a_0 и a_1 линейной модели $У_p = a_0 + a_1X$. Первоначально в этих ячейках значения должны отсутствовать. В дальнейшем в них будет занесен результат выполнения программы "Поиск решения".

6.3. Введите в ячейки C2:C11 результат расчетов по формуле $У_p = a_0 + a_1X$. В качестве значений a_0 и a_1 необходимо взять ячейки соответственно B14 и B15, значения X надо последовательно выбирать из массива A2:A11. Например, в ячейке C2 должна быть введена следующая формула = = $B$14 + $B$15*A2, в ячейке C3 должна быть введена формула = $B$14 + $B$15*A3 и так далее. Подробное описание расчетов $У_p$ смотрите в задаче 1. Так как первоначально значения a_0 и a_1 равны нулю, то и все значения $У_p$ тоже будут равны нулю.

6.4. Введите в массив ячеек D2:D11 квадраты остатков e^2. Квадраты остатков вычисляются по формуле $(У - У_p)^2$. Например, в ячейке D2 должна быть введена следующая формула =$(B2 - C2)^2$.

6.5. Введите целевую функцию. В ячейку C17 введите целевую функцию, которая будет равна сумме квадратов остатков, т. е. сумма ячеек массива D2:D11. Например, в ячейке C17 должна быть введена формула

$= D2 + D3 + D4 + D5 + D6 + D7 + D8 + D9 + D10 + D11.$

Все подготовительные операции произведены, теперь необходимо так изменять численные значения a_0 и a_1, чтобы значение целевой функции было минимальным. Эту процедуру выполнит программа “Поиск решения”.

6.6. Выполните программу “Поиск решения”. Нажмите кнопку “Сервис”, выберите пункт “Поиск решения”, на экране появится окно меню программы “Поиск решения”. Заполните параметры программы в соответствии с рис. 5.15. Нажмите кнопку “Выполнить” и получите результат, представленный на рис. 5.16.

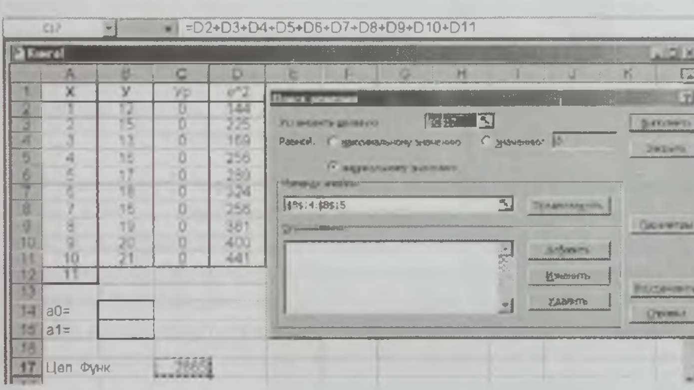
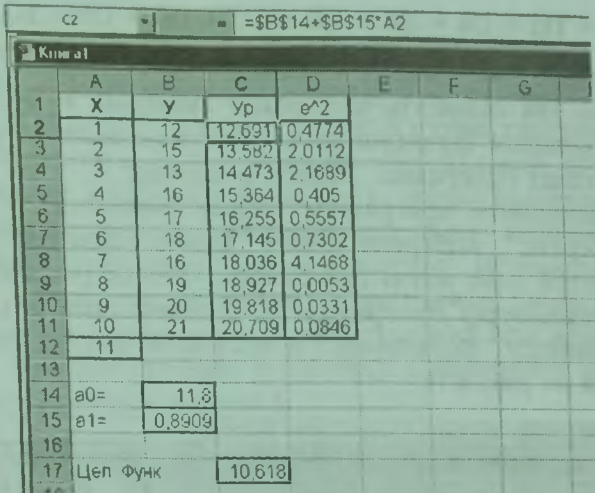


Рис. 5.15. Параметры программы “Поиск решения”

6.7. Проведите анализ полученных расчетов. Анализ результатов расчетов показывает, что коэффициенты линейной модели, найденные с помощью функции “Поиск решения” совпадают со значениями соответствующих коэффициентов, определенных с помощью функции “Линейн”.

6.8. Проверьте найденные численные значения коэффициентов на оптимальность. Постройте график зависимости У и $У_р = a_0 + a_1X$ от X. Измените оптимальные значения коэффициентов a_0 и a_1 в ячейках B14 и B15 и наблюдайте за тем,



C2 =$B$14+$B$15*A2

	A	B	C	D
1	X	У	Ур	e^2
2	1	12	12,691	0,4774
3	2	15	13,582	2,0112
4	3	13	14,473	2,1689
5	4	16	15,364	0,405
6	5	17	16,255	0,5557
7	6	18	17,145	0,7302
8	7	16	18,036	4,1468
9	8	19	18,927	0,0053
10	9	20	19,818	0,0331
11	10	21	20,709	0,0846
12	11			
13				
14	a0=	11,8		
15	a1=	0,8909		
16				
17	Цел Функ		10,618	

Рис. 5.16. Результат расчета коэффициентов линейной модели с помощью программы "Поиск решения"

как изменяются численные значения целевой функции в ячейке C17, а также изучите поведение графика $У_р = a_0 + a_1X$ относительно фактических значений У. Убедитесь в том, что при любом изменении оптимальных значений коэффициентов a_0 и a_1 значение целевой функции возрастает, а график $У_р$ оптимальным образом воспроизводит тенденцию зависимости У от X. Следовательно, найденные значения коэффициентов являются оптимальными.

5.1.8. Прогнозирование с помощью функции "ПРЕДСКАЗ"

Задача 7

Имеются данные зависимости У от X. Необходимо получить прогнозное значение У по линейной модели с использованием функции "ПРЕДСКАЗ".

Решение.

7.1. Введите данные в таблицу. Скопируйте весь лист, в котором решалась задача 1.

7.2. Назначьте ячейку, в которой будет находиться ожидаемое значение X. Например, в ячейке A12 будет находиться ожидаемое значение X.

7.3. Введите в ячейку A12 ожидаемое значение, равное 11.

7.4. Назначьте ячейку, в которой будет находиться прогнозное значение У, вычисленное по линейной модели. Например, в ячейке B12 будет прогнозное значение У.

7.5. Введите в ячейку B12 функцию "ПРЕДСКАЗ". Поставьте в ячейку B12 "=", щелкните по пиктограмме "вставка функции", выберите категорию "Статистические", выберите функцию "ПРЕДСКАЗ", на экране должно появиться изображение, представленное на рис. 5.17, нажмите на знак "?" и прочтите описание этой функции.

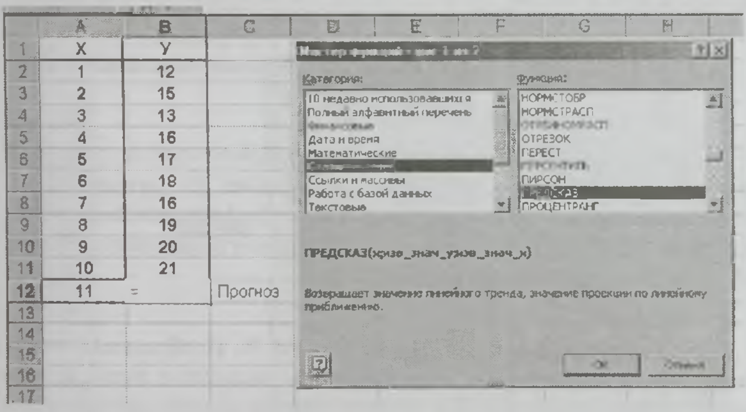


Рис. 5.17. Выбор функции "ПРЕДСКАЗ"

7.6. Получить прогнозное значение У по линейной модели. Нажмите кнопку "ОК" и введите параметры функции "ПРЕДСКАЗ", представленные на рис. 5.18. Нажмите кнопку "ОК" и получите в ячейке B12 прогнозное значение, вычисленное по линейной функции.

7.7. Сравнительный анализ полученных расчетов. Сравните полученный прогноз с предыдущими прогнозными зна-

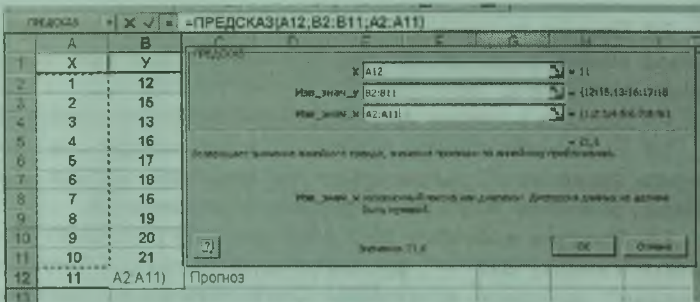


Рис. 5.18. Параметры функции "ПРЕДСКАЗ"

чениями, вычисленными с использованием других инструментальных средств.

5.2. Прогнозирование с использованием ППП Stadia

Для принятия ответственных управленческих решений на основе прогнозных расчетов желательно все расчеты проводить с использованием авторитетных профессиональных статистических пакетов. Одним из таких пакетов является Stadia. Описание пакета имеется в гл. 14.

Задача 8.

Имеются данные зависимости У от X. Необходимо получить прогноз значения У по линейной модели с использованием профессионального статистического пакета прикладных программ ППП Stadia 6.2/demo.

8.1. Загрузите Stadia.

8.2. Ввести данные в таблицу. В колонки X1 и X2 введите соответственно значения X и У, взятых из табл. 5.1. Результат действий показан на рис. 5.19.

8.3. Постройте график зависимости У от X. Нажмите кнопку "График=F6" в окне "Графики данных", изображенном на рис. 5.20; выберите тип графика "научные", вид графика

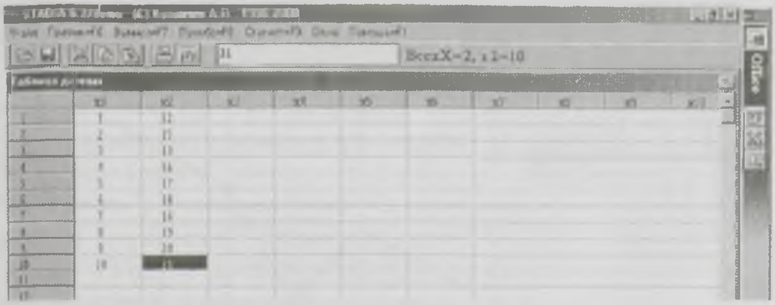


Рис. 5.19. Результат ввода данных

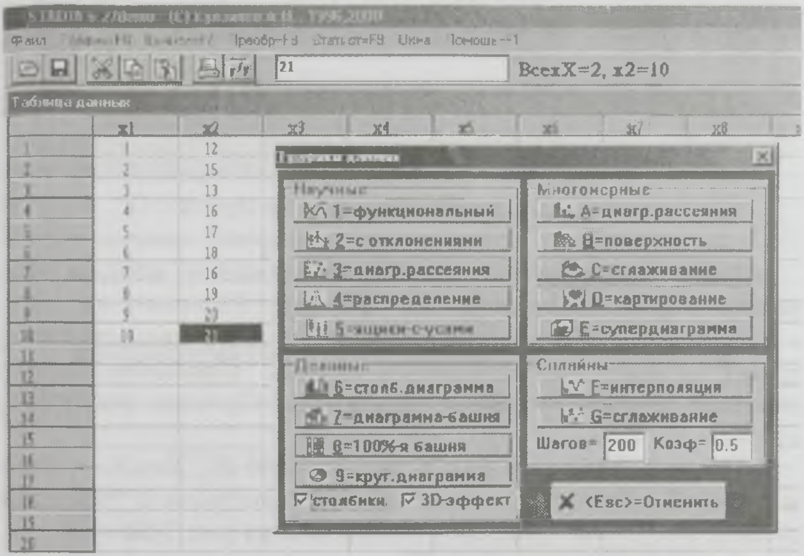


Рис. 5.20. Содержание окна "Графики данных"

"1-функциональный"; выберите для оси X переменную X_1, для оси У — переменную X2; нажмите кнопку "Утвердить"; получите график, изображенный на рис. 5.21.

8.4. Вычислите основные характеристики линейной модели зависимости У от X. Окно с графиком можно закрыть или оставить без изменения.

Рис. 5.21. График зависимости У от X

Этап 1. Откройте меню "Статистические методы". Нажмите кнопку главного меню: "Статист=F9".

Этап 2. Выберите вид функции "L=Простая регрессия/тренд". В окне "Статистические методы" в разделе "Регрессионный анализ" нажмите кнопку "L=Простая регрессия/тренд".

Этап 3. Выберите для X переменную X_1, для У — переменную X_2. Этот несложный процесс выполните самостоятельно. Результат действий изображен на рис. 5.22.

Этап 4. Получите протокол расчета основных характеристик линейной модели зависимости У от X. Нажмите кнопку "Утвердить", в окне "Регрессия" нажмите кнопку "1-линейная" линейной модели $У = a + bX$ и получите результат, изображенный на рис. 5.23.

Этап 5. Получите прогнозные значения У на дату $X_{ож} = 11$. В окне "Интерполяция" введите ожидаемое значение X равное 11, нажмите кнопку "Утвердить" и получите прогнозное значение У равное 21,6. Результат действий изображен на рис. 5.24.

Этап 6. Получите точечный и интервальный прогноз на 4 шага вперед и завершите анализ. В окошке “Интерполяция” нажмите кнопку “Отменить”, появится окно “Выдать график регрессии”. Можно нажать кнопку “Yes“ и посмотреть график фактических, расчетных и доверительных интервалов уравнения регресси. В окне “Посмотрите график” нажмите кнопку “Отменить”, в окне “Что дальше” установите количество шагов 4, нажмите кнопку “2=Прогнозирование”, получите результат, изображенный на рис. 5.25. В окне “Посмотрите график” нажмите кнопку “Отменить” и в окне “Что дальше” нажмите кнопку “Завершить анализ”.

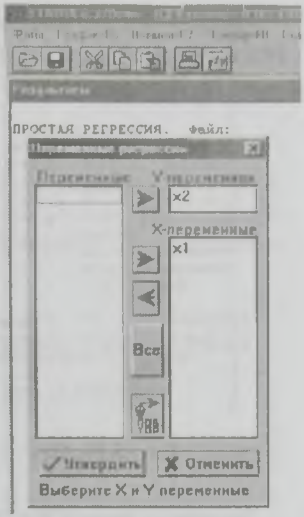


Рис. 5.22. Результат выбора переменных для определения зависимости У от Х

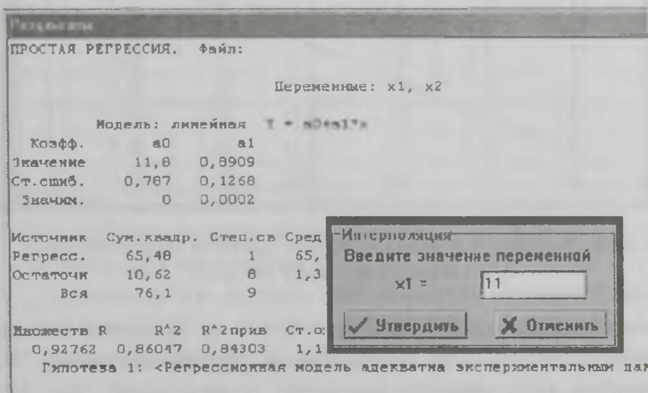
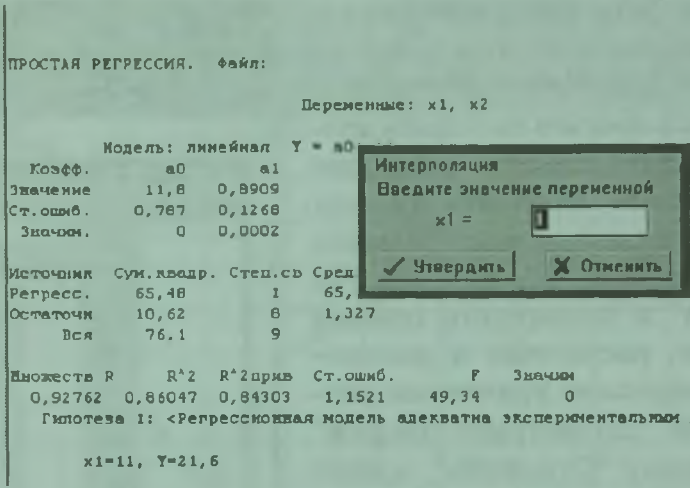


Рис. 5.23. Результат расчета основных характеристик линейной модели



ПРОСТАЯ РЕГРЕССИЯ. Файл:

Переменные: x1, x2

Модель: линейная Y = a0

Коэфф.	a0	a1
Значение	11,8	0,8909
Ст.ошиб.	0,787	0,1268
Значим.	0	0,0002

Источник	Сум.квадр.	Степ.св	Сред
Регресс.	65,48	1	65,
Остаточн	10,62	8	1,327
Вся	76.1	9	

Множеств R	R^2	R^2прив	Ст.ошиб.	F	Значим
0,92762	0,86047	0,84303	1,1521	49,34	0

Гипотеза 1: <Регрессионная модель адекватна экспериментальным данным>

x1=11, Y=21,6

Рис. 5.24. Протокол расчетов по линейной модели

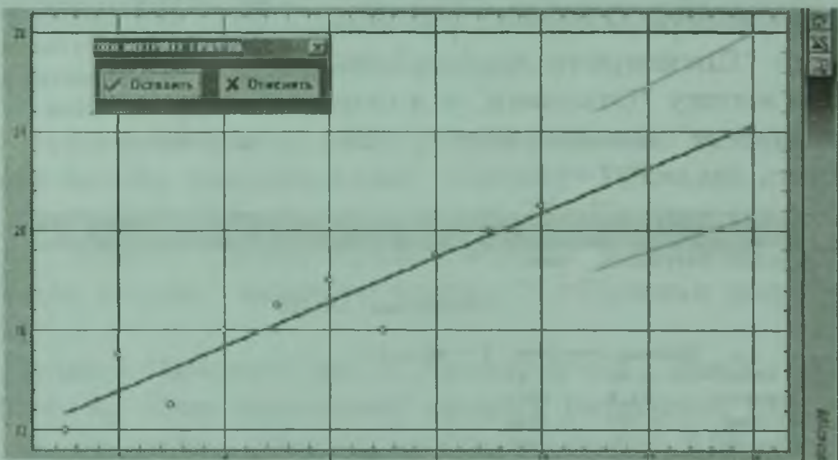


Рис. 5.25. Точечный и интервальный прогноз зависимости У от X на четыре шага вперед

Полученные протоколы расчетов и графики можно копировать и переносить в текстовые файлы Microsoft Word.

Анализ полученных расчетов показывает, что все характеристики линейной модели совпадают с результатами расчетов, выполненных в Excel.

5.3. Прогнозирование экономического состояния предприятия с использованием ППП Stadia

Задача 10.

Имеются условные данные основных показателей деятельности предприятий потребительской кооперации по областям за определенный временной интервал, представленных в табл. 5.3.

Таблица 5.3

Данные показателей деятельности предприятий потребсоюзов

Номер потребсоюза	ОТП	В том числе в млн руб. от ОТП					У
		X_1	X_2	X_3	X_4	X_5	
1	61,61	9,98	32,55	0,3	4,67	4,34	0,98
2	30,06	3,88	14,37	0,89	2,12	1,5	-0,42
3	14,05	2,13	8,91	0,12	1,52	0,58	-1,78
4	92,01	15,07	46,75	0,69	7,06	12,51	2,91
5	81,93	12,13	48,3	0,51	5,6	4,23	0,37
6	94,42	14,52	58,46	0,6	7,09	0,66	0,64
7	75,53	14,16	36,29	1,52	7,34	2,14	1,64
8	89,66	12,45	50,24	2,84	7,92	3,14	-2,43
9	33,63	4,77	20,35	0,24	2,99	0,63	-0,11
10	18,23	3,38	10,69	0,03	2,1	0,73	-0,14
11	40,32	7,01	19,36	0,6	3,54	2,28	0,81
12	22,83	3,95	11,67	0,05	2,96	0,94	-0,08
13	19,58	3,02	11,63	0,1	2,05	0,61	-0,45
14	74,35	9,13	38,22	2,02	3,52	0,02	0,53

ОТП — объем товарной продукции (млн руб.);

X_1 — расходы на оплату труда (млн руб.);

X_2 — расходы на сырье, основные и вспомогательные материалы (млн руб.);

X_3 — отчисления в фонд развития потребкооперации (млн руб.);

X_4 — расходы на топливо, энергию на технологические цели (млн руб.);

X_5 — расходы на прямые расходы (млн руб.);

У — размер прибыли (+) или убытков (-) (млн руб.).

Анализ данных показывает, что прибыль предприятий по потребсоюзам имеет два значения У: прибыльное и убыточное. Предположим, что выделенные показатели способны предсказать с определенной вероятностью, к какой группе — прибыльным или убыточным — относятся предприятия потребсоюза.

Тогда возникает вопрос, можно ли спрогнозировать, к какой группе — прибыльным или убыточным — относятся предприятия нового потребсоюза, имеющего такой же набор основных показателей?

Конкретизируем постановку задачи. В табл. 5.3 имеются показатели деятельности предприятий по 14 потребсоюзам. Располагая данными только по 13 потребсоюзам необходимо спрогнозировать к какой группе — прибыльных или убыточных — относится 14 потребсоюз.

Решение задачи.

Решение задачи можно осуществить с помощью ППП Stadia с использованием дискриминантного анализа.

Приводим содержание справки ППП Stadia по дискриминантному анализу.

Дискриминантный анализ позволяет:

- проверить гипотезу о непротиворечивости предполагаемой классификации заданного множества n объектов в m-мерном пространстве переменных;
- классифицировать новые объекты.

В ходе вычислений находится набор дискриминирующих функций, обеспечивающий заданную классификацию.

Исходные данные представляются в виде матрицы размером $(m + 1)n$, в которой первые m столбцов содержат значения m переменных для n объектов, а $(m + 1)$-я переменная в качестве своих значений содержит для каждого объекта номер его класса (натуральные числа от 1 до N, где N — число классов). Объекты (строки) в матрице могут располагаться произвольно относительно номеров классов.

Если, кроме вычисления дискриминантной функции, необходимо с ее помощью классифицировать ряд новых объектов, то такие объекты также исходно включаются в матрицу данных с номером класса.

Выдача результатов включает:

• суммарное межкластерное расстояние Махаланобиса D^2 (Mahalanobis) между классами с уровнем значимости Р для нулевой гипотезы "$D^2 = 0$" (для гипотезы о невозможности разбиения совокупности объектов на заданное число классов), а также скорректированное значение $D^2_{прив}$.

• коэффициенты дискриминантной функции, обеспечивающей отнесение объектов к данному классу, отдельно для каждого класса;

• таблицу, где для каждого объекта указывается: номер его класса i, расстояние Махаланобиса D_i^2 (от объекта до центра класса), уровень значимости Р нулевой гипотезы "$D_i^2 = 0$" (гипотезы о том, что объект может быть отнесен к данному классу) и вероятность отнесения объекта к этому классу.

Если $P > 0{,}05$, соответствующая нулевая гипотеза может быть принята.

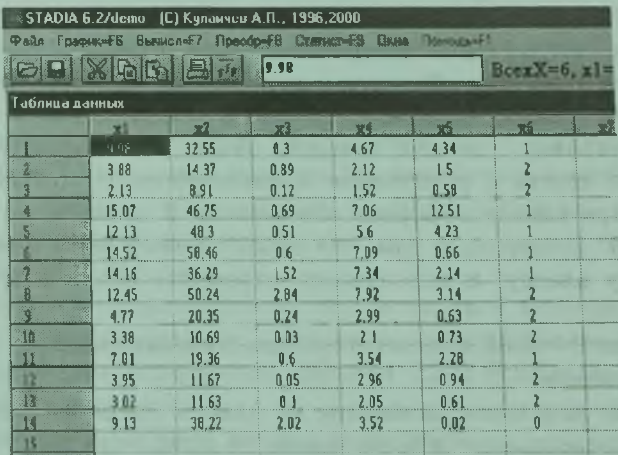
Если начальное разбиение на классы уверенно произвести не представляется возможным, то можно предварительно выполнить кластерный анализ с использованием дивизивной стратегии разбиения и испробовав несколько вариантов числа группировок.

Ограничения: число классов не может быть меньше двух, число переменных не должно превышать 140, 62, 20 при объеме матрицы данных в 20000, 4000 и 400 чисел.

Решение задачи выполним в такой последовательности:

10.1. Загрузите ППП Stadia 6.2/demo.

10.2. Введите данные табл. 5.3. Введите значения переменных X_1, X_2, X_3, X_4, X_5, взятые из табл. 5.3, в электронную таблицу ППП Stadia при соблюдения одного условия: дробная часть числа должна отделяться точкой. Введите в последнюю колонку преобразованные значения У по следующему правилу: если для первых 13 объектов значение У положительное (прибыльное), то присваиваем ему значение 1, если значение У отрицательное (убыточное), то присваиваем ему значение 2. Для последнего 14-го объекта переменной У присвоим значение 0. Код 0 будет означать, что с помощью дискриминантного анализа необходимо определить, к какой группе относится данный 14-й объект. Результат действий изображен на рис. 5.26.


STADIA 6.2/demo (C) Кулаичев А.П., 1996,2000

Файл График=F6 Вычисл=F7 Преобр=F8 Статист=F9 Окна

9.98

ВсехX=6, x1=

Таблица данных

	x1	x2	x3	x4	x5	x6
1	9.98	32.55	0.3	4.67	4.34	1
2	3.88	14.37	0.89	2.12	1.5	2
3	2.13	8.91	0.12	1.52	0.58	2
4	15.07	46.75	0.69	7.06	12.51	1
5	12.13	48.3	0.51	5.6	4.23	1
6	14.52	58.46	0.6	7.09	0.66	1
7	14.16	36.29	1.52	7.34	2.14	1
8	12.45	50.24	2.84	7.92	3.14	2
9	4.77	20.35	0.24	2.99	0.63	2
10	3.38	10.69	0.03	2.1	0.73	2
11	7.01	19.36	0.6	3.54	2.28	1
12	3.95	11.67	0.05	2.96	0.94	2
13	3.02	11.63	0.1	2.05	0.61	2
14	9.13	38.22	2.02	3.52	0.02	0
15						

Рис. 5.26. Результат ввода исходных данных в электронную таблицу ППП Stadia

10.3. Произведите дискриминантный анализ данных и определите, к какой группе — убыточной или прибыльной — относится объект № 14. Нажмите кнопку меню "Статист=F9" в окне "Статистические методы", нажмите кнопку "Р=дискриминантный " и получите результат, изображенный на рис. 5.27.

Анализ рис. 5.27 показывает, что объект № 14 отнесен к группе номер 1 с вероятностью 1. Это означает, что деятельность этого объекта является прибыльной. Фактическое состояние этого объекта является прибыльным, следовательно, ДИСКРИМИНАНТНЫЙ АНАЛИЗ. Файл:

Расстояние Махаланобиса=98,78, значимость=0

Класс	<--- Коэффициенты дискриминантной функции: a[0], a[1], ... --->					
1	-17,75	10,43	-0,6278	-4,778	-9,887	-0,9995
2	-2,781	-2,961	0,03387	-3,668	6,839	0,1949

Объект	Класс	D^2	Значим	Вероят.отнесения
1	1	0,6169	0,9872	1
2	2	6,309	0,2773	1
3	2	0,744	0,9804	1
4	1	8,276	0,1416	1
5	1	2,41	0,79	1
6	1	6,035	0,3028	1
7	1	8,259	0,1425	1
8	2	9,06	0,1067	1
9	2	0,6721	0,9844	1
10	2	1,13	0,9514	1
11	1	6,673	0,2461	[illegible]
12	2	4,397	0,4937	1
13	2	0,4195	0,9947	1
14	1	47,48	0	1

Рис. 5.27. Результат расчетов с помощью дискриминантного анализа

можно сделать два вывода. Первый вывод: выделенные показатели деятельности объекта способны идентифицировать предприятие на признак убыточности и прибыльности. Второй вывод: дискриминантный анализ способен произвести прогноз состояния предприятия по выделенным показателям.

5.4. Исходные данные индивидуальных заданий

Задача 11.

Имеются исходные данные, представленные в табл. 5.4.

Необходимо получить прогноз на 11 месяц по данным табл. 5.4 с использованием различных инструментальных средств Excel.

X — порядковый номер месяца 2002 г.;

$У_1 - У_3$ — прибыль соответствующих потребсоюзов, внедривших систему качества;

Таблица 5.4

Условные данные прибыли шести Райпотребсоюзов РФ на 2002 г.

X	$У_1$	$У_2$	$У_3$	$У_4$	$У_5$	$У_6$
1	12	21	31	5	9	8
2	13	23	33	6	5	9
3	15	22	34	2	8	7
4	14	24	32	3	7	6
5	16	25	35	1	6	4
6	17	26	36	2	5	5
7	18	28	38	3	4	6
8	19	27	37	4	3	3
9	18	29	39	2	2	1
10	20	30	41	1	1	2
11	?	?	?	?	?	?

$У_4 - У_6$ — прибыль соответствующих потребсоюзов, не внедривших систему качества.

Вопросы для самоконтроля

1. Укажите этапы построения графика “Точечная”.

2. Напишите расчетные формулы коэффициентов a_0 и a_1.

3. Укажите, в какой последовательности необходимо выполнить расчеты по формуле $\mathbf{A = (X'X)^{-1}X'У}$.

4. Укажите достоинства и недостатки функции “Линейн”.

5. Укажите назначение всех чисел в протоколе расчета по функции “Линейн”.

6. Назовите состав программ, которые можно выполнить с помощью пакета программ “Анализ данных”.

7. Укажите достоинства и недостатки пакета программ “Анализ данных”.

8. Укажите возможности функции “Поиск решения”.

9. Укажите шаги выполнения функции “Предсказ”.

10. Укажите этапы получения расчетов с помощью ППП Stadia 6.2/demo.

11. Укажите назначение дискриминантного анализа.

Глава 6
МУЛЬТИКОЛЛИНЕАРНОСТЬ

Мы приступаем к изучению второй части курса эконометрики, которая посвящена изучению нарушений предпосылок метода наименьших квадратов. Одним из этих нарушений является мультиколлинеарность.

6.1. Мультиколлинеарность и методы ее устранения

Определение, причины и последствия мультиколлинеарности.

При построении многофакторной модели необходимо соблюдение следующего правила.

Факторы, включаемые в модель, должны быть сильно связаны с зависимой переменной и не должны быть связаны между собой.

Необходимо запомнить это правило, так как на нем основано построение большинства многофакторных моделей.

Мультиколлинеарность — это свойство переменных объектов изучения, которое заключается в том, что существует хотя бы одна пара переменных, между которыми имеется существенная связь, определяемая по коэффициенту корреляции.

Причиной мультиколлинеарности является свойство экономической системы, в которой все переменные экономических объектов должны быть взаимосвязаны между собой. Иначе экономическая система не будет существовать. Поэтому мультиколлинеарность была, есть и будет, бороться с ней бесполезно, но учитывать ее при построении многофакторных моделей необходимо.

Мультиколлиниарность является свойством экономической системы, которое используется при построении систем одновременных уравнений.

Последствия мультиколлинеарности заключаются в том, что при построении многофакторной модели, в которой факторы мультиколлинеарны, возникают следующие проблемы в определении коэффициентов и качества модели.

а) Нельзя определить коэффициенты модели по формуле

$$\mathbf{A} = (\mathbf{X'X})^{-1}\mathbf{X'У},$$

так как нельзя найти обратную матрицу

$$(\mathbf{X'X})^{-1},$$

в которой матрица факторов X содержит неортогональные переменные.

Ортогональность в эконометрике означает независимость переменных между собой.

б) Если мультиколлинеарность незначительная, то удается определить коэффициенты модели, однако ошибки коэффициентов увеличиваются, так как ошибки коэффициентов модели зависят от диагональных элементов матрицы

$$(\mathbf{X'X})^{-1}.$$

При этом удается построить достоверную модель.

в) Если мультиколлинеарность является значительной, то коэффициенты модели теряют экономический смысл, ошибки коэффициентов становятся очень большими и модель, как правило, становится недостоверной.

г) При включении в модель нового мультиколлинеарного фактора предыдущие значения коэффициентов изменяются и их ошибки возрастают, достоверность модели снижается. Это свойство взамисвязанных факторов может быть средством обнаружения мультиколлинеарности.

Критерии проверки достоверности мультиколлинеарности

Критерий Феррара-Глобера определения мультиколлинеарности.

Критерий Феррара-Глобера позволяет определить факторы, которые можно вместе включать в модель с помощью последовательного приближения к обнаружению тех пар переменных, между которыми существует достоверная связь. Критерий Феррара-Глобера анализирует только объясняющие переменные и позволяет проверить три нулевые гипотезы[1].

Нулевая гипотеза 1.

H_0: "Мультиколлинеарность между всеми объясняемыми переменными отсутствует".

Нулевая гипотеза 2.

H_0: "Между переменной $X_к$ и всеми остальными переменными взаимосвязь отсутствует".

Нулевая гипотеза 3.

H_0: "Между перемененными X_i и X_j, мультиколлинеарность отсутствует".

Приступаем к изложению методики проверки выдвинутых гипотез.

Нулевая гипотеза 1.

H_0: "Мультиколлинеарность между всеми объясняемыми переменными отсутствует".

Если $\chi^2_ф > \chi^2_т$ ($\alpha = 0{,}05$; $m = d(d-1)/2$),

то нулевая гипотеза отвергается и утверждается, что между объясняемыми переменными существует достоверная мультиколлинеарность.

$\chi^2_ф = -[n - 1 - (2d + 5)/6] \ln |R|$,

где $\chi^2_ф$ — фактическое значение критерия Хи-квадрат;

$\chi^2_т$ ($\alpha = 0{,}05$; $m = d(d-1)/2$) — табличное значение критерия Хи-квадрат на уровне значимости a и числе степеней свободы v;

---

[1] *Клас А., Гергели К., Колек Ю., Шуян И.* Введение в эконометрическое моделирование. — М.: Статистика, 1978. С. 104–110.

n — объем выборки;

d — количество объясняемых переменных;

R — матрица парных коэффициентов корреляции между всеми d переменными;

|**R**| — определитель матрицы **R**.

$$r(X_1, X_2) = \frac{\sum (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\sqrt{\sum (X_{1i} - \bar{X}_1)^2 \sum (X_{2i} - \bar{X}_2)^2}},$$

где $r(X_1,X_2)$ — парный коэффициент корреляции между переменными X_1 и X_2.

Нулевая гипотеза 2.

H_0: "Между переменной $X_к$ и всеми остальными переменными взаимосвязь отсутствует".

Если $F > F_{кр}(\alpha = 0{,}05;\ m_1 = n - d;\ m_2 = d - 1)$,

то нулевая гипотеза отвергается и утверждается, что между переменной $X_к$ и всеми остальными переменными с вероятностью $1 - \alpha = 0{,}95$ существует достоверная мультиколлинеарность.

$$F = (C_{кк} - 1)(n - d)/(d - 1),$$

где F — фактическое значение критерия Фишера;

$C_{кк}$ — диагональный элемент матрицы $\mathbf{C} = \mathbf{R}^{-1}$;

к — номер переменной;

$\mathbf{R}^{-1}$ — обратная матрица от матрицы **R**.

$F_{кр}(\alpha = 0{,}05;\ m_1 = n - d;\ m_2 = d - 1)$ — табличное значение критерия Фишера на уровне значимости а и числа степеней свободы для большей дисперсии m_1 и степеней свободы для меньшей дисперсии m_2.

Нулевая гипотеза 3.

H_0: "Между переменными X_i и X_j, взятыми из совокупности d переменных, при объеме выборки n, мультиколлинеарность отсутствует".

Если $t(X_i,X_j)_ф > t_{\alpha/2}\ (\alpha = 0{,}05;\ m = n - d)$,

то нулевая гипотеза отвергается и утверждается, что между переменными X_i и X_j, взятыми из совокупности d переменных, мультиколлинеарность присутствует.

$$t(X_i, X_j)_ф = \frac{r_ч(X_i, X_j) \cdot \sqrt{n-d}}{\sqrt{1 - r_ч(X_i X_j)^2}},$$

где $t(X_i,X_j)_ф$ — фактическое значение критерия Стьюдента;

$r_ч(X_i,X_j)$ — частный коэффициент корреляции, показывает степень линейной зависимости между переменными X_i и X_j при условии, что все остальные переменные не изменяются и принимают средние значения.

$$r_ч(X_i, X_j) = \frac{-C_{ij}}{\sqrt{C_{ii} C_{jj}}},$$

где $\mathbf{C} = \mathbf{R}^{-1}$ — обратная матрица от матрицы $\mathbf{R}$;

$t_{\alpha/2}(\alpha = 0{,}05;\ m = n - d)$ — табличное значение критерия Стьюдента на уровне значимости $\alpha = 0{,}05$ и числе степеней свободы m.

Критерий Феррара-Глобера представляет интерес в теоретическом плане, однако практически ничего нового не дает по сравнению с матрицей парных и частных коэффициентов корреляции, поэтому он не получил широкого распространения (к тому же он имеет сложный алгоритм расчета).

Метод обнаружения мультиколлинарности

Известен простой метод обнаружения мультиколлинеарности, который основан на анализе парных коэффициентов корреляции между всеми переменными модели.

Если фактическое значение коэффициента корреляции будет больше критического, то мультиколлинеарность является достоверной, в противном случае мультиколлинеарность является незначительной. Более подробно о проверки достоверности коэффициента корреляции см. в параграфе 6.3.

Методы устранения мультиколлинеарности

Известно два направления устранения мультиколлинарности:

- методы, которые используют преобразования данных;

• методы, которые используют разные приемы построения моделей.

Различают следующие методы преобразований данных:

• метод ортогональных преобразований факторов, которые входят в состав модели;

• метод преобразований Тихонова, суть которого заключается в том, что необходимо так преобразовать матрицу **X'X**, чтобы можно было получить от нее обратную матрицу;

• вычисление ортогональных собственных векторов от матрицы факторов модели.

Преобразование данных для устранения мультиколлинеарности имеет следующие недостатки:

• после преобразований данных нет алгоритма перехода к исходным данным;

• сложно интерпретировать новые преобразованные данные.

По этим причинам методы преобразования факторов не получили широкого распространения.

Известны следующие часто используемые методы построения моделей, при наличии мультиколлинеарных факторов:

• шаговый метод построения модели;

• метод корреляционных плеяд.

Приступим к более подробному их изложению.

6.2. Шаговая регрессия

Шаговый метод является самым лучшим методом построения модели при наличии мультиколлинеарности между переменными.

Различают шаговую регрессию: включения, исключения, включения и исключения. Практическое распространение получила шаговая регрессия включения.

Метод включения состоит в том, что в модель включаются факторы по одному в определенной последовательности. Построение модели заканчивается, если модель перестает удовлетворять определенным условиям.

Метод исключения состоит в том, что в модель включаются все факторы, затем в определенной последовательности исключаются до тех пор, пока модель станет удовлетворять определенным условиям.

Метод включения и исключения состоит в том, что реализуется метод включения с возможностью исключения недостоверного фактора.

Практическую значимость получил метод включения по таким причинам:

- если имеется много факторов, количество которых больше объема выборки, то метод исключения реализовать не удается. Так как нарушается правило многофакторного моделирования — количество факторов должно быть меньше объема выборки. Метод включения в этой ситуации даст результат;
- метод включения позволяет постепенно улучшать модель до момента ее ухудшения, что соответствует логике научного исследования.

Шаговый метод включения построения модели заключается в том, что на каждом шаге в модель вводится по одному фактору и проверяется достоверность модели по критическому критерию Фишера, численное значение которого вводится пользователем. Если после ввода в регрессионное уравнение нового фактора критерий Фишера стал меньше критического, то построение модели заканчивается. Для реализации шаговой регрессии используется метод Эфроимсона[1].

Алгоритм шаговой регрессии, реализованный методом Эфроимсона, очень сложный. Но его можно выполнить достаточно просто по следующей упрощенной схеме. На первом шаге в модель вводится тот фактор, который имеет наибольший коэффициент корреляции с зависимой переменной. На втором шаге и последующих шагах в модель включается фактор, который имеет наибольший коэффициент корреляции с

---

[1] *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ. — М.: Статистика, 1973.

остатками модели. После каждого включения фактора в модель рассчитываются ее характеристики и проверяется на достоверность. Если после включения в модель фактора, модель стала недостоверной, то на этом построение модели заканчивают. Таким образом, шаговый метод включения можно реализовать с помощью электронных таблиц Excel[1].

6.3. Метод корреляционных плеяд

Если имеется матрица парных коэффициентов корреляции между всеми переменными, которые участвуют в построении модели, то можно упростить выбор факторов для включения в модель с помощью метода корреляционных плеяд.

Метод корреляционных плеяд заключается в том, что на круговой диаграмме размещают обозначения факторов, количество которых не должно превышать 20 (при большом количестве факторов график теряет наглядность), линией соединятся те факторы, между которыми существует достоверный коэффициент корреляции. При этом происходит фильтрация тех факторов, между которыми имеется существенная связь.

Плеяда — совокупность переменных, связь между которыми указана стрелками.

Достоверность коэффициента корреляции проверяется методом сравнения фактического коэффициента с критическим. Если фактический коэффициент корреляции по модулю больше критического, то линейная связь между двумя переменными является достоверной. В противном случае линейная связь между двумя переменными статистически является недоказанной.

Фактический и критический коэффициенты корреляции определяются по следующим формулам:

$$r(X_1X_2) = \frac{\sum(X_{1i} - \bar{X}_1)(X_2 - \bar{X}_2)}{\sqrt{\sum (X_{1i} - \bar{X}_1)^2 \sum(X_{2i} - \bar{X}_2)^2}},$$

[1] *Афифи А., Эйзен С.* Статистический анализ. Подход с использованием ЭВМ. — М.: Мир, 1982.

где $r(X_1,X_2)$ — фактический коэффициент корреляции между переменными X_1 и X_2.

Расчет критического значения коэффициента корреляции $r_{кр}$ на уровне значимости а = 0,05 и числе степеней свободы (n − 2) можно произвести с использованием следующих соотношений.

Известно, что проверка значимости выборочного коэффициента корреляции основана на вычислении критерия Стьюдента

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}.$$

После несложных преобразований можно получить критическое значение коэффициента корреляции $r_{кр}$.

$$r_{кр} = \frac{t_{\alpha/2}}{\sqrt{t_{\alpha/2}^{\ 2} + n - 2}}, \qquad (6.1)$$

где $t_{\alpha/2}(\alpha = 0{,}05;\ m = n - 2)$ — табличное значение критерия Стьюдента;

n — объем выборки.

Допустим, мы построили корреляционные плеяды для четырех переменных: У, X_1, X_2, X_3 (рис. 6.1).

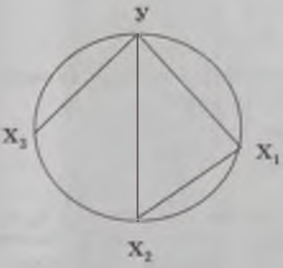


Рис. 6.1. Корреляционные плеяды

Анализ корреляционных плеяд, с использование основного правила построения модели позволяет сделать вывод, что в модель надо включить фактор X_3 и один из двух факторов X_1 или X_2, так как они тесно связаны между собой. Предпочтение можно отдать тому фактору, который теснее связан с У и более доступен для сбора и анализа данных.

Напоминаем основное правило построения модели:

Факторы, включаемые в модель, должны быть сильно связаны с зависимой переменной и не должны быть связаны между собой.

Метод корреляционных плеяд можно использовать для частных коэффициентов корреляции. Частный коэффициент корреляции учитывает связь между двумя переменными при условии, что все остальные факторы не изменяются.

Частные коэффициенты корреляции рассчитываются по формуле

$$r(X_i, X_j, X) = \frac{-C_{ij}}{\sqrt{C_{ii}C_{jj}}},$$

где частный коэффициент корреляции определяется между факторами X_i и X_J при условии, что остальные факторы не изменяются;

C_{ij} — элементы обратной матрицы, определенной от матрицы парных коэффициентов корреляции $\mathbf{R}(k)$ для k факторов, например для трех факторов.

$$\mathbf{R}(3) = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix}.$$

Вычисление обратной матрицы $\mathbf{R}(k)^{-1}$ можно выполнить в Excel.

Примечание. В большинстве эконометрической литературе частный коэффициент корреляции вычисляется по формуле

$$r(X_i, X_j, X) = \frac{-g_{ij}}{\sqrt{g_{ii}g_{jj}}},$$

где g_{ij}, g_{ii}, g_{jj} — алгебраические дополнения элементов r_{ij}, r_{ii}, r_{jj} матрицы выборочных коэффициентов корреляции [5, с. 128–29]. Расчет частного коэффициента корреляции по этой формуле можно проводить с помощью калькулятора. Но в Excel нет функции расчета алгебраического дополнения к матрице.

Методика выполнения метода корреляционных плеяд для частных и обычных коэффициентов корреляции является одинаковой. В качестве критического значения частного коэффициента корреляции можно использовать формулу (6.1).

Модели, построенные шаговым регрессионным анализом и методом корреляционных плеяд для частных коэффициентов корреляции, часто совпадают между собой.

6.4. Использование пакетов прикладных программ для проведения расчетов множественной регрессии

В настоящее время шаговая (step) регрессия имеется в большинстве статистических пакетах прикладных программ (Excel не имеет шаговой регрессии). Для иллюстрации расчетов шаговой регрессии воспользуемся доступной Российской программой Stadia.

Расчеты шаговой регрессии можно выполнить в такой последовательности:

Шаг 1. Выполняем файл Stadia.exe. Результат выполнения файла представлен на рис. 6.2.

Рис. 6.2. Результат запуска программы

Шаг 2. Входим в систему Stadia и вводим исходные данные. Результат выполнения шага 2 представлен на рис. 6.3.

Рис. 6.3. Результат ввода данных

Шаг 3. Нажимаем клавишу F9 и входим в меню статистических методов, представленных на рис. 6.4.

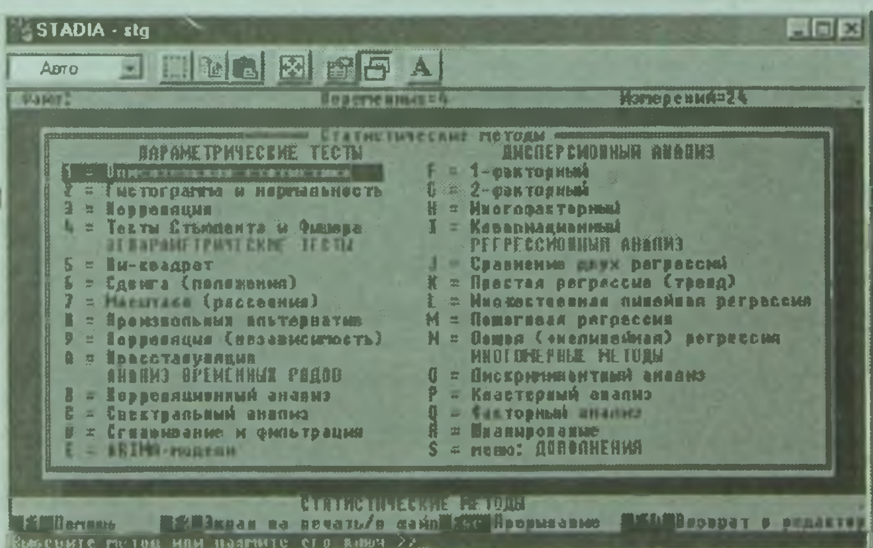


Рис. 6.4. Главное меню статистических методов

Шаг 4. Выбираем пункт "Пошаговая регрессия", программа предлагает записать корреляции в матрицу данных — отвечаем нет. Результат расчетов представлен на рис. 6.5.

Рис. 6.5. Протокол расчета

Протокол расчетов имеет следующие обозначения:

- Среднее — среднее арифметическое значение;
- Ст. отклон. — среднее квадратическое отклонение (S);
- X_1, X_2, X_3 — объясняемые факторы;
- X_4 — зависимая переменная $У$;
- по диагонали корреляционной матрицы стоят вариации переменных.

Шаг 5. Выбираем селекцию вперед. Эта операция означает, что мы выбираем шаговый метод включения. Выбираем критерий окончания селекции: 0 — F-уровень. Вводим F-уровень включения = 3,84. Вводим допустимый уровень сходства = 0,01. Результат расчетов представлен на рис. 6.6.

На рис. 6.6 изображен результат включения фактора X_3. Полученная модель имеет следующий вид: $У = 3{,}732 + 1{,}184X_3$. Нажимаем клавишу Enter. Результат расчетов представлен на рис. 6.7.

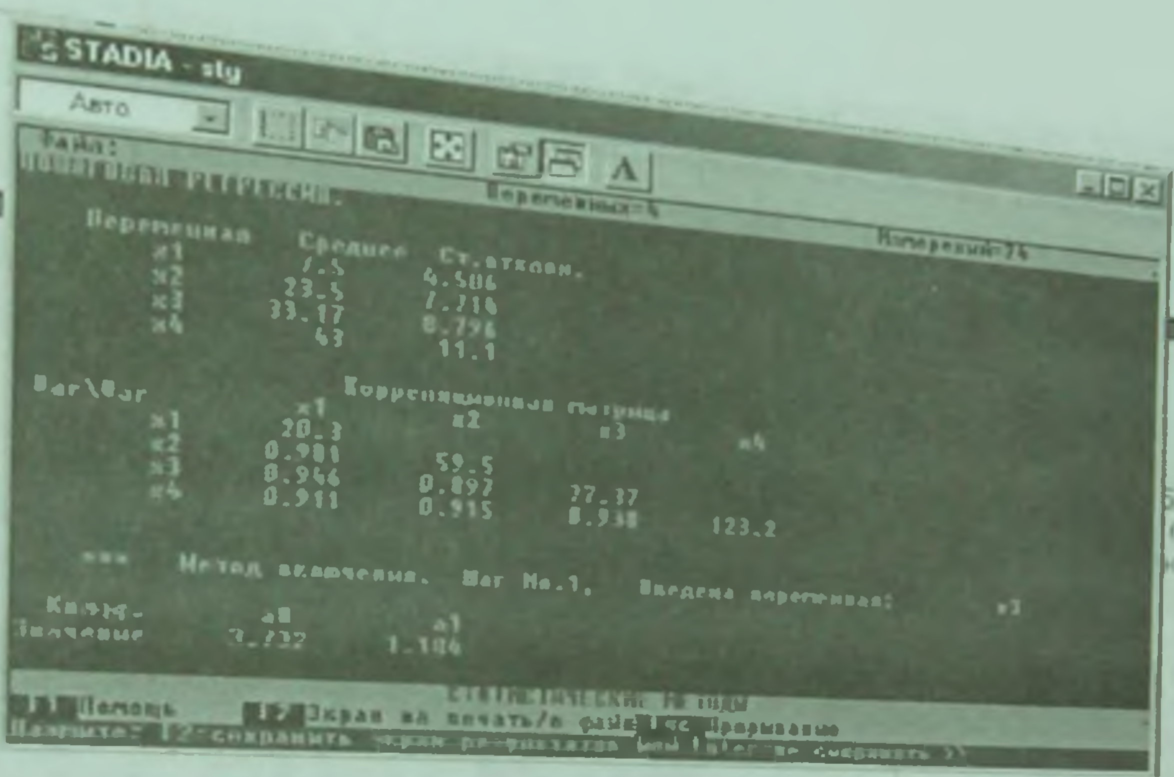


Рис. 6.6. Реализация включения первой переменной X_3

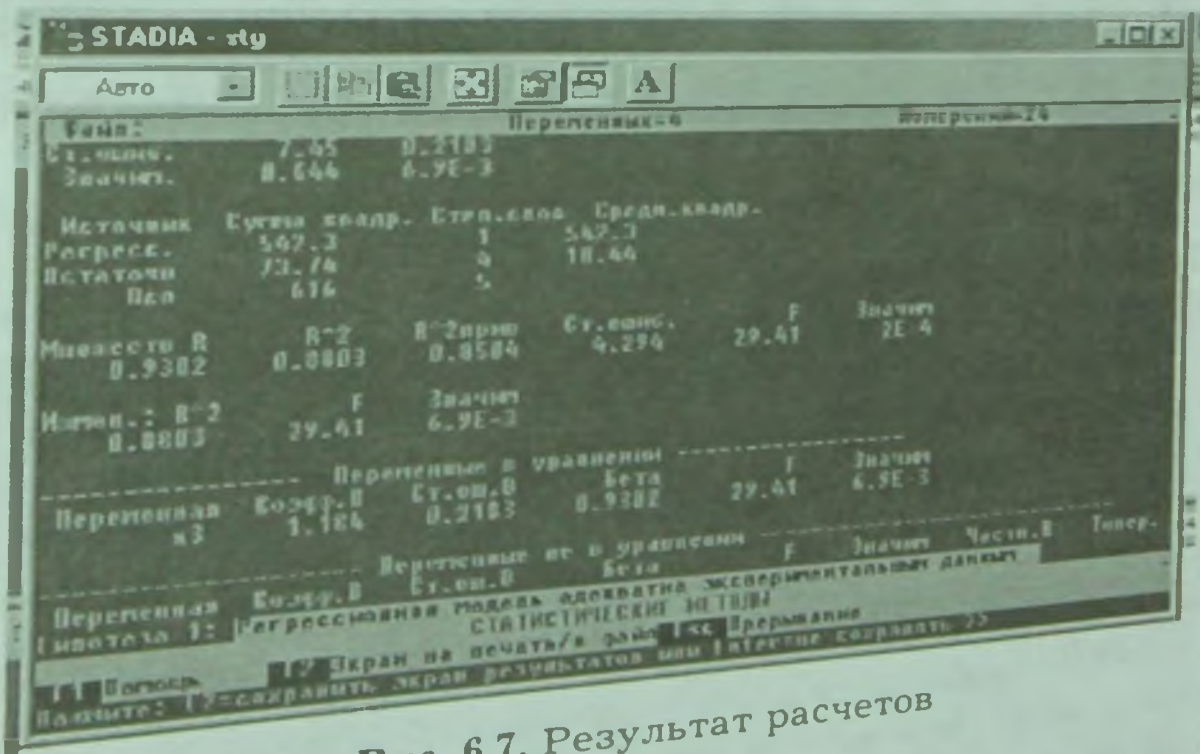


Рис. 6.7. Результат расчетов

На рис 6.7 имеется шесть таблиц.

Обозначения на рис. 6.7 имеют следующее содержание.

В первой таблице:

• Ст. ошиб. — стандартная ошибка коэффициента или среднее квадратическое значение коэффициента модели от своего математического ожидания;

• Значим. — значимость коэффициента модели. В нашем случае значимость свободного коэффициента, равная 0,644, больше, чем 0,05, следовательно этот коэффициент не отличается от нуля, значимость коэффициента a_1 равна 6,9Е−3 = 0,0069.

Во второй таблице (называемой дисперсионным анализом или Anova):

• Источник — источник вариации;

• Сумма квадр. — сумма квадратов или вариация соответствующей компоненты;

• Степ. своб. — степень свободы соответствующей компоненты;

• Средн. квадр. — среднее квадратическое отклонение.

В третьей таблице, называемой "Основные характеристики модели":

• Множеств. R — множественный коэффициент корреляции;

• R^2 — множественный коэффициент детерминации;

• R^2 прив. — приведенный множественный коэффициент детерминации;

• Ст. ошиб. — стандартная ошибка или ошибка модели Е;

• F — критерий Фишера;

• Значим. — значимость критерия Фишера. Значимость критерия Фишера равна 2Е-4 = 0,0002, что меньше 0,05, следовательно, модель является достоверной.

В четвертой и пятой таблицах имеются характеристики модели с бета-коэффициентами. Бета-коэффициентами называют коэффициенты модели, в которой переменные нормализованы. Нормализирование переменных X и У производится по формуле

$$X_н = \frac{X_i - \overline{X}}{S},$$

где S — среднее квадратическое отклонение для переменной X.

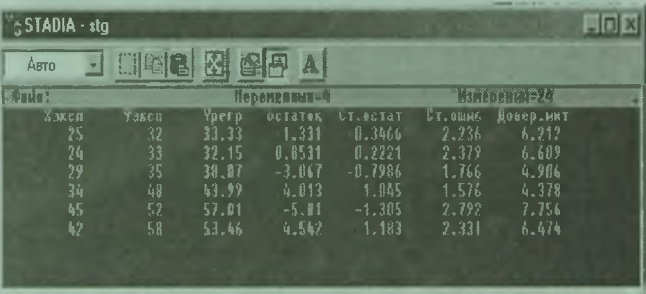
Бета-коэффициенты являются безразмерными, и их численные значения тесно связаны с долей влияния на зависимый признак.

В шестой таблице имеются переменные, которые не вошли в модель.

Гипотеза 1. Регрессионная модель адекватна экспериментальным данным — вывод, который делает программа.

Примечание. Слово “адекватна” следует заменить понятием “достоверна”. Понятие адекватности относится к планированию эксперимента, где она имеет особую специфику расчетов и проверку[1].

Шаг 6. Нажимаем клавишу Enter, задаем режим выдачи остатков, получаем результаты, представленные на рис. 6.8.



Хэксп	Уэксп	Урегр	остаток	Ст.остат	Ст.ошиб	Довер.инт
25	32	33.33	1.331	0.3466	2.236	6.212
24	33	32.15	0.8531	0.2221	2.379	6.609
29	35	38.07	-3.067	-0.7986	1.766	4.906
34	48	43.99	4.013	1.045	1.576	4.378
45	52	57.01	-5.01	-1.305	2.792	7.756
42	58	53.46	4.542	1.183	2.331	6.474

Рис. 6.8. Результат расчетов

На рис. 6.8 имеются следующие обозначения:

- Хэксп — экспериментальные значения X_3;
- Уэксп — экспериментальные значения У;
- Урег — расчетные значения У;
- остаток — е — остатки модели;
- Ст. остат — стандартизированные остатки;
- Ст. ошиб — стандартные ошибки;
- Довер. инт — доверительный интервал.

Если сохранить регрессионные остатки в матрице данных, то можно по этим данным построить графики с участием значений фактических, расчетных, доверительных интервалов.

---

[1] *Налимов В.В., Чернова Н.А.* Статистические методы планирования экстремальных экспериментов. — М.: Наука, 1965.

Шаг 8. Нажимаем клавишу Enter и возвращаемся в главное меню.

Вывод. Мы ознакомились с типичным диалогом проведения расчетов по методу шаговой регрессии с участием пакета прикладных программ Stadia. Диалоги по другим пакетам могут отличаться друт от друга, но суть остается одинаковой.

Вопросы для самоконтроля

1. Дайте определение мультиколлинеарности.
2. Приведите причины появления мультиколлинеарости в экономических системах.
3. Укажите последствия наличия мультиколлинеарности при построении многофакторных моделей.
4. Сформулируйте три нулевых гипотезы, которые проверяются с помощью метода Феррара-Глобера.
5. Укажите этапы проверки мультиколлинеарности с помощью матрицы парных коэффициентов корреляции.
6. Приведите формулу расчета частных коэффициентов корреляции.
7. Укажите экономический смысл частных коэффициентов корреляции.
8. Назовите способы устранения мультиколлинеарности.
9. Укажите сущность шаговой регрессии.
10. Раскройте суть метода корреляционных плеяд устранения мультиколлинеарности.
11. Укажите этапы построения шаговой регрессии с помощью ППП Stadia.

Глава 7
РЕГРЕССИОННЫЕ МОДЕЛИ С ПЕРЕМЕННОЙ СТРУКТУРОЙ

Имеется широкий вид моделей с переменной структурой, отличительной особенностью которых является отсутствие строго определенной математической функции или набора переменных, отражающих тенденцию зависимой переменной У.

Имеются три класса моделей, в которых может изменяться ее структура:

- регрессионная модель с гибкой структурой;
- метод группового учета аргументов;
- регрессионная модель с переменной структурой.

7.1. Линейные регрессионные модели с переменной структурой

Регрессионная модель с гибкой структурой — это такая модель, которая может самостоятельно изменять математическую функцию, отражающую тенденцию зависимости У от X. Первой моделью с гибкой структурой является функция Куликова[1], которая вследствие сложности алгоритма и неопределенности выбора параметров не получила широкого распространения. В теоретическом плане это направление представляет интерес, но в практическом плане

[1] *Саркисян С. А., Голованов Л. В.* Прогнозирование развития больших систем. — М.: Статистика, 1975. С. 58–72.

оно потеряло свою актуальность, так как имеются программы, которые используют более 2000 разных апроксимационных функций.

Метод группового учета аргументов (МГУА)[1] основан на предположении о том, что критерием достоверности модели является проверка ее на новых данных. На обучающей совокупности строятся функции различной сложности, но достоверность модели проверяется на новых данных. Для получения прогноза используется модель, которая является лучшей на новой совокупности. В этом методе изменялась структура модели, поэтому мы отнесли ее к классу модели с изменяющейся структурой. Метод МГУА в эконометрику не вошел, так как при подборе функции количество коэффициентов может быть больше объема выборки, что является нарушением предпосылок регрессионного анализа.

Однако идея подбора модели по контрольной совокупности является очень плодотворной и получила свое развитие в работе [1, с. 35–36, 379–380]. В настоящее время принципом ретроспективных расчетов (метод перекрестного анализа дееспособности модели) называют метод деления данных на обучающую и контрольную выборки, проведение спецификации, идентификации модели по обучающей выборке, а оценка качества модели проверяется на контрольной выборке, при этом все расчеты проводятся в рамках классического регрессионного анализа. Имеется пакет программ “Стат. эксперт”, где реализован принцип ретроспективных расчетов.

Регрессионная модель с переменной структурой — это такая модель, которая учитывает неоднородность количественной зависимой переменной с помощью качественного фактора (фиктивной переменной). Различают два типа неоднородностей зависимой переменной: функциональную и структурную.

---

[1] *Ивахненко А. Г. и др.* Справочник по типовым программам моделирования. — К.: Техника, 1980.

Функциональной неоднородностью называют такую тенденцию зависимой переменной, которая имеет четко выраженное изменение вида тенденции, обычно относящиеся к временным рядам

Структурной неоднородностью называют такую тенденцию зависимой переменной, которая имеет четко выраженные выбросы для временных рядов или расслоения для пространственных данных.

Пример 1. При изучении временного ряда прибыли за два года за каждый месяц было обнаружено, что в прошлом году прибыль изменялась около среднего значения, в текущем году прибыль стала монотонно возрастать. Функциональная неоднородность зависимой переменной объясняется изменением в текущем году величины налога на прибыль.

Пример 2. При изучении пространственной выборки было обнаружено, что с ростом дохода населения затраты на предметы роскоши возрастают, но расслаиваются по признаку принадлежности семьи к скромным или к честолюбивым. При этом остатки модели имеют структурную неоднородность с четко выраженной гетероскедастичностью.

В двух примерах наблюдается общая закономерность, которая заключается в том, что на зависимую переменную действует качественный фактор, имеющий две градации, называемый фиктивной переменной.

Фиктивная переменная — это объясняющая качественная переменная, имеющая два или более класса.

При изучении качественных переменных известны две группы моделей:

- зависимая и объясняемая переменная являются качественными. Модели, которые используют таблицы кросс табуляции или сопряженности называются моделями качественных переменных. Числовыми характеристиками таблиц сопряженности являются количество объектов, попавших в соответствующие классы. Известны критерии связи между качественными переменными;

• зависимая переменная является количественной, а некоторые объясняющие факторы являются качественными или фиктивными. Такие модели называются моделями с переменной структурой.

Фиктивной переменной в первом примере может быть фактор изменения налога на прибыль, который принимает значения 1 в предыдущем году и 0 в текущем году.

Фиктивной переменной во втором примере может быть фактор культуры семьи, который принимает значение 1 для скромной семьи и 0 для честолюбивой семьи.

В практике эконометрики часто используют фиктивную переменную для объяснения сильно выделяющихся значений зависимой переменной. При этом фиктивная переменная должна учитывать влияние тех факторов, которые не вошли в модель. Исследователь может самостоятельно проставлять значения фиктивной переменной из ограниченного диапазона. Наибольшему значению остатков должно соответствовать наибольшее значение фиктивной переменной. Для положительных остатков ставится положительное значение фиктивной переменной, для отрицательных значений остатков — отрицательные значения фиктивной переменной.

Можно на прогнозный период задавать ожидаемое значение фиктивной переменной. При этом точность прогнозирования резко возрастает.

В практике эконометрики только с помощью фиктивной переменной удается построить достоверную модель, в которой зависимая переменная содержит сильные выбросы.

Расчет коэффициентов модели с фиктивными переменными можно производить с помощью метода наименьших квадратов.

Основную сложность работы с фиктивной переменной составляет экономическая интерпретация численных значений фиктивной переменной.

Следует предупредить об одной ошибке — если сумма нескольких фиктивных переменных для каждых измерений имеет одно и то же значение, то фиктивные переменные и переменная, учитывающая свободный коэффициент, являются мультиколлинеарными и рассчитать коэффициенты модели методом наименьших квадратов не удается [1].

Модели связи между двумя качественными переменными

Количественные переменные имеют числовые значения. Числовые значения количественной переменной определяются с помощью шкалы измерения, имеющей цену деления. Цена деления отражает точность измерения.

Качественные переменные не имеют числового значения. Шкалой измерения качественной переменной служит перечень классов, в которых может находиться объект.

Если экономический объект характеризуется качественной переменной, то значением качественной переменной будет служить номер класса, к которому принадлежит объект.

Каждый экономический объект может характеризоваться несколькими качественными переменными. Если в каждой переменной расположить классы в порядке убывания их ранга, то можно утверждать о положительной или отрицательной связи между качественными переменными.

Если классы в качественных переменных не удается както систематизировать, то можно проверить только независимость переменных.

Объектами изучения взаимосвязи качественных переменных служат частоты или количество объектов, содержащихся в классах качественных переменных. В некоторых моделях объектом изучения служат вероятности попадания объекта в заданный класс. В настоящее время интенсивно развиваются лог-линейные модели, где объектами изучения служат натуральные логарифмы вероятности попадания объекта в заданный класс.

Таблицы, в которых собраны данные объектов, характеризующиеся двумя и более качественными переменными, носят названия: таблицы сопряженности, кросстабуляции, перекрестных данных.

Таблицы сопряженности активно изучаются с 1960 г. Значительный вклад в изучение взаимосвязи качественных переменных внесли ученые Гудман, Юл, Кендалл и др.

Проиллюстрируем методы анализа качественных переменных на конкретном примере.

Задача.

По условной области была взята выборка, состоящая из 18 предприятий, каждое предприятие характеризуется двумя качественными переменными:

(A) — состояние приватизации;

(B) — состояние рентабельности.

Переменная A имеет два класса:

(A_1) — приватизированные и (A_2) — неприватизированные.

Переменная B тоже имеет два класса:

(B_1) — рентабельные и (B_2) — нерентабельные.

Количество предприятий f_{ij}, обладающих свойствами A_iB_j, представлено в таблице сопряженности (табл. 7.1)

Таблица 7.1

Перекрестные данные двух качественных переменных

Состояние приватизации (A_i) $i = 1, m$	Состояние рентабельности (B_j) $j = 1, k$		Всего	Средние значения $\bar{A}_i$
	рентабельные (B_1)	нерентабельные (B_2)		
Приватизированные (A_1)	$f_{11} = 8$	$f_{12} = 4$	$f_{10} = f_{11} + f_{12} = 12$	$\bar{A}_1 = 6$
Неприватизированные (A_2)	$f_{21} = 2$	$f_{22} = 4$	$f_{20} = f_{21} + f_{22} = 6$	$\bar{A}_2 = 3$
Всего	$f_{01} = f_{11} + f_{21} = 10$	$f_{02} = f_{12} + f_{22} = 8$	$f_{00} = f_{10} + f_{20} = f_{01} + +f_{02} = 18$	
Средние значения $\bar{B}_j$	$\bar{B}_1 = 5$	$\bar{B}_1 = 4$		$\bar{f}_1 = 4,5$

Анализ качественных переменных основывается на двух простых утверждениях:

• изменения частот объектов можно объяснить отдельным влиянием переменных, а также их взаимовлиянием;

• если доля приватизированных предприятий среди прибыльных равна доле приватизированных предприятий среди убыточных предприятий, то приватизация не оказывает влияния на рентабельность.

Необходимо решить следующие задачи.

Задача 1.

Найти долю приватизированных предприятий среди рентабельных и долю приватизированных среди нерентабельных предприятий.

Доля приватизированных предприятий среди рентабельных составляет:

$$d_1 = \frac{f_{11}}{f_{01}} = \frac{8}{10} = 0{,}8.$$

Доля приватизированных среди нерентабельных предприятий составляет:

$$d_2 = \frac{f_{12}}{f_{02}} = \frac{4}{8} = 0{,}5.$$

Задача 2.

Найти вероятность рентабельности и нерентабельности приватизированного предприятия.

Вероятность того, что приватизированное предприятие будет рентабельным, составляет:

$$p_1 = \frac{d_1}{d_1 + d_2} = \frac{0{,}8}{1{,}3} = 0{,}615385.$$

Вероятность того, что приватизированное предприятие будет нерентабельным, составляет:

$$p_2 = \frac{d_2}{d_1 + d_2} = \frac{0{,}5}{1{,}3} = 0{,}384615.$$

Задача 3.

Определить, в каком случае приватизация оказывает положительное влияние на рентабельность.

Если вероятность рентабельности приватизированного предприятия больше вероятности его нерентабельности, то приватизация оказывает положительное влияние на рентабельность.

$P_1 > P_2$ или $d_1 > d_2$ или $f_{11}/f_{01} > f_{12}/f_{02}$;

$K = P_1/P_2 = 1{,}6$.

Вывод. Если предприятие приватизировано, то вероятность его рентабельности в 1,6 раза больше вероятности нерентабельности. Приватизация оказывает положительное влияние на рентабельность.

Задача 4.

Вывести соотношения, при которых связь между А и В отсутствует. Если между переменными А и В отсутствует связь, то должны соблюдаться соотношения:

$f_{11}/f_{01} = f_{12}/f_{02} = f_{10}/f_{00}$ — доля приватизированных предприятий среди рентабельных равна доле приватизированных предприятий среди нерентабельных и равна доле приватизированных предприятий среди всех предприятий,

$f_{21}/f_{01} = f_{22}/f_{20} = f_{20}/f_{00}$;

$f_{11}/f_{10} = f_{21}/f_{20} = f_{01}/f_{00}$;

$f_{12}/f_{10} = f_{22}/f_{20} = f_{02}/f_{00}$.

Задача 5.

Определить условия, при которых связь между А и В будет положительной, отрицательной или переменные не будут связаны.

Первое соотношение можно представить в виде равенства:

$f_{11} = f_{01}f_{10}/f_{00}$.

Если $f_{11} > f_{01}f_{10}/f_{00}$, то переменные А и В связаны положительно.

Если $f_{11} < f_{01}f_{10}/f_{00}$, то переменные А и В связаны отрицательно.

Если $f_{11} = f_{01}f_{10}/f_{00}$, то переменные А и В не связаны между собой.

$f_{11} = 8 > f_{01}f_{10}/f_{00} = 6{,}66666.$

Вывод: переменные А и В связаны между собой положительно[1].

Задача 6.

Рассчитать коэффициент корреляции между А и В при двух и более градациях переменных А и В.

Коэффициент корреляции между качественными переменными А и В рассчитывают по критерию Юла:

$$Q = \frac{f_{11}f_{22} - f_{12}f_{21}}{f_{11}f_{22} + f_{12}f_{21}}.$$

Достоверность критерия Юла проверяют по критерию Стьюдента.

Проверим нулевую гипотезу H_o: "Q= 0" или "между переменными А и В нет связи". Если фактический критерий Стьюдента

$$t_\phi = \frac{Q}{m_Q}$$

больше табличного $t_т$ ($\alpha = 0{,}05$; $m = n\ ,\ 1$), то нулевая гипотеза отвергается.

m — число степеней свободы;

n — количество всех объектов;

m_Q — ошибка Q.

$$m_Q = \frac{1-Q^2}{2}\sqrt{\frac{1}{f_{11}} + \frac{1}{f_{12}} + \frac{1}{f_{21}} + \frac{1}{f_{22}}}.$$

Произведем расчеты для нашего примера:

$$Q = \frac{f_{11}f_{22} - f_{12}f_{21}}{f_{11}f_{22} + f_{12}f_{21}} = 0{,}6;$$

$$m_Q = \frac{1-Q^2}{2}\sqrt{\frac{1}{f_{11}} + \frac{1}{f_{12}} + \frac{1}{f_{21}} + \frac{1}{f_{22}}} = 0{,}339411.$$

---

[1] *Кендалл М., Стьюард А.* Статистические выводы и связи. — М.: Наука, 1973. С. 721.

Так как фактическое значение критерия Стьюдента

$$t = \frac{Q}{m_q} = 1{,}767767$$

меньше табличного $t_{\alpha/2}(\alpha = 0{,}05,\ m = n - 1 = 18 - 1 = 17) = 2{,}11$, то нулевая гипотеза H_0: "$Q = 0$" принимается, и утверждается, что обнаруженная положительная связь между А и В является статистически недостоверной.

Гудмен и Краскал распространили критерий Юла на случай, когда количество классов у каждой переменной больше двух, и предложили критерий, который получил их имя — критерий Гудмана-Краскала.

$$G = \frac{S - D}{S + D},$$

где S — сумма произведений частоты каждой ячейки на сумму частот ячеек, лежащих ниже и правее выделенной ячейки.

D — сумма произведений частоты каждой ячейки на сумму частот ячеек, лежащих ниже и левее выделенной ячейки.

Напомним, что критерии G и Q так же, как коэффициент корреляции, изменяются от -1 до +1, и знак коэффициента имеет экономическую трактовку, если классы переменных расположены по степени уменьшения их ранга или качества объекта.

Если не удается упорядочить классы переменных, то знаки критериев Q и G не имеют экономической трактовки.

Задача 7.

Проверить независимость А и В на уровне значимости a=0,05.

Если градации переменных А и В не упорядочены, то мерой связи между А и В может быть критерий χ^2.

Проверим нулевую гипотезу H_0: "Переменные А и В не взаимосвязаны между собой".

Если $\chi^2_ф > \chi^2_т$, то нулевая гипотеза опровергается и утверждается, что между А и В существует взаимосвязь.

$$\chi^2_\phi = \sum \frac{(f_\phi - f_\tau)^2}{f_\tau},$$

где f_ϕ — фактические частоты;

f_τ — теоретические частоты, при которых связь между А и В отсутствует.

Используя соотношения частот, при которых связь между А и В отсутствует, можно получить расчетные значения f_τ.

$$f_{\tau 11} = \frac{f_{01} f_{10}}{f_{00}} = 6{,}666667;$$

$$f_{\tau 12} = \frac{f_{02} f_{10}}{f_{00}} = 5{,}333333;$$

$$f_{\tau 21} = \frac{f_{01} f_{20}}{f_{00}} = 3{,}333333;$$

$$f_{\tau 22} = \frac{f_{02} f_{20}}{f_{00}} = 2{,}666667;$$

$$\chi^2_\phi = \frac{(f_{11} - f_{\tau 11})^2}{f_{\tau 11}} + \frac{(f_{12} - f_{\tau 12})^2}{f_{\tau 12}} + \frac{(f_{21} - f_{\tau 21})^2}{f_{\tau 21}} + \frac{(f_{22} - f_{\tau 22})^2}{f_{\tau 22}} = 1{,}8.$$

Фактические значения критерия χ^2_ϕ можно посчитать по другой формуле:

$$\chi^2_\phi = \frac{f_{00}(f_{11} f_{22} - f_{12} f_{21})^2}{f_{10} f_{20} f_{01} f_{02}}.$$

χ^2_τ — табличное значение (ищется по таблицам);

χ^2_τ ($\alpha = 0{,}05$; $m = (p - 1)(k - 1)) = 2{,}71$;

p — число градаций переменной А;

k — число градаций переменной В.

Так как $\chi^2_\phi = 1{,}8$ меньше $\chi^2_\tau = 2{,}71$, то нулевая гипотеза H_0: “Переменные А и В не взаимосвязаны между собой” принимается.

Вывод. Связь между переменными A и B статистически не доказана.

Задача 8.

Разложить отклонение частоты f_{ij} от $\bar{f}$ на компоненты, обусловленные влиянием A, B и взаимовлиянием AB.

Представим фактические значения частот f_{ij} в виде линейной модели

$$f_{ij} = \bar{f} + a_i + b_j + ab_{ij} \text{ или}$$

$$f_{ij} - \bar{f} = a_i + b_j + ab_{ij},$$

где коэффициенты a_i, b_j и ab_{ij} численно равны тому влиянию, которое оказывают переменные A, B и их взаимовлияние AB на величину отклонения фактической частоты от средней частоты.

Коэффициенты a_i, b_j, ab_{ij} рассчитываются по формулам дисперсионного анализа[1].

$$a_i = \bar{A}_i - \bar{f};$$

$$b_j = \bar{B}_j - \bar{f};$$

$$ab_{ij} = f_{ij} - (\bar{f} + a_i + b_j) = f_{ij} - \bar{A}_i - \bar{B}_j + \bar{f}.$$

Определим коэффициенты a_i, b_j, ab_{ij} для каждой частоты f_{ij}.

$$f_{11} = 8 = \bar{f} + a_1 + b_1 + ab_{11} =$$

$$= \bar{f} + (\bar{A}_1 - \bar{f}) + (\bar{B}_j - \bar{f}) + (f_{11} - \bar{A}_1 - \bar{B}_j + \bar{f}) = +4{,}5 + 1{,}5 + 0{,}5 + 1{,}5.$$

Прокомментируем полученные расчеты. Отклонение $f_{11} - \bar{f} = 3{,}5$ вызвано влиянием переменной A на величину 1,5, переменной B на величину 0,5, взаимовлиянием переменных AB на величину 1,5.

Все переменные оказали положительное воздействие на величину f_{11}. Аналогично производим расчеты для всех оставшихся частот.

---

[1] *Длин А. М.* Факторный анализ в производстве. — М.: Статистика, 1975. С. 46.

$$f_{12} = 4 = \bar{f} + a_1 + b_2 + ab_{12} =$$
$$= \bar{f} + (\bar{A}_1 - \bar{f}) + (\bar{B}_2 - \bar{f}) + (f_{12} - \bar{A}_1 - \bar{B}_2 + \bar{f}) = +4{,}5 + 1{,}5 - 0{,}5 - 1{,}5.$$

$$f_{21} = 2 = \bar{f} + a_2 + b_1 + ab_{21} =$$
$$= \bar{f} + (\bar{A}_2 - \bar{f}) + (\bar{B}_1 - \bar{f}) + (f_{21} - \bar{A}_2 - \bar{B}_1 + \bar{f}) = +4{,}5 - 1{,}5 + 0{,}5 - 1{,}5.$$

$$f_{22} = 4 = \bar{f} + a_2 + b_2 + ab_{22} =$$
$$= \bar{f} + (\bar{A}_2 - \bar{f}) + (\bar{B}_2 - \bar{f}) + (f_{22} - \bar{A}_2 - \bar{B}_2 + \bar{f}) = +4{,}5 - 1{,}5 - 0{,}5 + 1{,}5.$$

Представим результат расчетов в виде табл. 7.2.

Таблица 7.2

Значения коэффициентов линейной модели $f_{ij} = \bar{f} + a_i + b_j + ab_{ij}$

Частоты f_{ij}	Коэффициенты			
	$\bar{f}$	a_i	b_j	ab_{ij}
f_{11}	8	1,5	0,5	1,5
f_{12}	4	1,5	-0,5	-1,5
f_{21}	2	-1,5	0,5	-1,5
f_{22}	4	-1,5	-0,5	1,5

Линейная модель частот позволяет получить модели расчетных значений частот, учитывающие раздельное и совместное влияние переменных.

Модель 1. $f_{pij} = \bar{f}$ не учитывает влияния переменных A и B, а также их взаимовлияние AB.

Модель 2. $f_{pij} = \bar{f} + a_i$ — учитывает влияние только переменной A.

Модель 3. $f_{pij} = \bar{f} + b_j$ — учитывает влияние только переменной B.

Модель 4. $f_{pij} = \bar{f} + ab_{ij}$ — учитывает взаимовлияние переменных AB.

Модель 5. $f_{pij} = \bar{f} + a_i + b_j$ — учитывает раздельное, невзаимосвязанное влияние переменных A и B.

Модель 6. $f_{pij} = \bar{f} + a_i + b_j + ab_{ij}$ имеет название насыщенной модели и точно воспроизводит фактические частоты.

Расчет теоретических частот, вызванных влиянием различных переменных и их взаимовлиянием, представлен в табл. 7.3.

Таблица 7.3

Расчет теоретических частот, определенных по моделям

Частоты f_{pij}	Модель 1	Модель 2	Модель 3	Модель 4	Модель 5	Модель 6
	$f_{pij} = \bar{f}$	$f_{pij} = \bar{f} + a_i$	$f_{pij} = \bar{f} + b_j$	$f_{pij} = \bar{f} + ab_{ij}$	$f_{pij} = \bar{f} + a_i + b_j$	$f_{pij} = \bar{f} + a_i + b_j + ab_{ij}$
f_{p11}	4,5	6	5	6	6,5	8
f_{p12}	4,5	6	4	3	5,5	4
f_{p21}	4,5	3	5	3	3,5	2
f_{p22}	4,5	3	4	6	2,5	4

Примечание. Модель 5 предполагает, что переменные А и В невзаимосвязаны между собой. Для этого случая мы уже получали теоретические частоты, которые отличаются от теоретических частот данной модели. Расхождения небольшие, но природа этой ошибки пока неизвестна.

Задача 9.

Рассчитать коэффициенты корреляции влияния А, В и АВ на изменения частот.

Разложение частот на компоненты позволяет рассчитать коэффициенты корреляции влияния каждого компонента на частоты, а также можно определить долю влияния каждого компонента на изменение частот.

Составим базу данных для расчета коэффициентов корреляции между частотами и их компонентами (табл. 7.4).

Таблица 7.4

База данных для расчета коэффициентов корреляции

Частота f_{ij}	Компоненты частот f_{ij}		
	a_i	b_i	ab_{ij}
8	1,5	0,5	1,5
4	1,5	-0,5	-1,5
2	-1,5	0,5	-1,5
4	-1,5	-0,5	1,5
$\bar{f} = 4,5$			

Коэффициенты корреляции между частотами f_{ij} и компонентами a_i, b_j и a_{ij} можно вычислить по следующим формулам:

$$r(f_{ij}, a_i) = \frac{\sum (f_{ij} - \bar{f}) \cdot a_i}{\sqrt{\sum (f_{ij} - \bar{f})^2 \sum a_i^2}} = 0{,}688247;$$

$$r(f_{ij}, b_j) = \frac{\sum (f_{ij} - \bar{f}) \cdot b_j}{\sqrt{\sum (f_{ij} - \bar{f})^2 \sum b_j^2}} = 0{,}229416;$$

$$r(f_{ij}, ab_{ij}) = \frac{\sum (f_{ij} - \bar{f})(ab_{ij})}{\sqrt{\sum (f_{ij} - \bar{f})^2 \sum (ab_{ij})^2}} = 0{,}688247.$$

Расчеты коэффициентов корреляции произведем с помощью табл. 7.5.

Таблица 7.5

Расчет коэффициентов корреляции

f_{ij}	a_i	b_j	ab_{ij}	$k_1 = f_{ij} - \bar{f}$	k_1^2
1	2	3	4	5	6
8	1,5	0,5	1,5	3,5	12,25
4	1,5	-0,5	-1,5	-0,5	0,25
2	-1,5	0,5	-1,5	-2,5	6,25
4	-1,5	-0,5	1,5	-0,5	0,25
$\bar{f} = 4{,}5$	Сумма				19

Продолжение табл. 7.5

a_{i2}	b_{j2}	ab_{ij}^2	k_1a_i	k_1b_j	k_1ab_{ij}
7	8	9	10	11	12
2,25	0,25	2,25	5,25	1,75	5,25
2,25	0,25	2,25	-0,75	0,25	0,75
2,25	0,25	2,25	3,75	-1,25	3,75
2,25	0,25	2,25	0,75	0,25	-0,75
9	1	9	9	1	9

Задача 10.

Определить долю влияния A, B и AB на изменение частот.

Доля влияния каждой переменной A, B и их взаимовлияние AB численно равны коэффициентам детерминации.

$(rf_{ij},a_i)^2 = 0{,}473684$ — доля влияния переменной A;

$(rf_{ij}.b_j)^2 = 0{,}052632$ — доля влияния переменной B;

$(rf_{ij},ab_{ij})^2 = 0{,}473684$ — доля взаимовлияния переменных AB[1].

Вывод.

Доля влияния переменной A и взаимовлияния переменных AB значительно превосходит долю влияния переменной B.

Примечание. Возникает естественный вопрос: можно ли отнести анализ качественных переменных к эконометрике?

Аргументы против:

• предложенные методы относятся к непараметрическим методам, т. е. нет ограничений на закон распределения анализируемых случайных величин. Непараметрические методы не являются традиционными для эконометрики;

• числовыми данными для расчетов служат не значения классов переменных, а количество объектов с изучаемыми качественными переменными, попавших в заданный класс;

• нет точечного и интервального прогноза для изучаемых величин.

Аргументы за:

• корреляционный анализ входит в состав эконометрики. Предлагаемые методы изучают корреляционную связь между качественными переменными, тем самым расширяют сферу применимости корреляционного анализа и позволяют системно изучать все виды переменных, как количественных, так и качественных;

• корреляционный анализ качественных переменных базируется на дисперсионном анализе. Дисперсионный анализ входит в состав эконометрики;

• анализ качественных переменных позволяет прогнозировать попадание зависимой переменной в определенный класс с определенной вероятностью, а чем это в принципе отличается от прогнозирования в регрессионном анализе?

• простота расчетов и ясность в интерпретации.

---

[1] *Аптон Г.* Анализ таблиц сопряженности. — М.: Финансы и статистика, 1982.

Вывод — анализ качественных переменных расширяет сферу корреляционного анализа, позволяет получать прогнозные возможные значения с определенной вероятностью, однако числовой базой для расчетов являются значения частот попадания объектов в заданный класс.

7.2. Нелинейные регрессионные модели и линеаризация

Метод наименьших квадратов применим к линейным (относительно коэффициентов) аддитивным регрессионным уравнениям следующего вида:

$$У = a_0 + a_1X_1 + a_2X_2 + \dots a_кX_к + e,$$

где коэффициенты регрессии являются линейными, переменные могут быть в любой степени и могут иметь любые математические преобразования, кроме одного ограничения — в степени переменной не должно быть определяемых коэффициентов.

Коэффициенты регрессии являются линейными, если выполняются два условия:

а) коэффициенты находятся в первой степени,

б) коэффициенты не являются степенью по отношению к другим коэффициентам или переменной.

Если к регрессионному уравнению не применим метод наименьших квадратов, то есть три пути решения данной проблемы:

1) предложить специально разработанные формулы расчетов коэффициентов уравнения регрессии;

2) использовать приближенные итеративные методы расчетов коэффициентов;

3) преобразовать его к линейному аддитивному виду.

Первый путь требует большого времени и изобретательности. Особенно много хлопот доставляет логистическая функция относительно оценки предела, к которому она стремится.

Второй путь можно реализовать с помощью пакетов прикладных программ. Известны программа “Эврика”[1], а также

[1] *Очков В. Ф., Пухначев Ю. В.* Уроки для пользователей IBM PC. — М.: Финансы и статистика, 1992. С. 87–130.

программа “Поиск решения”, входящая в состав Excel, которые с помощью итеративных методов могут рассчитать коэффициенты любых функций.

Третий путь реализуется с помощью использования несколько способов приведения функции к линейному виду.

7.3. Нелинейные зависимости, подчиняющиеся непосредственной линеаризации

Известно много функций, которые можно привести к линейному аддитивному виду с помощью методов линеаризации.

Линеаризация — процесс преобразования функции к линейному аддитивную виду.

Известны два метода линеаризации: логарифмирование функции и замена преобразованной переменной. Оба эти метода представлены в табл. 7.6.

Таблица 7.6

Список функций и их линеаризация

№ п/п	Вид функции	Название	Преобразование	Вид линейной функции
1	$У_p=a_0+a_1X$	Линейная	-	$У_p=a_0+a_1X$
2	$У_p=a_0+a_1X+a_2X^2$	Парабола	$Z=X^2$	$У_p=a_0+a_1X+a_2Z$
3	$У_p=a_0+a_1/X$	Гипербола 1	$Z=1/X$	$У_p=a_0+a_1Z$
4	$У_p=a_0+a_1\ln X$	Логарифмическая 1	$Z=\ln X$	$У_p=a_0+a_1Z$
5	$У_p=1/(a_0+a_1\ln X)$	Обратная логарифмическая 2	$Z_1=1/У$ $Z_2=\ln X$	$Z_1=a_0+a_1Z_2$
6	$У_p=a_0a_1^X$	Экспоненциальная 1	$Z_1=\ln У$, $Z_2=\ln a_0$, $Z_3=\ln a_1$	$Z_1=Z_2+Z_3X$
7	$У_p=a_0e^{a_1x}$	Экспоненциальная 2	$Z_1=\ln У$ $Z_2=\ln a_0$	$Z_1=Z_2+a_1X$
8	$У_p=e^{(a_0+a_2X)}$	Экспоненциальная 3	$Z_1=\ln У$	$Z_1=a_0+a_1X$
9	$У_p=a_0X^{a_1}$	Экспоненциальная 4. Кривая Энгеля	$Z_1=\ln У$ $Z_2=\ln X$	$Z_1=a_0+a_1Z_2$

№ п/п	Вид функции	Название	Преобразование	Вид линейной функции
10	$У_p = e^{(a_0 + a_2/X)}$	Экспоненциальная 5. S образная	$Z_1 = \ln У$ $Z_2 = 1/X$	$Z_1 = a_0 + a_1 Z_2$
11	$У_p = a_0 + a_1 e^X$	Экспоненциальная 6	$Z = e^X$	$У_p = a_0 + a_1 Z$
12	$У_p = a_0 e^{a_1/X}$	Экспоненциальная 7	$Z_1 = \ln У$ $Z_2 = \ln a_0$ $Z_3 = 1/X$	$Z_1 = Z_2 + a_1 Z_3$
13	$У_p = 1/(a_0 + a_1 e^{-X})$	Обратная экспоненциальная 8	$Z_1 = 1/У$ $Z_2 = e^{-X}$	$Z_1 = a_0 + a_1 Z_2$
14	$У_p = a_0 X^{a_1 X}$	Степенная 1	$Z_1 = \ln У$ $Z_2 = \ln a_0$ $Z_3 = X \ln X$	$Z_1 = Z_2 + a_1 Z_3$
15	$У_p = a_0 + a_1 X^{1/2}$	Степенная 2	$Z = X^{1/2}$	$У_p = a_0 + a_1 Z$
16	$У_p = 1/(a_0 + a_1 X)$	Гипербола 2	$Z = 1/У$	$Z = a_0 + a_1 X$
17	$У_p = X/(a_0 + a_1 X)$	Гипербола 3	$Z = X/У$	$Z = a_0 + a_1 X$
18	$У_p = a_0 X_1^{a_1} X_2^{a_2}$	Кобба-Дугласа	$Z_1 = \ln У$ $Z_2 = \ln a_0$ $Z_3 = \ln X_1$ $Z_4 = \ln X_2$	$Z_1 = Z_2 + a_1 Z_3 + a_2 Z_4$

Продемонстрируем технику расчетов коэффициентов степенной функции следующего вида: $У_p = a_0 X^{a_1 X}$ с помощью функции "Линейн".

Шаг 1. Логарифмирование степенной функции позволяет получить выражение

$$\ln y = \ln a_0 + a_1 X \ln X$$

Шаг 2. Произведем замену переменных и получим линейную модель с новыми переменными

$$Z_1 = Z_2 + a_1 Z_3,$$

где $Z_1 = \ln У$, $Z_2 = \ln a_0$, $Z_3 = X \ln X$.

Шаг 3. Создадим базу данных значений X и У, в среде Excel введем значения X, У в таблицу, добавим к исходной базе данных две колонки $\ln У$ и $X \ln X$, в которые введем преобразованные значения.

Шаг 4. Выполним функцию “Линейн”, в которой зависимой переменной будут значения колонки lnУ, объясняемой переменной — значения колонки XlnX. Функция “Линейн” позволяет получить коэффициенты Z_2, a_1 модели

$Z_1 = Z_2 + a_1 Z_3$.

Шаг 5. Вычисление коэффициентов a_0 и a_1. Коэффициент a_1 уже посчитан.

Определяем коэффициент a_0 из соотношения $Z_2 = \ln a_0$ при условии, что коэффициент Z_2 известен. Тогда $a_0 = e^{Z_2}$.

Шаг 6. Вычисление характеристик модели $У_p = a_0 X^{a_1 X}$ при условии, что коэффициенты a_0 и a_1 известны. Ошибку модели, критерий Фишера точечный и интервальный прогноз, доверительный интервал регрессии можно посчитать по формулам, предназначенным для линейной функции. Определение ошибок коэффициентов a_0 и a_1 требует научных исследований.

Вопросы для самоконтроля

1. Приведите три класса моделей, в которых может изменяться ее структура.
2. Приведите отличие функциональной от структурной неоднородности.
3. Дайте определение фиктивной переменной.
4. Назовите область использования фиктивных переменных.
5. Назовите способ повышения точности прогноза с использованием фиктивной переменной.
6. Раскройте трудности в экономической интерпретации фиктивной переменной.
7. Назовите ограничение на использование фиктивных переменных.
8. Назовите три способа расчета коэффициентов регрессионной модели при условии, что нельзя применить метод наименьших квадратов.
9. Назовите сущность линеаризации нелинейной функции.
10. Приведите шаги расчета коэффициентов степенной функции.

Глава 8
ОПРЕДЕЛЕНИЕ ВРЕМЕННОГО РЯДА

Формирование пространственных выборок и временных рядов принципиально отличаются между собой. Пространственная выборка представляется как совокупность реализаций случайной величины. Временной ряд является одной из возможных реализаций случайного процесса. Члены временного ряда не являются статистически независимыми и одинаково распределенными. Следовательно, методы анализа пространственных выборок и временных рядов должны отличаться между собой. Это отличие наблюдается в терминологии [1], [20].

Следует повторить материал параграфов 2,3, 2,4 и 3,2, где рассматривались свойства экономических временных рядов, классификация переменных и генерация эконометрических моделей.

8.1. Определение, структура, основные свойства и цели анализа временных рядов

Временной ряд — временная последовательность значений переменной У.

Временной ряд переменной обозначается как $У_t$,
где $t = \; = 1, ..., n$ — порядковый номер времени;

n — число уровней (длина, размер, количество значений, но не принято говорить объем выборки) временного ряда.

Численное значение временного ряда $У_t$ в момент времени, например $t = 2$, обозначается как $У_2$ которое является отдельным наблюдением временного ряда, называемое уровнем ряда.

Теоретической базой временных рядов является наука о случайных процессах.

Структура общей модели временного ряда.

Обычно временной ряд представляют в виде аддитивной модели, имеющей следующие компоненты:

$$У_t = f_{1t} + f_{2t} + f_{3t} + \varepsilon_{t,}$$

где f_{1t} — тренд, плавно изменяющаяся компонента, которая отражает влияние факторов, формирующих долговременную, как правило, монотонную, общую тенденцию в изменениях признака временного ряда $У_t$;

f_{2t} — сезонная компонента, которая отражает повторяемость экономических процессов в течение года;

f_{3t} — циклическая компонента, которая отражает повторяемость экономических процессов в течение длительных периодов, например, влияние волн научно-технической революции, демографических спадов, циклов солнечной активности и т. д.;

ε_t — случайное возмущение.

Известны два направления в изучении компонент временного ряда.

Первое направление (аналитическое) предполагает представление этих компонент в явном аналитическом виде с их экономической интерпретацией, как правило, с использованием текущих значений переменных.

Второе направление (алгоритмическое) связано с разработкой моделей, которые воспроизводили бы любые тенденции с использованием как прошлых, так и будущих значений переменных на основе их внутренней структуры, иногда без экономической интерпретации.

Первое направление.

Компонента f_{1t} (тренд) изучена очень хорошо. Имеется программа “TableCurve2D”, которая позволяет аппроксимировать тренд временного ряда с использованием более двух тысяч различных аналитических математических функций.

Компонента f_{2t} (сезонная) изучена достаточно хорошо. Известно много методов аналитического представления сезонной волны, в том числе с использованием различных индексов.

Компонента f_{3t} (циклическая) изучена слабо и требует решения трех проблем: 1) объяснения механизма появления периодических колебаний; 2) аналитического представления периодических колебаний; 3) экономической интерпретации периодических колебаний.

Причины периодических колебаний условно подразделяются на астрофизические и экономические. Астрофизические факторы, как правило, являются неуправляемыми, и их влияние носит глобальный характер. Известны периоды: суточные и годовые, связанные с вращением Земли вокруг своей оси и Солнца; месячные циклы, связанные с вращением Луны; 11-летние циклы солнечной активности, определяемые по числам Вольфа; долгосрочные периоды, связанные с ледниковыми периодами, смены полюсов Земли, прецессии оси вращения Земли [12].

Изучение экономических причин возникновения периодических колебаний требует специальных исследований. В эконометрической литературе таких исследований явно недостаточно. Известны пятилетние циклы, связанные с переизбранием президентов, десятилетние периоды смены устаревшего оборудования; 50-летние коньюнктурные циклы Кондратьева [24, с. 341–397] и др.

Аналитическое представление экономического периодического колебания в эконометрической, как в учебной, так и в научной литературе, чаще всего связывают с разложением его в ряд Фурье, которое имеется в большинстве статистических пакетов прикладных программ и в Excel (“Анализ данных”).

Разложение в ряд Фурье справедливо для воспроизведения физически существующих стоячих волн, которые возникают в замкнутых системах, например, струна; мембрана; колокол; электрический колебательный контур. Стоячие волны, или гармоники, имеют периоды, кратные основному периоду колебания.

Применение разложения временного экономического ряда в ряд Фурье приводит к появлению следующих проблем:

- экономический временной ряд за определенный период не является замкнутой системой, и там не могут возникнуть стоячие волны, так как нет отражения волны от концов среды ее существования и прошлое не зависит от текущего времени;
- выделенные гармоники не имеют экономического содержания;
- при изменении длины временного ряда изменяются длины периодов всех ее гармонических составляющих;
- если временной ряд достаточно гладкий, то разложение его в ряд Фурье приводит к точному его воспроизведению с ошибкой, равной нулю. Тогда прогнозирование сведется к точному воспроизведению исходного ряда на следующий период, что является абсурдным для экономических временных рядов при наличии случайной составляющей;
- может случиться, что гармоника будет равна действительно существующей периодической составляющей экономического временного ряда, но нет гарантии обнаружения всех периодических составляющих временного ряда.

Одним их методов аналитического представления периодических составляющих временного ряда является построение ее периодограммы и фильтрация существующих периодических колебаний [11].

Второе направление.

Второе направление изучения компонент временного ряда связано с разработкой моделей, которые воспроизводили бы любые тенденции с помощью простых средств с использованием как прошлых, так и будущих значений всех переменных регрессионной модели: Х, У, ε, δ, иногда без экономической интерпретации. Это направление включает разнообразные модели, которые необходимо классифицировать. О всех видах эконометрических моделей см. в параграфе 3.2.1. Выделим основные классы моделей и их комбинации.

Модели, учитывающие прошлые значения переменных:

• модель распределенных лагов РЛ(p) (distributed lags DL(p)), где p — максимальный лаг.

Модель РЛ(2) имеет следующий вид:

$$У_t = \alpha_0 + \alpha_1 X_t + \alpha_2 X_{t-1} + \alpha_3 X_{t-2} + \varepsilon_t;$$

• авторегрессионная модель АР(p) (autoregressive AR(p)) зависимой переменной,

где p — порядок модели.

Модель АР(2) имеет следующий вид:

$$У_t = \alpha_0 + \alpha_1 У_{t-1} + \alpha_2 У_{t-2} + \varepsilon_t;$$

• модель авторегрессии АР(p) возмущения или автокорреляция АК(p) (autocorrelation AC(p)) возмущения,

где p — порядок модели.

Модель АР(1) возмущения имеет следующий вид:

$$\varepsilon_t = \rho\varepsilon_{t-1} + \delta_t;$$

• модель авторегрессии и распределенных лагов АРРЛ(p, q) (autoregressive distributed lags ADL(p, q)),

где p и q — максимальные лаги зависимой и факторной переменных.

Модель АРРЛ(1, 2) имеет следующий вид:

$$У_t = \alpha_0 + \alpha_1 У_{t-1} + \alpha_2 X_t + \alpha_3 X_{t-1} + \alpha_4 X_{t-2} + \varepsilon_t;$$

• модель сосредоточенного лага СЛ(p) (concentrated lag CL(p)), где p — величина задержки.

Модель СЛ(2) имеет следующий вид:

$$У_t = \alpha_0 + \alpha_1 X_{t-2} + \varepsilon_t,$$

где $У_t$ и X_t переменные содержащие выбросы.

Модель СЛ(2) устанавливает зависимость между выбросами у причины и следствия со сдвигом на две даты;

• модель скользящей средней СС(р) (moving average MA(p)) [1, с. 261], [5, с. 148][1]. В скобках указан порядок модели.

Модель СС(2) имеет следующий вид:

$$\varepsilon_t = \delta_t + \alpha_1\delta_{t-1} + \alpha_2\,\delta_{t-2} + \nu_t,$$

где ε_t — случайные возмущения;

δ_t — белый шум;

• модель авторегрессии и скользящей средней АРСС(р, q) (ARMA(p, q)).

Модель АРСС(2,2) имеет следующий вид:

$$У_t = \alpha_0 + \alpha_1 У_{t-1} + \alpha_2\, У_{t-2} + \delta_t + \alpha_1\delta_{t-1} + \alpha_2\,\delta_{t-2} + \nu_t;$$

• модель авторегрессии-проинтегрированного скользящего среднего АРПСС(р, q, k) (AutoRegressive Integrated Moving Average ARIMA(p, q, k)),
где р, q, k — порядки авторегрессии, скользящей средней и последовательных разниц [1, с. 287]. Модель АРПСС(2, 2, 1) имеет следующий вид

$$z_t = \alpha_1 z_{t-1} + \alpha_2 z_{t-2} + \delta_t + \beta_1\delta_{t-1} + \beta_2\,\delta_{t-2} + \nu_t,$$

где $z_t = У_t - У_{t-1}$ — последовательные разности первого порядка, последовательные разности используются для исключения неслучайной составляющей, имеющий вид алгебраического полинома порядка р, с помощью вычисления последовательных разниц, повторенных р + 1 раз. Если имеется временной ряд $У_1$, $У_2$, ..., $У_n$, то последовательные разности первого порядка обозначаются с помощью $\Delta У_t = У_t - У_{t-1}$, $t = 2, ..., n$. Последовательные разности 2-го порядка — это разности от последовательных разностей первого порядка $\Delta(\Delta У_t) = \Delta У_t - \Delta У_{t-1}$;

• модели адаптивного прогнозирования Брауна, Хольта, Бокса-Дженкинса. Уинтерса, Тейло-Вейджа (будут изучаться позже).

---

[1] Модель скользящей средней не следует путать с методом скользящей средней сглаживания временных рядов.

Модели, учитывающие будущие значения переменных:

- среди моделей, связанных с будущими значениями переменных X, У, можно выделить модели адаптивных ожиданий и частичной корректировки (будут изучаться позже);
- пока не изучены и не получили применения модели с будущими значениями е (случайное возмущение) и δ (белый шум).

Модели, учитывающие прошлые и будущие значения переменных:

- метод скользящей средней МСС(3). Метод скользящей средней использует текущие прошлые и будущие значения зависимой переменной. Метод расчета скользящей средней третьего порядка МСС(3) реализуется по формуле

$У_{cct} = (У_{t-1} + У_t + У_{t+1})/3.$

Особенность этих методов заключается в том, что с помощью простых средств удается воспроизвести нелинейные тенденции. При этом от эконометриста иногда не требуется объяснения причин наличия во временном ряду выявленных тенденций и, в частности, периодических колебаний.

Дискуссия. В период с 1922 по 1991 гг. использование алгоритмических методов было очень актуальным, так как в то время не стимулировалось изучение механизма возникновения длиннопериодических волн по причине идеологической нецелесообразности. Так как считалось, что социалистическая экономика развивается планомерно и поступательно без периодических колебаний. Трагическая судьба Н. Д. Кондратьева не вдохновляла ученых продолжать его исследования. В настоящее время ситуация резко изменилась ввиду перехода России к рыночной экономике и возникшему спаду производства. Возникла острая необходимость объяснения причин спада производства и определения ориентировочного времени выхода на подъем производства и повышения благосостояния населения. Одним из направлений решения поставленной задачи является изучение длиннопериодических экономических циклов. Изучение циклов Кондратьева входит в Государственный общеобразовательный стандарт по дисциплине "Политология".

В зависимости от вида процесса можно выделить следующие временные ряды: экономические, технологические.

Экономические временные ряды отражают экономический процесс деятельности предприятия (организации).

Технологический временной ряд отражает процесс изготовления или испытания технического изделия.

Главное отличие экономического временного ряда от технологического заключается в том, что экономический временной ряд показателя деятельности предприятия нельзя повторить (нельзя предприятие вернуть в ту среду, которая уже прошла и не повторится), а технологический временной ряд можно повторить (например, многократные испытания самолета, изготовление одного вида деталей, сушка и обжиг партий кирпича).

Примечание. Анализ и управление технологическими процессами осуществляется с помощью методов SPC[1] (статистический контроль процессов) и пока не входит в программу эконометрики. Для более подробного ознакомления с методами SPC рекомендуем обратиться к специальной литературе: (ГОСТ Р 50779.40-96 (ИСО 7870-93). Многие методы SPC можно адаптировать к анализу экономических временных рядов. В научной литературе такие попытки уже успешно проводятся.

Экономический временной ряд — временная последовательность значений показателя экономического процесса.

$У_t$ — случайная величина, которая является характеристикой экономического процесса и принимает одно (из совокупности возможных значений) его значение (уровень) в фиксированный момент времени t.

$t_{к+1} - t_к$ — постоянный временной шаг, через который измеряется значение переменной. В эконометрике нет ограничений на величину временного шага, обычно им могут быть сутки, месяц, квартал, год, десятилетия.

Численное значение показателя $У_t$ в момент времени t является следствием воздействия на этот показатель системы управления, а также влияния случайных: обычных и особых причин (по терминологии SPC и системы качества).

Целенаправленные действия системы управления предприятия приводят к появлению в экономическом временном ряду определенной тенденции.

---

[1] Статистическое управление процессами (SPC). — Н. Новгород: АО НИЦ КД. СМЦ Приоритет, 1997.

Случайные обычные и особые причины могут исказить эту тенденцию.

Обычные случайные причины характеризуются тем, что они всегда присутствуют и действуют случайным образом, каждая в отдельности не оказывает существенного влияния на результирующий признак. Для устранения их влияния необходимо совершенствовать систему управления предприятием.

Особые причины (внезапные, разладочные см. [1, с. 202]) возникают эпизодически и вызваны нарушениями в деятельности предприятия. Для устранения действия особых причин необходимы меры оперативного управления.

Временное агрегирование (объединение). В эконометрике часто используется временное агрегирование, когда суточные данные агрегируются в месячные, месячные данные агрегируются в квартальные, квартальные данные агрегируются в годовые и т. д.

Последствия временного агрегирования. Суммирование временного ряда за определенный период устраняет варьирование переменной, которое имеется внутри этого периода.

Например, заведующего магазином "Перекресток" волнует проблема малой загруженности кассиров в определенные интервалы времени суток. Неоднородность загруженности кассиров можно выявить только по данным, определяемым через каждый час. Если агрегировать часовые данные в суточные, то влияние времени суток на загруженность кассиров обнаружить будет невозможно.

Основные свойства экономического временного ряда

1. Текущие состояние экономической системы испытывают влияния прошлых, настоящих и будущих значений переменных этой системы.

2. Для всех явлений в природе между причиной и следствием существует временной лаг (временная задержка).

3. Все временные экономические процессы происходят циклически, которые могут содержать периодические волны: короткие и длинные.

4. “Свежие” значения временного ряда оказывают большее влияние на его прогнозное значение, чем “старые” значения.

5. При построении доверительных интервалов прогноза и уравнения регрессии следует считать более точным не среднее значение временного ряда, а его последнее значение.

6. Возможные реализации экономического процесса в фиксированный момент времени t имеют определенный закон распределения вероятностей и соответствующие статистические характеристики: среднее арифметическое значение, дисперсия, ковариация и их математические ожидания. Если изучать две группы возможных реализаций в моменты времени t и t + к, то можно вычислить автоковариацию для пар случайных величин $У_t$ и $У_{t+к}$, где к = 1, 2,..., n − t;

7. В численных значениях временного ряда не должно быть пропусков. Если имеется пропуск, то он восстанавливается средним значением, обычно из ближайших к нему чисел.

Статистические характеристики временного ряда

Временные ряды могут иметь следующие статистические характеристики:

- среднее арифметическое значение;
- дисперсия;
- автоковариация;
- автокорреляционная функция (автокоррелограмма);
- периодограмма.

Среднее арифметическое значение временного ряда $У_t$ вычисляется по следующей формуле

$$\overline{У} = \frac{\sum У_t}{n}.$$

Дисперсия временного ряда $У_t$ вычисляется по формуле

$$S^2 = \frac{\sum (У_t - \overline{У})^2}{n-1}.$$

Автоковариация k-го порядка временного ряда $У_t$ вычисляется по формуле

$$Cov(У_t, У_{t+k}) = \frac{1}{n-k}\sum_{t=1}^{n-k}(У_t - \overline{У}_1)(У_{t+k} - \overline{У}_2).$$

Автокорреляция k-го порядка временного ряда $У_t$ — коэффициент корреляции $r(У_t, У_{t+k})$, рассчитанный между исходным временным рядом $У_t$ и этим же временным рядом, только сдвинутым вперед на k дат $У_{t+k}$. Автокорреляция показывает степень влияния предыдущих значений временного на их последующие значения с временным сдвигом, равным k датам.

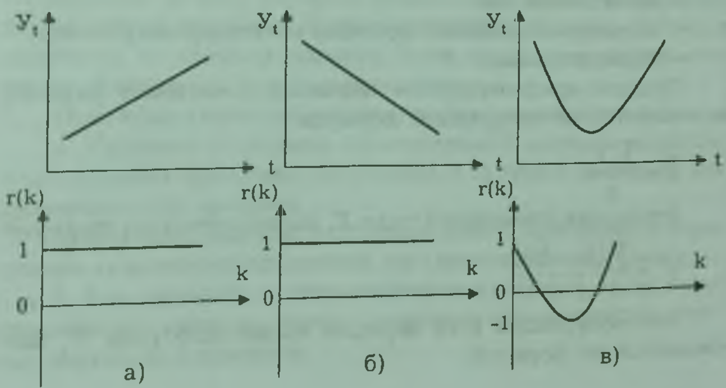
$$r(k) = r(У_t, У_{t+k}) = \frac{\sum_{t=1}^{n}(У_t - \overline{У}_1)(У_{t+k} - \overline{У}_2)}{\sqrt{\sum_{t=1}^{n-k}(У_t - \overline{У}_1)^2 \sum_{t=k}^{n}(У_{t+k} - \overline{У}_2)^2}}.$$

В расчетах автокорреляции k-го порядка должны участвовать только заполненные пары чисел анализируемых рядов. Это значит, что в расчетах не должны участвовать первые k чисел ряда $У_t$ и k последних чисел ряда $У_{t+k}$.

Автокорреляционная функция (автокоррелограмма) — зависимость коэффициентов автокорреляции от величины их порядка k.

Автокорреляция может изменяться от -1 до +1.

График автокорреляционной функции зависит от тенденции, присут-ствующей во временном ряду (рис. 8.1).



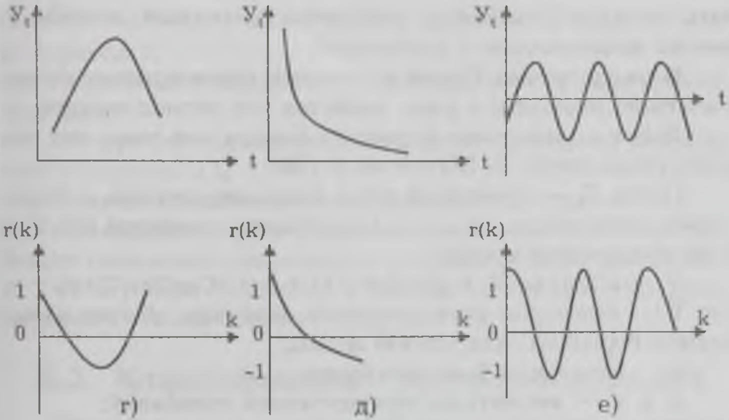


Рис. 8.1. Различные виды тенденций и их автокоррелограммы

Анализ рис. 8.1, а, б показывает, что если временной ряд имеет линейную тенденцию, то автокоррелограмма имеет вид горизонтальной линией с численными значениями, равными единице.

Анализ рис. 8.1, в, г, д, е позволяет утверждать следующее: автокоррелограмма повторяет тенденцию в нормированном виде, которая имеется во временном ряду.

Имеются определенные правила анализа автокоррелограмм для выявления тенденции во временном ряду. Большинство статистических пакетов позволяют рассчитать автокоррелограмму. Однако, как указывает ряд авторов, автокоррелограмма является вещью в себе, и требуется большой опыт по ее анализу, который трудно формализуется. Визуальный анализ временного ряда и изучение периодограммы позволяют часто заменять анализ автокоррелограммы. Более подробно со свойствами автокоррелограмм, частных автокоррелограмм (с указанием 95% интервала, в котором коэффициенты не отличаются от нуля) для различных видов временных трендов (линейные, линейно-аддитивный, сезонные) можно ознакомиться в источнике[1].

Периодограмма. Одной из важных характеристик экономического временного ряда является его периодограмма.

Дадим определение периодограммы в том виде, как оно было предложено А. Шустером в 1898 г.

Пусть $У_t$ — временной ряд с нулевым средним, а t пробегает целые числа от 1 до n. Представим временной ряд $У_t$ в виде следующей модели:

$У_t$ = aCos(2πt/T) + вSin(2πt/T) + e = cCos(2πt/T+φ) + e,
где T — некоторая фиксированная величина, обычно называемая периодом, или длиной волны;

φ — начальная фаза колебания;

а, в, с — амплитуды периодических колебаний;

$$a = \frac{2}{n}\sum_{t=1}^{n} У_t Cos(2\pi t/T);$$

$$в = \frac{2}{n}\sum_{t=1}^{n} У_t Sin(2\pi t/T);$$

$$c = \sqrt{a^2 + в^2};$$

$c^2(T) = a^2(T) + в^2(T).$

Периодограмма — график зависимости $c^2(T)$ от длины волны T, которая показывает зависимость квадрата амплитуды косинусоидального колебания от его периода для временного ряда с нулевым средним значением[2].

Примечание. Иногда для временных рядов вводят понятие спектрограммы, которая сложно строится, но для выделения периодических составляющих достаточно ограничиться периодограммой. (Более подробно о спектрограмме см. [1, с. 216–220]).

В практических целях можно упростить расчет периодограммы, если построить зависимость ошибки модели

---

[1] *Льюис К. Д.* Методы прогнозирования экономических показателей. — М.: Финансы и статистика, 1986. С. 67–74.

[2] *Тюрин Ю. Н., Макаров А. А.* Статистический анализ данных на компьютере. — М.: ИНФРА-М, 1998. С. 374.

$$У_t = a_0 + a_1t + a_2Sin(2\pi t/T) + a_3Cos(2\pi t/T) + e_t$$

от периода Т,

где Т принимает значения от 3 до удвоенной длины временного ряда.

Коэффициенты модели рассчитываются методом наименьших квадратов [12].

Целями анализа временных рядов являются:

- выявление структуры временного ряда для более глубокого понимания динамических экономических процессов;
- получение точечного и интервального прогноза динамического экономического показателя.

8.2. Классификация временных рядов

Временные ряды являются объектом изучения специального раздела математической статистики, посвященного анализу случайных процессов[1].

Упрощенно временные ряды можно классифицировать по степени зависимости основных характеристик временного ряда от времени. Основными характеристиками временного ряда обычно считают среднее значение, дисперсию и автокорреляционную функцию. Определение зависимости основных характеристик временного ряда от времени предполагает изучение численных значений этих характеристик при последовательном увеличении длины временного ряда.

Нестационарным временным рядом называется такой временной ряд, у которого среднее значение зависит от времени.

Слово "стационарный" происходит от латинского *stationarius* — постоянный, неизменяющийся.

Случайный процесс X_t называется стационарным в узком смысле (строго стационарным), если для любых n, t_1, t_2, ..., t_n и k распределения случайных величин $X(t_1)$, ..., $X(t_n)$ и $X(t_{1+k})$, ..., $X(t_{n+k})$ одинаковы.

[1] *Бриллинджер Д.* Временные ряды. — М.: Мир, 1980; *Яглом А. М.* Корреляционная теория стационарных случайных функций (с примерами из метеорологии). — М.: Гидрометеоиздат, 1981.

Это означает, что функция конечномерных распределений не меняется при сдвиге во времени, т. е.

$$F(X_{t_1},...,X_{t_n}) = F(X_{t_{1+k}},...,X_{t_{n+k}}).$$

В частности, образующие стационарную случайную последовательность случайные величины X_1, X_2, ..., X_t, ... распределены одинаково, но независимыми они не являются.

Дадим определение конечномерных распределений. Принято считать, что случайный процесс X_t задан, если для каждого t из T определена функция распределения величины X_t:

$$F_t(X) = P(X_t \le X), \tag{8.1}$$

для каждой пары элементов t_1 и t_2 из T определена функция распределения двумерной случайной величины (X_{t_1}, X_{t_2})

$$F_{t_1,t_2}(X_1,X_2) = P(X_{t_1} \le X_1, X_{t_2} \le X_2), \tag{8.2}$$

и вообще для любого конечного числа элементов t_1, t_2, ..., t_n из множества T определена n-мерная функция распределения величины $(X(t_1), X(t_2), ..., X(t_n))$

$$F_{t_1,t_2,...,t_n}(X_1,X_2,...,X_n) = \\ = P(X_{t_1} \le X_1, X_{t_2} \le X_2,..., X_{t_n} \le X_n). \tag{8.3}$$

Функции (8.1), (8.2), (8.3) называются конечномерными распределениями случайного процесса.

Определим условия независмости распределения случайных величин.

Случайные процесс X_t называется последовательностью независимо распределенных случайных величин, если для любых наборов чисел t_1, t_2, ..., t_n из T все конечномерные распределения определяются с помощью одномерных распределений:

$$F_{t_1,t_2,...,t_n}(X_1,X_2,...,X_n) = F_{t_1}(X_1), F_{t_2}(X_2)...F_{t_n}(X_n).$$

При этом все условные вероятности равны их безусловным значениям[1].

---

[1] *Тюрин Ю. Н., Макаров А. А.* Статистический анализ данных на компьютере. — М.: ИНФРА-М, 1998. С. 361–362.

Одним из видов стационарного случайного процесса является белый шум в узком смысле.

Белым шумом в узком смысле (строго белым шумом) называется случайный процесс, у которого среднее значение равно нулю, а составляющие его случайные величины $X(t)$ независимы и распределены одинаково при всех t.

Примечание. В определении случайных процессов не говорится о конкретном виде закона распределения случайных величин. Законов распределения случайных величин очень много. Обычно предполагается, что случайные величины подчиняются нормальному закону распределения (распределение Гаусса). Поэтому вводится понятие "гауссовский белый шум" — это последовательность независимых нормально распределенных случайных величин со средним, равным нулю, и общей дисперсией s^2.

Представим графически распределения случайного процесса $X(t)$ для $X(t_1)$ и $X(t_2)$ при независимых и зависимых распределениях (допустим, по нормальному закону распределения).

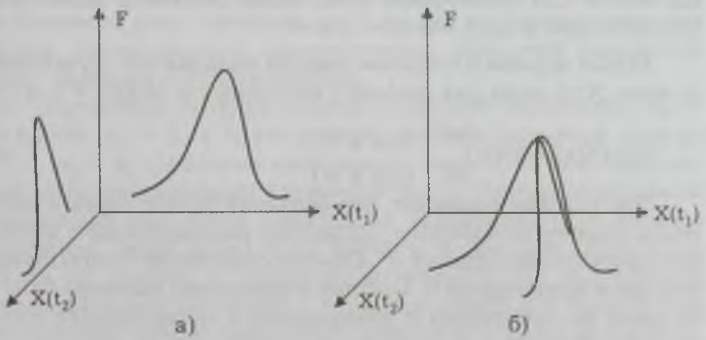


Рис. 8.2. Графики распределений независимого (а) и зависимого (б) случайного процесса для значений t_1 и t_2

Проверка на практике стационарности случайного процесса в узком смысле неосуществима. Поэтому для характеристики случайного процесса вводятся показатели: среднее значение, дисперсия и автоковариация. Этих показателей недостаточно для описания любых распределений, но с их помощью можно выразить основные свойства временных рядов. Поэтому вводится понятие стационарности временного ряда в широком смысле.

Стационарным временным рядом в широком смысле (слабо стационарным) называется такой случайный процесс, у которого среднее значение, дисперсия и автокорреляционная функция не зависят от времени. При этом автоковариационная функция зависит только от расстояния между аргументами.

Примечание. Если случайный процесс имеет распределение Гаусса, то среднего значения, дисперсии и автокорреляции достаточно, чтобы описать нормальный закон распределения. Поэтому если доказать, что временной ряд является стационарным в широком смысле, то он будет стационарным в узком смысле при условии нормального распределения случайного процесса. В общем случае стационарный временной ряд в широком смысле может не быть стационарным в узком смысле. Для практических целей вполне достаточно рассматривать временные ряды в широком смысле.

Белым шумом в широком смысле называется случайный процесс X(t), если для любого t выполняется MX(t) = 0 и

$$\mathrm{Cov}(X(к), X(t)) = \begin{cases} \sigma^2, & \text{при } к = t; \\ 0, & \text{при } к \neq t. \end{cases}$$

Эти условия означают, что значения белого шума в широком смысле независимы, одинаково распределены с нулевой средней и дисперсией σ^2. Обычно свойствами белого шума обладают возмущения ε. Поэтому сокращенно свойства белого шума по отношению к возмущению ε записываются следующим образом:

$\varepsilon \sim \mathrm{iid}\ (0, \sigma^2)$.

Это показывает, что значения возмущения ε являются независимыми, одинаково распределенными с нулевым средним и дисперсией σ^2 (independent, identically distributed) [6, с. 266].

Приведем пример, поясняющий введенные понятия (рис. 8.3).

Рис. 8.3. Графики разных видов временных рядов

Анализ рис. 8.3 показывает, что исходный временной ряд имеет четко выраженную линейную тенденцию с периодической составляющей, период которой равен 4. Если длину временного ряда изменять от 3 до 14, то среднее значение будет увеличиваться, следовательно, исходный временной ряд, является нестационарным.

Вычислим коэффициенты линейной тенденции: $a_0 = 0{,}2286$, $a_1 = 1{,}16$ и рассчитаем остатки линейной модели $У_t - (a_0 + a_1 t)$, которые изображены на рис. 8.3 под названием слабостационарный временной ряд 1. Слабостационарный временной ряд имеет следующие характеристики: среднее значение равно нулю и не зависит от времени, дисперсия тоже не зависит от времени, но имеет четко выраженное периодическое колебание с периодом, равным 4. Однако при увеличении длины временного ряда от 4 и далее до 14 автокорреляционная функция не будет изменять своего вида. Устраним в остатках периодическую составляющую с помощью модели $У_{pi} = a_0 + a_1 t + a_2 \mathrm{Sin}(2\pi t/4) + a_3 \mathrm{Cos}(2\pi t/4)$.

Вычислим остатки $e_1 = У - У_p$, изображенные на рис. 8.2 под названием слабостационарный временной ряд 2. Слабостационарный временной ряд 2 не содержит видимых четко выраженных тенденций, однако на любом участке временного ряда может наблюдаться некоторая зависимость каждого последующего значения от предыдущего значения временного ряда. Эта зависимость будет определяться расстоянием между значениями временного ряда, но не местом их расположения во временном ряду. Следовательно, слабостационарные временные ряды допускают наличие определенной автокорреляционной функции, которая будет одинаковой на любом участке временного ряда при заданной длине выделенного участка временного ряда. Попытаемся удалить эту зависимость с помощью авторегрессии остатков третьего порядка, примененных к слабостационарному временному ряду e_1 по следующей модели

$$e_{1pt} = a_0 + a_1 e_{1(t-1)} + a_2 e_{1(t-2)} + a_3 e_{1(t-3)}.$$

Вычислим остатки $e_{2t} = e_{1t} - e_{1pt}$, (начиная с $t = 4$, так как первые три значения временного ряда выпадают из расчетов), изображенные на рис. 8.3 под названием белый шум. Анализ белого шума показывает, что временной ряд не содержит как видимых, так и скрытых закономерностей на любом участке временного ряда.

8.3. Критерии проверки временного ряда на стационарность

Проверка гипотезы о неизменности среднего значения временного ряда.

Выдвигается нулевая гипотеза о том, что математическое ожидание временного ряда является константой и во временном ряду отсутствует периодическая составляющая.

Нулевую гипотезу проверим с помощью критерия восходящих и нисходящих серий.

Критерий восходящих и нисходящих серий.

Этот критерий улавливает постепенное смещение (по ходу выборочного обследования) не только монотонного, но и более общего, например периодического, характера.

Критерий восходящих и нисходящих серий выполняется в следующей последовательности.

Шаг 1. Вычисляются первые разности по формуле

$У_{t+1} - У_t$, $t = 1, ..., n - 1$.

Шаг 2. Численные значения первых разностей заменяются знаками плюс или минус по следующему соотношению:

если $(У_{t+1} - У_t) > 0$, то ставится плюс,

если $(У_{t+1} - У_t) < 0$, то ставится минус,

если два или несколько следующих друг за другом наблюдений равны между собой, то принимается во внимание только одно из них.

Последовательность подряд идущих плюсов (восходящая серия) соответствует возрастанию значений временного ряда, последовательность минусов (нисходящая серия) характеризует снижение значений временного ряда.

Критерий основан на том, что если выборка случайная, то в образованной нами последовательности знаков общее число серий не должно быть слишком малым, а их протяженность не слишком большой.

В частности, на уровне значимости $0{,}05 < \alpha < 0{,}0975$ количественное выражение этого правила имеет вид:

$$\left.\begin{aligned} &k_1(n) > \left[\frac{1}{3}(2n-1) - 1{,}96\sqrt{\frac{16n-29}{90}}\right]; \\ &k_2(n) < k_0(n), \end{aligned}\right| \tag{8.4}$$

где $k_1(n)$, $k_2(n)$ — соответственно общее число серий и количество подряд идущих плюсов или минусов в самой длинной серии;

[ ] — целая часть числа;

$k_0(n)$ — критическое значение величины $k_2(n)$, которое зависит от n и определяется из табл. 8.1

Таблица 8.1

Критические значения величины $k_2(n)$

n	$n \leq 26$	$26 < n \leq 153$	$153 < n \leq 1170$
$k_0(n)$	5	6	7

Шаг 3. Если хотя бы одно из неравенств (8.4) окажется нарушенным, то нулевую гипотезу следует отвергнуть и признать, что временной ряд содержит неслучайную составляющую, которая зависит от времени t [1, с. 223–225]

Вопросы для самоконтроля

1. Укажите отличие в формировании пространственной выборки и временного ряда.
2. Дайте определение временного ряда.
3. Приведите структуру общей модели временного ряда.
4. Укажите два направления в изучении компонент временного ряда.
5. Укажите отличие экономического от технологического временного ряда.
6. Укажите три вида причин изменения численных значений экономического временного ряда.
7. Укажите основные свойства экономического временного ряда.
8. Приведите формулы статистических характеристик временного ряда.
9. Приведите определения нестационарного и стационарного временного ряда.
10. Назовите шаги выполнения критерия восходящих и нисходящих серий.

Глава 9

МЕТОДЫ ПРЕОБРАЗОВАНИЯ НЕСТАЦИОНАРНОГО ВРЕМЕННОГО РЯДА В СТАЦИОНАРНЫЙ

9.1. Аналитические методы выделения неслучайной составляющей временного ряда

Моделирование компонент временного ряда:

$У_t = f_{1t} + f_{2t} + f_{3t} + e_t$.

Моделирование тренда f_{1t} .

Включение в модель f_{1t} тренда преследует две цели:

• первая — определить долговременную тенденцию временного ряда для получения долгосрочных прогнозов;

• вторая — получить стационарные остатки со средним значением, равным нулю, для выделения в них периодических составляющих.

Тренд f_{1t} обычно воспроизводится с помощью следующих математических функций, которые слабо искажают периодические составляющие:

$f_{1t} = a_0 + a_1 t$;

$f_{1t} = a_0 + a_1 \ln t$;

$f_{1t} = a_0 + a_1 (\ln t)^4$ (эта функция хорошо воспроизводит медленный постоянный рост, затем резкое постоянное возрастание);

$f_{1t} = a_0 + a_1/t.$

Коэффициенты модели $У_t = f_{1t} + e_t$ определяются методом наименьших квадратов.

Примечание. Н.Д. Кондратьев для исключения нециклической теденции использовал линейную функцию и очень редко параболу [24, с. 348, 353, 354]

Моделирование сезонной и циклических компонент f_{2t}, f_{3t}

Сезонная и циклическая компоненты являются периодическими с разными периодами. Обычно, в экономических временных рядах на длиннопериодическую циклическую составляющую накладывается сезонная волна или волны более высоких частот. При этом выдвигаем предположение о том, что эти волны не влияют друг на друга и можно использовать аддитивную модель их совместного существования.

Примечание. Следует заметить, что это предположение слишком упрощает действительность. Н. Д. Кондратьев указывает на то, что между волнами может быть взаимное влияние [24, с. 379–380].

Можно предложить методику определения всех периодических составляющих временного ряда, в том числе сезонных и длиннопериодических, определить их достоверность и затем включить в модель.

Процесс моделирования сезонной и циклической компонент выполняется в такой последовательности.

Шаг 1. Выбор модели для выявления периодических составляющих.

Для выявления периодических составляющих временного ряда используется следующая модель:

$$У_t = a_0 + a_1t + a_2\mathrm{Sin}(2\pi t/T) + a_3\mathrm{Cos}(2\pi t/T) + e_t,$$

где a_0, a_1, a_2, a_3 — коэффициенты, определяемые методом наименьших квадратов;

$a_0 + a_1t = f(t)$ — линейная функция (тренд), которая приводит временной ряд к стационарному виду с нулевым средним значением, без искажения периодических составляющих остатка;

$Sin(2\pi t/T)$, $Cos(2\pi t/T)$ — тригонометрические функции периодической составляющей временного ряда, сумма которых позволяет учесть начальную фазу колебания;

t — порядковый номер времени;

T — период или длина волны циклической составляющей временного ряда.

Длительность периода T исследователь задает в интервале от трех единиц до удвоенной длинны временного ряда. Дальнейшее увеличение длинны волны приводит к тому, что небольшие отрезки этой волны начинают описывать нециклические тенденции временного ряда;

$(2\pi t/T)$ — аргумент тригонометрической функции, выраженный в радианах;

$1/T$ — частота, численно равная количеству колебаний в единицу времени;

$a_2Sin(2\pi t/T) + a_3Cos(2\pi t/T)) = c \times Cos(2\pi t/T + \varphi)$,

где φ — начальная фаза колебания;

$\varphi = arctg(a_1/a_2)$;

c — коэффициент, определяемый по формуле

$$c = \sqrt{a_2^{\,2} + a_3^{\,2}}.$$

Шаг 2. Создается база данных.

$X_1 = t$ — время;

$X_2 = Sin(2\pi t/T)$;

$X_3 = Cos(2\pi t/T)$;

$У$ — значения показателя временного ряда;

где T — период циклической составляющей, значение которого вводит исследователь в пределах от 3 до удвоенной длины временного ряда.

Шаг 3. Вычисление ошибки E модели

$У_t = a_0 + a_1t + a_2Sin(2\pi t/T) + a_3Cos(2\pi t/T) + e_t$

с помощью статистической функции "Линейн".

Переход к шагу 2 с изменением периода T.

Шаг 4. Построение графика зависимости ошибки модели E временного ряда от периода T циклического фактора, включенного в модель.

Строится график периодограммы зависимости ошибки модели E от T.

Шаг 5. Определение периодических составляющих временного ряда.

Определяются такие значения периодов T, при которых ошибка модели E имеет глобальный и локальные минимумы. Обнаруженные периоды характеризуют периодические составляющие временного ряда, среди которых надо исключить ложные периоды.

Ложная периодичность относится к периодам (эхам), которые являются кратными величине основного периода[1]. Основной период имеет наименьшую ошибку модели, а кратные периоды или ложные периоды имеют увеличенную ошибку модели. Например, квартальная периодическая составляющая 3 мес. порождает эхо на периодах: $3 \times 2 = 6$ мес. или полугодовой период, $3 \times 3 = 9$ мес., $3 \times 4 = 12$ мес. или годовой период и т. д. [1].

Шаг 6. Определение достоверных периодических составляющих временного ряда.

Для определения достоверных периодических составляющих необходимо построить модель с участием всех обнаруженных периодических составляющих и по критерию Стьюдента определить достоверность коэффициентов, стоящих перед тригонометрическими функциями, с использованием статистической функции "Линейн" Excel. При этом можно предложить два способа построения модели: неавтоматизированный и автоматизированный.

Для определения достоверности периодических составляющих неавтоматизированным способом строится модель вида

$$У_t = a_0 + a_1 t + a_2 \mathrm{Sin}(2\pi t/T_1) + a_3 \mathrm{Cos}(2\pi t/T_1) + \\ + a_4 \mathrm{Sin}(2\pi t/T_2) + a_5 \mathrm{Cos}(2\pi t/T_2) + \ldots + e_t,$$

где T_1 и T_2 — периоды выделенных периодических составляющих.

---

[1] *Кендалл М.* Временные ряды. — М.: Финансы и статистика, 1981. С. 110.

Анализ характеристик многофакторной модели производится обычным образом. Коэффициенты модели, достоверные по критерию Стьюдента, укажут на достоверность наличия соответствующей периодической составляющей временного ряда.

Примечание. Предложенный алгоритм прост для реализации, но может возникнуть одна проблема (которая на практике анализа экономических временных рядов пока еще не встречалась), связанная с разрешающей способностью этого метода. Если во временном ряду будут находиться две волны с близкими частотами, то при какой разности этих частот предложенный метод позволит их выделить? Пока убедительного ответа нет. Мало того, эти близкие частоты, накладываясь друг на друга, создадут биения, которые имеют очень сложную форму, и как в этой форме выделить компоненты пока не ясно.

Автоматизированный способ обнаружения достоверных периодических составляющих выполняется в такой последовательности: создается база данных всех обнаруженных или предполагаемых периодических составляющих и линейной тенденции, затем шаговым регрессионным анализом методом включения с помощью программы Stadia включаются в модель все достоверные компоненты периодических колебаний. При этом возникают трудности, связанные с тем, что шаговый метод может включить в модель только одну достоверную составляющую периодической компоненты $Sin(2\pi t/T)$ или $Cos(2\pi t/T)$, но для учета начальный фазы колебания необходимо в модели присутствие двух составляющих периодической компоненты. Этот недостаток можно устранить с помощью неавтоматизированного метода определения достоверных периодических составляющих временного ряда.

Шаг 7. Построение модели временного ряда.

Окончательная модель временного ряда с включением всех компонент может иметь следующий вид

$$У_t = a_0 + a_1 t + a_2 Sin(2\pi t/T_1) + a_3 Cos(2\pi t/T_1) +$$
$$+ a_4 Sin(2\pi t/T_2) + a_5 Cos(2\pi t/T_2) + e_t,$$

где $a_0 + a_1 t = f_{1t}$ — тренд;

$a_2 Sin(2\pi t/T_1) + a_3 Cos(2\pi t/T_1) = f_{2t}$ — сезонная компонента;

T_1 = период сезонной компоненты;

$a_4 Sin(2\pi t/T_2) + a_5 Cos(2\pi t/T_2) = f_{3t}$ — циклическая компонента;

T_2 — период циклической компоненты;

e_t — остатки модели.

Коэффициенты модели рассчитываются методом наименьших квадратов.

Многолетний опыт применения изложенной методики моделирования временных рядов для различных экономических переменных позволяет судить о ее высокой эффективности.

Однако существуют следующие проблемы предложенной модели временного ряда:

• обнаруженные периодические составляющие временного ряда должны иметь экономическое обоснование и механизм их возникновения;

• если возникают трудности в экономической интерпретации тенденций временного ряда (периодических составляющих), то можно использовать алгоритмические методы. Если временной ряд имеет некоторую тенденцию, то эти методы способны ее воспроизвести. А вот почему существует такая тенденция во временном ряду, это вопрос к экономистам, а не к математикам;

• для построения доверительных интервалов регрессии и прогноза необходимо считать более точным не среднее значение временного интервала, а его последнее более свежее значение. К сожалению, в большинстве эконометрической литературы не делается различия в расчетах доверительных интервалов регрессии и прогноза для пространственных и временных рядов.

9.2. Алгоритмические методы выделения неслучайной составляющей временного ряда

Алгоритмические методы выделения неслучайной составляющей временного ряда могут быть следующих видов:

- метод скользящего среднего;
- метод экспоненциально взвешенного скользящего среднего (метод Брауна);
- метод последовательных разностей.

Метод скользящего среднего

В основе метода скользящего среднего лежит простая идея: если индивидуальный разброс значений временного ряда У(t) около своего среднего значения $\bar{У}$ характеризуется дисперсией S^2, то разброс среднего из n членов временного ряда $(У_1, У_2, ..., У_n)/n$ около того же среднего $\bar{У}$ будет характеризоваться гораздо меньшей величиной дисперсии, раной S^2/n. Уменьшение дисперсии случайного разброса означает сглаживание соответствующей траектории.

Алгоритм скользящего среднего реализуется в такой последовательности: во временном ряду выбирается начальный участок заданной длины (меньшей, чем длина всего временного ряда), длина выделенного участка называется длиной усреднения, затем этот выделенный участок аппроксимируют функцией, обычно алгебраическим полиномом определенной степени, вычисляют значение полинома в середине выбранного участка временного ряда. Данная процедура повторяется для следующего участка временного ряда, который смещается на одну дату. При этом выделяемые участки "скользят" по временному ряду.

Если в качестве сглаживающей функции выбрать алгебраический полином первой степени, то она будет представлять линейную функцию времени

$$У_{pt} = a_0 + a_1 t.$$

В этом случае метод скользящего среднего выполняется в такой последовательности:

Шаг 1. Выбирают нечетную длину усреднения $N = 2m + 1$, где $m < n/3$, обычно m не превышает 3; n — длина временного ряда.

Шаг 2. Вычисляются сглаженные значения Z_t по формуле

$$Z_t = \frac{1}{2m+1}\sum_{k=-m}^{m} У_{t+k},\ t = m+1, m+2, ..., n-m$$

или по формуле линейной функции

$$Z_t = a_0 + a_1 t,\quad t = m+1, m+2, ..., n-m,$$

где a_0, a_1 — коэффициенты уравнения $У_t = a_0 + a_1 t$, определенные МНК по данным временного ряда $У_t$, входящих в длину усреднения.

Шаг 3. Вычисляются сглаженные значения в краевых точках по следующей формуле:

$$Z_t = a_0 + a_1 t,\ t = 1, 2, ..., m,$$

где a_0, a_1 — коэффициенты уравнения $У_t = a_0 + a_1 t$, определенные МНК по данным временного ряда $У_1$, $У_2$, ..., $У_{2m+1}$.

$$Z_t = a_0 + a_1 t,\ t = n - m + 1, ..., n,$$

где a_0, a_1 — коэффициенты уравнения $У_t = a_0 + a_1 t$, определенные МНК по данным временного ряда $У_{n-2m}$, $У_{n-2m+1}$,..., $У_n$.

Скользящее среднее имеет несколько особенностей:

- для того, чтобы начать процесс скользящего среднего, необходимо иметь N предыдущих значений временного ряда;
- скользящее среднее находится в середине сглаживающего интервала, а это означает, что значения сглаживающего интервала, которые стоят после его середины, оказывают влияние на среднее, т. е. свершившиеся события влияют на прошлое, чего не может быть;
- данным, включенным в процесс расчета скользящего среднего, присваивается одинаковый вес. Вес отдельного наблюдения указывает на долю вклада его значения в значение среднего. В случае скользящего среднего эта доля равна 1/N для наблюдений, входящих в среднее. При этом более “свежие” данные имеют тот же вес, что и более “старые”;
- чувствительность скользящего среднего обратно пропорциональна N — числу точек, входящих в среднее;
- если в качестве сглаживающей функции выбирать алгебраический полином более высоких степеней, то траектория скользящей средней будет больше отражать колебания

временного ряда и степень сглаживания снижается. Более подробно см. [1, с. 227–238];

- соседние члены ряда скользящих средних сильно коррелированы, так как в их создании используются одни и те же члены временного ряда. Это может привести к появлению в ряде скользящих средних новых периодических колебаний, которых не было в исходном ряду. Это явление носит название эффекта Слуцкого-Юла[1].

Метод экспоненциально взвешенного скользящего среднего (метод Брауна)

Необходимо предложить модель, которая бы отражала следующие свойства временных рядов:

- среднее значение временного ряда в текущий момент времени должно зависеть от всех предыдущих значений временного ряда при условии, что доля их вклада уменьшается по мере удаления от текущего времени;
- среднее значение временного ряда в момент времени t можно считать его прогнозным значением на момент времени t+1.

В качестве весовых коэффициентов, учитывающих влияние на среднее значение, можно выбрать временной ряд весов, убывающих по экспоненциальному закону, сумма которых равна $1/\lambda$:

$$1 + (1 - \lambda) + (1 - \lambda)^2 + (1 - \lambda)^3 + ... + (1 - \lambda)^n = 1/\lambda,$$

где λ — коэффициент сглаживания, принимающий значения от 0 до 1.

Для того чтобы среднее значение сохранило свои свойства, необходимо, чтобы сумма долей вклада каждого значения, входящего в состав средней, равнялась единице. Разделим экспоненциальные веса на их сумму и получим следую-

[1] *Тюрин Ю. Н., Макаров А. А.* Статистический анализ данных на компьютере. — М.: ИНФРА-М, 1998. С. 395; *Кендалл М.* Временные ряды. — М.: Финансы и статистика, 1981.

щие доли вклада каждого значения временного ряда, входящего в состав среднего значения:

$$\lambda,\ \lambda\ (1-\lambda),\ \lambda\ (1-\lambda)^2,\ \lambda\ (1-\lambda)^3\ \ldots\ \lambda\ (1-\lambda)^n.$$

Экспоненциально взвешенное среднее вычисляется по формуле

$$Z_t = \lambda У_t + \lambda(1-\lambda)У_{t-1} + \lambda(1-\lambda)^2 У_{t-2} + \lambda(1-\lambda)^3 У_{t-3} + \ldots =$$
$$= \lambda У_t + (1-\lambda)[\lambda У_{t-1} + \lambda(1-\lambda)У_{t-2} + \lambda(1-\lambda)^2 У_{t-3} + \ldots],$$

где $\lambda У_{t-1} + \lambda(1-\lambda)У_{t-2} + \lambda(1-\lambda)^2 У_{t-3} + \ldots = Z_{t-1}$.

Экспоненциально взвешенное среднее значение вычисляется по формуле

$$Z_t = \lambda У_t + (1-\lambda)Z_{t-1}.$$

Экспоненциально взвешенное среднее обладает следующими свойствами:

• в экспоненциально взвешенном среднем значения весов убывают со временем, поэтому здесь нет точки, на которой веса обрываются;

• чем больше λ, тем выше чувствительность среднего, чем меньше λ, тем устойчивее становится экспоненциально взвешенное среднее;

• изменяя коэффициент λ можно найти его оптимальное значение по признаку ошибки модели. Если оптимальное значение $0 \le \lambda \le 0{,}3$, то временной ряд является стационарным, если $1 \ge \lambda > 0{,}3$, то временной ряд является нестационарным, и следует перейти к моделям, учитывающим тенденцию временного ряда;

• экспоненциально взвешенное среднее очень хорошо сглаживает как стационарные, так и нестационарные временные ряды;

• алгоритм просто реализуется в Excel, особенно определение оптимального значения λ.

Метод последовательных разностей

Метод исключения неслучайной составляющей, имеющий вид алгебраического полинома порядка p, с помощью вычисления последовательных разностей, повторенных p + 1 раз. Если имеется временной ряд $У_1$, $У_2$, ..., $У_n$, то последовательные разности первого порядка обозначаются с помощью $\Delta У_t = У_t - У_{t-1}$, $t = 2, \dots n$. Последовательные разности 2-го порядка — это разности от последовательных разностей первого порядка $\Delta(\Delta У_t) = \Delta У_t - \Delta У_{t-1}$. После вычисления последовательных разностей они анализируются на стационарность. Если тест на стационарность выполняется, то расчеты заканчиваются, если нет, то вычисляются последовательные разности более высокого порядка.

Метод последовательных разностей получил практическое применение для следующих целей: выявление периодических составляющих в постоянно возрастающих временных рядах. В плавно возрастающих временных рядах могут изменяться темпы роста. Первые разности от такого временного ряда будут положительными, но изменяющимися около своего среднего значения, к которому можно применить модель периодических колебаний. Метод последовательных разностей первого порядка можно заменить линейной регрессионной моделью.

Метод последовательных разностей используется в модели Бокса-Дженкинса.

Примечание. Первые разности могут устранить тенденцию, но приводят к тому, что в остатках появляется взаимосвязь возмущений.

$$\Delta У = У_t - У_{t-1} = \alpha_0 + \alpha_{1t} + \varepsilon_t - (\alpha_0 + \alpha_{1t-1} + \varepsilon_{t-1}) = \alpha_1 + \varepsilon_t - \varepsilon_{t-1},$$

где $\varepsilon_t \sim iid(0, \sigma^2)$;

$\varepsilon_t - \varepsilon_{t-1}$ — новые возмущения не обладают свойством независимости между собой[1].

Вопросы для самоконтроля

1. Приведите два метода выделения неслучайной составляющей временного ряда.

2. Приведите шаги моделирования компонент временного ряда.

---

[1] *Тюрин Ю. Н., Макаров А. А.* Статистический анализ данных на компьютере. - М.: ИНФРА — М, 1998. С. 383.

3. Приведите алгоритмические методы выделения неслучайной составляющей временного ряда.

4. Приведите алгоритм скользящего среднего.

5. Укажите достоинства и недостатки метода скользящего среднего

6. Приведите алгоритм взвешенного скользящего среднего.

7. Укажите достоинства и недостатки взвешенного скользящего среднего.

Глава 10
АВТОКОРРЕЛЯЦИЯ. АВТОРЕГРЕССИЯ. МОДЕЛИ ВРЕМЕННОГО ЛАГА

10.1. Определение, причины и последствия автокорреляции остатков модели

Определение понятия автокорреляции временного ряда.

Автокорреляция = авто (греч. *autos* — сам) + корреляция: корреляция с самим собой.

Автокорреляция является характеристикой временного ряда.

В эконометрике используются два понятия "автокорреляция временного ряда" и "автокорреляция возмущения".

Автокорреляция временного ряда порядка d равна коэффициенту корреляции, рассчитанному между исходным временным рядом и его лаговой переменной порядка d.

Лаговая переменная порядка d получается от исходного временного ряда методом перемещения его по времени вперед на d дат.

База данных для расчета коэффициента автокорреляции приведена в табл. 10.1.

Существуют две трактовки коэффициента автокорреляции: прямая и обратная.

Прямая трактовка — коэффициент автокорреляции показывает степень влияния текущих значений временного ряда на его будущее значение.

Таблица 10.1

База данных для расчета коэффициента автокорреляции

t	$У_t$	$У_{t-1}$
1	$У_1$	$У_{1-1} = У_0$
2	$У_2$	$У_{2-1} = У_1$
3	$У_3$	$У_{3-1} = У_2$
4	$У_4$	$У_{4-1} = У_3$
5		$У_{5-1} = У_4$

Обратная трактовка — коэффициент автокорреляции показывает степень влияния будущих значений временного ряда на его текущее значение.

В эконометрике подразумевается прямая трактовка коэффициента автокорреляции первого порядка. Для учета влияния будущих значений временного ряда на текущее значение используются модели адаптивных ожиданий и частичной корректировки. Логическим продолжением автокорреляции является авторегрессия, которую будем изучать в дальнейшем.

Автокорреляция остатков порядка d равна коэффициенту корреляции, рассчитанному между исходными остатками модели и лаговыми остатками порядка d. Коэффициент автокорреляции рассчитывается по обычной формуле коэффициента корреляции, только в качестве первой переменной являются исходные остатки, а в качестве второй переменной используются лаговые остатки порядка d.

Причиной появления в остатках достоверной автокорреляции первого порядка является наличие в остатках тенденций, вызванных плохой спецификацией модели.

Последствия наличия в остатках автокорреляции первого порядка

Если модель содержит остатки с достоверной автокорреляцией, то могут возникнуть следующие последствия:

• модель является плохо специфицированной и появляется возможность ее улучшения;

• ошибка модели является завышенной, а достоверность модели заниженной.

10.2. Критерии проверки достоверности автокорреляции остатков модели

Проверку достоверности коэффициента автокорреляции остатков первого порядка можно производить по упрощенной или по более точной методике.

Упрощенная методика проверки коэффициента автокорреляции совпадает с методикой проверки достоверности обычного парного коэффициента корреляции, которая выполняется в такой последовательности:

Шаг 1. Выдвигается нулевая гипотеза о том, что $r(d) = 0$, H_0: $r(d) = 0$.

Шаг 2. Вычисляется $r(d)$.

Шаг 3. Вычисляется ошибка $r(d)$.

$$S_{r(d)} = \sqrt{\frac{1-r(d)^2}{n-d}}.$$

Шаг 4. Вычисляется критерий Стьюдента

$$t = \frac{r(d)}{S_{r(d)}}.$$

Шаг 5. Проверяется условие:

если $|t| \geq t_{\alpha/2}(\alpha = 0{,}05;\ m = n - d - 2)$, то нулевая гипотеза отвергается с вероятностью $1 - \alpha$,
где t-критерий Стьюдента.

В противном случае нулевая гипотеза принимается.

Для более точной проверки достоверности коэффициента автокорреляции остатков используется критерий Дарбина-Уотсона (DW).

Шаг 1. Вычисляется критерий Дарбина-Уотсона

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \approx 2(1-r(1))\ \ [5,\ с.\ 170-171].$$

Шаг 2. Определяются по таблицам нижнее и верхнее пороговые значения соответственно $d_н$ и $d_в$, зависящие от числа измерений, уровня значимости и числа объясняемых факторов в модели.

Шаг 3. Проверяются следующие условия:

а) если $d_в < DW < 4 - d_в$, то нулевая гипотеза об отсутствии автокорреляции принимается;

б) если $d_н \le DW \le d_в$ или $4 - d_в \le DW \le 4 - d_н$, то нулевая гипотеза об отсутствии автокорреляции не принимается и не отвергается (область неопределенности критерия);

с) если $0 < DW < d_н$, то нулевая гипотеза об отсутствии автокорреляции отвергается и утверждается, что имеется положительная автокорреляция остатков;

д) если $4 - d_н < DW < 4$, то нулевая гипотеза об отсутствии автокорреляции отвергается и утверждается, что имеется отрицательная автокорреляция остатков.

Критерий DW рассчитывается практически во всех статистических пакетах прикладных программ.

Коэффициент автокорреляции первого порядка обладает следующими свойствами.

Коэффициент автокорреляции может принимать значения от -1 до +1.

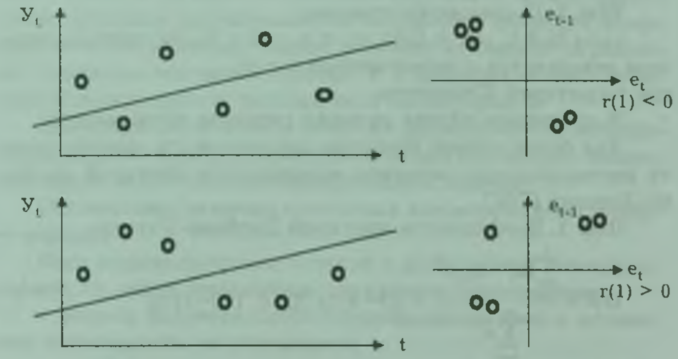


Рис. 10.1. Примеры отрицательного и положительного коэффициента автокорреляции остатков

Анализ рис. 10.1 показывает, что, если остатки часто меняют свой знак, коэффициент автокорреляции первого порядка будет отрицательным, а если имеются серии положительных и отрицательных значений остатков, — то положительным.

10.3. Модели, учитывающие автокорреляцию остатков

Если линейная модель

$У_t = a_0 + a_1X_t + e_t$

содержит автокорреляцию остатков первого порядка, то для ее устранения можно использовать следующую модель:

$У_t = a_0 + a_1X_t + e_t = a_0 + a_1X_t + pe_{t-1} + v_t,$

где р — коэффициент автокорреляции первого порядка;

e_{t-1} — лаговая переменная остатков первого порядка;

$e_t = pe_{t-1} + v_t;$

v_t — остатки новой модели.

Интрига предложенной модели заключается в том, что одновременно нельзя найти коэффициенты a_0, a_1 и р. Для определения этих коэффициентов предложено много разных методов, которые дают примерно одни и те же результаты. Для практических целей в первом приближении можно использовать более простой метод наименьших квадратов, для гурманов эконометрики можно рекомендовать более сложные.

10.4. Методы оценки параметров модели с автокоррелированными остатками

Эконометристами были предложены несколько методов определения коэффициентов модели

$У_t = a_0 + a_1X_t + pe_{t-1} + v_t.$

Первый метод. Метод наименьших квадратов

Методом наименьших квадратов определяются коэффициенты a_0, a_1 модели

$У_t = a_0 + a_1X_t + e_t.$

Вычисляются остатки модели

$e_t = У_t - (a_0 + a_1X_t).$

Методом наименьших квадратов определяется коэффициент р модели:

$e_t = pe_{t-1} + v_t.$

Второй метод. Метод Дарбина 1

Если в модели

$У_t = a_0 + a_1X_t + pe_{t-1} + v_t$

вместо e_{t-1} подставить его значение

$e_{t-1} = У_t - (a_0 + a_1X_{t-1}),$

то после несложных преобразований можно получить модель:

$$У_t = a_0(1 - p) + pУ_{t-1} + a_1X_t - a_1pX_{t-1} + v_t =$$
$$= b_0 + pУ_{t-1} + a_1X_t - a_1pX_{t-1} + v_t.$$

Коэффициенты полученной модели можно рассчитать методом наименьших квадратов с использованием следующей базы данных:

$X_{t-1},\ X_t,\ У_{t-1},\ У_t.$

Коэффициент a_0 можно определить по формуле

$a_0 = b_0/(1 - p)$

Метод Дарбина позволяет одновременно определить коэффициенты a_0, a_1, р.

Другие методы:

Метод Эйткена (в форме преобразований данных) вычисления коэффициентов модели

$У_t = a_0 + a_1X_t + pe_{t-1} + v_t,$

где $e_t = pe_{t-1} + v_t$ — возмущение, подчиняющееся авторегрессионной модели первого порядка, и для всех t выполняются условия:

$M(v_t) = 0,$

$M(v_t\ v_{t+d}) = S^2$ для $d = 0;$

$M(v_t\ v_{t+d}) = 0$ для $d \neq 0$;

$У_t$ — зависимая переменная временного ряда;

X_t — объясняемая переменная;

t — порядковый номер времени;

$e_t = У_t - (a_0 + a_1X_t)$.

Необходимо вычислить a_0, a_1, p при условии

$v'v \to \min$.

Метод Эйткена выполняется в такой последовательности:

Шаг 1. Вводится предполагаемый коэффициент p.

Шаг 2. Вычисляются методом наименьших квадратов характеристики модели

$$У_{нt} = b_0X_{н1t} + b_1X_{н2t} + z_t,$$

где новые переменные вычисляются по следующим формулам:

для $t = 1$ (поправка Прайса-Уинстена)[1]

$$У_{нt} = \sqrt{1-p^2}\,У_t;$$

$$X_{н1t} = \sqrt{1-p^2}\,X_t;$$

$$X_{н2t} = \sqrt{1-p^2}\,X_t;$$

для $t \geq 2$ (преобразования Бокса-Дженкинса)[2]

$$У_{нt} = У_t - pУ_{t-1};$$

$$X_{н1t} = 1 - p;$$

$$X_{н2t} = X_t - pX_{t-1};$$

b_0, b_1 — коэффициенты Эйткена.

Шаг 3. Вычисляются характеристики модели

$$У = b_0 + b_1X + pe_{t-1} + v_t.$$

Метод Дарбина 2 вычисления коэффициентов модели

$$У_t = a_0 + a_1X_t + pe_{t-1} + v_t.$$

---

[1] *Доугерти К.* Введение в эконометрику. — М.: ИНФРА-М, 1997. С. 223.

[2] *Замков О. О., Тостопятенко А. В., Черемных Ю.Н.* Матемаитческие метды в экономике. — М.: МГУ им. М. В. Ломоносова, Дис, 1997. С. 361.

Необходимо рассчитать коэффициенты a_0, a_1, p при условии

$v'v \to \min$.

Метод Дарбина выполняется в такой последовательности.

Шаг 1. Вычисляется коэффициент p, входящий в состав модели

$$У_t = a_0(1 - p) + pУ_{t-1} + a_1X_t - pa_1X_{t-1} + z_t.$$

Шаг 2. Вычисляются коэффициенты Эйткена модели

$$У_{нt} = d_0X_{н1t} + d_1X_{н2t} + z_t,$$

где новые переменные преобразуются по следующим формулам:

для t = 1:

$$У_{нt} = \sqrt{1-p^2}\,У_t;$$

$$X_{н1t} = \sqrt{1-p^2}\,X_t;$$

$$X_{н2t} = \sqrt{1-p^2}\,X_t;$$

для $t \geq 2$:

$У_{нt} = У_t - pУ_{t-1};$

$X_{н1t} = 1 - p;$

$X_{н2t} = X_t - pX_{t-1};$

d_0, d_1 — коэффициенты Эйткена.

Шаг 3. Вычисляются характеристики модели

$$У = d_0 + d_1X + pe_{t-1} + v_t.$$

Метод Кочрена-Оркатта вычисления коэффициентов модели:

$$У_t = a_0 + a_1X_t + pe_{t-1} + v_t.$$

Необходимо вычислить коэффициенты a_0, a_1, p при условии

$v'v \to \min$.

Метод Кочрена-Оркатта использует метод Эйткена и выполняется в такой последовательности.

Этап 1. Задается произвольное значение p_1, выполняется метод Эйткена.

Этап 2. Вычисляется p_2 при использовании коэффициентов Эйткена.

Этап 3. Если p_1 и p_2 отличаются на заданную величину, то повторяется этап 1 метода Эйткена, где в качестве p вводится численное значение p_2, полученное на 2 этапе. Затем повторяются этапы 2, 3 до тех пор, пока коэффициенты p и p_2 не станут равными в пределах заданной ошибки.

Метод Хилдрета-Лу вычисления коэффициентов модели

$$У_t = a_0 + a_1 X_t + pe_{t-1} + v_t.$$

Необходимо вычислить коэффициенты a_0, a_1, p при условии

$$\mathbf{v'v} \to \min.$$

Метод Хилдрета-Лу выполняется в такой последовательности.

Шаг 1. Вводится значение p из диапазона от -1 до +1 с заданным шагом.

Шаг 2. Выполняется метод Эйткена и определяется ошибка модели S_v.

Шаг 3. Определяется при каком значении p ошибка модели S_v будет минимальной[1].

10.5. Особенность прогнозирования с учетом автокорреляции остатков

Особенность прогнозирования с учетом автокорреляции остатков первого порядка заключается в том, что удается улучшить прогноз, но только на одну дату. В эконометрике особо ценятся методы, способные улучшить прогноз.

Прогнозирование по модели с коэффициентом автокорреляции производится по формуле

$$У_{пр(n+1)} = a_0 + a_1 X_{n+1} + pe_n,$$

---

[1] *Доугерти К.* Введение в эконометрику. — М.: ИНФРА-М, 1997. С. 224.

где $У_{пр(+1)}$ — прогнозное значение зависимой переменной на прогнозный период;

e_n — значение остатка в предшествующий период;

$a_0 + a_1 X_{n+1}$ — обычный прогноз по линейной функции;

pe_n — поправка прогноза на автокорреляцию остатка.

Улучшение прогноза с использованием автокорреляции остатков первого порядка возможно только при прогнозировании на одну дату.

10.6. Авторегрессионные модели

Авторегрессионная модель p-го порядка имеет вид:

$У_t = a_0 + a_1 У_{t-1} + a_2 У_{t-2} + ... + a_p У_{t-p} + e_t$, $t = 1, 2, ..., n$.

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

В первом приближении коэффициенты модели можно определить методом наименьших квадратов.

Модель частичной корректировки

В модели частичной корректировки предполагается, что поведенческое уравнение определяет не фактическое значение зависимой переменной $У_t$, а ее желаемый уровень $У_{жt}$:

$$У_{жt} = a + bX_t + e_t. \quad (10.1)$$

Предполагается также, что фактическое приращение зависимой переменной $У_t - У_{t-1}$ пропорционально разнице между ее желаемым уровнем и фактическим значением в предыдущий период, т. е. $У_{жt} - У_{t-1}$:

$$У_t - У_{t-1} = \lambda(У_{жt} - У_{t-1}) + v_t \ (0 \le \lambda \le 1),$$

где v_t — случайная составляющая.

Это выражение можно переписать как

$$У_t = \lambda У_{жt} + (1 - \lambda)У_{t-1} + v_t \ (0 \le l \le 1). \quad (10.2)$$

Откуда видно, что $У_t$ получается как взвешенное среднее желаемого уровня и фактического значения этой переменной в предыдущем периоде. Чем больше значение l, тем быстрее происходит процесс корректировки.

Подставив выражение (10.1) в формулу (10.2), можно получить:

$$У_t = al + b\lambda X_t + (1 - \lambda)У_{t-1} + v_t + \lambda e_t.$$

Так как стохастическая переменная $У_{t-1}$ не коррелирует со случайными составляющими v_t и λe_t (они рассчитываются после определения $У_{t-1}$), то расчет коэффициентов полученной модели можно произвести с помощью метода наименьших квадратов, используя в качестве зависимой переменной $У_t$ в качестве объясняемых переменных X_t и $У_{t-1}$. Коэффициент перед $У_{t-1}$ характеризует $(1 - \lambda)$, а следовательно, и λ; коэффициент при X_t, деленный на λ, позволяет вычислит b; постоянный коэффициент, деленный на λ, равен коэффициенту a[1].

Пример использования модели частичной корректировки дивидентов (модель Литнера) см. [5, с. 206–207].

Модель адаптивных ожиданий

Предположим, что зависимая переменная $У_t$ связана с ожидаемым значением фактора $X_{ож(t+1)}$ моделью адаптивных ожиданий:

$$У_t = a + bX_{ож(t+1)} + e_t, \qquad (10.3)$$

где ожидаемое значение фактора $X_{ож(t+1)}$ на период t+1 зависит от фактических X_t и ожидаемых $X_{ожt}$ значений этого фактора на текущий момент времени t. Чем больше разница между фактическим значением X_t и $X_{ожt}$, тем больше должна быть разница между $X_{ож(t+1)}$ и $X_{ожt}$.

$$X_{ож(t+1)} - X_{ожt} = \lambda(X_t - X_{ожt}).$$

$$X_{ож(t+1)} = \lambda X_t + (1 - \lambda)X_{ожt} \ (0 \le \lambda \le 1). \qquad (10.4)$$

---

[1] *Доугерти К.* Введение в эконометрику. — М.: ИНФРА-М, 1997. С. 291–292.

Чем больше величина λ, тем быстрее ожидаемое значение адаптируется к предыдущим реальным значениям переменной.

В модели адаптивных ожиданий (10.3) $X_{ож(t+1)}$ является неизвестной величиной. Выразим ее через известные переменные. Для этого потребуется следующее утверждение: если справедливо уравнение (10.3) в момент времени t, то это соотношение будет справедливо в момент времени t-1:

$$У_{t-1} = a + bX_{ожt} + e_{t-1},\ bX_{ожt} = У_{t-1} - a - e_{t-1}. \qquad (10.5)$$

Используя уравнения (10.3), (10.4), (10.5), можно получить модель адаптивных ожиданий, содержащую известные значения переменных:

$$\begin{aligned} У_t &= a + bX_{ож(t+1)} + e_t = a + b[\lambda X_t + (1 - \lambda)X_{ожt}] + e_t = \\ &= a + b\lambda X_t + (1 - \lambda)bX_{ожt} + e_t = \\ &= a + b\lambda X_t + (1 - \lambda)(У_{t-1} - a - e_{t-1}) + e_t. \end{aligned}$$

Окончательно получим модель адаптивных ожиданий, представленной в виде модели с распределенными лагами:

$$У_t = a\lambda + b\lambda X_t + (1 - \lambda)У_{t-1} + e_t - (1 - \lambda)e_{t-1},$$

в которой лаговая переменная $У_{t-1}$ и остатки e_{t-1} коррелируют между собой, следовательно, оценки метода наименьших квадратов не будут состоятельными.

Однако в первом приближении, можно использовать метод наименьших квадратов, где зависимой переменной будет $У_t$, объясняемыми переменными должны быть X_t и $У_{t-1}$. Коэффициент перед $У_{t-1}$ дает оценку $(1 - \lambda)$, а следовательно, и λ; коэффициент при X_t, деленный на оценку λ, дает оценку b; постоянный коэффициент, деленный на оценку λ, дает оценку a.

Примечание. Для получения состоятельных оценок необходимо применить обратное преобразование Койка и затем нелинейный метод наименьших квадратов. Пример решения см. [5, с. 208–210].

10.7. Определение, причины, последствия и примеры появления временных лагов между причиной и следствием

Лаг (от англ. *lag* — отставать, задерживать) — задержка.

Временной лаг — временная задержка между причиной и следствием.

Временной лаг — это свойство связи между причиной и следствием, которое заключается в том, что между причиной и следствием существует временная задержка.

Временной лаг существует для всех явлений в природе.

Причинами появления временного лага между экономическими явлениями являются: инерционность экономической системы и инерционность во времени действия причины. Если возникла причина, которая может привести к некоторым последствиям в экономической системе, то необходимо время, для того чтобы причина вызвала следствие у экономического объекта.

Временной лаг может состоять из следующих составляющих:

- время передачи информации от источника причины к экономическому объекту;
- время обработки информации;
- время для принятия решения;
- время для выполнения принятого решения;
- время для получения видимого эффекта от принятого решения.

Инерционность действия причины проявляется в том, что прошлые события продолжают оказывать воздействие на текущие и будущие явления.

Последствия наличия временного лага между причиной и следствием.

Если между причиной и следствием существует достоверный лаг, то появляется возможность улучшить точечный и интервальный прогноз на время действия лагового фактора.

Если иметь в виду, что между причиной и следствием может быть временная задержка, то раскрываются дополнительные возможности более достоверно моделировать экономические процессы и уменьшения ошибки прогноза.

Примеры временных лагов. Заморозки весной приводят осенью к повышению цен на сельскохозяйственные продукты.

Неурожай сахарной свеклы и зерновых приводит через два года к снижению товарооборота Полтавского облпотребсоюза (по данным 1960–1991 гг.).

Эмиссия денег через 3 мес. приводит к снижению уровня жизни населения.

Дополнительные инвестиции в производство приводят к постепенному росту прибыли, затем происходит снижение роста прибыли, после того как будут использованы все положительное влияние, полученное от инвестиций.

Классическим историческим примером временного лага является строительный лаг. Основной задачей является определение оптимального распределения по времени денежных средств, которые были отпущены на строительство.

Изобретения в науке и технике реализуются через некоторое время на практике. Время задержки может изменяться от нескольких месяцев до десятков лет.

Все причинно-следственные явления проходят с временной задержкой, только обнаружить эту задержку удается в случае, если интервал времени, в котором изучается процесс, будет больше, чем временной лаг и амплитуда реакции следствия на причину будет больше случайной составляющей.

Анализ приведенных примеров позволяет сделать вывод о том, что существуют два вида лагов — сосредоточенный и распределенный.

10.8. Виды лагов

Сосредоточенный лаг характеризует время задержки между причиной и следствием.

Распределенный лаг характеризует переходной процесс воздействия причины на следствие. В соответствии с этой классификацией лагов используются модели сосредоточенного и распределенного лага.

Примечание. В эконометрической литературе принято рассматривать только распределенные лаги, и понятие сосредоточенного лага пока не входит в терминологию эконометрики.

10.9. Модель сосредоточенного лага

Модели лагов используются только для временных рядов.

Постановка задачи.

Предположим, что между причиной и следствием имеется сосредоточенный временной лаг неизвестной длительности. Необходимо построить модель сосредоточенного лага и определить ее длительность.

Для определения сосредоточенного лага необходимо выполнение необходимых и достаточных условий.

Необходимое условие — длительность лага должна быть больше шага единицы измерения времени.

Достаточное условие — амплитуда воздействия лаговой переменной должна быть в два раза больше амплитуды случайной составляющей.

Примечание. Достаточное условие требует пояснения. При анализе связи между двумя случайными величинами — причиной и следствием — необходимо учитывать следующее обстоятельство: варьирование следствия обусловлено влиянием причины и случайной составляющей с возможным лагом. Если амплитуда случайной составляющей больше реакции следствия на причину, то обычным коэффициентом корреляции с учетом временной задержки обнаружить не удается. Если предположить, что сильные изменения причины (выбросы) могут привести с задержкой сильные изменения (выбросы) у следствия, то анализ этих выбросов с задержкой позволят обнаружить их связь.

Если выполняются необходимые и достаточные условия обнаружения сосредоточенного лага, то используется следующая модель сосредоточенного лага:

$$У_t = a_0 + a_1X_t + a_2X_{t-d} + e_t,$$

где d — максимальная длительность лага, которая должна быть меньше длины временного ряда n;

X_{t-d} — лаговая переменная порядка d.

Для вычисления коэффициентов модели можно использовать метод наименьших квадратов.

Для определения длительности лага d можно использовать три способа:

- последовательно изменять длительность лага и определить оптимальную длительность лага, при котором ошибка модели будет минимальной;
- вычислить характеристики модели

$$У_t = a_0 + a_1X_t + a_2X_{t-1} + a_3X_{t-2} + \dots + a_{d+1}X_{t-d} + e_t,$$

определить по критерию Стьюдента достоверный лаговый фактор, который укажет на величину определяемого лага;

- проанализировать матрицу парных коэффициентов корреляции между зависимой переменной и всеми возможными лаговыми факторами. Достоверный лаговый фактор укажет на величину лага между причиной и следствием.

Если достаточное условие не выполняется, то необходимо анализировать временные задержки между выбросами причины и следствия методом наложения графиков двух переменных и определения процентов совпадающих выбросов при последовательном перемещении одного графика относительно другого. Если количество совпадений выбросов между причиной и следствием будет больше 70%, то связь между этими переменными с данной лаговой задержкой будет достоверной с вероятность 0,95 при длине временного ряда, равного 40 [14].

Пример сосредоточенного лага между причиной и следствием показан на рис. 10.2.

Определение степени связи между пиками.

Модель сосредоточенного лага имеет следующий вид:

$$У_t = a_0 + a_1X_t + a_2X_{t-s} + e_t,$$

где s — величина лага;

Рис. 10.2. Пример сосредоточенного лага между причиной Х и следствием У

e — случайная составляющая;

t — порядковый номер времени.

Для обнаружения величины лага s должны соблюдаться необходимые и достаточные условия. Необходимое условие заключается в том, что интервал времени между последовательными измерениями должна быть меньше длины лага s.

Достаточные условия обнаружения лага заключается в том, что лаговый эффект от действия фактора X_{t-s} должен быть больше случайной составляющей e_t.

В научной литературе имеются методы определения лага s. Однако так как часто нарушаются достаточные условия обнаружения лага, то многие лаговые факторы не удается выявить.

Если колебания Х и У не выходят за пределы случайного шума, то их взаимосвязь не удается обнаружить.

Можно предположить, что если между Х и У имеется лаговая связь, то сильные изменения (выбросы) Х могут вызвать с лаговой задержкой сильные изменения (выбросы) У. Следовательно, следует изучить степень взаимосвязи между выбросами. В качестве простейшего критерия степени связи между выбросами можно выбрать процент их совпадения с лаговой задержкой s. Поставим перед собой следующую задачу: необходимо определить достоверный процент совпадения выбросов между Х и У с различной временной задержкой на уровне значимости $\alpha = 0{,}05$. В качестве выбросов будем рассматривать пики значений переменных.

Выдвигаем нулевую гипотезу о том, что пики переменных Х и У не связаны между собой.

Для проверки нулевой гипотезы можно предложить следующие этапы исследования:

1. Определить вероятность появления пика у Х и У при условии, что появление пиков является случайным событием.

2. Определить математическое ожидание и дисперсию количества пиков у Х и У. При этом если фактическое количество пиков будет достоверно больше или меньше их математического ожидания, то можно утверждать о неслучайности количества пиков.

3. Определить математическое ожидание и дисперсию совпадений пиков у Х и У при условии, что эти совпадения являются случайными событиями.

4. Определить процент достоверного совпадения пиков на уровне значимости а = 0,05.

Решение задачи.

1. Определим вероятность появления пика у Х и У при условии, что появление пиков является случайным событием.

Уточним понятие "пик". Пик — это величина, которая больше двух соседних.

Определим, каково распределение пиков в случайном ряду.

Рассмотрим конечный ряд из n величин X_1, X_2, ..., X_n. Начальное значение нельзя считать пиком, так как X_0 неизвестно, нельзя рассматривать в качестве пика последнее значение, так как неизвестно X_{n+1}. Для определения пика требуются три последовательных значения. Если ряд случаен, то эти три значения могут следовать в любом из шести возможных порядков с равной вероятностью (1 2 3, 1 3 2. 2 3 1. 2 1 3, 3 1 2, 3 2 1). Только в двух из них (1 3 2, 2 3 1) будет пик.

Примечание. Количество перестановок трех элементов равно 3! = 6.

Следовательно, вероятность появления пика в любой группе из трех значений или вероятность того, что величина X_2 является пиком, равна

$$P(X_2 \text{ — пик}) = 2/6 = 1/3^1.$$

Определим вероятность того, что величина X_3 является пиком. Рассмотрим ряд, содержащий четыре значения 1, 2, 3, 4. Четыре элемента позволяют получить 4! = 24 перестановок, в которых только в 8 случаях третий элемент является пиком. Следовательно, вероятность того, что величина X_3 является пиком? равна $P(X_3 \text{ — пик}) = 8/24 = 1/3$. Аналогичные расчеты позволяют утверждать следующее: вероятность того, что величина X_t является пиком равна

$$P(X_t \text{ — пик}) = 1/3, \text{ при } t = 2, ..., n - 1.$$

2. Определим математическое ожидание и дисперсию количества пиков у X и У.

Пусть имеются два временных ряда X_t и $У_t$, $t = 1, ..., n$, где X_t является лаговым фактором по отношению к $У_t$.

Зависимость У от X можно представить в виде модели

$$У_t = a_0 + a_1X_{t-s} + e_t,$$

где s — величина лага и заранее не известна.

Предполагается, что случайная компонента e_i позволяет обнаружить лаговый эффект не от каждых значений $У_t$ и X_{t-s}, а только от локальных максимумов (пиков) этих переменных.

Введем счетные переменные:

$$K_t = \begin{cases} 1, \text{ при } X_{t-1} < X_t < X_{t+1}, \; t = 1, ..., n; \\ 0, \text{ в остальных случаях;} \end{cases}$$

$$L_t = \begin{cases} 1, \text{ при } У_{t-1} < У_t < У_{t+1}, \; t = 1, ..., n; \\ 0, \text{ в остальных случаях.} \end{cases}$$

---

[1] *Кендалл М.* Временные ряды. — М.: Финансы и статистика, 1981. С. 28.

Необходимо определить достоверность процента совпадений 1 между $К_t$ и L_t с лаговой задержкой s.

Если временные ряды X_t и $У_t$ являются случайными, то вероятность появления пика у X и У одинакова и равна 1/3, или

$P(К_t = 1) = P(L_t = 1) = 1/3$ при $t = 2, ..., n - 1$.

Определим C — количество пиков для случайной величины X_t.

$$C = \sum_{t=1}^{n-2} К_t.$$

Математическое ожидание количества пиков для случайной переменной X_t будет равно

$$M(C) = M\sum_{t=1}^{n-2} К_t = \sum_{t=1}^{n-2}(К_t P(К_t)) = \sum_{t=1}^{n-2}[(К_t = 1)\frac{1}{3} + (К_t = 0)\frac{2}{3})] =$$

$$= \frac{1}{3}(n-2),$$

где M — обозначение математического ожидания;

$P(К_t)$ — вероятность появления события $К_t$;

$P(К_t = 0) = 2/3$ — вероятность того, что $К_t$ примет значение, равное 0.

Пояснение. События $К_t = 1$ и $К_t=0$ составляют полную группу, тогда $P(К_t = 1) + P(К_t = 0) = 1$, если $P(К_t = 1) = 1/3$, то $P(К_t = 0)= 1 - 1/3 = 2/3$.

Очевидно, что математическое ожидание количества пиков для случайной переменной $У_t$ тоже будет равно $(n - 2)/3$.

Определим математическое ожидание дисперсии количества пиков для переменной X_t.

$MD(C) = M(C^2) - [M(C)]^2$.

Вычислим составляющие: $[M(C)]^2$ и $M(C^2)$

$[M(C)]^2 = [(n - 2)/3]^2 = n^2/9 - 4n/9 + 4/9$.

Вычислим величину математического ожидания квадрата количества пиков для переменной X_t.

$$M(C^2) = M(\sum_{t=1}^{n-2} K_t)^2 = M(K_1 + K_2 + \ldots + K_{n-2})^2 =$$

$$= M(\sum_{n-2} K_t^2 + 2\sum_{n-3} K_t K_{t+1} + 2\sum_{n-4} K_t K_{t+2} + \sum_{(n-4)(n-5)} K_t K_{t+j}), \tag{10.4}$$

где $j \neq 0,\ 1,\ 2$;

индексы под знаком суммы указывают на количество слагаемых в сумме.

Определим каждое слагаемое выражения (10.4) в отдельности.

Определим вероятности событий K_t^2, $K_t K_{t+1}$, $K_t K_{t+2}$, $K_t K_{t+j}$, $j \neq 0,\ 1,\ 2$, при $K_t = 1$.

$P(K_t^2) = P(K_t) = 1/3$.

$P(K_t K_{t+1}) = 0$, так как события $K_t = 1$ и $K_{t+1} = 1$ несовместимы (так как пики могут следовать минимум через один интервал).

$P(K_t K_{t+2}) = 18/120 = 3/20$, так как если взять пять последовательных чисел:

1, 2, 3, 4, 5, то из $5! = 120$ перестановок в 18 наблюдается сочетание $K_t = 1$ и $K_{t+2} = 1$.

$P(K_t K_{t+j}) = (1/3) \times (1/3)$, $j \neq 0,\ 1,\ 2$, (так как события, K_t и K_{t+j} независимые, потому, что находятся на расстоянии больше трех значений).

$$M\left(\sum_{t=1}^{n-2} K_t^2\right) = \sum_{t=1}^{n-2} M\left(K_t^2\right) = \sum_{t=1}^{n-2} M(K_t) = \frac{1}{3}(n-2).$$

$$M\left(2\sum_{t=1}^{n-3} K_t K_{t+1}\right) = 0.$$

$$M\left(2\sum_{t=1}^{n-4} K_t K_{t+2}\right) = 2\sum_{t=1}^{n-4} M(K_t K_{t+2}) = 2\frac{3}{20}(n-4).$$

$$M\left[\sum_{t=1}^{(n-4)(n-5)} K_t K_{t+j}\right] = \sum_{t=1}^{(n-4)(n-5)} M(K_t K_{t+j}) = \frac{1}{9}(n-4)(n-5).$$

$M(C^2) = (n - 2)/3 + 3(n - 4)/10 + (n - 4)(n - 5)/9 =$
$= -33n/90 + n^2/9 + 32/90.$

$$[M(C)]^2 = [(n-2)/3]^2 = n^2/9 - 4n/9 + 4/9.$$

Математическое ожидание дисперсии количества пиков для переменной X_t будет равно:

$$M[D(C)] = M(C^2) — [M(C)]^2 = -33n/90 + n^2/9 + 32/90 - \\ - n^2 + 4n/9 - 4/9 = (7n - 8)/90^1.$$

Из литературы известно, что распределение количества пиков и впадин быстро стремится к нормальному. Логично предположить, что распределение пиков тоже стремится к нормальному закону распределения.

Отсюда следует вывод, что наблюдаемое количество пиков можно сравнивать с математическим ожиданием количества пиков нормального закона распределения

$$MC = \frac{n-2}{3}$$

со стандартным отклонением

$$S = \sqrt{\frac{7n-8}{90}}.$$

3. Определим математическое ожидание и дисперсию совпадений пиков у X и У при условии, что эти совпадения являются случайными событиями.

Рассмотрим пики K_t и L_t при $t = 1, ..., n - 2$ для двух временных рядов X и У.

Математические ожидания количества пиков будут равны

$$M\left(\sum_{t}^{n-2} K_t\right) = M\left(\sum_{t=1}^{n-2} L_t\right) = \frac{n-2}{3}.$$

Дисперсии количества пиков будут равны

$$M\left(D\sum_{t}^{n-2} K_t\right) = M\left(D\sum_{t=1}^{n-2} L_t\right) = \frac{7n-8}{90}.$$

---

[1] *Кендалл М.* Временные ряды. — М.: Финансы и статистика, 1981. С. 28–29.

Определим математические ожидания произведения пиков двух временных рядов и математическое ожидание дисперсии произведения пиков этих рядов.

$$M\sum_{t}^{n-2} K_t L_t.$$

$$M\left(D\sum_{t}^{n-2} K_t L_t\right)$$

Определим вероятность совпадения пиков у двух временных рядов K_t и L_t

$P[(K_t = 1)(L_t = 1)] = (1/3) \times (1/3) = 1/9$, так как события K_t и L_t являются независимыми.

Математическое ожидание совпадений пиков у двух временных рядов K_t и L_t будет равно

$$M\left(\sum_{t}^{n-2} K_t L_t\right) = \frac{n-2}{9}.$$

Определим математическое ожидание дисперсии произведения рядов K_t и L_t или дисперсию количества совпадения пиков у двух временных рядов K_t и L_t.

$$M\left(D\sum_{t}^{n-2} K_t L_t\right) = M\left[D\sum_{t=1}^{n-2} Z_t\right] = M\left(\sum_{t=1}^{n-2} Z_t\right)^2 - \left[M\left(\sum_{t=1}^{n-2} Z_t\right)\right]^2,$$

где $Z_t = K_t L_t$.

Произведем расчеты по аналогии с формулой (10.4), учитывая вероятности следующих событий, при условии, что $Z_t = 1$:

$P(Z_t^2) = P(Z_t) = 1/9;$

$P(Z_1 Z_2) = P(Z_t Z_{t+1}) = 0;$

$P(Z_1 Z_3) = P(Z_t Z_{t+2}) = (1/9)(1/9).$

$$M\left(\sum_{t=1}^{n-2} Z_t\right)^2 = \frac{n-2}{9} + 0\times(n-3) + \frac{2(n-4)}{81} + \frac{(n-4)(n-5)}{81} =$$

$$= \frac{n^2}{81} + \frac{2n}{81} - \frac{6}{81}.$$

$$\left[M\left(\sum_{t}^{n-2} Z_t\right)\right]^2 = \left(\frac{n-2}{9}\right)^2 = \frac{n^2-4n+4}{81}.$$

Вычислим математическое ожидание дисперсии совпадений пиков у X и У

$$M\left[D\left(\sum_{t=1}^{n-2} Z_t\right)\right] = \frac{n^2}{81} + \frac{2n}{81} - \frac{6}{81} - \left(\frac{n^2}{81} - \frac{4n}{81} + \frac{4}{81}\right) = \frac{6n-10}{81}.$$

4. Определим процент достоверного совпадения пиков на уровне значимости $\alpha = 0{,}05$ при условии, что длина временного ряда составляет $n = 33$.

Если изменения временных рядов X и У являются случайными, то математические ожидания количества пиков у X и У равны между собой и составят величину

$$A = M\left(\sum_{t=1}^{n-2} K_t\right) = M\left(\sum_{t=1}^{n-2} L_t\right) = \frac{n-2}{3} = \frac{33-2}{3} = 10{,}33.$$

Если пики у переменных X и У не связаны между собой, то математическое ожидание случайных совпадений пиков будет равно

$$M\left(\sum_{t=1}^{n-2} Z_t\right) = \frac{n-2}{9} = \frac{33-2}{9} = 3{,}44.$$

Математическое ожидание дисперсии случайных совпадения пиков равно

$$M\left[D\left(\sum_{t}^{n-2} Z_t\right)\right] = \frac{6n-10}{81} = \frac{6\times 33-10}{81} = 2{,}32.$$

Среднее квадратическое отклонение случайных совпадений пиков равно

$$S = \sqrt{M\left[D\left(\sum_{t}^{n-2} Z_t\right)\right]} = \sqrt{2{,}32} = 1{,}52.$$

Если количество совпадений пиков имеет нормальный закон распределения, то критическое значение количества совпадений пиков равно

$$K_{кр} = M\left(\sum_{t=1}^{n-2} Z_t\right) + t_{\alpha/2}(\alpha = 0{,}05; m = A - 2)S,$$

где $\alpha = 0{,}05$ — уровень значимости критерия;

m — число степеней свободы для критерия Стьюдента t;

A — количество минимальных пиков в каждом из анализируемых рядов;

$t_{\alpha/2}$ — критерий Стьюдента.

Табличное значение критерия Стьюдента на уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $A - 2 = 10 - 2 = 8$ равно

$t_{\alpha/2}(\alpha = 0{,}05,\ m = 10 - 2 = 8) = 2{,}31.$

$K_{кр} = 3{,}44 + 2{,}31 \times 1{,}52 = 6{,}95.$

Критическое значение количества совпадений пиков составляет

$6{,}95 \times 100/10{,}33 = 67{,}3\%$ от минимального числа пиков во временном ряду.

Вывод. Если процент совпадения пиков превысит 67%, то с вероятностью 0,95 можно утверждать, что эти совпадения являются достоверными при длине временного ряда, равном 33. Этот вывод будет справедлив для анализа пиков временных рядов с лаговым сдвигом. Различия будут в числе степеней свободы для критерия Стьюдента.

Составим табл. 10.2 зависимости критического значения совпадений пиков от длины временного ряда n.

Таблица 10.2

Зависимость $K_{кр}$ от длины временного ряда n

n	$K_{кр}$(%)	n	$K_{кр}$(%)
20	87,25863	80	52,45481
30	70,03679	90	51,22601
40	62,97492	100	50,20632
50	58,69752	150	46,85044
60	56,01787	200	44,93508
70	54,038		

$$K_{кр}(\%) = \frac{\frac{n-2}{9} + t_{\alpha/2}(\alpha = 0{,}05; m = \frac{n-2}{3} - 2)\sqrt{\frac{6n-10}{81}}}{\frac{n-2}{3}} \times 100,$$

где n — длина временного ряда (при n > 20);

$t_{\alpha/2}(\alpha = 0{,}05, m = (n - 2)/3 - 2)$ — критерий Стьюдента;

$m = (n - 2)/3 - 2$ — число степеней свободы для критерия Стьюдента;

$(n - 2)/3$ — математическое ожидание количества пиков;

$(n - 2)/9$ — количество случайных совпадений пиков;

$(6n - 10)/81$ — дисперсия случайных совпадений пиков.

Зависимость критического процента совпадения пиков $K_{кр}$ от длины временного ряда n можно аппроксимировать с помощью уравнения регрессии при n = 20, ..., 1000.

$K_{кр} = 17{,}701009 + 141{,}14993/Lnn + 111{,}76907/n +$
$+ 6698{,}8072/n^2$;

Данное уравнение имеет следующие характеристики:

$R^2 = 0.99989$ — коэффициент детерминации;

$F = 47364$ — критерий Фишера.

Функция рассчитывалась по программе "TableCurve2D". Результаты расчетов представлены на рис. 10.3.

10.10. Модель распределенного лага

Модель распределенного лага можно представить в следующем виде

$У_t = a_0 + a_1X_t + a_2X_{t-1} + a_3X_{t-2} + a_4X_{t-3} + e_t,$

где X_t — инвестиции в производство;

$У_t$ — прибыль.

Ограничения: значения X не должны изменяться с равным интервалом.

База данных для модели с распределенным лагом приведена в табл. 10.3.

Динамика воздействия причины на следствие представлена на рис. 10.4.

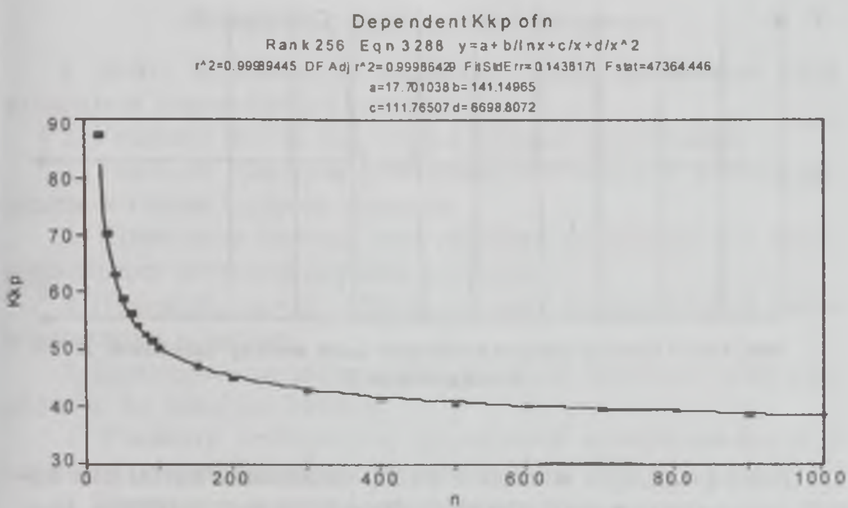


Рис. 10.3. График зависимости критического процента совпадения пиков от объема выборки n.

Таблица 10.3

База данных для модели с распределенным лагом

t	X_t	X_{t-1}	X_{t-2}	$X_{(t-3)}$	$У_t$
1	1				1
2	0	1			2
3	0	0	1		3
4	0	0	0	1	2
5	1	0	0	0	1
6	0	1	0	0	2
7	0	0	1	0	3
8	0	0	0	1	1
9		0	0	0	
10			0	0	
11				0	

Анализ рис. 10.4. показывает, что после вложения инвестиций в производство прибыль в течение двух интервалов времени растет, затем уменьшается.

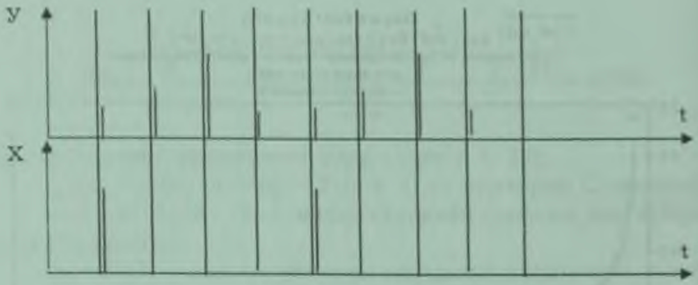


Рис. 10.4. Пример распределенного лага между причиной X и следствием У

Коэффициенты модели с распределенным лагом для целей предварительного анализа определяются методом наименьших квадратов. Для более точных расчетов используются специальные методы [1].

Имеется несколько методов оценки коэффициентов модели с распределенным лагом. Среди них метод Ширли Алмон. Этот метод очень красиво выводится и представляет интерес с точки зрения теории, однако практическое его применение показало, что с точностью до четвертого знака коэффициенты, вычисленные методом Ширли Алмон и методом наименьших квадратов, совпадают. Метод Ширли Алмон описан в книге [1].

10.11. Прогнозирование с учетом временного лага

Прогнозирование будет улучшено на интервале действия лагового фактора.

$$У_{пр(n+1)} = a_0 + a_1X_{n+1} + a_2X_n + a_3X_{n-1} + a_4X_{n-2}.$$

К обычному прогнозу добавляется поправка на действие лагового фактора.

Вопросы для самоконтроля

1. Дайте определение автокорреляции временного ряда зависимой переменной и остатков.

2. Укажите метод получения лаговой переменной.

3. Укажите причины появления достоверной автокорреляции остатков первого порядка.

4. Приведите методы определения достоверности автокорреляции остатков первого порядка.

5. Поясните, на что указывают знак коэффициента автокорреляции остатков.

6. Приведите модель автокорреляции остатков и методы расчета их коэффициентов.

7. Укажите особенность улучшения прогнозирования с помощью автокорреляции остатков.

8. Приведите модель авторегрессии.

9. Приведите модель частичной корректировки и адаптивных ожиданий.

10. Приведите виды лагов.

11. Дайте определение сосредоточенного и распределенного лага.

12. Приведите критерий связи между пиками.

Глава 11

ГЕТЕРОСКЕДАСТИЧНОСТЬ. АДАПТИВНОЕ ПРОГНОЗИРОВАНИЕ

11.1. Определение, критерии наличия, последствия и методы устранения гетероскедастичности

Гетероскедастичность — неоднородность дисперсий или "неодинаковый разброс".

Гомоскедастичность — однородность дисперсий или "одинаковый разброс".

Гомоскедастичность остатков — однородность остатков.

Пример гомоскедастичных остатков приведен на рис. 11.1.

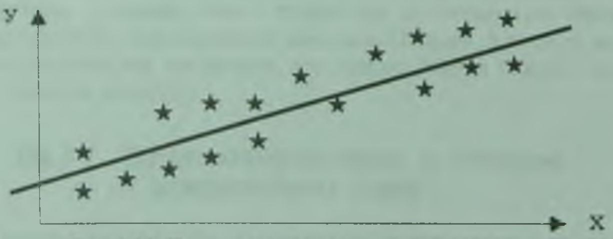


Рис. 11.1. Пример гомоскедастичности остатков

Анализ рис. 11.1 показывает, что зависимость У от X имеет линейный вид, разброс остатков около линии тренда примерно одинаковый. Дисперсии остатков для каждого фиксированного значения X являются примерно одинаковыми или однородными. Однородность дисперсий остатков является их свойством, которое имеет название "гомоскедастичность остатков". Следовательно, гомоскедастичность остатков — это свойство остатков, которое заключается в том, что их дисперсии или разбросы для каждого фиксированного значения X являются однородными или одинаковыми.

Гетероскедастичность остатков — это свойство остатков, которое заключается в том, что их дисперсии или разбросы для каждого фиксированного значения X являются неоднородными или неодинаковыми.

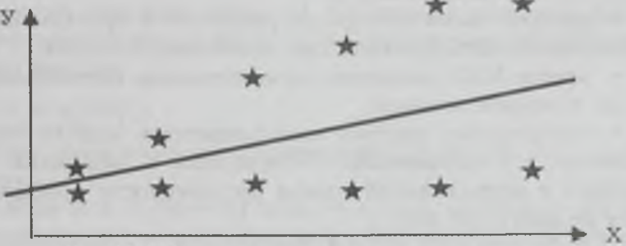


Рис. 11.2. Пример гетероскедастичности остатков

Пример гетероскедастичных остатков приведен на рис. 11.2.

Свойство гетероскедастичности может относиться не только к остаткам, но и к любой переменной, например к зависимой переменной У. Пример гетероскедастичности переменной У приведен на рис. 11.3.

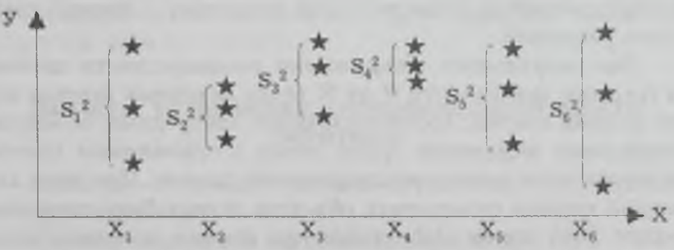


Рис. 11.3. Пример гетероскедастичности переменной У: S_i^2 — дисперсия У при значении объясняемой переменной X_i

Анализ рис. 11.3 показывает, что дисперсии У для каждого значения X не равны между собой $S_1^2 \neq S_2^2 \neq \ldots \neq S_6^2$.

Причины появления гетероскедастичности остатков:

• плохая спецификация модели;

• нарушение принципа однородности данных, который проявляется включением в базу данных двух и более групп объектов с сильно отличающимися свойствами.

Последствия гетероскедастичности остатков:

• уравнение регрессии может описывать свойства тех объектов, которых вообще нет;

• доверительные интервалы регрессии и прогноза могут не соответствовать фактическим значениям остатков;

• оценки МНК являются несмещенными, состоятельными, но неэффективными;

• стандартные ошибки коэффициентов модели будут занижены, а, следовательно, t-статистика — завышена, что приведет к неправильной оценке достоверности коэффициентов модели [1, с. 203].

Тесты проверки достоверности гетероскедастичности остатков.

Для обнаружения гетероскедастичности остатков используется визуальный анализ графика зависимости У от X, линии тренда и остатков.

Плохая спецификация визуально обнаруживается с помощью сравнения существующей тенденции с линией уравнения регрессии.

Для визуального обнаружения неоднородности данных на графике зависимости У от X точки исходных данных метят разным цветом, соответствующие однородным объектам. Визуальное выделение групп точек с одинаковым цветом позволяет обнаружить неоднородность данных. При этом для каждой группы однородных объектов может быть своя тенденция. Этот прием эффективен при анализе пространственных выборок, когда объекты анализа относятся к разным сферам деятельности или территории.

Для проверки достоверности гетероскедастичности остатков используется тест Голдфелда-Квандта.

Тест Голдфелда-Квандта предполагает, что дисперсия возмущения возрастает пропорционально квадрату одной из объясняющих переменных, возмущения распределены нормально и не подвержены автокорреляции.

Все n наблюдений в выборке упорядочиваются по величине X; отбрасываются "c" наблюдений, оказавшиеся в центре; оцениваются МНК отдельные регрессии для первых $n_1 = (n - c)/2$ и последних $n_2 = (n - c)/2$ наблюдений. Обозначая дисперсии остатков регрессий для первых и вторых наблюдений соответственно как S_1^2 и S_2^2, рассчитаем значение критерия Фишера

$$F = S_2^2/S_1^2,$$

которое сравнивается с критическим $F_{кр}(\alpha = 0{,}05;\ m_1 = n_2 - k;\ m_2 = n_1 - k)$.

Если $F \geq F_{кр}(\alpha = 0{,}05;\ m_1 = n_1 - k;\ m_2 = n_2 - k)$, то с вероятностью 1 — а утверждается, что гетероскедастичность остатков является достоверной, в противном случае наличие гетроскедастичности является не доказанной.

Мощность критерия зависит от выбора значения n_1 и n_2 по отношению к n. Обычно выбираются $n_1 = n_2$ таким образом, чтобы вся совокупность разделилась на три равные части. Однако С. Голдфелд и Р. Квандт уточняют это правило и рекомендуют брать значения $n_1 = n_2 = 11$, если $n = 30$ и $n_1 = n_2 = 22$, если $n = 60$ [1].

Методы устранения гетероскедастичности остатков

Известно несколько методов устранения гетероскедастичности остатков:

- метод выделения однородных объектов,
- обобщенный метод наименьших квадратов, или метод Эйткена.

Метод выделения однородных объектов

Самый простой метод — это выделение группировкой однородных объектов, для которых остатки модели будут одинаковыми.

Обобщенный метод наименьших квадратов и взвешенная регрессия

Допустим имеется линейное регрессионное уравнение

$$У_i = a_0 + a_1X_1 + e_i,$$

в котором остатки гетероскедастичны.

Если все уравнение разделить на модуль остатков e_i, то для новых переменных остатки новой модели будут принимать значения +1 или -1.

$$У_i/|e_i| = a_0/|e_i| + a_1X_1/|e_i| + e_i/|e_i| .$$

Следовательно, остатки новой модели будут гомоскедастичны. Однако возникают трудности в экономической интерпретации новых переменных.

Коэффициенты модели, определенные по преобразованным данным МНК, называются коэффициентами обобщенного метода наименьших квадратов, или коэффициентами Эйткена.

Коэффициенты Эйткена обладают следующим свойством: если коэффициенты Эйткена использовать для непреобразованных данных, то уравнение регрессии пройдет ближе к тем значениям, которые имеют меньшие остатки.

Примечание. Имеется алгоритм, с помощью которого любая функция может проходить ближе к тем значениям, вес которых будет большим. Например, известно, что последнее значение временного ряда является наиболее информативным для прогноза. Следовательно, можно на аппроксимационную функцию наложить ограничение — она должна проходить как можно ближе к последнему значению временного ряда. Допустим, для прогнозирования была выбрана линейная функция (можно выбрать любую функцию)

$$У = a_0+a_1t + e,$$

а также имеются данные временного ряда У. Введем новую переменную Z с численными значениями, равными единице для всех значений У, кроме последнего, которому дадим значение, равное 10. Умножим линейную функцию на Z, получим новую модель

$УZ = в_0Z + в_1tZ + e$.

Можно построить график зависимости УZ от tZ. Последнее значение временного ряда будет сильно выделяться среди всех остальных значений. При этом уравнение регрессии будет стремиться пройти ближе именно к этой точке, чего и хотелось исследователю.

Коэффициенты $в_0$ и $в_1$ модели $УZ = в_0Z + в_1tZ + e$ можно вычислить методом наименьших квадратов с помощь функции Excel "Линейн" без свободного коэффициента.

Если вычислить расчетные значения

$У_p = в_0 + в_1t$,

то они пройдут ближе к последнему значению временного ряда.

Предложенный алгоритм был использован для получения прогноза курса доллара с помощью взвешенной параболы. Наибольшие веса давались трем последним значениям временного ряда. Прогнозы подтверждались с очень малой ошибкой, хотя модель получилась недостоверной, с большой ошибкой на обучающей совокупности.

Приводим подробное описание методики расчета матрицы P, предназначенной для преобразований данных, а также прием получения доверительных интервалов уравнения регрессии для исходных данных при наличии гетероскедастичности остатков.

Постановка задачи.

Изучается модель

$\mathbf{У} = \mathbf{Xb} + \mathbf{u}$,

где **У** — вектор-столбец зависимой переменной;

X — матрица, которая состоит из столбцов-факторов, включая вектор-столбец единиц для учета свободного коэффициента;

b — вектор-столбец коэффициентов модели;

u — вектор-столбец остатков.

Предполагается, что

$M(\mathbf{u}) = \mathbf{0}$;

$M(\mathbf{uu'}) = \mathbf{V}$,

где M — знак математического ожидания;

V — симметрическая, положительно определенная матрица.

Необходимо: изучить свойства матрицы **V**, предложить методы устранения гетероскедастичности и построения интервального прогноза в условиях гетероскедастичности остатков.

Изучим свойства матрицы **V**.

Если количество измерений составит $n = 3$, то матрица **V** будет иметь следующий вид:

$$\mathbf{V} = \begin{pmatrix} M(u_1u_1) & M(u_1u_2) & M(u_1u_3) \\ M(u_2u_1) & M(u_2u_2) & M(u_2u_3) \\ M(u_3u_1) & M(u_3u_2) & M(u_3u_3) \end{pmatrix},$$

где диагональные элементы матрицы **V** означают дисперсию остатков, недиагональные элементы — ковариацию остатков между соответствующими измерениями и характеризуют автокорреляцию остатков разного порядка.

Свойство 1.

У симметрической матрицы $v_{ij} = v_{ji}$.

Например.

$$\mathbf{V} = \begin{pmatrix} 2 & 3 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix}.$$

Свойство 2.

От симметрической матрицы можно найти обратную матрицу.

Например,

$$\mathbf{V} = \begin{pmatrix} 2 & 3 & 0 \\ 3 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix}; \quad \mathbf{V}^{-1} = \begin{pmatrix} 4 & 3 & 0 \\ 3 & -2 & 0 \\ 0 & 0 & 0{,}1667 \end{pmatrix}.$$

Свойство 3.

Матрица **V** является положительно определенной, если ее собственные значения и определитель будут положительными.

Следовательно, дисперсии остатков должны быть больше ковариаций остатков.

Собственные значения **k** матрицы **V** можно определить из уравнения

$$\mathbf{Vh} = k\mathbf{h},$$

где **h** — собственный (характеристический) вектор матрицы **V**;

k — собственное (характеристическое) значение матрицы **V**.

Для каждого характеристического значения имеется свой характеристический вектор. Например,

$$\mathbf{V} = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 4 & 8 \\ 3 & 8 & 6 \end{pmatrix}.$$

Определим по одной из программ “Stat” или “Mat Lab” собственные вектора и собственные значения матрицы **V**.

$$\mathbf{k} = \begin{pmatrix} 13{,}78669 & 0 & 0 \\ 0 & -3{,}3764 & 0 \\ 0 & 0 & 1{,}58971 \end{pmatrix};$$

$$\mathbf{h} = \begin{pmatrix} 0{,}24147 & -0{,}246140 & 0{,}93867 \\ -0{,}62884 & -0{,}69703 & -0{,}34454 \\ 0{,}73909 & 0{,}67247 & -0{,}01353 \end{pmatrix}.$$

Матрица **V** не является положительно определенной, так как второе ее собственное значение отрицательное.

Проверим соотношение

$$\mathbf{Vh} = \mathbf{kh} \text{ или } \mathbf{h'Vh} = \mathbf{k},$$

$$\mathbf{h'Vh} = \begin{pmatrix} 13{,}787 & 0 & 0 \\ 0 & -3{,}372 & 0 \\ 0 & 0 & 1{,}5897 \end{pmatrix} = \mathbf{k}.$$

Свойство 4.

Если **V** положительно определенная матрица, то можно найти такую невырожденную матрицу **P**, позволяющую составить выражение

$$\mathbf{V} = \mathbf{PP}',$$

где $\mathbf{P} = [(\mathbf{hD})^{-1}]'$;

$$\mathbf{D} = \mathbf{I}\frac{1}{\sqrt{\mathbf{h}'\mathbf{Vh}}} = \mathbf{I}\frac{1}{\sqrt{\mathbf{k}}};$$

I — единичная матрица[1].

Примечание. Матрица **P** используется для преобразований данных.

Например,

$$\mathbf{V} = \begin{pmatrix} 10 & 1 & 2 \\ 1 & 20 & 3 \\ 2 & 3 & 25 \end{pmatrix}; \quad \mathbf{k} = \begin{pmatrix} 26{,}70426 \\ 18{,}59523 \\ 9{,}70051 \end{pmatrix};$$

$$\mathbf{h} = \begin{pmatrix} 0{,}13264 & 0{,}0062 & 0{,}99115 \\ 0{,}42126 & 0{,}90482 & -0{,}06204 \\ 0{,}89719 & -0{,}425750.67247 & -0{,}1174 \end{pmatrix};$$

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad \mathbf{D} = \begin{pmatrix} \frac{1}{\sqrt{k_1}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{k_2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{k_3}} \end{pmatrix};$$

$$\mathbf{D} = \begin{pmatrix} 0{,}193513 & 0 & 0 \\ 0 & 0{,}231899 & 0 \\ 0 & 0 & 0{,}321072 \end{pmatrix}.$$

Определим $\mathbf{P} = [(\mathbf{hD})^{-1}]'$.

---

[1] *Джонстон Д.* Эконометрические методы. — М.: Статистика, 1980. С. 112–113.

$$\mathbf{hD} = \begin{pmatrix} 0{,}025668 & 0{,}001438 & 0{,}318231 \\ 0{,}081519 & 0{,}209827 & -0{,}01992 \\ 0{,}173618 & -0{,}09873 & -0{,}03769 \end{pmatrix};$$

$$(\mathbf{hD})^{-1} = \begin{pmatrix} 0{,}685425 & 2{,}176854 & 4{,}636337 \\ 0{,}02676 & 3{,}901762 & -1{,}83596 \\ 3{,}086969 & -0{,}19321 & -0{,}36566 \end{pmatrix};$$

$$\mathbf{P} = [(\mathbf{hD})^{-1}]' = \begin{pmatrix} 0{,}685425 & 0{,}02676 & 3{,}086969 \\ 2{,}176865 & 3{,}901762 & -0{,}19321 \\ 4{,}636337 & -1{,}83596 & -0{,}36566 \end{pmatrix}.$$

Проверим условие **V=PP'**.

$$\mathbf{PP}' = \begin{pmatrix} 9{,}999901 & 1{,}000056 & 1{,}99995 \\ 1{,}000056 & 19{,}99982 & 2{,}99985 \\ 1{,}99995 & 2{,}99985 & 25{,}00008 \end{pmatrix}.$$

$$\mathbf{V} = \begin{pmatrix} 10 & 1 & 2 \\ 1 & 20 & 3 \\ 2 & 3 & 25 \end{pmatrix}.$$

Условие **V=PP'** (с учетом ошибок округления) соблюдается.

Свойство 5.

Если **V** диагональная матрица, то расчет матрицы $\mathbf{P}^{-1}$ упрощается.

Если матрица **V** диагональная, то справедливо выражение

$$\mathbf{V} = \mathbf{I}u^2.$$

Диагональная матрица **V** обладает свойствами:

1) обратная матрица $\mathbf{V}^{-1}$ имеет такой вид

$$\mathbf{V}^{-1} = \mathbf{I}\left(\frac{1}{v_{ii}}\right) = \mathbf{I}\left(\frac{1}{u_{ii}^2}\right).$$

где u_{ij}^2 — квадраты возмущения;

2) если выполняется условие

PP'=V,

то справедливы соотношения:

$$(\mathbf{P}^{-1})'\mathbf{P}^{-1} = \mathbf{P}^{-1}\mathbf{P}^{-1} = \mathbf{V}^{-1}.$$

3) если использовать первое и второе свойства диагональной матрицы, то можно получить выражение для матрицы $\mathbf{P}^{-1}$

$$\mathbf{P}^{-1} = \sqrt{\mathbf{V}^{-1}} = \mathbf{I}\left(\frac{1}{|u_{ij}|}\right).$$

Полученное выражение является теоретическим обоснованием устранения гетероскедастичности остатков способом деления уравнения регрессии на модуль остатков.

Примечание. Все расчеты проводились в Excel с использованием матричных операций.

Определим разными способами коэффициенты линейной модели обобщенным методом наименьших квадратов или методом Эйткена.

Расчет коэффициентов модели методом Эйткена с использованием матрицы $\mathbf{V}^{-1}$

Расчет коэффициентов линейной модели производится методом Эйткена по следующей формуле

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{У}$$

Например. Имеется база данных (табл. 11.1).

Таблица 11.1

Исходные данные

№	X_0	X_1	У	e
1	2	3	4	5
1	1	1	10	0,1966
2	1	3	12	-4,538
3	1	4	16	-3,906
4	1	6	40	13,359
5	1	10	35	-5,111

Рассчитаем характеристики линейной модели с помощью функции "Линейн" (Excel).

$У = a_0 + a_1X + e = 6{,}4359 + 3{,}3675X + e,$

где $a_0 = 6{,}4359$; $a_1 = 33{,}3675$.

Вычислим остатки **e** и составим матрицу **V**.

$$\mathbf{V} = \mathbf{I}\mathbf{e}^2 = \begin{pmatrix} 0{,}0386 & 0 & 0 & 0 & 0 \\ 0 & 20{,}598 & 0 & 0 & 0 \\ 0 & 0 & 15{,}257 & 0 & 0 \\ 0 & 0 & 0 & 178{,}46 & 0 \\ 0 & 0 & 0 & 0 & 26{,}123 \end{pmatrix}; \mathbf{У} = \begin{pmatrix} 10 \\ 12 \\ 16 \\ 40 \\ 35 \end{pmatrix};$$

$$\mathbf{X}' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 10 \end{pmatrix};$$

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{У} = \begin{pmatrix} 7{,}3019 \\ 2{,}6913 \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}.$$

Полученные коэффициенты обладают следующими свойствами:

• коэффициентов модели, вычисленные МНК и методом Эйткена, отличаются;

• расчетные значения $У_р$ зависимой переменной, определенной по уравнению $\mathbf{У}_р = \mathbf{Xb}$, проходят ближе к тем значениям У, для которых остатки модели были меньше. Можно подобрать такие значения матрицы **V**, при которых уравнение регрессии будет проходить ближе к более информативным измерениям (например, к последним значениям временного ряда), где **X** — матрица исходных данных, **b** — коэффициенты модели, определенные методом Эйткена;

• с помощью коэффициентов **b** можно найти точечный прогноз обычным способом $\mathbf{У}_{пр} = \mathbf{X}_{ож}\mathbf{b}$ (где $\mathbf{X}_{ож}$ — ожидаемые значения исходных данных), однако нет формул для расчета интервального прогноза.

Известен второй способ расчета коэффициентов модели методом Эйткена, суть которого заключается в том, что исходные данные преобразуются так, что в них отсутствует гетероскедастичность и автокорреляция остатков. Для новых данных рассчитываются коэффициенты модели обычным методом наименьших квадратов. Все характеристики для модели с новыми данными рассчитываются по стандартным формулам. Если для точечных и интервальных прогнозов, определенных по преобразованным данным, произвести обратные преобразования, то можно получить интервальные прогнозы, учитывающие гетероскедастичность остатков для исходных данных.

Представим матрицу $\mathbf{V}$ в виде произведения двух матриц

$$\mathbf{V} = \mathbf{PP}'.$$

Умножим уравнение модели $\mathbf{У} = \mathbf{Xa} + \mathbf{e}$ слева на матрицу $\mathbf{P}^{-1}$, получим

$$\mathbf{У}_н = \mathbf{X}_н\mathbf{a} + \mathbf{e}_н,$$

где $\mathbf{У}_н = \mathbf{P}^{-1}\mathbf{У}$; $\mathbf{X}_н = \mathbf{P}^{-1}\mathbf{X}$, $\mathbf{e}_н = \mathbf{P}^{-1}\mathbf{e}$;

$\mathbf{P}^{-1}$ — обратная матрица, которую можно найти по формулам свойств 4 и 5 матрицы $\mathbf{V}$.

Если матрица $\mathbf{V}$ — недиагональная, то матрица $\mathbf{P}^{-1}$ вычисляется по формуле

$$\mathbf{P}^{-1} = \{[(\mathbf{hD})^{-1}]'\}^{-1}.$$

Если матрица $\mathbf{V}$ — диагональная, то матрица $\mathbf{P}^{-1}$ вычисляется по формуле

$$\mathbf{P}^{-1} = \sqrt{\mathbf{V}^{-1}} = \mathbf{I}\left(\frac{1}{|e_{ij}|}\right).$$

База данных расположена в табл. 11.1.

Допустим матрица $\mathbf{V}$-диагональная (предполагается, что автокорреляция возмущения отсутствует) и состоит из квадратов остатков, вычисленных по уравнению регрессии с коэффициентами, найденными МНК по исходным данным.

Вычислим диагональную матрицу **V**, элементами которой будут квадраты остатков, расположенных в колонке 5 табл. 11.1.

$$\mathbf{V} = \mathbf{I}\mathbf{u}^2 = \begin{pmatrix} 0{,}038644 & 0 & 0 & 0 & 0 \\ 0 & 20{,}59763 & 0 & 0 & 0 \\ 0 & 0 & 15{,}2567 & 0 & 0 \\ 0 & 0 & 0 & 178{,}4622 & 0 \\ 0 & 0 & 0 & 0 & 26{,}12346 \end{pmatrix}$$

Если матрица **V** диагональная, то на основании ее свойства 4 матрица $\mathbf{P}^{-1}$ вычисляется по формуле

$$\mathbf{P}^{-1} = \sqrt{\mathbf{V}^{-1}} = \mathbf{I}\left(\frac{1}{\left|\mathbf{c}_{ij}\right|}\right)$$

$$\mathbf{V}^{-1} = \begin{pmatrix} 25{,}877 & 0 & 0 & 0 & 0 \\ 0 & 0{,}0485 & 0 & 0 & 0 \\ 0 & 0 & 0{,}0655 & 0 & 0 \\ 0 & 0 & 0 & 0{,}0056 & 0 \\ 0 & 0 & 0 & 0 & 0{,}0383 \end{pmatrix}$$

$$\mathbf{P}^{-1} = \sqrt{\mathbf{V}^{-1}} = \begin{pmatrix} 5{,}087 & 0 & 0 & 0 & 0 \\ 0 & 0{,}2203 & 0 & 0 & 0 \\ 0 & 0 & 0{,}256 & 0 & 0 \\ 0 & 0 & 0 & 0{,}0749 & 0 \\ 0 & 0 & 0 & 0 & 0{,}1957 \end{pmatrix}$$

Умножим уравнение модели $\mathbf{У} = \mathbf{Xa} + \mathbf{e}$ слева на матрицу $\mathbf{P}^{-1}$, получим

$$\mathbf{У}_н = \mathbf{X}_н \mathbf{b} + \mathbf{e}_н,$$

где $\mathbf{У}_н = \mathbf{P}^{-1}\mathbf{У}$; $\mathbf{X}_н = \mathbf{P}^{-1}\mathbf{X}$; $\mathbf{e}_н = \mathbf{P}^{-1}\mathbf{e}$.

$$У_н = P^{-1}У = \begin{pmatrix} 50{,}87 \\ 2{,}6441 \\ 4{,}0963 \\ 2{,}9942 \\ 6{,}8478 \end{pmatrix}; \quad X_н = P^{-1}X = \begin{pmatrix} 5{,}087 & 5{,}087 \\ 0{,}2203 & 0{,}661 \\ 0{,}256 & 1{,}0241 \\ 0{,}0749 & 0{,}4491 \\ 0{,}1957 & 1{,}9565 \end{pmatrix}.$$

Вычислим коэффициенты **b** по новым переменным с помощью функции “Линейн” (Excel):

$b_0 = 7{,}3019$; $b_1 = 2{,}6913$.

Сравним полученные результаты с расчетами по формуле

$$\mathbf{b} = \left(X'V^{-1}X\right)^{-1} X'V^{-1}У = \begin{pmatrix} 7{,}3019 \\ 2{,}6913 \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}.$$

Анализ полученных результатов показывает, что коэффициенты b_0 и b_1, определенные разными способами, равны между собой.

Вычислим расчетные значения $У_{рн}$ и остатки $e_н$ для новых данных.

$$У_{рн} = b_0 X_{0н} + b_1 X_{1н} = \begin{pmatrix} 50{,}835 \\ 3{,}38779 \\ 4{,}6255 \\ 1{,}7553 \\ 6{,}6942 \end{pmatrix}; \quad e_н = У_н - У_{рн} = \begin{pmatrix} 0{,}0347 \\ -0{,}744 \\ -0{,}529 \\ 1{,}2389 \\ 0{,}1536 \end{pmatrix}.$$

Анализ остатков $e_н$ показывает, что они по модулю значительно отличаются и их квадраты тоже имеют значительный разброс, поэтому очевидно гетероскедастичность полностью не устранена.

Если матрица **V** является диагональной и численно равной квадратам остатков, то самый простой метод устранения неоднородности остатков заключается в том, что в уравнении **У = Xa + e** все его переменные необходимо разделить на модуль остатков е. При этом новые остатки будут иметь значения -1 или +1. Этот метод является следствием свойства 5 матрицы **V**. Для преобразованных данных применяется обычный регрессионный анализ с получением всех характеристик модели и доверительных интервалов прогноза и уравнения регрессии. Для получения доверительных интервалов прогноза и уравнения регрессии для исходных данных необходимо их значения, полученные по преобразованным данным, умножить на модуль остатков е.

Метод Эйткена позволяет получить доверительные интервалы уравнения регрессии с учетом гетероскедастичности остатков.

Расчетные формулы метода Эйткена:

$У_{рас.Э.i} = в_1 + в_2X_i$, — расчетные значения У, определенные методом Эйткена;

$У_{пр.Э.} = У_{н.пр.} |e_{ож.}|$, — прогнозное значение У, определенное методом Эйткена;

$У_{н.пр(i=n+1)} = в_1X_{1н(i=n+1)} + в_2X_{2н(i=n+1)}$;

$У_{пр.мин.Э.} = У_{н.пр.мин.} |e_{ож.}|$ — нижний прогнозный доверительный интервал;

$У_{пр.мак.Э.} = У_{н.пр.мак} |e_{ож.}|$ — верхний прогнозный доверительный интервал;

$У_{н.пр.мин.} = У_{н.пр(i=n+1)} - t_{\alpha/2}E$;

$У_{н.пр.мак.} = У_{н.пр(i=n+1)} + t_{\alpha/2}E$;

где $в_1$, $в_2$ — коэффициенты, определенные по модели

$У_{нi} = в_1X_{1нi} + в_2X_{2нi}$;

$У_{нi} = У_i/|e_i|$, $X_{1нi} = 1/|e_i|$, $X_{2нi} = X_i/|e_i|$;

$У_i$, $У_{нi}$ — фактические и преобразованные значения зависимой переменной;

X_i, $X_{нi}$ — фактические и преобразованные значения объясняемой переменной;

$e_i = У_i - У_{pi}$, $У_{pi} = a_0 + a_1X_i$;

a_0, a_1 — коэффициенты, определенные методом наименьших квадратов;

$e_{ож}$ — ожидаемое значение остатка;

$t_{\alpha/2}$ ($\alpha = 0{,}05$, $m = n - к$) — критерий Стьюдента;

Е — ошибка модели;

n — объем выборки;

к — количество коэффициентов в модели.

Приводим результаты расчетов для условных данных (рис. 11.4).

Из рис. 11.4 видно, что чем больше остатки модели, тем больше доверительный интервал уравнения регрессии, что и требовалось получить.

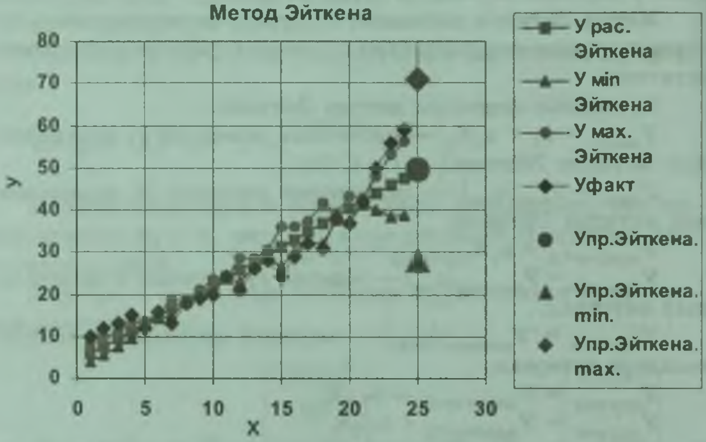


Рис. 11.4. Результат использования метода Эйткена для построения доверительных интервалов.

11.2. Прогнозирование при наличии гетероскедастичности остатков

В эконометрической литературе пока не стоит вопрос о прогнозировании зависимой переменной при наличии гетероскедастичности остатков. Большинство исследований заканчивается на этапе преобразования данных и устранения гетероскедастичности для преобразованных данных. Однако нам необходимо знать прогнозные значения для фактических данных, а не для преобразованных данных. Пока на этот вопрос четкого ответа нет. Предлагается приблизительное решение данной проблемы.

Если произвести полный эконометрический анализ для преобразованных данных, деленных на модуль остатков, а затем полученные точечный прогноз, доверительные интервалы для прогноза и уравнения регрессии умножить на модуль остатков e_i, то новые точечный и интервальный прогнозы и доверительные интервалы уравнения регрессии будут учитывать свойства остатков для фактических непреобразованных значений зависимой переменной. Данное утверждение является интуитивным и требует доказательства. Практическая реализация данного метода дает хороший результат.

11.3. Адаптивные модели прогнозирования Брауна, Хольта, Бокса-Дженкинса, Уинтерса, Тейла-Вейджа

Адаптивная модель прогнозирования Брауна

Характерной чертой адаптивных методов прогнозирования является их способность непрерывно учитывать эволюцию динамических характеристик изучаемых процессов, приспосабливающихся под эту эволюцию, придавая больший вес тем значениям, которые ближе к текущему моменту прогнозирования.

Все адаптивные методы хорошо воспроизводят плавно изменяющиеся значения временного ряда и плохо аппроксимируют резко изменяющиеся данные и выбросы. Аппроксимированный временной ряд с помощью адаптивных методов повторяет резко изменяющиеся данные, но сдвинутыми во времени на одну дату, что значительно ухудшает качество модели. Это негативное свойство адаптивных методов прогнозирования следует учитывать при подборе коэффициентов сглаживания.

Экспоненциально взвешенное среднее значение вычисляется по формуле

$Z_t = \lambda У_t + (1 - \lambda)Z_{t-1}$.

(обоснование формулы см. в разд. 9.2. — Метод экспоненциального скользящего среднего (Метод Брауна)).

Если в качестве "наивного" прогноза принять следующий алгоритм: "прогноз спроса на некоторый товар в следующем месяце равен спросу на товар в этом месяце", то можно утверждать следующее: прогноз временного ряда на дату $t + 1$ равен экспоненциальной взвешенной средней на момент времени t. В соответствии с этим и при условии, что

$У_t = a_0 + e_t$,

где a_0 — неизвестный параметр, не зависящий от времени;

e_t — случайный остаток со средним, равным нулю, и конечной дисперсией, адаптивная модель прогнозирования Брауна имеет следующий вид:

$П_{t+1} = Z_t = \lambda У_t + (1 - \lambda)Z_{t-1} = \lambda У_t + (1 - \lambda)П_t$;

$П_{t+1} = П_t + \lambda (У_t - П_t)$,

где $П_{t+1}$ — прогноз временного ряда на момент времени $t+1$;

Z_t — экспоненционально взвешенное среднее на момент времени t;

$У_t - П_t$ — ошибка прогноза;

$П_1$ — первое прогнозное значение определяется экспертным способом, рассчитывают его с помощью регрессионного анализа или считают равным первому значению временного ряда;

λ — коэффициент сглаживания.

Адаптивный метод прогнозирования Брауна обладает следующими свойствами:

• для построения прогноза по экспоненциально взвешенному среднему необходимо знать только начальную оценку прогноза, дальнейшее прогнозирование возможно по поступлении свежих данных;

• в экспоненциально взвешенном среднем значения весов убывают со временем, поэтому здесь нет точки, на которой веса обрываются;

• чем больше λ, тем выше чувствительность среднего, чем меньше λ, тем устойчивее становится экспоненциально взвешенное среднее;

• изменяя коэффициент λ, можно найти его оптимальное значение по признаку ошибки модели. Если оптимальное значение $0 \le \lambda \le 0{,}3$, то временной ряд является стационарным, если $1 \ge \lambda > 0{,}3$, то временной ряд является нестационарным и следует перейти к моделям, учитывающим тенденцию временного ряда;

• прогнозы, основанные на экспоненциально взвешенном среднем хорошо воспроизводят гладкие временные ряды, но очень плохо воспроизводят выбросы;

• если временной ряд имеет линейную тенденцию роста, то экспоненциальная средняя приводит к смещенным прогнозам[1].

Адаптивная модель прогнозирования Хольта

Экспоненциальная средняя приводит к смещенным прогнозам, т. е. дает систематическую ошибку, когда временной ряд имеет тенденцию линейного роста. Для этого случая разработано несколько вариантов адаптивных моделей, также использующих процедуру экспоненциального сглаживания. В основе моделей лежит гипотеза о том, что прогноз может быть получен по уравнению

$$\Pi_{t+k} = a_{0,t} + da_{1,t},$$

где $a_{0,t}$, $a_{1,t}$ — текущие оценки коэффициентов адаптивного полинома первого порядка;

d — глубина прогноза.

Одной из первых моделей этого типа была двухпараметрическая модель Ч. Хольта, в которой оценка коэффициентов производится следующим образом:

$$\Pi_{t+1} = a_{0,t} + a_{1,t},$$

[1] *Льюис К. Д.* Методы прогнозирования экономических показателей. — М.: Финансы и статистика, 1986; *Лукашин Ю. П.* Адаптивные методы краткосрочного прогнозирования. — М.: Статистика, 1979.

$$a_{0,t} = \lambda_1 У_t + (1 - \lambda_1)(a_{0,t-1} + a_{1,t-1}) = \lambda_1 У_t + (1 - \lambda_1)П_t;$$

$$a_{1,t} = \lambda_2(a_{0,t} - a_{0,t-1}) + (1 - \lambda_2)a_{1,t-1},$$

где λ_1, λ_2 — параметры экспоненциального сглаживания ($0 < \lambda_1, \lambda_2 < 1$).

Оптимальное значение параметров λ_1, λ_2 можно определить по минимальной ошибке модели.

Адаптивные модели прогнозирования Бокса-Дженкинса

Если модель Хольта усовершенствовать путем включения разности ошибок, то получится полная трехпараметрическая модель Дж. Бокса и Г. Дженкинса:

$$П_{t+1} = a_{0,t} + a_{1,t};$$

$$a_{0,t} = \lambda_1 У_t + (1 - \lambda_1)(a_{0,t-1} + a_{1,t-1}) + \lambda_3(e_t - e_{t-1});$$

$$a_{1,t} = \lambda_2(a_{0t} - a_{0,t-1}) + (1 - \lambda_2)a_{1,t-1},$$

где λ_1, λ_2, λ_3 — параметры экспоненциального сглаживания ($0 < \lambda_1, \lambda_2, \lambda_3 < 1$).

Адаптивные модели прогнозирования Уинтерса

Модель Хольта была развита Уинтерсом так, чтобы она охватывала помимо линейного тренда еще и сезонные эффекты. Прогноз, сделанный в момент t на k тактов времени вперед, равен:

$$П_{t+k} = (a_{0,t} + ka_{1,t}) + w_{t+k-T};$$

где w_{t+k-T} — коэффициент сезонности;

T — число временных тактов, содержащихся в полном сезонном цикле (для месячных данных $T = 12$).

Модель Уинтерса содержит три параметра сглаживания λ_1, λ_2, λ_3 ($0 < \lambda_1, \lambda_2, \lambda_3 < 1$).

Коэффициенты модели вычисляются по следующим формулам:

$$a_{0,t} = \lambda_1 \frac{У_t}{w_{t-T}} + (1 - \lambda_1)(a_{0,t-1} + a_{1,t-1});$$

$$w_t = \lambda_2 \frac{У_t}{a_{0,t}} + (1-\lambda_2) w_{t-T};$$

$$a_{1,t} = \lambda_3(a_{0,t} - a_{0,t-1}) + (1-\lambda_3)a_{1,t-1},$$ [1, с. 325–327].

Адаптивная аддитивная модель сезонности Тейла-Вейджа

Предположим, что исходный временной ряд был преобразован так, чтобы его можно аппроксимировать аддитивной моделью:

$У_t = a_{0,t} + w_t + \delta_t$;

$a_{0,t} = a_{0,t-1} + a_{1,t}$,

где $a_{0,t}$ — уровень процесса после выделения сезонных колебаний;

$a_{1,t}$ — аддитивный коэффициент роста;

w_t — аддитивный коэффициент сезонности;

δ_t — белый шум.

Прогноз, сделанный на момент t на k временных тактов вперед, вычисляется по следующей формуле:

$П_{t+k} = a_{0,t} + ka_{1,t} + w_{t+k-T}$,

где $a_{0,t} = a_{0,t-1} + a_{1,t-1} + \lambda_1(У_t - П_t)$;

$a_{1,t} = a_{1,t-1} + \lambda_1\lambda_2(У_t - П_t)$;

$w_t = w_{t-T} + (1 - \lambda_1)l_3(У_t - П_t)$;

$\lambda_1, \lambda_2, \lambda_3$ $(0 < \lambda_1, \lambda_2, \lambda_3 < 1)$ — параметры адаптации;

k = 1 — глубина прогноза;

T — число временных тактов, содержащихся в полном сезонном цикле [1, с. 326–327].

Вопросы для самоконтроля

1. Дайте определение гетероскедастичности и гомоскедастичности.

2. Укажите причины возникновения гетероскедастичности остатков.

3. Укажите последствия гетороскедастичности.

4. Укажите критерий проверки достоверности гетероскедастичности.

5. Укажите методы устранения гетероскедастичности.

6. Укажите на особенность прогнозирования при наличия гетероскедастичности остатков.

7. Укажите характерные черты адаптивных методов прогнозирования.

8. Приведите виды моделей прогнозирования Брауна, Хольта.

9. Приведите вид модели прогнозирования Бокса-Дженкинса. Уинтерса, Тейла-Вейджа.

Глава 12
СИСТЕМА ОДНОВРЕМЕННЫХ УРАВНЕНИЙ

12.1. Определение и назначение системы одновременных уравнений

В предыдущих темах изучались парные и множественные уравнения регрессии изолированно друг от друга. Однако в экономической системе многие переменные связаны между собой и зависимые переменные могут входить в состав другого уравнения как объясняемые переменные. При этом зависимые переменные получают новое свойство, которое заключается в том, что между ними установится обратная связь. Наличие обратной связи между зависимыми переменными требует решения следующих вопросов.

1. Как будут вести себя зависимые переменные в системе с обратной связью?

2. Как получить прогнозное значение зависимой переменной при известных значениях объясняемых переменных (прямая задача)?

3. Как оценить параметры уравнения регрессии, если объясняемая переменная и остатки будут связаны между собой?

4. Как вычислить значения объясняемых переменных для получения желаемого значения зависимой переменной (обратная задача)?

Для решения поставленных задач необходимо ввести новые понятия.

Функционирование экономической системы предполагает взаимосвязь ее переменных. Системой одновременных уравнений (СОУ) называют такую систему уравнений, которая учитывает взаимовлияние основных переменных экономической системы.

Пример. Приведем схему связей переменных магазина (рис. 12.1).

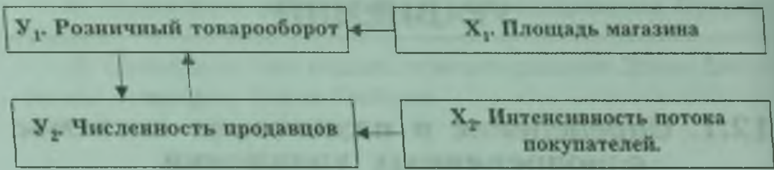


Рис. 12.1. Структурная схема взаимосвязи переменных

Анализ рис. 12.1 показывает, что имеются две группы переменных, которые отличаются направленностью воздействия на переменные. Переменные первой группы $У_1$ и $У_2$ влияют и зависят от других переменных, они получили название внутренних переменных, или эндогенных переменных. Переменные второй группы X_1 и X_2 только влияют на переменные, но не зависят от других переменных, они получили название внешних переменных, или экзогенных переменных.

Изменение одной переменной приводит к изменению других до перехода системы в динамическое равновесное состояние.

Эндогенные переменные зависят от переменных системы и могут влиять на остальные переменные. Эндогенные переменные принято обозначать буквой У.

Экзогенные переменные не зависят от деятельности системы, но могут влиять на эндогенные переменные. Экзогенные переменные принято обозначать буквой X. Экзогенными переменными можно условно считать инвестиции в экономику, гуманитарную помощь. Истинными экзогенными переменными можно считать лаговые эндогенные переменные, природные, космические факторы.

Свойства эндогенных переменных

Свойство 1. Эндогенные переменные имеют обратные связи, которые порождают проблему устойчивости экономической системы. Предположим, что в исходный момент времени эндогенные переменные находятся в среднединамическом равновесном состоянии (ярким примером служат устоявшиеся значения спроса, цены и предложения, изучаемые в микроэкономике). При изменении эндогенной переменной возможны следующие сценарии ее изменения:

- она плавно монотонно или циклически вернется к своему исходному значению, такое поведение эндогенной переменной характерно для устойчивой экономической системы;
- она плавно монотонно или циклически примет минимальное или максимальное возможные значения, такое поведение эндогенной переменной характерно для неустойчивой экономической системы.

Изучение условий устойчивости экономической системы относится к отдельному разделу экономической теории и в эконометрике пока не рассматривается.

На рис. 12.2 показаны возможные траектории поведения эндогенной переменной при ее изменении.

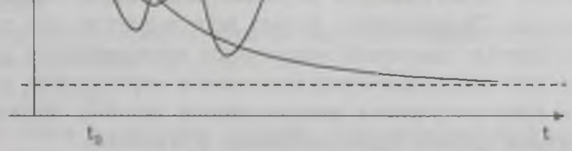


Рис. 12.2. Возможные варианты переходных процессов эндогенной переменной при ее изменении в момент времени t_0:
- центральная линия — равновесное значение переменной У;
- пунктирные линии — верхние и нижние допустимые значения У;
- сплошные линии — возможные траектории переходных процессов переменной У.

Свойство 2. Если экономическая система находится в устойчивом состоянии, то изменение эндогенной переменной не способно изменить средние динамические равновесные значения эндогенных переменных.

Свойства экзогенных переменных

Свойство 1. Предположим, что экономическая система находится в устойчивом состоянии и эндогенные переменные имеют средние динамические равновесные значения. Изменение одной или нескольких экзогенных переменных приводит через несколько циклов к изменению средних динамических значений эндогенных переменных.

Свойство 2. Прогнозные значения эндогенных переменных можно получить только с помощью экзогенных переменных.

Взаимосвязь экзогенных и эндогенных переменных описывается системой одновременных уравнений, которые принято называть эконометрической моделью.

Известны две формы одновременных уравнений: структурная и приведенная.

Структурная форма одновременных уравнений содержит в качестве объясняющих переменных как эндогенные, так и экзогенные переменные, которые отражают реальную структуру взаимосвязи переменных.

Приведенная форма одновременных уравнений содержит в качестве объясняющих переменных только экзогенные переменные. Приведенная форма используется для получения прогнозных значений эндогенных переменных и для получения расчетных значений эндогенных переменных, используемых для получения несмещенных оценок параметров структурной формы одновременных уравнений.

Приведем структурную систему одновременных уравнений для переменных, изображенных на рис. 12.1.

$$\begin{cases} У_1 = a_0 + a_1 У_2 + a_2 X_1 + e_1, \\ У_2 = в_0 + в_1 У_1 + в_2 X_2 + e_2. \end{cases}$$

Если в первом уравнении вместо $У_2$ подставить второе уравнение, а во втором уравнении вместо $У_1$ подставить первое уравнение, то после несложных преобразований можно получить приведенную систему одновременных уравнений.

$$\begin{cases} У_1 = c_0 + c_1X_1 + c_2X_2 + e_3; \\ У_2 = d_0 + d_1X_1 + d_2X_2 + e_4. \end{cases}$$

Однозначный переход от структурной системы к приведенной системе одновременных уравнений можно было произвести при условии идентифицируемости структурной системы одновременных уравнений.

Структурная система одновременных уравнений обладает такими свойствами:

- количество уравнений равно количеству эндогенных переменных;
- структурная система одновременных уравнений может быть: неидентифицируемой, идентифицируемой, сверхидентифицируемой. Поясним более подробно эти свойства системы одновременных уравнений. Система одновременных уравнений неидентифицируема, если для какого-нибудь уравнения системы выполняется неравенство:

$$n < (n_i + d_i),$$

где n — общее число всех экзогенных переменных системы;

n_i — число экзогенных переменных i-го уравнения;

d_i — число объясняющих эндогенных переменных i-го уравнения.

Модель идентифицируема, если для каждого уравнения системы выполняется равенство:

$$n = (n_i + d_i).$$

Из этого равенства следует, что количество всех эндогенных переменных должно равняться количеству экзогенных переменных. В каждом уравнении количество объясняющих переменных должно равняться количеству эндогенных переменных. Комбинации объясняющих переменных в каждом уравнении не должны повторяться.

Как правило, экзогенных переменных не хватает для того, чтобы система стала идентифицируемой, поэтому часто в качестве экзогенной переменной выбирают лаговую эндогенную переменную.

Модель сверхидентифицируемая, если хотя бы для одного уравнения системы выполняется неравенство

$n > (n_i + d_i).$

12.2. Методы определения коэффициентов системы одновременных уравнений

Для расчета коэффициентов структурной и приведенной системы одновременных уравнений можно использовать метод наименьших квадратов.

Эндогенные переменные могут быть случайными величинами, поэтому, находясь в системе одновременных уравнений в качестве объясняемой переменной, они могут быть связаны с остатком. Взаимосвязь объясняемых переменных с остатками модели является нарушением предпосылок Гаусса-Маркова для метода наименьших квадратов, что делает оценки параметров модели смещенными и неэффективными.

Докажим смещенность коэффициента a_1 линейной модели

$У_i= a_0+ a_1X_i + e_i,$

при условии, что X_i и e_i связаны между собой.

Коэффициент a_1 вычисляется методом наименьших квадратов по следующей формуле:

$$a_1 = \frac{\sum(У_i - \bar{У})(X_i - \bar{X})}{\sum(X_i - \bar{X})^2}.$$

Если вместо $У_i$ подставить их значения, вычисленные по регрессионной модели генеральной совокупности $У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$, раскрыть скобки, то получится следующее выражение:

$$a_1 = \alpha_1 + \frac{\sum (X_i - \bar{X})\, \varepsilon_i}{\sum (X_i - \bar{X})^2}.$$

Коэффициент a_1 будет равен своему математическому ожиданию, если

$\frac{\sum (X_i - \bar{X})\, \varepsilon_i}{\sum (X_i - \bar{X})^2} = 0$ при условии $\sum(X_i \varepsilon_i) = 0$, что означает отсутствие связи между X_i и ε_i.

Следовательно, если объясняемая переменная X не связана с остатками, то коэффициент a_1, вычисленный по методу наименьших квадратов, будет несмещенным. Если $\sum(X_i \varepsilon_i) \neq 0$, то значение a_1 будет смещенным.

При этом появляется возможность улучшить характеристики модели с помощью нескольких методов, учитывающих указанное нарушение предпосылки Гаусса-Маркова метода наименьших квадратов.

Двухшаговый метод наименьших квадратов

Задача. Необходимо получить несмещенные значения коэффициентов структурного уравнения регрессии

$У_{1t} = a_0 + a_1 У_{2t} + a_2 X_{1t} + e_t.$

Необходимо проверить предпосылку метода наименьших квадратов о независимости объясняемых переменных и остатков. Для этой цели вычислим коэффициенты структурного уравнения методом наименьших квадратов, вычислим остатки, вычислим суммы: $\sum У_{2t} e_t$, $\sum X_{1t} e_t$. Если эти суммы равны нулю, то полученные значения коэффициентов структурного уравнения будут несмещенными и дальнейшее улучшение коэффициентов можно не проводить. Если обе суммы или одна из них раны нулю, то для соответствующей переменной можно получить несмещенную оценку следующим образом.

Расчет коэффициентов модели

$У_{1t} = a_0 + a_1 У_{2t} + a_2 X_{1t} + e_t,$

при наличии связи между $У_{2t}$ и e_t (если $\sum У_{2t}e_t \neq 0$), производится в два шага.

На первом шаге устраняется зависимость $У_{2t}$ от e_t с помощью уравнения приведенной системы одновременных уравнений

$$У_{2pt} = b_0 + b_1X_{1t} + b_2X_{2t}.$$

Переменная $У_{2pt}$ не содержит случайной составляющей e_t.

На втором шаге рассчитываются коэффициенты модели

$$У_{1t} = a_0 + a_1У_{2pt} + a_2X_{1t} + e_t$$

методом наименьших квадратов. Так как $У_{2pt}$ не зависит от e_t, то коэффициенты a_0, a_1, a_2, определенные методом наименьших квадратов, будут эффективными из класса линейных и несмещенных коэффициентов.

Трехшаговый метод наименьших квадратов

Трехшаговый метод наименьших квадратов используется в качестве метода статистического оценивания параметров одновременно всех уравнений структурной системы одновременных уравнений с учетом возможной взаимной коррелированности регрессионных остатков различных уравнений этой системы. В трехшаговом методе наименьших квадратов сохраняются два первых шага, коэффициенты, полученные после выполнения второго шага, используются для получения остатков. Эти остатки используются для получения ковариационной матрицы случайных остатков, которая, в свою очередь, используется для вычисления всех структурных коэффициентов с помощью обобщенного метода наименьших квадратов в рамках соответствующим образом построенной обобщенной линейной модели множественной регрессии. Более подробно см. [1, с. 360–364].

Косвенный метод наименьших квадратов

При переходе от структурной к приведенной форме системы одновременных уравнений было отмечено, что в приведенной системе одновременных уравнений регрессионные остатки удовлетворяют условиям классической модели регрессии. Следовательно, оценки, полученные методом наименьших квадратов, будут несмещенными и эффективными. Если эти оценки использовать при обратном переходе к структурной системы одновременных уравнений, то полученные оценки будут несмещенными и эффективными. Данный метод получил название косвенного метода наименьших квадратов.

12.3. Прямая задача

Величину средних динамических равновесных эндогенных переменных можно получить по приведенной системе одновременных уравнений.

Если экзогенные переменные не изменяются, то спонтанное отклонение одной или нескольких эндогенных переменных от своих средних динамических равновесных значений приведет к появлению циклических волн, которые восстановят прежние среднединамические равновесные значения эндогенных переменных.

Если экзогенные переменные изменятся, то это приведет к появлению циклических волн и возникновению других средних динамических равновесных значений эндогенных переменных.

Определим прогнозные значения $У_1$ и $У_2$ при ожидаемых значениях X_1 и X_2.

Для прогноза $У_1$ и $У_2$ используют приведенную форму системы одновременных уравнений.

$$\begin{cases} У_1 = a_0 + a_1 У_2 + a_2 X_2; \\ У_2 = b_0 + b_1 У_1 + b_2 X_2. \end{cases}$$

Подставим в приведенную форму системы одновременных уравнений ожидаемые значения $X_{1ож}$, $X_{2ож}$ и получим прогнозные значения $У_{1п}$ и $У_{2п}$.

12.4. Обратная задача

Предположим, имеется приведенное уравнение регрессии

$У_1 = a_0 + a_1X_1 + a_2X_2$,

в котором коэффициенты a_0, a_1, a_2 являются известными числами. Необходимо определить такие значения X_1 и X_2, при которых $У_1$ примет желаемое значение (обратная задача). Поставленная задача имеет множество решений, которые получаются следующим методом. При известном значении $У_1$ присваивают переменной X_1 допустимое значение и из уравнения регрессии определяют значение X_2. Если значение X_2 приняло допустимое значение, то полученное решение является допустимым. При этом может получиться множество допустимых решений. Для получения единственного решения необходимо на допустимые решения наложить ограничения по какому либо-признаку (сумма затрат на допустимое решение). Задача сводится к линейному программированию.

$X_1 > X_{1min}$; $X_1 < X_{1max}$; $X_2 > X_{2min}$; $X_2 < X_{2max}$;

$a_1X_1 + a_2X_2 = У_1 - a_0$;

$c_1X_1 + c_2X_2 \rightarrow min$,

где X_{1min}, X_{1max}, X_{2min}, X_{2max} — минимальные и максимальные допустимые значения переменных X_1 и X_2;

a_0, a_1, a_2 — известные коэффициенты регрессионного уравнения;

c_1, c_2 — коэффициенты целевой функции, учитывающие удельные расходы на получение значений переменных X_1 и X_2.

Задача решается помощью программы “Поиск решения” (Excel).

Выводы.

В результате изучения систем одновременных уравнений можно коротко ответить на поставленные в начале главы вопросы.

а) Как будут вести себя зависимые переменные в системе с обратной связью?

Предполагается, что экономическая система является устойчивой и при отклонении эндогенной переменной от своего равновесного значения она асимптотически монотонно вернется к своему равновесному значению.

б) Как получить прогнозное значение зависимой переменной при известных значениях объясняемых переменных (прямая задача)?

Прогнозные значения эндогенных переменных можно получить с помощью приведенной системы одновременных уравнений.

в) Как оценить коэффициенты уравнения регрессии, если объясняемая переменная и остатки будут связаны между собой?

Для получения несмещенных и эффективных оценок при наличии нарушений предпосылки Гаусса-Маркова метода наименьших квадратов, заключающейся в наличии связи между объясняемой переменной и остатками необходимо использовать двухшаговый метод наименьших квадратов или косвенный метод.

г) Как вычислить значения объясняемых переменных для получения желаемого значения зависимой переменной (обратная задача)?

Решение задачи имеет множество допустимых решений. Если на них наложить ограничения в виде минимизации затрат на каждое решений, то получится одно решение. Решение задачи сводится к линейному программированию.

Вопросы для самоконтроля

1. Укажите проблемы, которые необходимо решить при наличии обратной связи между зависимыми переменными.
2. Дайте определение экзогенной и эндогенной переменной.
3. Укажите свойства эндогенных и экзогенных переменных.
4. Дайте определение структурной и приведенной форм одновременных уравнений.
5. Укажите свойства структурной системы одновременных уравнений.

6. Укажите последствия связи объясняемой переменной с остатками.

7. Укажите этапы двухшагового метода наименьших квадратов.

8. Укажите этапы трехшагового метода наименьших квадратов.

9. Укажите этапы косвенного метода наименьших квадратов.

10. Назовите сущность прямой и обратной задачи определения средних динамических эндогенных переменных.

Глава 13
ЭКОНОМЕТРИКА И УПРАВЛЕНИЕ КАЧЕСТВОМ

13.1. Основные понятия системы качества

Напомним основные положения системы качества.

Эконометрическое моделирование не является самоцелью, а служит инструментом решения проблем в деятельности предприятий или организаций.

Любая деятельность предприятий или организаций связана с производством продукции или услуги. Известны двенадцать стадий жизненного цикла продукции, которые располагаются в такой последовательности: маркетинг и изучение рынка, проектирование и разработка продукции, планирование и разработка процессов, закупки, производство или предоставление услуг, проверка, упаковка и складирование, сбыт и продажа, монтаж и сдача в эксплуатацию, техническая помощь и обслуживание, эксплуатация, утилизация либо вторичная переработка после окончания срока службы (ISO 9004-1-95). Разработка жизненного цикла продукции относится к стратегическому планированию. Реализация каждой стадии жизненного цикла продукции является процессом, относящимся к оперативному планированию.

Любой процесс выполняется по циклу Деминга: планируй, делай, анализируй, воздействуй. Проблемы у предприятия могут возникать на различных стадиях стратегического и оперативного планирования.

Пояснение. Любые проблемы экономического объекта связаны с низкой эффективностью его деятельности. Любая деятельность предприятия связана с реализацией процессов. Рассмотрение любой деятельности предприятия как реализация процессов позволяет формализировать процессы, предложить методы их улучшения, а следовательно и улучшения качества продукции, повышения конкурентоспособности предприятия и его эффективности.

Рассмотрение деятельности предприятия как процесса является четвертым принципом системы качества.

Общим условием эффективной работы любой организации и предприятия является соблюдение принципов системы качества.

Перечислим восемь принципов системы качества.

1. Ориентация на потребителя.

Организация зависит от своих потребителей и поэтому должна понимать их текущие и будущие потребности, выполнять их требования и стремиться превзойти их ожидания.

2. Лидерство руководителя.

Руководители обеспечивают единство цели и направления деятельности организации. Им следует создавать и поддерживать внутреннюю среду, в которой работники могут быть полностью вовлечены в решение задач организации.

3. Вовлечение работников.

Работники всех уровней составляют основу организации, и их полное вовлечение дает возможность организации с выгодой использовать их способности.

4. Процессный подход.

Желаемый результат достигается эффективнее, когда деятельностью и соответствующими ресурсами управляют как процессом.

5. Системный подход к менеджменту.

Выявление, понимание и менеджмент взаимосвязанных процессов как системы содействуют результативности и эффективности организации при достижении ее целей.

6. Постоянное улучшение.

Постоянное улучшение деятельности организации в целом следует рассматривать как ее неизменную цель.

7. Принятие решений, основанных на фактах.

Эффективные решения основываются на анализе данных и информации.

8. Взаимовыгодные отношения с поставщиками.

Организация и ее поставщики взаимозависимы, и отношения взаимной выгоды повышает способность обеих сторон создать ценности [17, с. 4]

Соблюдение всех принципов системы качества является необходимым условием высокоэффективной работы предприятия. Любое нарушение перечисленных принципов системы качества приводит к резкому снижению эффективности деятельности предприятий и реализации эконометрических моделей.

Взаимодействие между подпроцессами должно основываться на принципах "пяти нельзя" (иногда называют их принципами "пяти нулей"):

- нельзя принимать дефектную продукцию от предыдущего подпроцесса;
- нельзя передавать дефектную продукцию на следующий подпроцесс;
- нельзя второй раз совершать ошибку;
- нельзя изменять утвержденные технологии (совершенствовать технологии надо, но применять их можно после утверждения);
- нельзя создавать условия, при которых может появиться дефект.

Реализация этих принципов позволяет получить качественную продукцию.

Дискуссия. Анализ практической реализации этих принципов на предприятиях и в учебных заведениях показывает, что общий эффект от использования принципов определяется не их суммированием, а умножением. Это означает, что если какой-то принцип не используется, то общий эффект воздействия всех принципов сводится к нулю. Если предприятия используют все принципы системы качества, то они становятся близнецами братьями — высокоэффективными и конкурентоспособными на рынке производителей продукции и услуг. Если у предприятия появляются серьезные проблемы, то в первую очередь надо проверить соблюдение всех принципов системы качества. Например, если какой-то принцип не соблюдается, то организация или отдел работает малоэффективно. Это означает, что работники вынуждены перерабатывать за счет личного времени, вознаграждение за переработанное время не ожидается, так как фактически работники устраняют брак, накапливается остаточная усталость, система становится перегруженной. Любая перегруженная система разрушается. Это явление, к сожалению, очень часто приходится наблюдать в действительности. Если это касается человека, то он увольняется, если предприятия, то оно или разоряется или уходит в теневую — криминальную экономику. Соблюдение принципов системы качества является необходимой составляющей культуры любого руководителя на всей вертикали управления [26].

Эконометрические модели позволяют воспроизвести основные свойства экономических процессов для получения прогноза, который служит основой для принятия управленческих решений по улучшению качества продукции. Качество продукции возможно повышать только при улучшении качества процессов. Для улучшения качества процессов нужно выполнить необходимые и достаточные условия. Необходимыми условиями улучшения качества процессов являются подготовка менеджментом заинтересованной среды, так как известно, что **любой статистический метод потерпит неудачу, если руководством не подготовлена заинтересованная среда**[1].

Достаточными условиями улучшения процессов являются знание и умение использования средств улучшения качества.

В соответствии со стандартом ISO 9001−2001 используется следующая модель системы управления качеством, изображенной на рис. 13.1.

Анализ рис. 13.1 показывает, что удовлетворенность потребителей является определяющим в конкурентной борьбе производителей продукции.

Эконометрическое моделирование и система качества предназначены для решения проблем на предприятии, сле-

---

[1] Статистическое управление процессами (SPC). — Н. Новгород: АО НИЦ КД. СМЦ Приоритет, 1997. С. 29, 91.

довательно, они должны иметь области общих интересов как по методам и моделям, так и по средствам их реализации.

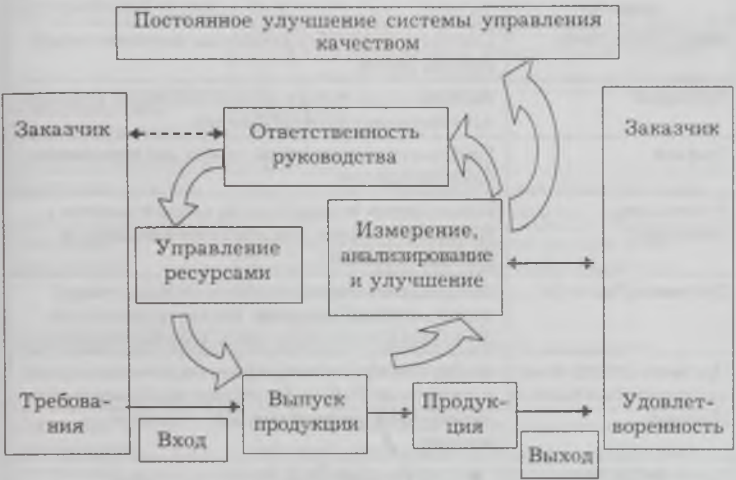


Рис. 13.1. Модель системы управлением качеством

Приводим перечень средств и методов улучшения качества и их назначений (см. табл. 13.1).

Анализ таблицы показывает, что многие средства и методы системы качества входят в состав эконометрики, остальные можно использовать на различных этапах эконометрического моделирования.

Очевидно, что средства и методы улучшения качества и эконометрики взаимно пересекаются и могут дополнить друг друга.

Подробное описание средств и методов улучшения качества имеется в следующей литературе: [9], [18], [25], [26].

Таблица 13.1

Назначение средств и методов улучшения качества

Простые средства и методы улучшения качества	Назначение
Форма сбора данных	Систематический сбор данных для получения четкой картины фактов
Расслоение	Выделение однородных групп по заданному признаку для определения влияния признака
Графики	Графическое представление данных для упрощения принятия решения
Установление ориентиров	Сопоставление процесса с теми, которые имеются у признанных лидеров, для выяснения возможности улучшения качества
Диаграмма Парето	Отображение в порядке значимости вклада каждой причины в общий результат. Оценка возможностей улучшения с их расстановкой по уровню значимости
Причинно-следственная диаграмма (Диаграмма Исикавы)	Анализ и отображение причинно-следственных связей по диаграмме. Упрощение решения проблемы за счет отслеживания цепочки: симптом — причина — решение
Контрольные карты	Диагностика: оценка стабильности процесса. Управление: определение необходимости в наладке процесса. Подтверждение: подтверждение улучшение процесса
Гистограмма	Отображение модели рассеивания данных. Визуальный съем информации о поведении процесса. Принятие решений о точках сосредоточения усилий по улучшению качества
Диаграмма разброса	Установление и подтверждение зависимостей между двумя связанными совокупностями данных. Подтверждение прогнозируемых зависимостей между двумя связанными совокупностями данных
Мозговой штурм	Выявление возможных решений проблем и потенциальных возможностей улучшения качества
Корреляционный анализ	Определение линейной степени связи между причиной и следствием
Регрессионный анализ	Определение формы связи между причиной и следствием

Окончание табл. 13.1

Простые средства и методы улучшения качества	Назначение
Дисперсионный анализ	Разложение общего варьирования признака на составляющие части, вызванные влиянием факторов и случайной составляющей, для определения статистически достоверного влияния учитываемых факторов
Новые средства и методы улучшения качества	Назначение
Диаграммы сродства	Группирование большого массива идей, мнений или соображений по конкретному вопросу.
Диаграммы зависимостей	Диаграмма зависимостей служит для того, чтобы показать стрелками причинно-следственные связи причин возникновения проблем, зафиксированные в диаграмме сродства
Матричные диаграммы	Назначение — выявить причины, оказывающие наибольшее влияние на появление дефектов, а также влияние причин на процессы, вызывающие дефекты
Системные (древовидные) диаграммы	Системная диаграмма используется в качестве метода системного определения оптимальных средств решения проблемы и строится в виде многоступенчатой древовидной структуры, элементами которой являются различные средства и способы решения
Сетевые графики	Сетевые графики (модели сетевого планирования) используется на этапе составления оптимальных планов тех или иных мероприятий после того, как определены проблемы, требующие решения, намечены необходимые меры, определены сроки и размечен ход осуществления запланированных мер, т. е. после составления первых четырех диаграмм
Диаграммы планирования оценки процессов	Эта диаграмма используется для оценки сроков и правильности осуществления программ и возможности корректирования тех или иных мероприятий в ходе их выполнения в соответствии со стрелочной диаграммой в случае решения сложных проблем
Метод анализа матричных данных	Преобразование исходных факторов в обобщенные таким образом, чтобы несколько обобщенных факторов заменили большое количество исходных факторов

13.2. Цикл Деминга улучшения процессов и этапы эконометрического моделирования

Следует признать, что эконометрическое моделирование производится в конечном итоге с целью повышения эффективности и конкурентоспособности деятельности организации или предприятия и их продукции. Повышение качества продукции связано с улучшением процессов. Эффективное улучшение любого процесса должно проводиться в определенной последовательности, называемой циклом Деминга: **планируй, делай, анализируй, воздействуй** (в честь американского ученого Э. Деминга, который помог Японии занять лидирующее место в мировой экономике с помощью использования цикла улучшения процессов, 14 принципов деятельности предприятия и статистических методов менеджмента качества).

Методика реализации цикла Деминга изложена в стандарте ИСО 9004-4-93, 1995 и включает следующие восемь пунктов:

1) привлечение всех работников организации и признание возможности улучшения качества;

2) инициирование проектов или мероприятий, направленных на улучшение качества;

3) исследование возможных причин;

4) установление причинно-следственных связей;

5) проведение предупреждающих и корректирующих действий;

6) подтверждение достигнутого уровня;

7) закрепление достигнутых результатов;

8) последующие улучшения.

Все восемь шагов хорошо вписываются в цикл Деминга, изображенный на рис. 13.2.

Этапы эконометрического моделирование входят в состав цикла Деминга по улучшению качества процессов. Следовательно, эконометрика получает четко определенное место в цикле улучшения деятельности предприятия.

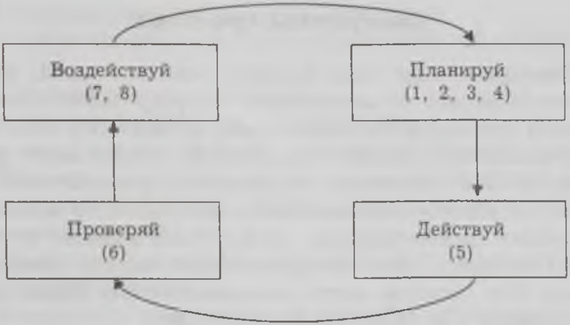


Рис. 13.2. Цикл Деминга и восемь шагов улучшения качества

13.3. Область использования средств и методов управления качеством в этапах эконометрического моделирования

Приводим область эффективного применения средств менеджмента качества в этапах эконометрического моделирования.

Этап 1. Постановочный. Формирование цели исследования. Выявление экономических проблем и выделение из них наиболее существенной.

Средства системы качества.

На этапе формирования цели исследования и проблем, которые надо решать, важно знать стратегические направления деятельности предприятия. Для этого собирается разнообразная и разнородная информация о внешней и внутренней среде существования экономического объекта по всем направлениям его деятельности. Важно разобраться в этом тумане идей, тенденций, перспектив путем их систематизации, т. е. построить диаграмму сродства.

Диаграммы сродства

Описание. При сборе больших массивов идей, мнений или соображений по конкретному вопросу данное средство помогает группировать информацию по принципу естественных взаимосвязей между ними. Данный процесс имеет целью стимулировать творческие способности и участие всего персонала. Он наиболее эффективен в небольших по численности группах (рекомендуется, чтобы в нее входило не более восьми человек), члены которых имеют навыки совместной работы. Это средство часто используется для общих идей, высказанных при "мозговом штурме" [22].

Порядок применения:

а) формулируют в общих чертах подлежащий изучению вопрос (подробности могут заранее предопределять ответы);

б) записывают как можно больше отдельных идей, мнений и соображений на карточках (по одному предложению на каждой карточке);

в) перемещают карточки и раскладывают их произвольно на большом столе;

г) группируют родственные карточки следующим образом:

• рассортировывают карточки, которые оказались родственными, по группам;

• ограничивают число групп (не более десяти), не присовокупляя к ним отдельные карточки, не относящиеся к этим группам;

• выбирают или создают основную карточку, отражающую содержание каждой группы;

• эту основную карточку помещаю сверху;

д) переносят информацию с карточек на бумагу по группам [22].

Области эффективного использования:

• составление стратегического планирования предприятий и организаций;

• определение перспективного направления деятельности предприятия;

• группировка требований к изделию;
• группировка тенденций окружающей среды предприятия.

Результат исследования:

• формирование цели исследования;
• совокупность проблем, которые необходимо решить.

Пример реализации. В одном из предприятий за декабрь месяц 2001 г. были собраны данные по несоответствиям в количестве 274 шт.

Необходимо построить несколько диаграмм сродства со следующими целями:

• сгруппировать данные по признаку, кто выявил несоответствия с целью оказания помощи в их деятельности, если в этом будет необходимость;

• внутри каждой группы произвести вторичную группировку по признаку № заказа, названия участка, подразделения, в работе которых появляются несоответствия с целью определения источников несоответствий для определения основного направления по их устранению;

• сгруппировать данные по признаку наименований подразделений, которые допускают несоответствия с целью выявления основного подразделения, в работе которого наблюдается наибольшее количество несоответствий и для более углубленного изучения причин, которые вызвали эти несоответствия.

В результате построения диаграмм сродства все подразделения были проранжированы по признаку количества несоответствий, которые были в них обнаружены. Результаты ранжировки представлены в виде диаграммы Парето (рис. 13.3)

Анализ диаграммы Парето показывает, что в критической зоне А находятся четыре подразделения, которые допускают более 80% всех несоответствий. Явным лидером по несоответствиям является первое подразделение. Следует более подробно изучить деятельность первого подразделения.

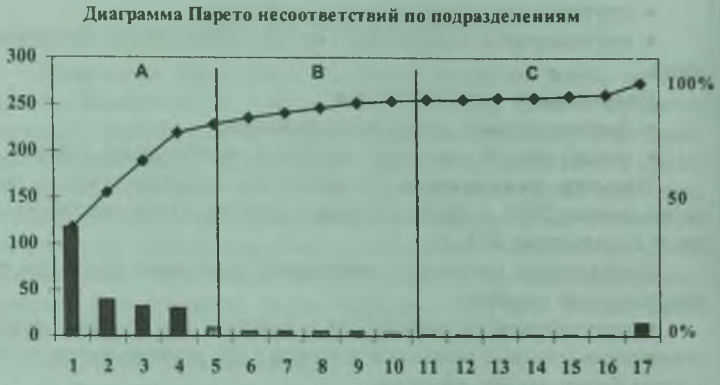


Рис. 13.3. Диаграмма Парето несоответствий по подразделениям: каждому номеру соответствует подразделение; номеру 17 соответствует признак "прочие подразделения".

Выявление проблем по методикам самооценки

Обычно экономические проблемы выявляются самооценкой по различным методикам, соответствующим премиям в области качества: премия имени Э. Деминга, премия имени М. Болдриджа, Европейская премия по качеству, Российская премия качества [26, с. 153–162]. Каждая премия предполагает, что организация работает по определенной модели. В частности Европейская премия по качеству предполагает модель организации, содержащей 9 критериев: лидерство; персонал; политика и стратегия; партнерство и ресурсы; процессы; результаты, которые относятся к персоналу; результаты, которые относятся к потребителям; результаты, которые относятся к обществу; основные деловые результаты.

Каждый критерий имеет от двух до шести подкритериев. Каждый подкритерий имеет от 10 до 30 вопросов. Общее количество вопросов по всем критериям составляет около 400 шт., которые тестируют все стороны деятельности организации. После проведения самооценки выявляются сильные и слабые стороны в деятельности организации и обнаруживаются проблемы, которые надо решать. Проблем может быть очень много, однако надо выбрать наиболее существенные (с точки зрения цели деятельности предприятия), которые следует устранять в первую очередь.

По каждой проблеме вычисляются потери предприятия в денежном выражении. Ранжировка проблем по величине потерь позволяет выделить несколько наиболее существенных. Выделение существенных проблем реализуется с помощью диаграммы Парето по проблемам (результатам) см. рис. 13.4.

Анализ рис. 13.4 показывает, что первая и вторая проблемы дают более 75% от всех потерь. Следовательно, в первую очередь следует решать первую или вторую проблему.

Однако предпочтение следует отдать той проблеме, у которой больше величина эффективности ее решения. Для этой цели необходимо вычислить эффективность решения первой и второй проблем: отношение отдачи от решения проблемы к затратам на решение этой проблемы в денежном выражении. Предположим, эффективность решения второй проблемы выше, чем первой, следовательно, в первую очередь следует решать вторую проблему.

Примечание. В практической деятельности улучшения процессов на предприятии после самооценки по модели Европейской премии по качеству приступают к разработке системы качества по международному стандарту ICO 9000–2001 параллельно с решением первоочередных проблем, выделенных диаграммой Парето по проблемам. По этой схеме работают многие сертификационные организации Украины и России.

Если на предприятии имеется несколько проблем, которые надо решать, то на интуитивном уровне надо выбрать такую, от решения которой значительно улучшится эффективность производства. Интуитивное положение можно формализовать с помощью диаграммы Парето.

Диаграмма Парето

Назначение. Диаграмму Парето используют для:

Рис. 13.4. Диаграмма Парето по проблемам (данные условные)

• отображения в порядке значимости каждой причины в общем результате;

• оценки возможности улучшения с их расстановкой по уровню значимости [22].

История.

Итальянский экономист Парето (1845–1923) впервые высказал закономерность: "20% населения имеют 80% денежных средств всего населения". Эта закономерность была распространена Демингом Э. на свойства причин дефектов — "20% причин создают 80% всех дефектов".

Примечание. В литературе начинает встречаться большое количество примеров действия правила 20/80, относящихся к различным видам деятельности как человека, так и птиц, животных.

Теоретическое обоснование. Теоретическим обоснованием найденной закономерности могут служить следующие факты:

• большинство характеристик человека, в частности его способности, подчиняются нормальному закону распределения, а это означает, что среди группы людей будет 20% неспособных, 60% со средними способностями, 20% способных. Однако, если взять выделенную группу, допустим способных людей, их способности тоже распределятся по нормальному закону: 20% будет способных, 60% — очень способных, 20% — талантливых;

- установлено, что если всю качественную продукцию принять за 100%, то 80% ее производят 20% способных сотрудников, а также если всю дефектную продукцию принять за 100%, то 80% ее производят 20% неспособных сотрудников.

Примечание. Когда высшему руководству предприятия приходится говорить об этих закономерностях, то следует моментальная реакция — покажите этих 20% бракоделов, они тут же будут уволены. Уволить не трудно, но это проблемы не решит. Так как проблема заключается в том, что эти бракоделы работают не на своем месте. Дело в том, что сложность выполнения работ тоже подчиняется нормальному закону распределения. Тогда необходимо поставить в соответствие способности и сложность выполнения работы. Это значит, что 20% малоспособных сотрудников должны выполнять 20% работы низкой сложности.

Описание.

Диаграмма Парето позволяет в простой графической форме расположить причины в определенном порядке: от наиболее часто встречающихся до наименее часто встречающихся. В основе диаграммы Парето лежит принцип Парето, который гласит, что, как правило, на результат влияет в большей степени лишь несколько причин. Отделив наиболее существенные причины от несущественных, можно добиться наибольшего улучшения при наименьших затратах.

Диаграмма Парето показывает в порядке убывания относительный вклад каждой причины в общий результат. Относительный вклад может определяться на основании повторяемости, стоимости, связанной с каждой причиной, или других факторов влияния на результат. Для отображения относительного влияния каждой причины используются прямоугольники. Для отображения совокупного влияния всех причин используется кумулятивная кривая [22].

Порядок применения:

а) выбирают причины, подлежащие анализу;

б) выбирают единицу измерения, используемую при анализе, например повторяемость, стоимость или другую меру влияния;

в) выбирают промежуток времени, в течение которого будет проводиться анализ данных;

г) перечисляют причины слева направо вдоль горизонтальной оси в порядке уменьшения величины единицы измерения. Категории, содержащие самые незначительные причины, могут быть объедены в одну категорию "прочие". Эту категорию размещают крайней справа;

д) строят две вертикальные оси по обе стороны горизонтальной линии. Левая ось должна быть размечена в единицах измерения, и ее высота должна равняться сумме величин всех причин. Правая ось должна иметь ту же высоту и быть размечена от 0 до 100%;

е) над каждой причиной вычерчивают прямоугольник, высота которого соответствует величине единицы измерения данной причины;

ж) строят кумулятивную кривую, суммируя соответствующие каждой причине величины слева направо;

и) при помощи диаграммы Парето определяют важнейшие объекты улучшения качества.

Метод ABC заключается в том, что все дефекты диаграммы Парето делят на три примерно одинаковые части и обозначают их А, В, С. В группу А войдет наибольшее количество дефектов, в группу В — небольшое число дефектов, а группа С будет включать незначительное количество дефектов.

Дефекты группы А требуют первоочередного решения.

После проведения корректирующих и предупреждающих мер по устранению дефектов в группу А войдут уже другие дефекты.

Область применения — анализ деятельности предприятия по результатам и по причинам.

Ограничения:

• единицы измерения характеристик объектов должны быть одинаковыми;

• количество факторов должно быть от 7 до 10, последним должен быть фактор "прочие".

Возможные виды объектов: продукция, процесс, организация.

Единицы измерения характеристик объектов: штуки, денежная единица.

Виды диаграмм Парето

Известны два вида диаграмм Парето: по результатам деятельности и по причинам.

Диаграмма Парето по результатам деятельности предприятия за отчетный период предназначена для выявления главной проблемы и отражает следующие проблемы:

• нежелательные результаты деятельности или несоответствия (количество случаев несоответствий или количество денег, направленных на устранение несоответствий, или потери предприятия в денежном выражении, обусловленные несоответствиями);

• несоответствия качества продукции: дефекты, поломки, ошибки, отказы, рекламации, ремонты, возвраты продукции;

• несоответствия по себестоимости продукции: объем потерь, затраты;

• несоответствия по срокам поставок: нехватка запасов, ошибки в составлении счетов, срыв сроков поставок;

• несоответствия по безопасности: несчастные случаи, трагические ошибки, аварии;

Пример. Зависимость суммы затрат на устранение несоответствий от видов несоответствий: дефектов, поломок, ошибок, рекламаций и т. д.

Диаграмма Парето по причинам отражает причины проблем, возникающих в ходе производства, и используется для выявления главной из них:

• сырье: изготовитель, тип и торговая марка, завод-поставщик, партия;

- оборудование: станки, агрегаты, инструменты, оснастка, организация использования, модели, штампы, тип и форма, конструкция, срок службы, расположение;
- рабочий: бригада, возраст, опыт работы, квалификация, индивидуальные характеристики;
- метод работы: условия производства, заказы-наряды, приемы работы, последовательность операций, система сдачи продукции;
- время: смена, время суток, день недели, дата;
- среда:
  - физическая среда: место нахождения (континент, государство, область, район, климат), рабочее место оператора;
  - абстрактная среда: законы, правила, традиции, моральные принципы.

Пример. Зависимость аварий от номеров смен.

Возможная сфера применения диаграммы Парето.

Финансовая сфера:

- анализ себестоимости изделий отдельно по видам изделий;
- анализ сбыта;
- анализ соотношения затрат на деятельность по контролю по факторам контроля;
- анализ прибыли отдельно по видам изделий;
- анализ процента прибыли;
- анализ “мертвого капитала” в обороте денежных сумм отдельно по назначению сумм.

Сфера сбыта:

- анализ прогноза потребностей по видам изделий;
- анализ выручки от продажи изделий отдельно по продавцам и по магазинам;
- анализ случаев получения рекламаций отдельно по содержанию рекламаций и анализ суммы потерь от рекламаций;
- анализ числа возвращенных изделий отдельно по видам изделий;

- анализ выручки отдельно по сумме выручки, отдельно по видам изделий.

Сфера материально-технического обеспечения:

- анализ числа случаев специального отбора по видам сырья и материалов;
- анализ числа дней задержки поставок отдельно по видам сырья и материалов;
- анализ денежных потерь в результате бесполезной задержки на складах;
- отдельно по видам сырья и материалов;
- анализ расходов на хранение на складах отдельно по видам сырья и материалов.

Сфера производства:

- анализ числа переделок отдельно по рабочим участкам;
- анализ числа неполадок отдельно по станкам;
- анализ времени последовательного выполнения операций;
- анализ процента брака отдельно по дням недели;
- анализ числа случаев остановки процесса отдельно по процессам;
- анализ потерь времени отдельно по процессам;
- анализ числа дней хранения на складах и денежных средств на это отдельно по видам изделий;
- анализ числа случаев поломок отдельно по рабочим участкам;
- анализ эффективности от принятых мер (анализ двух диаграмм Парето до принятия мер и после).

Сфера делопроизводства:

- анализ числа предложений отдельно по сотрудникам;
- анализ числа дней обработки документов отдельно по предложениям;
- анализ количества нереализованных материалов и процента их реализации
- отдельно по рабочим участкам;
- анализ числа ошибок в накладных отдельно по видам накладных;

• анализ процента выполнения плана отдельно по подразделениям.

Определение эффекта от принятых мер по устранению несоответствий.

Если построить диаграммы Парето до и после принятия мер по устранению несоответствий то по разности общего количества несоответствий можно определить эффект от принятых мер.

Этап 2. Априорный. Анализ сущности изучаемого объекта. Определяют главные и второстепенные факторы, влияющие на проблему.

Для анализа сущности изучаемого процесса и определения факторов, влияющих на его продукцию, используются следующие методы:

• причинно-следственная диаграмма, которая иногда называется “рыбий скелет”, или диаграмма Исикава (в честь японского ученого, который впервые предложил данный метод);

• метод “мозгового штурма” или “мозговой атаки”;

• технологическая схема (структурная схема процесса);

• метод “Почему? Почему?”;

• расслоение;

• корреляционный и регрессионный анализ.

Корреляционный и регрессионный анализы входят в состав эконометрики, поэтому их рассматривать не будем.

Причинно-следственная диаграмма

Определение. Причинно-следственная диаграмма, или диаграмма Исикавы, представляет собой графическое изображение влияния ресурсов предприятия на качество любого процесса.

Назначение. Причинно-следственная диаграмма применяется для:

• анализа причинно-следственных связей;

• отображения причинно-следственных связей;

• упрощения решения задачи по цепочке симптом — причина — решение.

Диаграмма Исикавы предназначена для системного анализа всех факторов, оказывающих влияние на качество продукции, и выявления важнейших из них.

Цель построения диаграммы Исикавы — обнаружить и устранить причины, которые вызвали нарушение процесса производства продукции.

Область применения: любые процессы.

Возможности. Диаграмма Исикавы позволяет произвести очень глубокий системный анализ всех факторов неограниченного порядка, при этом удается найти истинные причины, которые относятся к факторам высокого порядка и которые нельзя обнаружить при анализе только факторов первого порядка.

Практическое применение диаграммы Исикавы показало ее высокую эффективность и позволяет не упустить даже малозначимого фактора.

Примечание. Использование диаграммы Исикавы позволило выявить около 300 факторов, влияющих на успеваемость студентов. Одним из этих факторов является посещаемость студентов, которая часто связана с необходимостью зарабатывать на обучение. Приятно было узнать, что некоторые учебные заведения ввели возможность учиться на дневном отделении в удобное для студентов время с 8 до 21 часа.

Описание. Причинно-следственная диаграмма служит средством установления и отображения связей между следствием (например, отклонением характеристики качества) и его потенциальными причинами. Весь массив потенциальных причин разбивается по основным категориям, подкатегориям таким образом, чтобы схема имела вид рыбьего скелета. Поэтому этот способ еще называют диаграммой “рыбий скелет” [22].

Порядок применения:

а) точно и сжато формулируют следствие;

б) определяют основные категории возможных причин. В число рассматриваемых факторов входят: системы данных и информационные данные; условия окружающей среды; оборудование; материалы; измерения; методы; люди; время;

в) приступая к построению диаграммы, заносят следствие в прямоугольник справа, а перед ним располагают основные категории причин как входы в этот прямоугольник;

г) наращивают диаграмму, продумывая и записывая все причины следующего уровня, и продолжают для уровней более высокого порядка. Хорошо построенная диаграмма не должна иметь ответвлений менее чем с двумя уровнями, а многие ее ответвления должны иметь три и больше;

д) отбирают и определяют небольшое число (от 3 до 5) причин высшего уровня, которые, как ожидается, в наибольшей степени влияют на следствие и требуют последующих действий, например, сбора данных, попыток оперативного управления и т. д.

Примечания: 1. Альтернативный метод построения причинно-следственной диаграммы заключается в выявлении путем "мозгового штурма" всех всзможных причин с последующим упорядочением их по категориям и подкатегориям при помощи диаграммы объединения по общему признаку.

2. В ряде случаев может оказаться целесообразным перечисление основных этапов процесса в виде категорий, например, когда протекание технологического процесса является следствием, рассматриваемым с точки зрения возможности его улучшения. При определении таких этапов часто полезно построить структурную схему.

3. Построенная диаграмма становится "активным инструментом", поскольку по мере получения новой информации и новых данных, в нее могут вноситься дальнейшие усовершенствования.

4. В построении такой диаграммы, как правило, участвует группа специалистов, но ее могут составить и отдельные лица, обладающие достаточным знанием процесса и соответствующим опытом.

Пример причинно-следственной диаграммы Исикавы приведен на рис. 13.5.

"Мозговой штурм"

Определение. "Мозговой штурм" — это эвристический метод определения возможных решений и потенциальных возможностей улучшения качества.

Рис. 13.5. Диаграмма Исикавы

Назначение. К "мозговому штурму" прибегают для определения возможных решений проблем и потенциальных возможностей улучшения качества.

Описание. "Мозговой штурм" — это метод побуждения возможностей творческого мышления группы, имеющей целью выработку и уточнение ряда идей, проблем и процессов.

Порядок применения. Метод состоит из двух стадий:

а) стадия выработки идей. Руководитель группы дает необходимые указания и разъясняет цель проведения мозгового штурма, после чего члены группы начинают высказывать свои идеи. Цель состоит в выработке как можно большего числа идей;

б) стадии уточнения. Группа изучает весь перечень высказанных идей с тем, чтобы каждому члену было понятны все идеи. По окончании мозгового штурма оценивают все идеи.

Указания по организации мозгового штурма:

- должен быть определен руководитель;
- должна быть четко сформулирована цель проведения мозгового штурма;
- все члены группы выступают по очереди, высказывая по одной идее;

• при возможности члены группы развивают идеи, высказанные другими;

• на этой стадии идеи не критикуются и не обсуждаются;

• идеи регистрируются открыто, чтобы все члены группы могли следить за их регистрацией;

• эта процедура продолжается до тех пор, пока ее участниками не будут высказаны все имеющиеся у них идеи;

• все идеи анализируются с целью их уточнения.

Технологическая схема (структурная схема процесса).

Определение. Технологическая схема — графическое представление процессов.

Назначение. Технологическая схема применяется при:

• описании имеющегося процесса;

• проектировании нового процесса.

Описание. Технологическая схема представляет собой графическое изображение этапов процесса и удобна для изучения возможностей его улучшения за счет глубокого вникания в суть процесса. Изучение взаимозависимостей между различными этапами процесса часто позволяет определить потенциальные причины недостатков. Технологические схемы применимы ко всем аспектам любого процесса, от поступления материалов до этапов сбыта или технического обслуживания изделия.

При построении технологических схем используются легко распознаваемые обозначения. Наиболее часто применяются следующие обозначения (рис. 13.6):

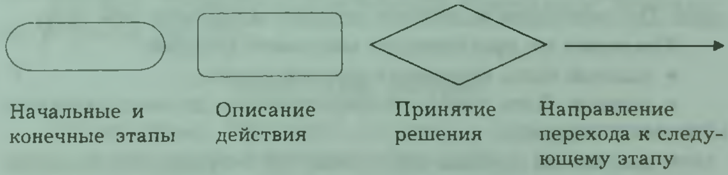


Рис. 13.6. Обозначения, применяемые в технологической схеме

Порядок применения. Описание имеющегося процесса:

а) определяют начало и конец процесса;

б) наблюдают ход процесса от начала до конца;

в) определяют этапы процесса (действия, решения, входы, выходы);

г) строят черновой вариант структурной схемы, отображающей ход процесса;

д) анализируют черновой вариант технологической схемы вместе с людьми, привлеченными к выполнению данного процесса;

е) по результатам анализа вносят в технологическую схему необходимые поправки;

ж) сверяют технологическую схему с реальным протеканием процесса;

и) проставляют дату составления технологической схемы для того, чтобы

обращаться к ней и использовать ее в дальнейшем (в технологической схеме зарегистрирован реальный ход процесса, и она может использоваться для выполнения возможностей его улучшения).

Проектирование нового процесса:

а) определяют начало и конец процесса;

б) предварительно намечают этапы процесса (действия, решения, входы, выходы);

в) определяют этапы процесса (действия, решения, входы, выходы);

г) строят черновой вариант структурной схемы, отображающей ход процесса;

д) анализируют черновой вариант технологической схемы вместе с работниками, которые будут привлекаться к выполнению данного процесса;

е) по результатам анализа вносят в технологическую схему необходимые поправки;

ж) проставляют дату составления технологической схемы для того, чтобы обращаться к ней и использовать ее в дальнейшем (в технологической схеме зарегистрирован реальный ход процесса, и она может использоваться для выполнения возможностей его улучшения).

Метод "Почему? Почему?"

Определение. Метод "Почему? Почему?" — очень простой и эффективный эвристический метод определения причин и следствий.

Назначение. Метод "Почему? Почему?" предназначен для того, чтобы:

- логически прийти к основной причине и понять взаимоотношение причин и следствий.
- структурировать мышление команды.

Порядок применения:

а) ставится проблема, согласовывается и записывается, чтобы все могли ее видеть. Очень важно иметь четкое определение проблемы (показатель, направление);

б) повторите определение в форме "почему?";

в) запишите все возможные ответы. Не критикуйте возможные ответы;

г) на каждый возможный ответ повторите ответ в форме вопроса "почему?";

д) запишите все возможные ответы на втором уровне;

е) повторите процесс, продолжая задавать вопрос почему?, пока у вас больше не останется вопросов;

ж) возле каждой причины запишите, как она может быть проверена, чтобы подтвердить, что это основная причина. Это можно сделать посредством разработки эксперимента, проверки данных и т. д. Если после устранения причины несоответствие больше не стало наблюдаться, то это была основная причина.

Расслоение

Расслоение (стратификация) — выделение однородных групп.

Назначение — выделение несколько групп объектов в соответствии с градациями факторов, свойства которых сильно отличались бы между группами и мало отличались внутри группы.

Ограничения — свойства объектов должны иметь одинаковую единицу измерения.

Цель — выявить влияние градаций фактора на показатель качества.

Примеры.

Если мы желаем определить влияние освещения на количество несоответствий при электромонтаже объектов в электро-монтажном участке 2 (ЭМУ2), то необходимо расслоить данные о несоответствиях по степени удаленности объекта от источника света.

Если мы желаем определить влияние стажа работы на количество несоответствий при электромонтаже объектов в электромонтажном управлении ЭМУ2, то необходимо расслоить данные о несоответствиях по электромонтажникам на несколько групп, в каждой из которой работники имеют примерно одинаковый стаж работы.

Если мы желаем определить влияние времени суток на количество несоответствий при электромонтаже объектов в ЭМУ2, то необходимо расслоить данные о несоответствиях по времени свершения несоответствий на несколько групп, допустим на первую половину смены и вторую половину смены. Можно произвести двумерную группировку, допустим, определить, как влияют стаж работы и освещенность на количество несоответствий. При этом данные по несоответствиям надо расслоить по стажу работы, затем данные в каждой полученной группе расслоить по степени освещенности.

Если количество несоответствий в первой и второй группах примерно одинаковое, то можно считать, что фактор освещенности не оказывает влияние на количество несоответствий.

Если возникнет необходимость в получении более обоснованных выводов, то можно использовать статистический критерий Стьюдента проверки равенства двух средних значений.

Сравнение двух выборочных средних из нормальных совокупностей.

Проверяется H_0: $\mu_1 = \mu_2$ о равенстве средних значений генеральных совокупностей.

Для практических целей пригодна статистическая проверка критерия:

Если

$$t = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} > t_{\alpha/2}(\alpha = 0{,}05,\ m = n_1 + n_2 - 2),$$

то с вероятностью $1 - \alpha$ H_0: $\overline{X}_1 = \overline{X}_2$ отвергается, в противном случае принимается,
где $t_{\alpha/2}$ — критерий Стьюдента;

H_0 — нулевая гипотеза;

μ_1, μ_2 — математические ожидания для генеральных совокупностей;

S_1^2, S_2^2 — выборочные дисперсии для двух выборок из соответствующих генеральных совокупностей;

$\overline{X}_1$, $\overline{X}_2$ — средние значения для двух выборок из соответствующих генеральных совокупностей;

n_1, n_2 — объем выборки для двух выборок;

m — число степеней свободы.

Этап 3. Информационный. Сбор данных и статистической информации по всем переменным причинно-следственных связей зависимости проблемы от предполагаемых причин.

Средства системы качества.

Сбор данных осуществляется с помощью регистрации информации в форму сбора данных или контрольный листок.

Форма сбора данных — это шаблон произвольной формы, предназначенный для сбора и регистрации необходимой и достаточной информации результатов контроля качества продукции, с помощью которой будет решаться поставленная задача. Он позволяет упорядочить сбор данных и упрощает их анализ.

Форма сбора данных используется для всех объектов качества.

Форма бланка зависит от задачи, а также от инструмента проверки качества, с помощью которого надо решить поставленную задачу.

Чаще всего форма сбора данных представляет собой таблицу с колонками и строчками. Форма сбора данных позволяет ответить на вопрос: как часто случается данное событие? С него начинается превращение мнений и предположений в факты.

Область применения. Формы сбора данных используются в качестве носителя информации для различных инструментов контроля качества.

Ограничения:

• форма сбора данных должна содержать данные для контроля информации: суммы по строкам и столбцам, итоговую сумму, а также указание на время и место получения информации, подпись исполнителя;

• при вводе данных в ЭВМ необходимо ввести защиту от несанкционированных действий оператора, которая заключается в следующем: вводимая информация должна анализироваться на тип данных — числовой или текстовый, на нахождение значений в заданном интервале;

• информация в форме сбора данных должна удовлетворять условиям 3Д: достоверность, доступность, достаточность.

Этапы составления формы сбора данных.

По ISO 9004-4-98 предлагается такой порядок применения формы сбора данных.

а) Устанавливают конкретную цель сбора данных (решаемые на их основе вопросы).

Цель сбора данных должна соответствовать цели отдела, где эти данные собираются. Цель отдела должна соответствовать цели организации.

Во всех подразделениях организации должна быть установлена цель улучшения качества. Она должна быть тесно согласована с общей целью деятельности организации и направлена на повышение удовлетворенности потребителя и совершенствование эффективности и результативности процессов. Цель улучшения качества должна ставиться так, чтобы было возможно определение показателей достигнутого прогресса. Она должна быть понятной, масштабной и соответствовать стоящим перед организацией задачам. Стратегия достижения этой цели должна быть разъяснена всем, кому предстоит работать в этом направлении, и согласована с ними. Цель улучшения качества должна регулярно пересматриваться с тем, чтобы она отражала изменение ожиданий потребителя [22].

Уровень выполнения целей зависит от трех "должно":

- цели должны устанавливаться публично;
- исполнитель должен иметь внутренний контроль;
- цели должны быть самостоятельно установлены быстрее, чем подписано наверху [18, с. 235].

б) Определяют, какие данные необходимы для достижения поставленной цели (решения вопросов).

г) Составляют форму регистрации данных. Предусматривают графы для регистрации данных:

- о лице, собравшем данные;
- о месте, времени и способе сбора данных.

д) Проверяют форму путем сбора и регистрации некоторых данных.

е) Пересматривают, анализируют и в случае необходимости изменяют форму.

Формы учета данных могут быть двух видов: первичные, собранные на рабочем месте, и вторичные или обобщенные, выполненные на основе первичных.

Источниками сбора данных могут быть следующие мероприятия:

1. Инспекционный контроль:

- регистрация данных входного контроля исходного сырья и материалов;
- регистрация данных контроля готовых изделий;

• регистрация данных инспекционного контроля процесса (промежуточного контроля).

2. Производство и технологии:

• регистрация данных контроля процесса;

• повседневная информация о применяемых операциях, регистрация данных;

• контроля оборудования (неполадки, ремонт, техническое обслуживание);

• патенты и статьи из периодической печати.

3. Поставка материалов и сбыт продукции:

• регистрация данных через склады (входная и выходная нагрузка);

• регистрация сбыта продукции (данные о получении и выплате денежных сумм, контроль срока поставок).

4. Управление и делопроизводство:

• регистрация прибыли;

• регистрация возвращенной продукции;

• регистрация обслуживания постоянных клиентов;

• журнал регистрации продажи;

• регистрация обработки рекламаций;

• материалы анализа рынка.

5. Финансовые операции:

• таблица сопоставления дебета и кредита;

• регистрация подсчета потерь;

• экономические расчеты.

6. Материалы внутреннего аудита.

Этап 4. Спецификация математической модели. Определение вида математической функции, которая описывает влияние объясняемых переменных на зависимую переменную.

Средства и методы системы качества

Для изучения экономических процессов, представленных временными рядами, можно эффективно использовать контрольные карты.

Определения.

Изменчивость — неизбежные различия среди индивидуальных результатов процесса, их источники могут расслаиваться на два основных класса: обычные причины и особые причины.

Обычные причины — многочисленные источники изменчивости в процессе, которые имеют стабильное и повторяемое распределение во времени. Обычные причины ведут себя как стабильная система случайных причин.

Стабильность (управляемость, контролируемость, нормальность, стационарность) — свойство процесса; если действуют только обычные причины, среднее значение процесса и его разброс не меняются. (Термин "стабильный" взят из [22], определение взято из ненормативного словаря.)

Особые причины (неслучайные причины) — это любые вызывающие изменения факторы, которые действуют на процесс не всегда. Если особые причины присутствуют, то процесс становится непредсказуемым и нестабильным во времени.

Контрольная карта — это графическое представление характеристики процесса, показывающее нанесенные значения этой характеристики, центральную линию и контрольные границы (см. рис. 13.7).

Пределы регулирования (контрольные границы, доверительные интервалы) процесса обозначают границы регулирования процесса.

Признаки действия особых причин:

1) выход за контрольные границы, если на графике имеются точки, которые выходят за пределы контрольных границ;

2) серия — это проявление такого состояния, когда точки неизменно оказываются по одну сторону от средней линии; число таких точек называется длиной серии. Серия длиной в 7 точек, а иногда меньше 6 точек рассматривается как признак нестабильности процесса.

Процесс считается нестабильным, если наблюдаются такие признаки:

1) не менее 10 из 11 точек оказываются по одну сторону от центральной линии;

2) не менее 12 из 14 точек оказываются по одну сторону от центральной линии;

3) не менее 16 из 20 точек оказываются по одну сторону от центральной линии;

4) имеется тренд (дрейф). Если точки образуют непрерывно повышающуюся или понижающуюся кривую, то имеет место тренд;

5) имеется приближение к контрольным границам. Рассматриваются точки, которые приближаются к 3-сигмовым контрольным пределам, причем 2 или 3 точки оказываются за 2-сигмовыми линиями;

6) имеется приближение к центральной линии. Если большинство точек концентрируются внутри центральных полуторасигмовых линий, делящих пополам расстояние между центральной линией и каждой из контрольных границ, то это обусловлено неподходящим способом разбиения данных на группы. Приближение к центральной линии вовсе не означает, что достигнуто стабильное состояние, напротив, это значит, что в группах смешиваются данные из различных распределений, что делает размах контрольных границ слишком широким. В таком случае надо изменить способ разбиения данных на группы;

7) имеется периодичность. Кривая повторяет структуру волны “то подъем, то спад” с примерно одинаковыми интервалами времени.

Анализ 7 признаков нестабильности процесса поможет выявить особые причины их появления.

Имеются следующие признаки нестабильности процесса.

Если контрольная карта разделена на зоны, кратные среднеквадратическому значению S показателя процесса (соответственно зоны А, Б, С), то имеется признак действия особой причины и необходимо сделать соответствующую корректировку процесса при условии, что:

а) две точки из трех находятся по одну сторону от центральной линии в зоне А и далее;

б) четыре точки из пяти расположены по одну сторону от центральной линии в зоне Б или далее;

в) девять точек находятся по одну сторону от центральной линии;

г) шесть последовательных точек возрастают или уменьшаются;

д) четырнадцать точек в ряду колеблются вверх или вниз;

е) пятнадцать точек в ряду находятся в зоне Б (ниже или выше центральной линии)[1].

Описание. Контрольная карта является средством, позволяющим отличить отклонения вследствие устанавливаемых или особых причин от случайных отклонений, вносимых самим процессом. Случайные отклонения повторяются произвольно в прогнозируемых пределах.

Отклонения вследствие устанавливаемых или особых причин свидетельствуют о том, что ряд влияющих на процесс факторов подлежит идентификации, исследованию и учету при управлении.

Построение контрольных карт основывается на математической статистике. В контрольных картах для установления пределов, в которых могут ожидаться последующие наблюдения при условии, что на процесс не будут воздействовать установленные или особые причины, используют текущие рабочие данные.

Назначение. Контрольную карту (рис. 13.7) используют в следующих целях:

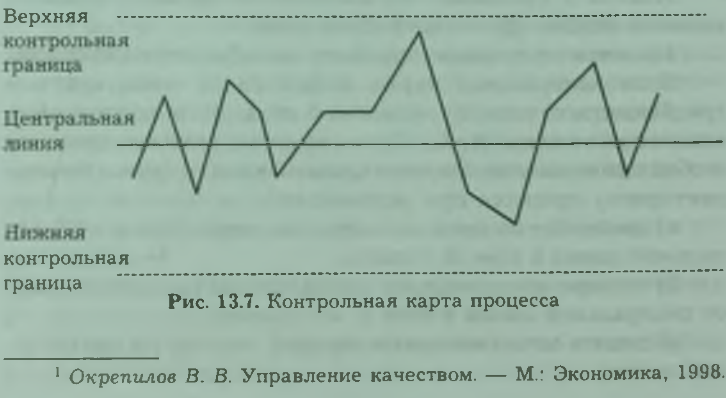


Рис. 13.7. Контрольная карта процесса

[1] *Окрепилов В. В.* Управление качеством. — М.: Экономика, 1998.

а) диагностика: для оценки стабильности процесса;

б) управление: для определения того, нуждается ли процесс в наладке или его следует оставить без изменений;

в) подтверждение: подтверждение улучшения процесса.

Порядок применения:

а) отбирают характеристики для построения контрольной карты;

б) выбирают нужный тип контрольной карты;

в) принимают решение в отношении подгруппы (небольшой совокупности объектов, в пределах которой, согласно допущению, отклонения носят только случайный характер), ее размера и частоты выборки из нее;

г) снимают и регистрируют данные не менее чем для 20–25 подгрупп или же используют данные, зарегистрированные ранее;

д) вычисляют статистические характеристики выборок из каждой подгруппы;

е) подсчитывают пределы регулирования исходя из статистических характеристик выборок из подгрупп;

ж) создают контрольную карту, строя график статистических характеристик подгрупп;

и) проверяют, имеются ли в графике точки, выходящие за пределы регулирования, и свидетельства наличия устанавливаемых (особых) причин;

к) принимают решение о последующих действиях. (ISO 9004-4-98)

По этапам 5–7 эконометрического моделирования пока нет эффективных методов средств системы качества.

Этап 5. Идентификация модели. Статистический анализ модели и оценка параметров регрессионной модели.

Этап 6. Определение качества модели по обучающей выборке.

Этап 7. Верификация модели. Определение адекватности модели по контрольной выборке

Этап 8. Выводы и предложения.

Диаграммы Парето до и после эконометрического моделирования можно использовать для определения эффективности планируемых мероприятий.

Вопросы для самоконтроля

1. Назовите основные принципы эффективной работы предприятия.

2. Назовите пять принципов взаимодействия между процессами.

3. Приведите перечень основных средств и методов улучшения качества.

4. Назовите четыре этапа цикла Деминга.

5. Назовите область эффективного использования диаграммы сродства.

6. Назовите область эффективного использования диаграммы Парето.

7. Укажите методики выявления проблем.

8. Укажите средства менеджмента качества для определения факторов, влияющих на проблему.

9. Приведите этапы составления формы сбора данных.

10. Укажите область эффективного использования контрольных карт.

Глава 14
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ЭКОНОМЕТРИЧЕСКИХ ИССЛЕДОВАНИЙ

14.1. Структурная схема информационных технологий эконометрических исследований

Эконометрические исследования связаны со сбором, обработкой экономической информации, предназначенной для принятия управленческих решений по улучшению процессов. Информационные технологии эконометрических исследований состоят из двух частей: функциональная часть эконометрических исследований и инструментальные средства, предназначенные для выполнения функциональной части.

Функциональная часть эконометрических исследований совпадает с этапами эконометрического моделирования. Инструментальные средства представляют собой набор технических средств и пакетов прикладных программ, с помощью которых выполняются этапы эконометрического моделирования.

14.2. Функциональная часть эконометрических исследований

Функциональная часть эконометрических исследований включает следующие этапы моделирования с перечнем всех эконометрических моделей.

1. Выявление проблем, которые имеются на предприятии.

2. Выявление и анализ проблем наиболее важных для предприятия.

3. Выявление показателя У деятельности предприятия, требующего улучшения.

4. Определение факторов (материалы, машины, методы, операторы среда), оказывающих влияние на У.

5. Определение степени связи факторов с У.

6. Определение тенденций зависимости У от факторов и проведение эконометрического анализа полученных моделей.

7. Шаговый метод включения факторов в модель.

8. Анализ остатков на автокорреляцию и гетероскедастичность.

9. Построение моделей, учитывающих автокорреляцию и гетероскедастичность остатков.

10. Получение точечных и интервальных прогнозов.

11. Построение системы одновременных уравнений.

12. Оценка коэффициентов структурной системы одновременных уравнений методами: косвенным, двухшаговым МНК, трехшаговым МНК.

13. Анализ временных рядов.

14. Вычисление основных характеристик временного ряда.

15. Проверка временного ряда на стационарность.

16. Построение моделей нестационарного временного ряда.

17. Построение моделей стационарного временного ряда.

18. Построение адаптивных моделей временного ряда.

19. Получение точечных и интервальных прогнозных значений временного ряда с учетом выбранной модели.

20. Составление предложений по улучшению процессов.

14.3. Инструментальные средства выполнения функционального блока эконометрических исследований

Для выполнения функциональной части эконометрических исследований можно использовать два вида технических средств: микрокалькуляторы и ПК.

Вычислительные возможности микрокалькуляторов очень ограничены и практически не позволяют эффективно проводить расчеты. Программное обеспечение ПК включает электронные таблицы Excel и пакеты прикладных программ.

Excel обладает следующими достоинствами:

- является доступным по месту и времени, так как входит в состав Microsoft Office любого ПК;
- имеется возможность проведения расчетов с использованием большого количества статистических функций. С помощью статистической функции "Линейн" можно получить расчеты основных характеристик всех эконометрических моделей;
- имеется доступное обширное методическое обеспечение с примерами использования в эконометрике;
- обязательно изучается в курсе информатики;
- очень эффективен на этапе обучения эконометрики;
- электронные таблицы согласуются с большинством статистических пакетов прикладных программ;
- имеется пакет анализа данных, содержащий некоторые эконометрические методы;
- имеется возможность составления программ реализации эконометрических моделей.

Excel обладает следующими недостатками:

- не реализована функция вычисления собственных значений и собственных векторов матрицы;
- очень ограничены возможности трехмерной графики;
- малоэффективен при расчетах с изменяющимся объемом выборки;
- слабая защищенность электронных таблиц от копирования.

Для устранения недостатков Excel можно предложить следующие рекомендации:

- вычисление собственных значений и собственных векторов матрицы можно реализовать в среде Excel по соответствующим алгоритмам;
- можно добавить в Excel двухмерную и трехмерную графику следующим образом. При инсталяции статистичес-

кого пакета S-PLUS 2000 можно добавить в Excel очень обширный двухмерный и трехмерный (2D, 3D) графический модуль S-PLUS, который остается там постоянно;

◆ при проведении расчетов с изменяющимся объемом выборки надо пользоваться статистическими пакетами прикладных программ;

◆ защитить программный продукт от копирования в Excel практически невозможно. Существуют разного рода “уловки” (скрытые файлы, файлы с расчетами располагают в разных директориях, расчетные формулы располагают на рабочем листе значительно ниже основного текста, расчетные формулы делают невидимыми с помощью белого цвета и др.), но специалист их всегда обнаружит.

Статистические пакеты прикладных программ имеют очень много методов и моделей анализа экономической информации, большинство из которых можно использовать в эконометрике. Проблема заключается в том, что мало отечественной литературы, которая описывала бы все методы, имеющиеся во всех статистических пакетах. Справочная система зарубежных пакетов написана на английском языке и во многом случае является ограниченной. В сложившейся ситуации не специалисту следует предпочтение отдавать отечественному пакету с достаточной справочной системой.

14.4. Классификация и обзор пакетов прикладных программ, используемых в эконометрических исследованиях

Для успешного функционирования в условиях жесткой конкуренции западные фирмы, банки, страховые компании и т. д. нуждаются в тщательном анализе имеющейся информации и получении надежных и обоснованных выводов. Поэтому потребность в средствах статистического анализа данных на Западе очень велика, что и послужило причиной для развития рынка статистических программ, на котором предлагаются более тысячи программ. Различные по объему и качеству реализованной статистики, области возможного применения, пользовательскому интерфейсу, цене, требованиям к оборудованию, они отражают многообразие потребностей обработки данных в различных областях человеческой деятельности.

Даже справочники, содержащие очень краткие описания пакетов, составляют солидные тома. Информацию о новых версиях пакетов можно найти в популярных компьютерных журналах и газетах "PC Magazin", "PC Word", "Byte", "PC Week" и др. Некоторые рекомендации по выбору статистических пакетов периодически публикует "Мир ПК".

Число статистических пакетов, получивших распространение в России, тоже достаточно велико (несколько десятков), и спрос на них заметно возрос в середине 90 гг. XX в. Из зарубежных пакетов это STATGRAPHICS, SPSS, SYSTAT, BMDP, SAS, STATISTICA, S-plus и др. (кстати, большинство из этих пакетов занимают по качеству лидирующие места в мире). Из отечественных можно назвать такие пакеты, как STADIA, ЭВРИСТА, МЕЗОЗАВР, ОЛИМП: СтатЭксперт, Статистик-Консультатнт, Matrixer, САНИ, КЛАСС-МАСТЕР и др. Имеется проблема выбора наиболее подходящего пакета для данной категории пользователей, круга решаемых задач, типа и возможностей компьютеров и т. д.

Виды статистических пакетов

Основную часть имеющихся статистических пакетов составляют специализированные пакеты и пакеты общего назначения, а также неполные пакеты общего назначения.

Специализированные пакеты обычно содержат методы из одного-двух разделов статистики или методов, используемых в конкретной предметной области (контроль качества промышленной продукции, расчет страховых сумм и т. д.). Программа TableCurve2D служит для расчета характеристик для более двух тысяч различных аналитических математических функций.

Чаще всего встречаются пакеты для анализа временных рядов (например, Эвриста, МЕЗОЗАВР, ОЛИМП:СтатЭксперт, Forecast Expert), регрессионного и факторного анализа, кластерного анализа, многомерного шкалирования. Обычно такие пакеты содержат весьма полный набор традиционных методов в своей области, а иногда включают также оригинальные методы и алгоритмы, созданные разработчиками пакета. Как правило, пакет и его документация ориентированы на специалистов, хорошо знакомых с соответствующими методами. Применять такие пакеты целесообразно в тех случаях, когда требуется систематически решать задачи из той области, для которой предназначен специализированный пакет, а возможностей пакетов общего назначения недостаточно.

Пакеты общего назначения. Особое место на рынке занимают так называемые статистические пакеты общего назначения. Отсутствие прямой ориентации на специфическую предметную область, широкий диапазон статистических методов, дружелюбный интерфейс пользователя привлекает в них не только начинающих пользователей, но и специалистов. Универсальность этих пакетов особенно полезна:

- на начальных этапах обработки, когда речь идет о подборе статистической модели или метода анализа данных;
- в процессе обучения основам статистики.

Пакеты общего назначения составляют большинство продаваемых на рынке статистических программ. К таким пакетам относятся системы STADIA, STATGRAPHICS, SPSS, SYSTAT, S-plus и др.

Неполные пакеты общего назначения, как правило, это либо недоработанные первые версии вновь создаваемых пакетов, либо вынесенные на рынок программы для внутреннего, узкоспециализированного использования. Они характеризуются ограниченностью статистических методов, недоработанными интерфейсами, скудностью сервисных возможностей, отсутствием или слабой методической проработкой документации.

Требования к статистическим пакетам общего назначения

Для того чтобы статистический пакет общего назначения был удобен и эффективен в работе, он должен удовлетворять следующим требованиям:

- должен содержать достаточно полный набор стандартных статистических методов;
- быть достаточно простым для быстрого освоения и использования;
- отвечать высоким требованиям к вводу, преобразованиям и организации хранения данных, а также к обмену с широко распространенными базами данных (Excel, dBase и др.);
- иметь широкий набор средств графического представления данных и результатов обработки: картинка порой отражает суть дела лучше, чем любые статистические показатели;
- предоставлять удобные возможности для включения в отчеты таблиц исходных данных, графиков, промежуточных и окончательных результатов обработки;
- иметь подробную документацию, доступную для начинающих и информативную для специалистов-статистиков.

Отечественными разработчиками статистических пакетов являются НПО "Информатика и компьютеры" (пакеты STADIA, CONAN, SIGN); "СТАТ-ДИАЛОГ" (пакеты МЕЗОЗАВР, КЛАСС-МАСТЕР, САНИ), Центр статистических исследований МГУ (пакеты Эвриста, Сократ, Ананасс), "Росэкспертиза" (пакет ОЛИМП:СтатЭксперт)[1].

Пакеты по эконометрике

Matrixer 3.4. В настоящее время имеется отечественный специализированный эконометрический пакет Matrixer 3.4, доступный по Интернету (автор А. Циплаков). Справка (на русском языке) содержит редактируемую информацию по эконометрике. Меню включает следующие разделы:

---

[1] *Тюрин Ю. Н., Макаров А. А.* Статистический анализ данных на компьютере. — М.: ИНФРА-М, 1998. С. 468–513.

- линейная регрессия;
- линейная регрессия с гетероскедастичностью;
- метод инструментальных переменных;
- модель Бокса-Дженкинса (ARIMA);
- система одновременных уравнений;
- регрессия с AR(1) процессом в остатках;
- регрессия с MA(1) процессом в остатках;
- регрессия с АРМА процессом в остатках.

Econometric Views 3.1. Среди зарубежных специализированных эконометрических пакетов в России получил широкое распространение как в учебных заведениях, так и в научных институтах. Благодаря стройной и логичной идеологии построения Windows-интерфейса очень прост в освоении. Содержит развитую подсказку (на английском языке), являющуюся справочником по эконометрическим методам. Более широкую информацию см. [5, с. 279–287].

Имеется компьютерный практикум по пакету EViews 3.1 (Молчанов И. Н. 2001, htth//molchanov.narod.ru/ucheb_posob/econommetr_pract_2000.html)

STADIA 6. Среди отечественных статистических пакетов общего назначения следует выделить STADIA 6, который имеет дружественный русскоязычный интерфейс, содержит небольшой (по сравнению с мощными пакетами SPSS) набор статистических методов, но достаточный для проведения эконометрического анализа с использованием средств системы качества, очень прост в освоении, имеет развитую подсказку на русском языке, продолжает использоваться во многих отраслях народного хозяйства, научных и учебных заведениях с 1989 г.

Обучающая версия STADIA и новости свободно доступны в Internet: statsoft.msu.ru. Обучающая версия отличается от профессиональной только отсутствием возможности чтения и записи файлов базы данных. Для учебного процесса этот недостаток не является существенным.

Общее представление о статистических процедурах дает меню статистики, представленное в табл. 14.1.

Таблица 14.1

Меню статистических методов пакета STADIA

Статистические методы	
Параметрические тесты	**Дисперсионный анализ**
Описательная статистика	1-факторный
Гистограмма и нормальность	2-факторный
Корреляция	Многофакторный
Тесты Стьюдента и Фишера	Ковариационный
Непараметрические тесты	**Регрессионный анализ**
Хи-квадрат	Сравнение двух регрессий
Сдвига (положения)	Простая регрессия (тренд)
Масштаба (рассеяния)	Множественная линейная регрессия
Корреляция (независимость)	Пошаговая регрессия
Кросстабуляция	Общая (нелинейная) регрессия
Анализ временных рядов	**Многомерные методы**
Корреляционный анализ	Дискриминантный анализ
Спектральный анализ	Кластерный анализ
Сглаживание и фильтрация	Факторный анализ
ARIMA — модели	Шкалирование
	Меню: дополнения

Подробное описание пакета содержится в книге[1].

Имеется литература, где дается теория статистического анализа и решение практических задач, с использованием в основном пакетов STADIA, STATGRAPHICS, а также описание применения пакетов SPSS и Эвриста[2].

STATGRAPHICS. Среди зарубежных статистических пакетов общего назначения следует выделить STATGRAPHICS, получивший распространение среди российских и украинских вузов. Общее представление о статистических процедурах дает меню представленного в табл. 14.2.

---

[1] *Кулаичев А. П.* Методы и средства анализа данных в среде Windows Stadia. — М.: Информатика и компьютеры, 1999.

[2] *Тюрин Ю. Н., Макаров А. А.* Статистический анализ данных на компьютере. — М.: ИНФРА-М, 1998.

Таблица 14.2

Меню статистических методов пакета STATGRAPHICS

STATGRAPHICS Statistical Graphics System	
DATA MANAGEMENT SYSTEM UTILITIES	TIME SERIES PROCEDURES
Data Management	Forecasting
System Environment	Quality Control
Report Writer and graphics Replay	Smoothing
Graphics Attributes	Time Series Analysis
PLOTTING AND DESCRIPTIVE STATISTICS	ADVANCED PROCEDURES
Plotting Function	Categorical Data Analysis
Discriptive Methods	Multivariate Methods
Estimation and Testing	Nonparametric Method
Distribution Function	Sampling
Exploratory Data Analysis	Experimental Design
ANOVA AND REGRESSION ANALYSIS	MATHEMATICAL AND USER PROCEDURES
Analysis of Variance	Matematical Function
Regression Analysis	Suppleementary Operations

Пакет STATGRAPHICS превосходит STADIA в таких разделах статистики, как планирование эксперимента, лог-линейный анализ, прогнозирование, использование различных методов сглаживания и др. В системе STADIA полнее, чем в STATGRAPHICS, реализованы методы дисперсионного анализа, кластерного анализа, многомерного шкалирования и критерии проверки согласия для сложных гипотез. Начиная с пятой версии STATGRAPHICS содержит подробное описание используемых процедур.

Анализ временных рядов с помощью пакетов прикладных программ

Методы анализа временных рядов широко представлены во многих универсальных статистических пакетах. Но анализ временных рядов — это очень специфическая область статистики, отличающаяся по кругу задач и методов их решения, а также по кругу пользователей, применяющих эти методы. Поэтому для анализа временных рядов имеются также и специализированные статистические пакеты. Рассмотрим возможности специализированного статистического пакета ЭВРИСТА и универсального статистического пакета SPSS. Выбор данных пакетов обусловлен следующими причинами.

ЭВРИСТА является одним из лучших специализированных отечественных пакетов для анализа временных рядов. Его функциональные возможности значительно шире стандартных процедур анализа временных рядов универсальных статистических пакетов. Пакет постоянно совершенствуется и пополняется, он хорошо зарекомендовал себя во многих организациях[1].

Анализ временных рядов в пакете ЭВРИСТА

Меню статистических методов пакета включает следующие процедуры:

1. Предварительный анализ — вычисление основных описательных статистик ряда, различные преобразования ряда (удаление среднего, преобразование Бокса-Кокса, сезонные и несезонные разности), проверка гипотезы о случайности и нормальности выборки (критерии поворотных точек, хи-квадрат, Колмогорова-Смирнова) и др.

2. Анализ тренда — оценка тренда методом простой линейной или нелинейной регрессии, методом скользящего среднего, а также удаление тренда из временного ряда и предоставление других сервисных возможностей.

3. Прогнозирование — построение прогнозов с помощью сглаживания методом Брауна и сезонного сглаживания Хольта-Уинтерса, а также прогнозов на базе подобранных моделей простой и полиномиальной регрессии и моделей типа авторегрессии скользящего среднего.

---

[1] *Баласанов Ю. Г., Дойников А. Н., Королев М.Ф., Юровский А. Ю.* Прикладной анализ временных рядов с программой ЭВРИСТА. — Центр СП Диалог: МГУ, 1991.

4. Спектральный анализ — оценка спектральной плотности и автокорреляционной функции с помощью различных периодограмм.

5. АРСС модели — идентификация сезонной и несезонной моделей авторегрессии скользящего среднего с помощью выборочных автокорреляционных и частных автокорреляционных функций. Автоматический подбор порядка этих моделей с помощью различных критериев, подгонка сезонных моделей АРСС и выявление соответствующей ей спектральной плотности, автокорреляционной функции или таблицы параметров.

6. Кепстральный анализ — оценка кепстра для исследования ряда на наличие в нем эхо-эффекта, а также удаление эха в кепстральной или временной области.

7. Кросс-спектр — оценка параметрической и непараметрической кросс-спектральной плотности, кросс-корреляционной функции и когерентности, а также модели передаточных функций.

8. Регрессионные модели — оценка линейных регрессионных моделей в различных режимах, включая шаговую регрессию.

9. Факотрный анализ — оценка модели факторного анализа методом главных компонент.

10. Анализ интервенций — оценка параметрой динамической модели интервенции и удаление интервенции из ряда. Порядок и вид интервенции могут задаваться вручную или вычисляться автоматически.

11. Гармонический анализ — оценка параметров гармонической модели и их статистических характеристик.

12. Моделирование данных — моделирование сезонных и несезонных временных рядов типа авторегрессии скользящего среднего, а также функции интервенции.

Этот далеко не полный список возможностей пакета наглядно иллюстрирует многообразие методов и инструментов анализа временных рядов.

Анализ временных рядов в SPSS

Универсальный пакет SPSS занимает одно из первых мест в мире среди программ статистической обработки данных. Другим аргументом выбора данного пакета является желание познакомить читателя с англоязычной терминологией в области анализа временных рядов.

Современные версии пакета имеют модульную структуру, в которой анализ временных рядов выделен в отдельный модуль SPSS Trend. При отсутствии этого модуля возможности в области анализа временных рядов будут ограничены только процедурами регрессионного анализа базового модуля пакета SPSS Base.

Кратко перечислим основные процедуры пакета SPSS в области анализа временных рядов:

- Regression (Регрессионный анализ) — позволяет выделять широкий набор моделей трендов;
- ARIMA (Модели авторегрессии проинтегрированного скользящего среднего или модель) — вычисляет оценки параметров для сезонной и несезонной моделей, а также строит доверительные интервалы для прогноза;
- EXSMOOTH (Экспоненциальное сглаживание) — включает широкий набор моделей экспоненциального сглаживания для сезонных и несезонных рядов с трендом;
- SEASON (Сезонные составляющие) — оценивает мультипликативные или аддитивные сезонные составляющие для сезонных временных рядов;
- SPECTRA (Спектральный анализ) — производит разложение временного ряда на гармонические составляющие. Вычисляет и выводит на график одномерную и двумерную периодограмму и оценку спектральной плотности. Позволяет использовать различные спектральные окна;
- AREG (Авторегрессионный анализ) — оценивает регрессионную модель, когда ошибки близких по времени значений ряда коррелируют между собой.

Пакет также выполняет широкий круг других процедур, например генерацию временных рядов, вычисление автокорреляционной и частной автокорреляционной функции, построение различных типов графиков временных рядов и т. д.

Электронные обучающие комплексы по эконометрике

Одним из электронных обучающих комплексов по эконометрике является Econ3 (автор Валентинов В. А.).

Обучающий комплекс Econ3 предназначен для проведения лабораторных занятий и самостоятельного изучения курса эконометрики студентами очной, заочной форм обучения, слушателями факультетов при получении второго высшего образования в режиме обычного аудиторного или дистанционного обучения.

Обучающий комплекс Econ3 содержит: введение, 17 лабораторных работ (в расчете на один семестр) и приложение.

Введение включает такие элементы:

- введение (описание комплекса);
- меню лабораторных работ;
- требования для получения зачета;
- перечень умений и навыков для получения зачета;
- методика проведения занятия;
- презентация всех 17 лабораторных работ.

Каждая лабораторная работа имеет следующую структуру:

- **введение** содержит уточнение цели лабораторной работы, пояснения к основным понятиям, краткое изложение сути методов, которые используются в лабораторной работе;
- **реферат** включает основные сведения о лабораторной работе (тема, объект и предмет изучения, актуальность темы, рабочая гипотеза, метод, способ, задачи, ожидаемый результат, а также конечный результат выполненной работы);
- **вопросы самостоятельной работы** переносятся в тетрадь, на которые составляются ответы по рекомендованной литературе;

• **входной обучающий — контролирующий тест** предназначен для проверки степени подготовки студентов по вопросам самостоятельной работы;

• **опорный конспект** содержит расчетные формулы тех моделей, которые используются в лабораторной работе;

• **задания четырех уровней сложности** предназначены для студентов, обладающих разными уровнями математической и компьютерной подготовки, а также учитывают индивидуальные особенности обучающихся.

• **задания первого уровня сложности** предназначены для выполнения всеми студентами и содержит решение контрольного примера с возможностью самостоятельного выполнения задания;

• **задания второго уровня сложности** предлагают студентам повторить расчетные формулы существующих блоков;

• **задания третьего уровня сложности** предлагают студентам: разработать новые блоки, обладающие дополнительными возможностями; реализовать другую модель изучаемого экономического процесса; проявить творческие способности по более глубокому изучению теоретического и программного обеспечения курса эконометрики. Обычно приводятся алгоритмы и примеры реализации этих заданий;

• **задания четвертого уровня сложности** предлагают студентам найти в литературе описание метода решения задачи и создать блок, реализующий данный метод, оформленный в виде реферата. После выполнения этого задания оно переходит на третий уровень сложности и приобретает имя разработчика.

• **дополнительные задания** включает виды работ, которые появились в процессе совершенствования данного обучающего комплекса. Например, в дополнительные задания включены выполнение несложных расчетов с помощью профессиональных пакетов прикладных программ;

• **биографии ученых-эконометристов** содержат основные биографические данные выдающихся ученых в области использования математических методов в экономических исследованиях;

• **выходной тест** предназначен для проверки умений и навыков, полученных в процессе выполнения лабораторной работы.

Приложение имеет следующий состав:

• рабочая учебная программа;
• рабочая тетрадь по эконометрии;
• биографии ученых эконометристов;
• благодарности.

Опубликованные статьи по эконометрики:

• генерация эконометрических моделей;
• определение степени влияния факторов в регрессионной модели.

Приводим меню комплекса Econ3 (табл. 14.3.)

Таблица 14.3

Меню комплекса Econ3

№	Темы лабораторных работ
1	Введение (презентация учебного комплекса)
2	Визуальный анализ динамики экономического показателя
3	Расчет характеристик линейной модели
4	Эконометрический анализ результатов расчета
5	Спецификация модели
6	Контрольная работа по теме " Линейная модель"
7	Расчет характеристик модели матричным способом в EXCEL
8	Производственная функция
9	Мультиколлинеарность
10	Гетероскедастичность (неоднородность) остатков
11	Анализ временных рядов. Автокорреляция остатков
12	Модель лаговых процессов
13	Спектральный анализ экономических временных рядов
14	Система одновременных уравнений
15	Методы исследования качественных показателей
16	Изучение пакетов прикладных программ
17	Контрольная работа по основным темам курса
	Приложения

MINITAB 13.31. Англоязычный интерфейс. Имеются практически все инструменты качества, имеется тест на выявление регулярностей. Некоторые возможности пакета изображены на рис. 14.1, 14.2.

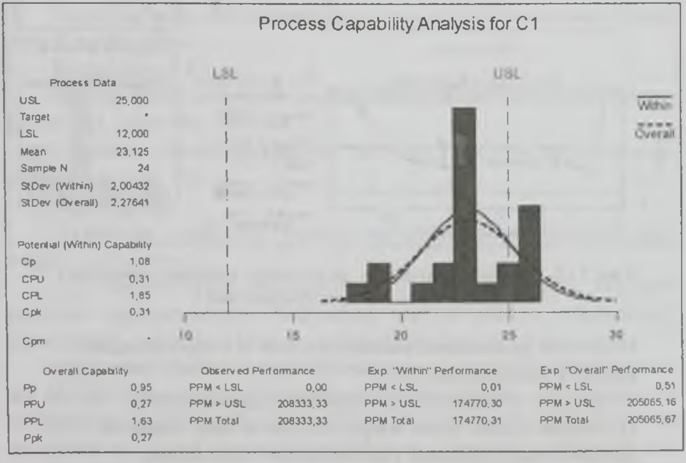


Рис. 14.1. Гистограмма и воспроизводимость процесса, нижние и верхние допуски

Stadia 6. Пакет имеет следующие инструменты качества: гистограмма, диаграмма Парето, контрольные карты разного вида.

Объединение "Приоритет" (Нижний Новгород) имеет следующий набор специализированных программ по инструментам качества на русском языке.

Attestator 2.0:

Контрольные карты.

Гистограммы.

Регрессия.

Расчет характеристик распределений.

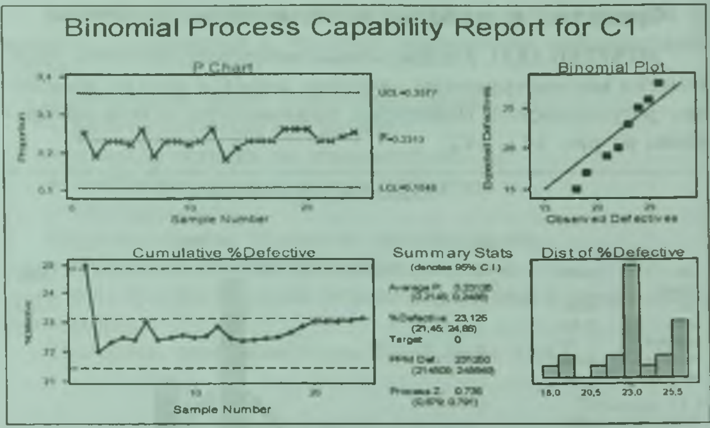


Рис. 14.2. Анализ количества дефектных изделий, имеющих биномиальное распределение

Индексы возможностей процессов и оборудования.
Quality Informator 2.1:
Описание конкретных условий производства.
Периодический ввод информации в базу данных.
Анализ информации различными способами.
Plank 2.0:
Выбор типового плана контроля.
Проверка произвольного плана на допустимость.
Получение плана по произвольным исходным данным.
Графический анализ характеристик плана.
QStat 3.0:
Подбор наилучшего плана контроля.
Проведение анализа статистических характеристик.
Анализ вероятности арбитражной ситуации.
Шкала 2.0:
Ввод базы данных по конкурирующей продукции.

Сравнение аналогов продукции по различным показателям качества.

Изучение тенденций изменения требований к потребительским свойствам

продукции.

Нормирование и прогнозирование требования к потребительским свойствам модернизируемой и вновь создаваемой продукции, а также ее цены.

Определение соответствия цены потребительским свойствам продукции.

Regulator 2.0:

Выбор удобного количества зон выборочного контроля в пределах допуска.

Задание удобных границ зон или количества изделий, пропускаемых без

контроля.

Указание удобного объема выборки при очередной проверке состояния.

По контактному телефону (8312) 46-63-37 вам вышлют каталог программных продуктов, посвященных управлению качеством, или см. Интернет http://prioritet.tripod.com.

Мультимедийные программные средства К программному обеспечению эконометрики следует отнести средства подготовки мультимедийных учебных пособий по курсу. Эта тема выходит за рамки курса эконометрики. Поэтому приведем возможности только программы Camcorder, которых было достаточно, чтобы подготовить деморолики для электронного обучающего комплекса Econ3.

Camcorder. Программа Camcorder входит в состав лицензионного пакета программ Microsoft Office и предназначена для создания видеофильма изображений монитора при выполнении действий оператора. Частота кадров позволяет сделать плавным движение мышки по экрану. Имеется возможность добавить изображение звуковым сопровождением. Для сжатия полученного файла служит процедура уплотнения информации. В среднем для демонстрации деморолика в течение 1 мин создается файл размером 100 кбайт. Программа Camcorder имеет англоязычный справочник, инсталляция проходит без проблем, гиперссылка является самым устойчивым методом обращения в Excel к файлу деморолика.

Программные средства создания динамических моделей

К программному обеспечению эконометрики можно отнести средства создания динамических моделей.

Можно выделить два вида динамических моделей, создаваемых с помощью программы Sketchpad и средствами Excel.

Динамические модели — это такие модели, которые изменяют свои характеристики при изменении управляемых факторов или элементов. Эти модели предназначены для более углубленного изучения свойств модели от изменяемых величин.

Программа **Sketchpad** имеет встроенный язык для программирования, в котором можно разобраться с помощью приводимых готовых разработок. Интерфейс англоязычный. Приводим одну из готовых динамических моделей, созданные с помощью программы Sketchpad (рис. 14.3).

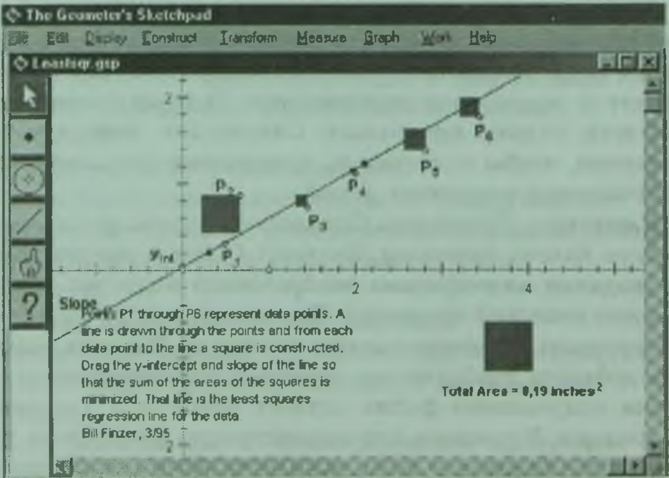


Рис. 14.3. График динамической модели

Для изменения управляемого элемента предусматриваются следующие действия пользователя: курсором мышки выделяется уравнение регрессии и изменяют ее положение на графике, одновременно с этим изменяются значения остатков и общая сумма квадратов остатков. Необходимо найти такое положение уравнения регрессии относительно фактических данных, при котором сумма квадратов остатков будет минимальна.

Программа Sketchpad располагает большим набором управляемых элементов и широкими возможностями. Однако следует выделить три недостатка: англоязычный интерфейс, недостаточно доступно описан язык программирования, большая трудоемкость при создании динамических моделей. В пакет входит набор готовых динамических моделей в основном по математике.

Динамические модели, созданные в Excel. Структура динамической модели показана на рис. 14.4.

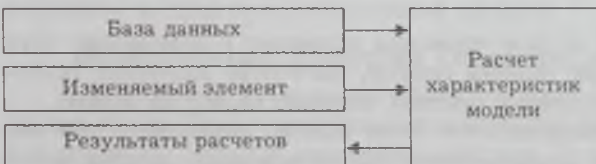


Рис. 14.4. Структура динамической модели

База данных может находиться в любом месте электронной таблицы.

Расчет характеристик модели, как правило, достаточно сложный и выносится на отдельный лист.

Изменяемый элемент и результаты расчетов должны быть расположены на одном экране дисплея и доступны для пользователя.

Приводим пример динамической модели адаптивного прогнозирования Брауна, представленный на рис. 14.5.

Динамическая модель приводится в действие кнопкой изменения коэффициента дисконтирования. Изменение ко-

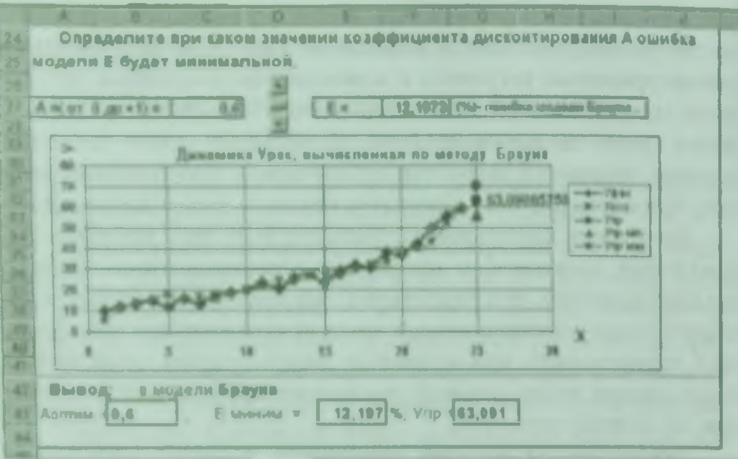


Рис. 14.5. Динамическая модель адаптивного прогнозирования Брауна, реализованная в Excel

эффициента дисконтирования приводит к изменению всех характеристик: ошибка модели, расчетные значения, точечный и интервальный прогноз, которые видны на экране.

Перед пользователем стоит задача: определить, при каком значении коэффициента дисконтирования ошибка модели будет минимальной, расчетные значения более точно пройдут через фактические данные, интервальный прогноз будет наименьший. При этом пользователю ясно, куда нажимать, куда смотреть, что надо получить — необходимые составляющие при выполнении любых работ.

Использование динамических моделей, реализованных в Excel, показали высокую эффективность, особенно для студентов заочного обучения, по следующим темам: метод наименьших квадратов (определение зависимости ошибки модели от ее коэффициентов), спецификация модели (зависимость графика математической функции от ее коэффициентов); автокорреляция, авторегрессия, адаптивный метод прогнозирования Брауна (зависимость расчетных значений зависимой переменной, ошибки модели, точечного и интервального прогноза от соответствующих элементов: коэффициента автокорреляции остатков, коэффициентов авторегрессии, коэффициента дисконтирования); анализ временных рядов (при изучении периодических составляющих временного ряда с помощью построения периодограммы зависимости ошибки модели $У_t = a_0 + a_1 t + a_2 Sin(2\pi t/T) + a_3 Cos(2\pi t/T) + e_t$ от периода T); взвешенная регрессия (зависимость ошибки модели, точечного и интервального прогноза от весов значений временного ряда); многомерный анализ (при построении четырехмерного графика для определения изменения формы трехмерного графика под воздействием четвертой переменной); система одновременных уравнений (при определении среднединамических равновесных значений эндогенных переменных при изменении любой из эндогенных или экзогенных переменных в системе одновременных уравнений). Все перечисленные динамические модели реализованы в Econ3 и вызывают у студентов повышенный интерес.

Программные средства создания тестов

К программному обеспечению следует отнести средства создания тестов. Существует большое количество программ разной стоимости, которые позволяют создавать тесты. Для контроля знаний можно использовать профессиональные программы тестов, реализующие следующие функции: случайный выбор вопроса, ограничение времени ответов, промежуточные и итоговые оценки, составление протокола тестирования и др. Однако следует указать на возможности Excel для эффективного создания обучающих тестов по курсу эконометрика.

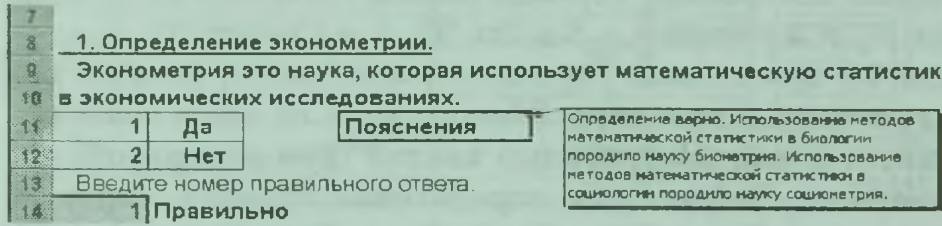
В среде Excel можно реализовать тесты следующего содержания:

- среди набора утверждений указать правильные;
- на приведенном графике указать место нахождения его элементов;
- выполнить расчеты разной сложности, от использования простых операций до статистических функций и мат-

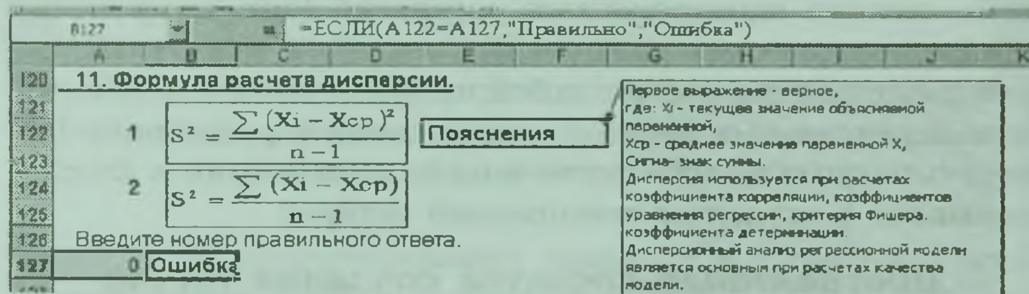
ричных операций, результат расчетов занести в ячейку для проверки его правильности;

- создать сводный отчет по тестам.

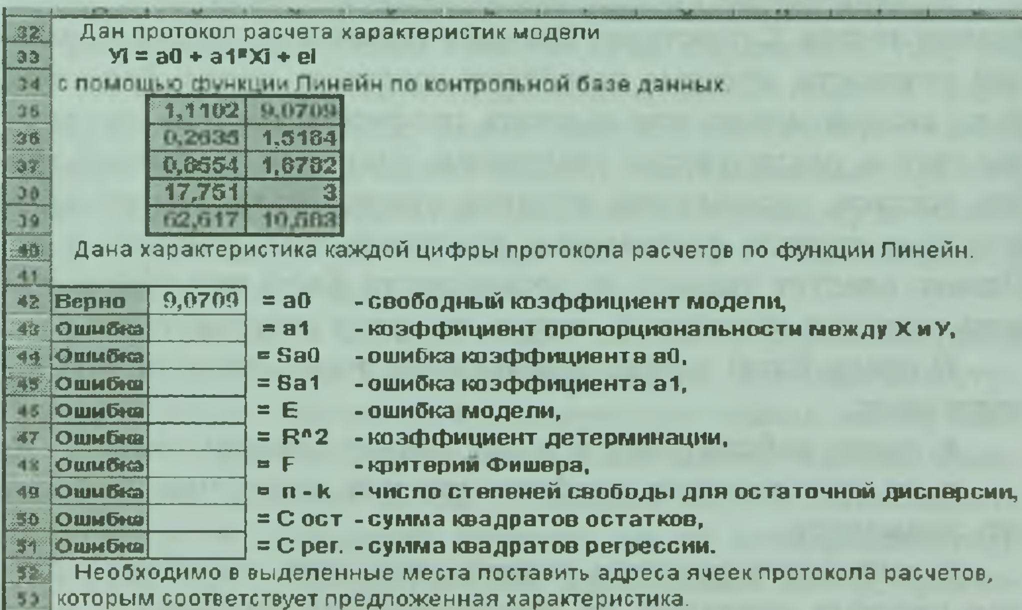
Приводим примеры тестов, созданных в Excel (рис. 14.6, А, Б, В, Г, Д).



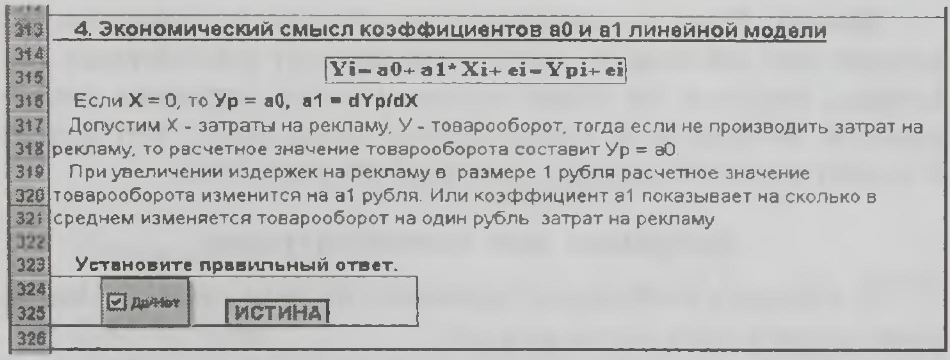
А)



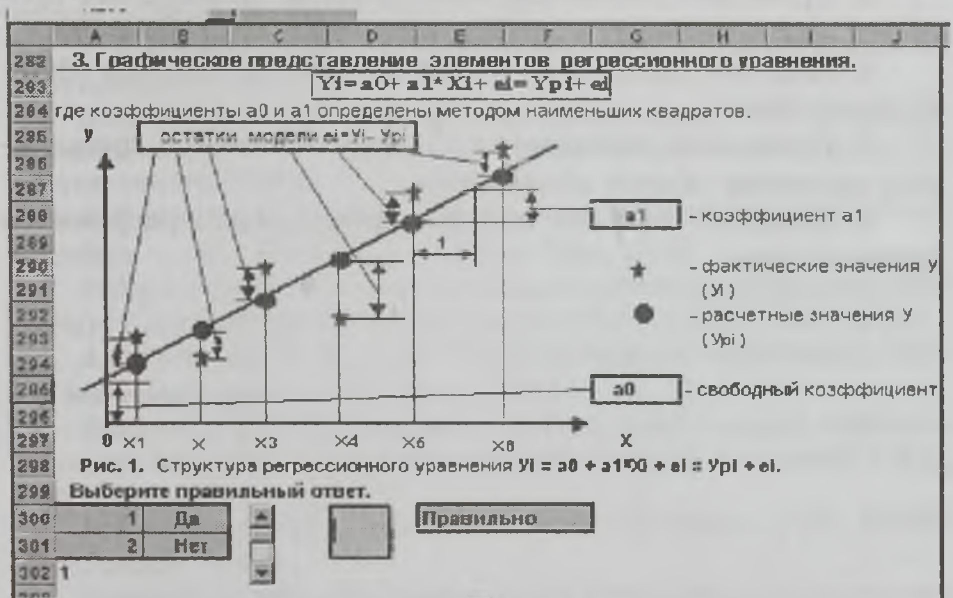
Б)



В)



Г)



Д)

Рис. 14.6. Примеры тестов, созданных в Excel

Возможности Excel при создании обучающих тестов, связанных с построением графиков и проведения расчетов с использованием статистических функций или программ “Пакет анализа”, часто превосходят мощные тестирующие профессиональные программы.

Вывод. Если пользователь не владеет в совершенстве английским языком, то надо пользоваться российскими пакетами, которые по своим возможностям уступают иностранным, но при этом пользователь точно знает, что делает и может точно объяснить полученный результат.

Вопросы для самоконтроля

1. Укажите этапы моделирования функциональной части эконометрических исследований.
2. Укажите достоинства и недостатки Excel.
3. Проведите классификацию пакетов прикладных программ, используемых в эконометрических исследованиях.
4. Укажите требования к статистическим пакетам общего назначения.
5. Приведите основные возможности пакетов прикладных программ общего назначения.
6. Укажите свойства пакетов прикладных программ по эконометрике.

Литература

Основная

1. *Айвазян С. А.* Основы эконометрики. Т. 2. — М.: ЮНИТИ-ДАНА, 2001.

Учебник полностью соответствует стандарту по курсу эконометрики для всех экономических специальностей, требует хорошего знания математической статистики.

2. *Елисеева И. И. и др.* Эконометрика. — М.: Финансы и статистика, 2001.

В доступной и простой форме изложены основные понятия курса эконометрики, но неполностью соответствует стандарту для специальности 351400.

3. *Елисеева И. И. и др.* Практикум по эконометрике: Учеб. пособие. — М.: Финансы и статистика, 2002.

Содержит примеры решения только некоторых основных задач по курсу эконометрика с использованием Excel и ППП Statgraphics.

4. *Катышев П. К. и др.* Сборник задач к начальному курсу эконометрики. — М.: Дело, 2002.

Приведены примеры решения основных задач по курсу эконометрики, примеры доказательства и выводы основных положений и формул, используемых по курсу.

5. *Кремер Н. Ш., Путко Б. А.* Эконометрика. — М.: ЮНИТИ-ДАНА, 2002.

Учебник написан в доступной форме с необходимыми выводами и решением сквозного примера, соответствует стандарту для экономических специальностей.

6. *Магнус Я. Р. и др.* Эконометрика. Начальный курс. 6-е изд., перераб. и доп. — М.: Дело, 2004.

Учебник содержит основы курса эконометрика с использованием теоретического обоснования формул.

Дополнительная

7. *Айвазян С. А., Мхитарян В. С.* Теория вероятностей и прикладная статистика. Т. 1. — М.: ЮНИТИ-ДАНА, 2001.

Учебник содержит основные сведения из теории вероятностей и математической статистики, которые используются в эконометрике.

8. *Айвазян С. А., Мхитарян В. С.* Практикум по прикладной статистике и эконометрике. — М.: МЭСИ, 2000.

9. *Адлер Ю. П., Полховская Т. М., Шпер В. Л., Нестеренко П. А.* Управление качеством. Ч. 1. Семь простых методов. — М.: МИСИС, 2002.

Контрольный листок, диаграмма Парето, гистограмма, диаграмма Исикавы, стратификация, корреляционный и регрессионный анализы, контрольные карты, диаграмма рассеивания.

10. *Бородич С. А.* Эконометрика. — Минск: Новое знание, 2001.

Учебник содержит в простой и доступной форме основные положения эконометрики в соответствии с белорусским и украинским стандартами. Однако нет модели периодических колебаний и адаптивных методов прогнозирования. Данный учебник можно рассматривать как конспект лекций по эконометрике.

11. *Валентинов В. А.* Генерация и классификация эконометрических моделей // Региональные перспективы. 2000. № 5 (12). С. 48–51.

Приводится механизм генерации эконометрических моделей, метод моделирования временных рядов с периодической составляющей, алгоритм расчета коэффициентов логистической функции.

12. *Валентинов В. А.* Влияние цикличности солнечной активности на уровень жизни людей. Сб. трудов. — Весник Ровненского государственного технического университета. Выпуск 2 (9), 2001. С. 200–206.

Устанавливается синхронность изменений 11-летнего цикла в рядах уровня жизни россиян и чисел Вольфа солнечной активности с 1913 по 1995 г.

13. *Валентинов В. А.* Эконометрия. Толково-терминологический словарь. — Полтава: ПКИ, 2001.

Приводятся основные понятия эконометрики.

14. *Валентинов В. А.* Определение связи между пиками. Региональные перспективы // Научно-практичный журнал. 2001. № 4(17). С. 26–28.

Выводится критерий связи между пиками двух временных рядов с лаговой задержкой. Приводится критическое значения крите-

рия на заданном уровне значимости в зависимости от размеров временных рядов.

15. *Виноградов С. Н.* Открытие Шаталова. — М.: ГУП ЦРП, 2003.

Коротко и емко изложены основные дидактические принципы, используемые Шаталовым. Особо следует выделить оптимальное число 7 количества одновременно воспринимаемых идей, элементов. Роль визуального восприятия информации и т. д.

16. ГОСТ Р 50779.40-96 (ИСО 7870-93) — Контрольные карты. Общее руководство и введение.

17. ГОСТ Р ИСО серии 9000 — 2001. Системы менеджмента качества. Основные положения и словарь. — М.: Госстандарт России, 2001.

18. *Глудкин О. П. и др.* Всеобщее управление качеством. — М.: Горячая линия — Телеком, 2001.

Один из базовых учебников по системе качества.

19. *Дорохина Е. Ю., Преснякова Л. Ф., Тихомиров Н. П.* Сборник задач по эконометрике. — М.: Экзамен, 2003.

Имеются примеры доказательств; задачи для расчетов показателей предполагают использовать Excel и STATGRAPHICS.

20. *Ежеманская С. Н.* Эконометрика. — Ростов н/Д: Феникс, 2003.

Просто и в доступной форме изложены все разделы эконометрики.

21. *Замков О. О.* Эконометрические методы в макроэкономическом анализе. — М.: ГУ ВШЭ, 2001.

Курс лекций по классической эконометрике. Приводятся примеры расчетов макроэкономических моделей с использованием реальных данных России и США. Нет анализа временных рядов.

22. ИСО 9004-4-93 Административное управление качеством и элементы системы качества. Часть 4. Руководящие указания по улучшению качества. — М.: ГУП ЦПП, 1995.

Изложены принципы качества, этапы и средства улучшения процессов.

23. *Колемаев В. А.* Эконометрика: Учебник. — М.: ИНФРА-М, 2004.

Автору удалось достичь баланса между научностью и доступностью. Добротный учебник для экономических специальностей.

24. *Кондратьев Н. Д.* Большие циклы конъюнктуры и теория предвидения. Избранные труды. — М.: ЗАО "Экономика", 2002.

Имеется доклад "Большие циклы экономической конъюнктуры" (стр. 341–397), в котором приводится методика выявления 50-летних циклов (стр. 348–349) и делается попытка их объяснения на основе приведенных статистических данных (стр. 350, 353, 354, 356, 357, 359, 360, 363–366) за 150 лет Англии, Франции, США.

25. *Лапидус В. А. Рекшинский А. Н.* Диплог консультанта с руководителем компании. — Н. Новгорол.: СМЦ "Приоритет", 2000.

Книга полезна для руководителей любого уровня, особенно для сторонников репрессивного менеджмента.

26. *Лапидус В. А.* Всеобщее качество (TQM) в российских компаниях. — М.: ОАО Типография "Новости", 2000.

Монография, написанная на основе большого практического и теоретического материала, изложенной в доступной форме. Настольная книга руководителя любого уровня.

27. *Мардас А. Н.* Эконометрика. — СПб: Питер, 2001.

Особый интерес представляют главы 5 и 6, посвященные эконометрическому эксперименту и моделированию в маркетинговых исследованиях.

28. *Молчанов И. Н., Герасимова И. А.* Компьютерный практикум по начальному курсу эконометрики (реализация на Eviews). — Ростов-н/Д, 2001.

Интернет: htth//molchanov.narod.ru/ucheb_posob/ econommetr_pract_2000.html

Имеется несколько уроков по изучению пакета Econometric Views 3.1

29. *Мхитарян В. С., Архипова М. Ю.* Эконометрика: Учеб. пособие, руководство по изучению дисциплины, методические указания и задания для контрольных работ для студентов заочников, тесты по дисциплине / Московский государственный университет экономики, статистики и информатики. — М.: МЭСИ, 2004.

Имеются ссылки на интернет-ресурсы по каждому разделу курса. Нет главы, посвященной временным рядам.

30. *Новиков А. И.* Эконометрика. — М.: ИНФРА-М, 2003.

Просто и доступно изложены все основные разделы курса.

31. *Образцова О. И., Назарова О. В., Канторович Г. Г.* Экономическая статистика. Эконометрика (Программы, тесты, задачи, решения). — М.: ГУ-ВШЭ, 2000.

Методические материалы содержат подробную программу курса, материалы для проверки знаний, ответы, решения. Объем материала рассчитан на два семестра и содержат дополнительные статистические критерии. В качестве основной рекомендуется переводная литература.

32. *Орлов А. И.* Эконометрика. — М.: Экзамен, 2002.

Материал учебного пособия выходит за рамки обычного курса эконометрики. Особый интерес вызывает глава 13, посвященная эконометрическим методам управления качеством и сертификации продукции. Вводится понятие эконометрика качества. Имеется обширная библиография.

33. *Писарева О. М.* Базовые эконометрические методы и модели. Учеб. пособие. — М.: ГУУ, 2002.

Основное внимание уделено временным рядам.

34. *Просветов Г. И.* Эконометрика. Задачи и решения. — М.: РДЛ, 2004.

Даны примеры решения эконометрических задач по всему курсу эконометрики с использованием Excel. Соответствует программе для экономических специальностей.

35. *Салманов О. Н.* Математическая экономика с применением Mathcad и Excel. — СПб.: БХВ-Петербург, 2003.

Пособие по математическим моделям, имеется раздел, посвященный решению эконометрических задач. Введена тема "Типологическая регрессия". Нет анализа временных рядов.

36. *Тихомиров Н. П., Дорохина Е. Ю.* Эконометрика. — М.: Экзамен, 2003.

Учебник, полностью соответствует Государственному образовательному стандарту.

37. *Фишер Р.* Новые методы торговли по Фибоначчи. — М.: ИК Аналитика, 2002.

38. ISO 9004-1-95. Управление качеством и элементы системы качества. Ч. 1. Руководящие указания.

Приложения

Приложение 1

Информационные ресурсы

Описание Украинской эконометрической модели можно найти в Институте экономического прогнозирования Национальной Академии Наук Украины по адресу: Киев (01011), ул. Панаса Мирного 26.

http://www.cemi.rssi.ru — Исходные данные и описание Российской эконометрической модели, разрабатываемой в ЦЭМИ РАН. Информация о еженедельном проведение научного семинара “Многомерный статистический анализ и вероятностное моделирование реальных процессов”, который работает по средам с 10 до 12 часов (телефон для справок 129-1233, 332-4491, 134-3147).

Matrixer — эконометрический пакет находится на сайте Циплакова А. На этом сайте имеется большая подборка обновляемой разнообразной информации по эконометрике.

http://www.statsoft.msu.ru — Обучающая версия пакета “Stadia 6”.

htth//molchanov.narod.ru/ucheb_posob/econommetr_pract_2000.html — компью-терный практикум по начальному курсу эконометрики (реализация на Eviews).

Rambler ключевые слова: Нобель, экономика — информация о лауреатах Нобелевской премии в области экономики.

Программа семинаров по преподаванию эконометрики, которыми руководит А. А. Пересецкий, расположена на сайте Российской экономической школы:

http://www.nes.ru/russian/outreach/workshops/econometrics.htm

Упражнения по эконометрики в соответствии с учебником [6]:

http://econometrics.nes.ru/mkp/.

Определение эконометрики, общие вопросы:

http://upereslavl.botik.ru/UP/ECON/econometrics/topl/tsld006.htm

http://www.nsu.ru/ef/tsy/ecmr/study.htm

http://www.nsu.ru/ef/tsy/ecmr/index.htm

http://www.statsoft.ru/home/textbook/glossary/default.htm

http://www.dataforce.net/~antl/article/econometric

http://www.tvp.ru/vnizd/mathem4.htm

Обобщенная линейная множественная регрессия, мультиколлинеарность, автокорреляция остатков:

http://www.kgtu.runnet.ru/WD/TUTOR/textbook/modules/stmulreg.html

http://www.shpargalka.ru/statis.ru/doc/shpr_e3l.htm

http://dsmu.donetsk.ua/~statbook/modules/stmulred.html#cunique

http://www3.unicor.ac.ru/d024/p011993.htm

http://www.gauss.ru/educat/systemat/butenkov/.asp

http://crow.academy.ru/econometrics/seminars/sem_08_/sem_08/htm

http://crow.academy.ru/econometrics/lestures_/lect_03/index.htm

http://upereslavl.botik.ru/UP/ECON/econometrics/

http://crow.academy.ru/econometrics

Регрессионные модели с переменной структурой:

http://www.econ.msu.ru/kaf/DEI/books/prognoz/lec10.html

http://www.econ.msu.ru/dei/books/prognoz/lec10.html

3http://soc-gw.univ.kiev.ua/RDUCAT/BASIC/MMPS/LABS/LOGRES.htm

http://www.econ.nsu.ru/ef/tsy/ecmr/some_mle/contents.htm

Нелинейные регрессии и их линеаризация:

http://www.doktor.ru/doctor/biometr/sp/contents4.htm http://www.kgtu.runnet.ru/WD/TUTOR/textbook/modules/stnonlin.htm

http://dsmu.donetsk.ua/~statbook/modules/stnonlin.html

http://softline.perm.ru/statistica/www-page_STATISTICA_for_Windows.html

http://eco.rea.ru/Lspace/EconTh.nsf/136acc8cc4a429f5c325654d004b4fc2

http://lanserv2.kemsu.ru/departs/matekon/Chapter4/par4_4.html

http://www.dvgu.ru/pin/math/for_students/eco/node4.html

http://www.hse.ru/rectorat/grebnev/economics/glava13.htm

http://ecfor.rssi.ru/0497_r_k.htm

http://www.socionet.ru:8100/RuPEc/data/Articles/rusrssicf4_1997article4.html

http://www.mstu.edu.ru/publish/conf/section5/section5/section5_7.html

Производственная функция:

http://www.kgtu.runnet.ru/WD/TUTOR/textbook/modules/stnonlin.htm

http://softline.perm.ru/statistica/www-page_STATISTICA_for_Windows.html

http://ecfor.rssi.ru/0497_r_k.htm

http://www.socionet.ru:8100/RuPEc/data/Articles/rusrssicf4_1997article4.html

http://www.mstu.edu.ru/publish/conf/section5/section5/section5_7.html

Система одновременных уравнений:

http://subscribe.ru/frchive/sciencr.humanity.econometrika/200007/17050500/html

http://www.cemi.rssi.ru/rus/publicat/e-pubs/ep97001/1.htm

http://www.softlist/ru/cgi-bin/program.cgi?id=1988-7K

http://web/ido/ru/www/Courses.nsf/CoursesList?Open&About=055-2K

http://www.freeware32.ru/download.php3?id=1355-17K

http://www.nes.ru/Acad_year_2001/Prob_Stat.htm

(Мхитарян В.С., 2004)

Инструменты системы качества.

http://prioritet.tripod.com — Реклама набора специализированных програм по инструментам системы качества на русском языке.

Приложение 2

Глоссарий

Автокорреляция — корреляция между временной переменной и лаговой переменной, составленной от той же переменной.

Авторегрессия — регрессия зависимости временной переменной от лаговой переменной, составленной от той же переменной.

Агрегировать — объединять, суммировать какие-либо однородные показатели с целью получения более общих, обобщенных, совокупных показателей.

Адаптация — (лат. *adaptatio* — приспособлять) — приспособление строения и функций организма к условиям существования.

Аддитивный (лат. *additio* — прибавление) — полученный путем сложения.

Аппроксимация (лат. *approximare* — приближаться) — приближенное выражение каких-либо величин через другие более простые величины.

Варьирование — изменения численных значений переменной. Для измерения степени варьирования используются среднее квадратическое отклонение, коэффициент вариации, вариация.

Выборочная совокупность — часть генеральной совокупности. Выборочная совокупность должна обладать свойством репрезентативности.

Генеральная совокупность — вся совокупность объектов исследования, объем выборки которой должен быть равен бесконечности.

Гетероскедастичность — неоднородность относительно дисперсии.

Гомоскедастичность — однородность относительно дисперсии.

Главные миноры матрицы — определители различных подматриц, образованные в результате вычеркивания строк и столбцов исходной матрицы, имеющих одинаковые номера.

Детерминированный — причинно обусловленный, строго определенный.

Идентификация (ср.-лат. *idenficare* — отождествлять) — отождествление, установление совпадения чего-либо с чем-либо.

Идентификация в эконометрике — выделение главных и второстепенных факторов, определение коэффициентов модели.

Изолиния — линия на графике, численные значения которой имеют постоянное значение.

Имитация — воспроизведение.

Интервальный прогноз — интервал, в котором с определенной вероятностью находится фактическое значение прогнозной переменной экономического объекта.

Интерполяция (лат. *interpolatio* — изменение) — получение расчетных значений функции при условии, что значения аргумента входят в область определения функции.

Итеративный (лат. *iterativus* — часто повторяемый). Известны следующие итеративные методы вычисления коэффициентов модели: градиентный, крутого восхождения и др. Пакет прикладных программ “Эврика” позволяет производить расчет любых функций итеративными методами. В Excel имеется программа “Поиск решения”, которая определяет оптимальные значения коэффициентов итеративным методом.

Корректный (лат. *correctus* — выправленный) — соответствующий действительности.

Коэффициент корреляции — числовая характеристика совместного распределения двух случайных величин, выражающая их взаимосвязь.

Коэффициент корреляции выборочный r(X,У) рассчитывается по результатам (X_i, $У_i$), i= 1, ..., n наблюдений двумерной величины (X,У) по следующей формуле:

$$r(X,У)=\frac{\sum(X_i-\bar{X})(У_i-\bar{У})}{\sqrt{\sum(X_i-\bar{X})^2\sum(У_i-\bar{У})^2}},$$

где $\bar{X}=\frac{\sum X_i}{n}$ $\bar{У}=\frac{\sum У_i}{n}$.

Величина r измеряет степень линейной зависимости результатов X_i и $У_i$, i = 1, ..., n наблюдений. Целью корреляционного анализа является определение степени связи между переменными для того, чтобы определить, какие факторы следует включить в многофакторную модель.

Примечание. Практически во всех статистических пакетах и пособиях по корреляционному анализу предлагается вычислять коэффициент корреляции по эквивалентной упрощенной формуле

$$r(XУ)=\frac{\sum X_iУ_i-\frac{\sum X_i\sum У_i}{n}}{\sqrt{\left(\sum X_i^2-\frac{(\sum X_i)^2}{n}\right)\left(\sum У_i^2-\frac{(\sum У_i)^2}{n}\right)}}.$$

Следует предупредить, что при значительном n и многоразрядных значениях X и У эта формула может давать существенные ошибки и были случаи, когда коэффициент корреляции, вычисленный по данной формуле, имел численное значение больше единицы.

Для коэффициентов корреляции двух случайных переменных X и У справедливо:

1) $-1 = r \leq +1$;

2) при $r = \pm 1$ имеется функциональная зависимость, все точки графика зависимости У от X лежат на прямой;

3) если $r = 0$, то X и У линейно не связаны между собой или некоррелированы;

4) для двумерной нормально распределенной случайной переменной из равенства $r = 0$ следует стохастическая независимость X и У.

Кумуляция (лат. *cumulatio* — увеличение) — накопление численных значений переменной за определенный период или по отдельным объектам.

Лаг — задержка. Обычно лаг рассматривают между причиной и следствием.

Линеаризация — процедура приведения математической функции к линейному виду.

Математическая статистика — раздел математики, посвященный математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов.

Метод эконометрики — выборочный метод. Сущность его заключается в том, что закономерности, определенные по выборочной совокупности, распространяются на всю генеральную совокупность с определенной вероятностью.

Мультиколлинеарность (лат. *multum* — много + коллинеарность — параллельность) означает множественная связь.

Мультипликативный (лат. *multiplicator* — умножающий) — полученный путем умножения.

Несоответствие — невыполнение установленным требованиям.

Объект эконометрики — экономические процессы, происходящие в экономической системе общества.

Объясняющая переменная (причина, независимая переменная, объясняющий фактор, фактор) — это характеристика объекта, которая вызывает следствие у зависимой переменной.

Ортогональный (гр. *orthogonio* — прямоугольный) — прямоугольная система координат, перпендикулярный.

Плеяда — группа выдающихся лиц, связанных между собой общими целями и задачами.

Плеяды (гр. *Pleias*) — в древнегреческой мифологии семь дочерей Атланта и Плейоны, после смерти помещены Зевсом на небо в виде созвездия.

Положительно определенная матрица. Матрица **V** является положительно определенной, если выполняются следующие условия:

• матрица **V** является невырожденной (определитель матрицы не равен нулю);

• матрица **V** имеет положительные собственные значения и положительный определитель;

• матрица **V** имеет положительные значения для всех главных миноров.

Правило Парето (20/80). Парето в начале XX в. обнаружил закономерность, которая заключалась в том, что 20% богатых людей имеют 80% денежных средств, находящихся у населения. В дальнейшем это правило было распространено на несоответствия: 20% причин дают 80% несоответствий, 20% работников делают 80% брака. Правило Парето (20/80) отразило существующую закономерность о том, что способности человека (быстрота реакции, скорость решаемых задач, степень активности), его физические характеристики (вес, рост, длина стопы) и сложность выполняемых работ подчиняются нормальному закону распределения. Это значит, что в коллективе будет 20% с большими способностями, 20% с низкими способностями и 60% со средними способностями людей. Все виды работ на предприятии имеют 20% большую сложность, 20% — низкую сложность и 60% — среднюю сложность. При этом важно правильно распределить работников по видам работ в соответствии с их способностями. При оказании различных услуг человеку необходимо учитывать, что их потребности имеют нормальный закон распределения.

Предмет эконометрики — количественная оценка взаимосвязи между случайными событиями, признаками, показателями, факторами переменных экономических объектов.

Приведенная система одновременных уравнений — система уравнений, которая отражает зависимость эндогенных переменных только от экзогенных переменных.

Проблема — разница между тем, что хочется человеку, и тем, что имеется.

Проверка наличия корреляции. Наличие корреляции, т. е. гипотеза о том, может ли выборочный коэффициент корреляции иметь случайные отклонения от нуля при генеральной совокупности с параметрами $r = 0$, проверяется по критерию Стьюдента:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n}}} \text{ при } n > 100.$$

Если $|t| > t_{\alpha/2}(\alpha,\ k = n - 2)$, то H_0: $r = 0$ отвергается с вероятностью $1 - \alpha$, в противном случае нулевая гипотеза H_0: $r = 0$ принимается.

Прогноз (гр. *prognosis* — предвидение) — предвидение, предсказание чего-либо. Различают два вида прогнозов: точечный и интервальный.

Ранжировка — расположение численных значений переменной по возрастанию или убыванию.

Регрессионный анализ — раздел аналитической статистики, изучающий форму зависимости характеристик стохастического процесса от одного или нескольких факторов. Термин "**регрессия**" (лат. *regressio* — движение назад) ввел Гальтон. Изучая статистическим методом наследования количественные признаки, он обнаружил, что количество высокорослых и низкорослых родителей отклоняется (регрессирует) от них на 1/3 в сторону среднего уровня этого признака в данной популяции[1].

Признаки элементов, образующих систему, должны быть непосредственно связанными между собой на основании определения системы. Эта связь может быть непосредственной между рядом расположенными элементами; опосредованной между элементами разделенными несколькими элементами.

Сигнал, поступивший на вход системы, изменяет численные значения признаков элементов. Волна изменений, идущая от входа системы представляет динамический про-

---

[1] *Лакин Г.Ф.* Биометрия. — М.: Высшая школа, 1980. С. 181.

цесс, происходящий в пространстве и во времени при наличии обратных связей носит колебательный характер двух видов: затухающий (устойчивый), который приводит к равновесию системы; усиливающийся (неустойчивый), который может разрушить систему.

В предложенной модели элемент системы должен иметь несколько входов и несколько выходов, связывающих его с другими элементами.

В скалярном виде регрессионную модель можно представить следующим образом;

$$У_i = \sum_{i=1}^{k} f(X_{ji}) + e_i,$$

где $i = 1, ..., n$ — номер наблюдения;

$У_i$ — фактические значения зависимой переменной величины i-го наблюдения;

$j = 1, ..., k$ — номер фактора независимой переменной величины;

X_{ji} — численные значения j-го фактора i-го наблюдения;

$f(X_{ji})$ — вид математической функции;

e_i — случайная составляющая, включающая влияние неучтенных факторов, ошибки измерения переменных X и У, ошибки неадекватности используемой математической функции.

Репрезентативность — свойство выборочной совокупности, состоящее в том, что выборка должна быть случайной и объем выборки должен обеспечить заданную ошибку характеристик выборочной совокупности.

Симметрическая матрица. Матрица **V** является симметрической, если выполняется условие $\mathbf{V} = \mathbf{V}'$.

Собственные характеристические значения и векторы. Матрица **V** может иметь собственные характеристические значения и соответствующие им собственные характеристические вектора, которые находятся в следующем соотношении:

$$\mathbf{Vh} = k\mathbf{h},$$

где **V** — исходная матрица размером $n \times n$;

k — число, собственное характеристическое значение матрицы **V** и соответствующий для него **h** — собственный характеристический вектор.

Спецификация (лат. *specificatio* — вид, разновидность + *facere* — делать) — перечисление специфических особенностей чего-либо; документ с перечислением условий, которым должен отвечать производственный заказ.

Спецификация математической модели — определение такой математической функции, которая воспроизводит определенное количество регулярностей (закономерностей) зависимой переменной.

Статистические данные — сведения о числе объектов в какой-либо более или менее обширной совокупности, обладающей теми или иными признаками.

Стохастический — случайный, вероятностный.

Структура корреляционного анализа — проверка наличия корреляции и проверка отсутствия всех типов ложной корреляции.

Структурная система одновременных уравнений — система уравнений, которая отражает зависимость эндогенных переменных как от эндогенных, так и от экзогенных переменных.

Тенденция (лат. *tendere* — направляться, стремиться) — направление, в котором совершается развитие какого-либо явления.

Теоретическая база эконометрики — математическая статистика, экономическая кибернетика.

Типы ложных корреляций. Коэффициент корреляции будет корректным, если соблюдаются две предпосылки, которые относятся к объекту исследования:

1) переменные X и У — случайные величины;

2) каждая переменная X и У в отдельности имеет нормальный закон распределения и является однородной.

Нарушение этих предпосылок порождает ложную корреляцию. Существуют следующие типы ложных корреляций:

• **ложная корреляция стратификации** (неоднородности) возникает в том случае, если одна или обе переменные X и У неоднородны и на координатной плоскости наблюдается несколько облаков точек. Коэффициенты корреляции внутри облаков отличаются от коэффициента корреляции для объединенных облаков;

• **ложная корреляция временных рядов** проявляется в том, что два временных ряда испытывают влияние со стороны третьего общего фактора времени;

• **ложная корреляция Пирсона** возникает в том случае, если X и У являются расчетными величинами, содержащими общую переменную, например

$$X = \frac{X_1}{X_2}; \quad У = \frac{X_3}{X_2};$$

• **ложная корреляция процентных чисел** проявляется в том случае, если X и У — являются процентными числами, дополняющими друг друга, например, X — процент розничного товарооборота от всего товарооборота, У — процент оптового товарооборота от всего товарооборота, X + У = 100;

• **ложная корреляция неслучайных переменных** проявляется в том, что значения переменной задаются исследователем по его воле[1].

Точечный прогноз — среднее прогнозное значение изучаемой переменной экономического объекта.

Требование — потребность или ожидание, которое установлено, обычно предполагается или является обязательным.

Тренд (очевидно, происходит от слова "трен" — шлейф женского белья) — изменение, определяющее общее направление развития, основную тенденцию временного ряда.

Цель эконометрики — получение точечных и интервальных прогнозов деятельности экономических систем.

---

[1] *Четвериков Н. С.* Статистические исследования. — М.: Наука, 1975. С. 316–317.

“**Эконометрика**” состоит из двух слов: эконо (гр. *oikonomia* — управление хозяйством) и метрия (гр. *metreo* — измеряю), что означает измерение в экономике.

Эконометрика как дисциплина — наука, которая использует методы математической статистики в экономических исследованиях.

Экстраполяция (приставка экстра — лат. *extra* — вне) — прогноз, получение расчетных значений при условии, что значения аргумента выходят за пределы области определения функции.

Приложение 3

Тесты по дисциплине

Вопрос 1. Определение эконометрики.

Эконометрика — это наука, которая использует для описания теоретических моделей реальных хозяйственных экономических процессов методы:

Варианты ответа:

1) статистики;

2) математической статистики.

Вопрос 2. Предмет эконометрики.

Предметом эконометрики является определение зависимости между:

Варианты ответа:

1) детерминированными;

2) случайными

событиями, признаками, показателями, факторами, переменными экономических объектов, используемыми при построении теоретических моделей реальных экономических процессов для получения прогнозов деятельности экономических систем.

Вопрос 3. Цель эконометрики.

Целью эконометрики является получение:

Варианты ответа:

1) точечного и интервального;

2) точного

прогноза деятельности всех объектов экономической системы на основании расчетов по данным выборочной совокупности.

Пояснения. *Целью* эконометрики является оценка точечных и интервальных прогнозов деятельности всех объектов экономической системы на основании расчетов по данным выборочной совокупности.

Вопрос 4. Задачи эконометрики.

Задачами эконометрики являются:

Варианты ответа:

1) идентификация, спецификация, оценка параметров, проверка достоверности прогнозной модели;

2) долгосрочное экспертное прогнозирование.

Вопрос 5. Метод эконометрики.

Методом эконометрики является:

Варианты ответа:

1) изучение всей генеральной совокупности;

2) выборочный метод, при котором по выборочной совокупности оцениваются свойства генеральной совокупности.

Пояснения. *Методом* эконометрики является выборочный метод. Сущность выборочного метода заключается в том, что закономерности, определенные по выборочной совокупности, распространяются на всю генеральную совокупность с определенной вероятностью.

Вопрос 6. Теоретическая база эконометрики.

Теоретической базой эконометрики являются:

Варианты ответа:

1) экономическая кибернетика, математическая статистика;

2) статистика.

Пояснения. *Теоретической базой* эконометрики является математическая статистика и экономическая кибернетика.

Вопрос 7. Определение модели.

Модель это:

Варианты ответа:

1) условный образ объекта изучения;

2) точный образ объекта изучения.

Пояснения. *Модель* — условный образ объекта исследования. Модель конструируется субъектом так, чтобы изобразить взаимосвязи переменных, существенные для цели исследования.

Вопрос 8. Классификация экономико-математических моделей.

В состав абстрактных моделей входят:

Варианты ответа:

1) графические, словесно-описательные, математические модели;

2) физические модели.

Пояснения. Все модели можно подразделить по средствам моделирования на материальные и абстрактные.

Материальные модели воспроизводят характеристики объекта с помощью материальных средств. Например, используя объект магазина, можно искать оптимальный вариант размещения оборудования, перемещения покупателей, путей движения товаров.

Абстрактные модели воспроизводят характеристики объекта с помощью умозаключений, которые являются плодом человеческого мышления.

По способу моделирования абстрактные модели подразделяются на графические, математические, словесно-описательные.

Графические модели — это представление взаимосвязи переменных в виде графиков, диаграмм. Визуальный анализ графических моделей является одним из самых мощных средств обнаружения вида тенденции и прогнозной оценки динамики экономического процесса, степени взаимосвязи переменных. Графические модели позволяют провести качественный анализ взаимосвязи переменных.

Математическая модель — это совокупность уравнений, отображающих взаимосвязи и зависимости экономических показателей. Математические модели позволяют количественно измерить степень взаимосвязи переменных и получить расчетные значения признаков объектов.

Словесно-описательные модели — это совокупность умозаключений, которые качественно характеризуют логическую взаимозависимость экономических показателей экономической системы.

Все три вида абстрактных моделей взаимно дополняют друг друга и широко используются в эконометрическом анализе.

Вопрос 9. Регрессионное уравнение.

Зависимость следствия У от причины X моделируется с помощью линейного регрессионного уравнения:

Варианты ответа:

1) $У_i = a_0 + a_1X_i + e_i$;

2 $У_i = a_0 + a_1X_i$.

Пояснения. Общий вид линейной регрессионной модели можно представить следующим математическим выражением :

$У_i = f(X_{1i}, X_{2i}, X_{3i}, ...) + e_i = У_{pi} + e_i$,

где $У_i$ — случайная зависимая переменная;

$У_{pi}$ — расчетные значения зависимой переменной;

$f(X_{1i}, X_{2i}, X_{3i},...)$ — математическая функция, содержащая параметры в первой степени, отражающая детерминированную составляющую $У_i$, которая диктуется законодательной, правовой средой общества в сочетании с потребностями членов общества;

X_{1i}, X_{2i}, X_{3i} — независимые переменные, оказывающие существенное влияние на $У_i$;

$e_i = У_i - У_{pi}$ — случайная составляющая (остатки) $У_i$, которая определяется свободой выбора членами общества в принятии решения на воздействие переменных, а также учитывает влияние неучтенных факторов;

i — порядковый номер измерения i = 1, ..., n;

n — объем выборки.

В зависимости от количества переменных в модели различают однофакторные и многофакторные модели.

Наличием случайной составляющей регрессионные модели отличаются от моделей, описывающих детерминированные процессы.

Вопрос 10. Линейная регрессия.

Известна следующая регрессия:

$У_i = a_0 + a_1X_i + e_i$, которая называется линейной регрессией.

Варианты ответа:

1) да;

2) нет.

Пояснения. В эконометрике часто используются следующие виды математических функций, отражающих тенденции зависимости У от X:

$У_{pi} = a_0 + a_1X_i$ — линейная;

$У_{pi} = a_0 + a_1X_i + a_2(X_i)^2$ — параболическая;

$У_{pi} = a_0 + a_1/X_i$ — гиперболическая;

$У_{pi} = a_0 + a_1\ln(X_i)$ — логарифмическая;

$У_{pt} = a_0 + a_1t + a_2\mathrm{Sin}(2\pi t/T) + a_3\mathrm{Cos}(2\pi t/T)$ — периодическая,

где t — время (1, 2, ..., n);

T — период колебания.

Все эти функции являются линейными относительно параметров модели, так как параметры модели находятся в первой степени, однако переменные могут находиться в степени, отличной от единицы, например степенная функция

$У_{pi} = a_0X_i^{a1}$.

Степенная функция приводится к линейному виду путем логарифмирования левой и правой частей уравнения.

Приводим все виды математических функций, которые используются в курсе эконометрики.

$У_{pi} = a_0 + a_1X_{1i} + a_2X_{2i} + a_3X_{3i}$ — многофакторная линейная модель;

где X_1, X_2, X_3 — факторы, оказывающие влияние на У;

$У_{pt} = a_0 + a_1X_{1t} + a_2X_{1(t-1)} + a_3X_{1(t-2)}$ — модель распределенного лага;

$У_{pt} = a_0 + a_1X_t + a_2e_{t-1}$ — модель, учитывающая автокорреляцию остатков;

$У_{pt} = a_0 + a_1X_t + a_2У_{t-1} + a_3У_{t-2}$ — авторегрессионная модель, учитывающая влияние на У их же предыдущих значений У;

$(У_i/X_i)_{pi} = a_0 + a_1(1/X_i)$ — модель, устраняющая гетероскедастичность остатков.

Система одновременных уравнений, учитывающая взаимовлияние переменных $У_1$ и $У_2$:

$$\left.\begin{aligned} У_{1i} &= a_0 + a_1У_{2i} + a_2X_{1i} + e_{1i}; \\ У_{2i} &= b_0 + b_1У_{1i} + b_2X_{2i} + e_{2i}. \end{aligned}\right\}$$

Выбор вида функции производится на основе экономической теории изучаемого процесса при соблюдении основных принципов построения модели: простота, ясность экономической интерпретации модели.

Вопрос 11. Словесно-описательная модель общества.

Экономическая система действует в пространстве и времени в правовой и законодательной среде в сочетании с потребностями членов общества, при соблюдении законов сохранения ресурсов: трудовых, денежных, материальных, духовных, интеллектуальных и богатства, которое можно взять с заданной территории.

Варианты ответа:

1) да;

2) нет.

Вопрос 12. Цикл Деминга.

Цикл Деминга улучшения процессов включает следующие этапы:

Варианты ответа:

1) планируй, делай, анализируй, совершенствуй;

2) делай, анализируй, совершенствуй.

Вопрос 13. Свойства экономической системы.

Экономическая система является

Варианты ответа:

1) саморегулирующей;

2) не саморегулирующей.

Вопрос 14. Свойства экономических процессов.

Все экономические процессы происходят в пространстве и во времени

Варианты ответа:

1) да;

2) нет.

Вопрос 15. Свойства экономических процессов.

Экономическая система обладает инерционными свойствами.

Варианты ответа:

1) да;

2) нет.

Вопрос 16. Свойства экономических процессов.

Экономическая система — самонастраиваемая система, которая может находиться в состоянии динамического равновесия.

Варианты ответа:

1) да;

2) нет.

Вопрос 17. Свойства экономических процессов.

Текущие состояние экономической системы испытывают влияния прошлых, настоящих и будущих значений переменных этой системы

Варианты ответа:

1) да;

2) нет.

Вопрос 18. Свойства экономических процессов.

Для всех явлений в природе между причиной и следствием существует временной лаг, или временная задержка.

Варианты ответа:

1) да;

2) нет.

Вопрос 19. Свойства экономических процессов.

Все временные экономические процессы происходят циклически.

Варианты ответа:

1) да;

2) нет.

Вопрос 20. Свойства экономических процессов.

Прошлые значения показателя временного ряда оказывает влияние на его текущее значение, но не зависит от него.

Варианты ответа:

1) да;

2) нет.

Вопрос 21. Процесс моделирования.

Этапы экономического моделирования:

Этап 1. Постановочный.

Этап 2. Априорный

Этап 3. Информационный.

Этап 4. Спецификация

Этап 5. Идентификация модели.

Этап 6. Определение качества модели.

Этап 7. Верификация модели.

Этап 8. Выводы и предложения.

Варианты ответа:

1) да;

2) нет.

Вопрос 22. Регрессионная модель.

Регрессионная модель имеет следующий вид:

$У_i = f(X_i) + e_i$.

Варианты ответа:

1) да;

2) нет.

Вопрос 23. Оценка параметров модели.

Коэффициенты линейной модели

$У_i = a_0 + a_1X_i + e_i$

рассчитываются по следующим формулам:

$$a_1 = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2};$$

$$a_0 = \overline{Y} - a_1\overline{X}.$$

Варианты ответа:

1) да;

2) нет.

Вопрос 24. Остатки модели.

Остатки модели вычисляются по следующей формуле

$e_i = У_i - (a_0 + a_1X_i)$.

Варианты ответа:

1) да;

2) нет.

Вопрос 25. Ошибка модели.

Стандартная ошибка модели вычисляется по формуле

$$E = \sqrt{\frac{\sum e_i^2}{n-k}}.$$

Варианты ответа:

1) да;

2) нет.

Вопрос 26. Критерий Фишера.

Статистический критерий Фишера вычисляется по формуле

$$F = S^2_{рег} / S^2_{ост} = \frac{\sum (У_{pi} - \bar{У})^2}{k-1} : \frac{\sum (У_i - У_{pi})^2}{n-k}.$$

Варианты ответа:

1) да;

2) нет.

Вопрос 27. Нулевая гипотеза.

Нулевая гипотеза — это:

- предположение о том, что между двумя переменными нет связи;
- дисперсия регрессии не отличается от дисперсии остатков;
- коэффициенты модели не отличаются от нуля.

Варианты ответа:

1) да;

2) нет.

Вопрос 28. Уровень значимости.

Уровень значимости — это вероятность совершить ошибку при отклонении нулевой гипотезы.

Варианты ответа:

1) да;

2) нет.

Вопрос 29. Проверка достоверности модели.

Известно, что

$F_{кр}(\alpha = 0{,}05;\ m_1 = 1;\ m_2 = 8) = 5{,}12$; $F = 3{,}1$; H_0: “модель недостоверна”.

Так как $F = 3{,}1 < F_{кр} = 5{,}12$, то

H_0: "модель недостоверна" принимается.

Варианты ответа:

1) да;

2) нет.

Вопрос 30. Запись математической модели.

Известно, что линейная модель $У_{pi} = a_0 + a_1X_i$ имеет следующие коэффициенты: $a_0 = 16{,}5$; $a_1 = 2{,}3$. Необходимо записать математическую модель с численными значениями коэффициентов.

Варианты ответа:

1) $У_{pi} = 2{,}3 + 16{,}5X_i$;

2) $У_{pi} = 16{,}5 + 2{,}3X_i$.

Вопрос 31. Ошибка коэффициента модели.

Стандартное отклонение (ошибка) коэффициента a_1 модели

$У_{pi} = a_0 + a_1X_i$ вычисляется по формуле

$$S_{a_1} = E\sqrt{\frac{1}{\sum(X_i - \bar{X})^2}}.$$

Варианты ответа:

1) да;

2) нет.

Вопрос 32. Проверка достоверности коэффициента a_1

Если $a_1 = 2{,}6$; $S_{a1} = 1{,}1$; $n = 8$; $У_{pi} = a_0 + a_1X_i$, то проверка достоверности коэффициента a_1 выполняется в следующей последовательности:

Шаг 1. Выдвигается нулевая гипотеза H_0: $\alpha_1 = 0$.

Шаг 2. Определяется критическое двухстороннее значение критерия

Стьюдента $t_{\alpha/2}(\alpha = 0{,}05,\ m = n - 2 = 8 - 2 = 6) = 2{,}45$.

Шаг 3. Определяется t_{a1} — критерий Стьюдента для коэффициента a_1.

а) $t_{a_1} = \dfrac{a_1}{S_{a_1}} = \dfrac{2{,}6}{1{,}1} = 2{,}36.$

Шаг 4. Проверяется нулевая гипотеза H_0: $\alpha_1 = 0$.

Определяется зона, где находится t_{a1}, если

$t_{a1} > t_{кр}$,

то t_{a1} попало в критическую зону, где нулевая гипотеза H_0: $\alpha_1 = 0$ отвергается и утверждается с вероятностью $1 - \alpha$, что коэффициент a_1 достоверно отличается от нуля.

Варианты ответа:

1) да;

2) нет.

Вопрос 33. Экономический смысл коэффициента a_1 модели

$У_{pi} = a_0 + a_1X_i$.

Если переменная X изменится на единицу, то значение $У_p$ изменится на величину a_1.

Варианты ответа:

1) да;

2) нет.

Вопрос 34. Точечный прогноз.

Имеется модель $У_{pi} = a_0 + a_1X_i$,
где $a_0 = 2,6$; $a_1 = 1,6$.

Оценить прогнозное значение математического ожидания У при $X_{ож} = 10$.

$У_{пр} = a_0 + a_1X_{Ож} = 2,6 + 1,6 \times 10 = 18,6$.

Варианты ответа:

1) да;

2) Нет.

Вопрос 35. Интервальный прогноз математического ожидания У.

Имеется модель

$У_{pi} = a_0 + a_1X_i$,
где $a_0 = 2,6$; $a_1 = 1,6$;

$X = 5$; $\sum(X_i - \overline{X})^2 = 5$;

$n = 10$; ожидаемое значение $X_{ож} = 10$; $E = 2,1$; $У_{пр} =$ $= 18,6$;

$$S_{MУ} = E\sqrt{\frac{1}{n} + \frac{(X_{ож} - \bar{X})^2}{\sum (X_i - \bar{X})^2}};$$

где $S_{MУ}$ — стандартная ошибка прогнозного значения математического ожидания У. $S_{MУ} = 4{,}75$.

Варианты ответа:

1) да;

2) нет.

Вопрос 36. Интервальный прогноз математического ожидания.

Если $У_{пр} = 18{,}6$; $t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k = 10 - 2 = 8) = 2{,}31$; $S_{MУ} = 4{,}75$, то с вероятностью 0,95 можно утверждать, что математическое ожидание У будет находиться в интервале

$У_{пр} \pm t_{\alpha/2}\, S_{MУ} = 18{,}6 \pm 2{,}31 \times 4{,}75$ или от 7,62 до 29,57;

Варианты ответа:

1) да;

2) нет.

Вопрос 37. Интервальный прогноз для индивидуального значения У.

Известно, что $S_{Уи}$ — стандартная ошибка прогнозного индивидуального значения У, вычисляется по формуле

$$S_{Уи} = E\sqrt{1 + \frac{1}{n} + \frac{(X_{ож} - \bar{X})^2}{\sum (X_i - \bar{X})^2}}.$$

Для нашего примера $S_{Уи} = 5{,}19$; $У_{пр} = 18{,}6$;

$t_{\alpha/2}(\alpha = 0{,}05;\ m = n - k = 10 - 2 = 8) = 2{,}31$.

С вероятностью 0,95 можно утверждать, что индивидуальное значение У будет находиться в интервале

$У_{пр} \pm t_{\alpha/2}\, S_{Уи} = 18{,}6 \pm 2{,}31 \times 5{,}19$ или от 6,61 до 30,59.

Варианты ответа:

1) да;

2) нет.

Вопрос 38. Эконометрический анализ.

Этапы эконометрического анализа результатов расчета характеристик модели.

Шаг 1. Представление протокола результатов расчета характеристик модели.

Шаг 2. Запись математического вида модели и ее характеристик.

Шаг 3. Проверка достоверности модели.

Шаг 4. Проверка достоверности коэффициента модели a_1.

Шаг 5. Анализ ошибки модели E.

Шаг 6. Экономическая интерпретация коэффициентов a_0 и a_1.

Шаг 7. Точечный и интервальный прогноз.

Шаг 8. Выводы и предложения.

Варианты ответа:

1) да;

2) нет.

Вопрос 39. Пример эконометрического анализа линейной модели.

Необходимо специфицировать модель линейной функцией вида

$У_p = a_0 + a_1 X$.

Расчеты характеристик линейной модели с помощью функции "Линейн".

1,8717	3,4783
0,1408	2,0121
0,8893	4,7754
176,67	22
4028,9	501,71

Запишем результаты расчетов в общепринятом обозначении.

Протокол расчетов.

$У_i = 3{,}47826 + 1{,}87174\ X_i + e_i$;

$S_{a0} = 2{,}01213$; $S_{a1} = 0{,}14082$;

$t_{a0} = 1{,}72864$; $t_{a1} = 13{,}2917$; $t_{\alpha/2} = 2{,}07388$;

$E = 4{,}77544$; $E = 17{,}7691$ (%);

$R^2 = 0{,}88926$;

$F = 176{,}669$; $F_{кр} = 4{,}30094$,

где S_a — ошибка коэффициента а;

E — ошибка модели;

R^2 — коэффициент детерминации;

t_{a0} — критерий Стьюдента для коэффициента a_0;

t_{a1} — критерий Стьюдента для коэффициента a_1;

$t_{\alpha/2}$ — табличное значение критерия Стьюдента;

F — критерий Фишера;

$F_{кр}$ — табличное значение критерия Фишера.

Представляем результаты эконометрического анализа.

1. Вычисление доли объясненной вариации.

88,92633% исходных данных имеют тенденцию выбранной функции.

2. Проверка достоверности модели.

По критерию Фишера проверяется достоверность модели.

Так как $F = 176{,}669 > F_{кр}(\alpha = 0{,}05;\ m_1 = k - 1;\ m_2 = n - k) = 4{,}30094$, то модель является достоверной с вероятностью 0,95.

3. Проверка достоверности коэффициентов модели.

По критерию Стьюдента проверяется достоверность коэффициентов модели.

Так как $|t_{a0}| = 1{,}72864 < t_{\alpha/2} = 2{,}07388$, то коэффициент a_0 статистически не отличается от нуля.

Так как $|t_{a1}| = 13{,}2917 > t_{\alpha/2} = 2{,}07388$, то коэффициент a_1 с вероятностью 0,95 отличается от нуля.

4. Вычисление точечного и интервального прогноза.

Среднее значение зависимой переменной будет равно $У_{пр} = 50{,}2717$.

Фактическое прогнозное значение зависимой переменной с вероятностью 0,95 будет находиться в интервале от 40,3681 до 60,1754.

Вывод. Линейная регрессия имеет следующие характеристики:

ошибка модели равна $E = 17{,}769\%$; точечный прогноз равен $У_{пр} = 50{,}272$.

Варианты ответа:
1) да;
2) нет.

Вопрос 40. Транспонирование матрицы.
Дана матрица **A**, представленная как вектор-столбец

$$\mathbf{A} = \begin{vmatrix} 2 \\ 1 \\ 3 \end{vmatrix}.$$

Необходимо транспонировать матрицу **A**.
Варианты ответа:

1) $\mathbf{A}' = \begin{vmatrix} 3 \\ 1 \\ 2 \end{vmatrix};$

2) $\mathbf{A}' = \begin{vmatrix} 2 & 1 & 3 \end{vmatrix}.$

Вопрос 41. Произведение матриц.
Даны две матрицы:

$$\mathbf{A} = \begin{vmatrix} 2 \\ 1 \\ 3 \end{vmatrix}; \quad \mathbf{B} = \begin{vmatrix} 1 \\ 2 \\ 3 \end{vmatrix}.$$

Необходимо получить произведение матриц **A'B**.
Варианты ответа:
1) **A'B** = 17;
2) **A'B** = 13.

Пояснения. При произведении двух матриц элементы строки первой матрицы умножаются на соответствующие элементы столбца второй матрицы, результаты умножения складываются.

Вопрос 42. Свойства матричных операций.
Транспонирование произведения двух матриц
(AB)'.

Варианты ответа:
1) $(AB)' = B'A'$;
2) $(AB)' = A'B'$.

Пояснения. $(AB)' = B'A'$.

Вопрос 43. Свойства матричных операций.
Транспонирование произведения трех матриц $(ABC)'$.
Варианты ответа:
1) $(ABC)' = C'B'A'$;
2) $(ABC)' = A'B'C'$.

Пояснения. $(ABC)' = C'B'A'$

Вопрос 44. Свойства матричных операций.
Произведение матриц $(AB)C$.
Варианты ответа:
1) $(AB)C = A(BC) = ABC$;
2) $(AB)C = CAB$.

Пояснения. $(AB)C = A(BC) = ABC$.

Вопрос 45. Свойства матричных операций.
Транспонирование суммы двух матриц.
Варианты ответа:
1) $(A + B)' = A' + B'$;
2) $(A + B)' = (A')' + B'$.

Пояснения. $(A + B)' = A' + B'$.

Вопрос 46. Свойства обратных матриц.
Получить обратную матрицу от произведения матриц.
Варианты ответа:
1) $(AB)^{-1} = B^{-1}A^{-1}$;
2) $(AB)^{-1} = A^{-1}B^{-1}$.

Пояснения. $(AB)^{-1} = B^{-1}A^{-1}$

Вопрос 47. Свойства обратных матриц.
Получить произведение обратной матрицы на исходную $A^{-1}A$.
Варианты ответа:
1) $A^{-1}A = I$;

2) $\mathbf{A^{-1}A = A}$.

Пояснения. $\mathbf{A^{-1}A = I}$, где $\mathbf{I}$ — единичная матрица.

Вопрос 48. Расчет коэффициентов модели.

Вычисление коэффициентов линейной модели в матричном виде:

Варианты ответа;

1) $\mathbf{A = (X'X)^{-1} X'У}$;

2) $\mathbf{A = (X'X) X'У}$.

Пояснения. $\mathbf{A = (X'X)^{-1} X'У}$.

Вопрос 49. Бетта-коэффициенты.

$У_{нр} = b_0 + b_1Z_1 + b_2Z_2,$

где $Z_{1i} = \dfrac{X_{1i} - \overline{X}_1}{S_{x1}};$

$$Z_{2i} = \frac{X_{2i} - \overline{X}_2}{S_{x2}};$$

$$Z_{нрi} = \frac{У_i - \overline{У}}{S_y};$$

Z_{1i}, Z_{2i}, $У_{нрi}$ — нормированные значения переменных.

Бетта-коэффициенты позволяют по их значениям определять степень влияния.

Варианты ответа:

1) в сопоставимом виде;

2) в несопоставимом виде.

Вопрос 50. Анализ многомерной модели.

Имеются следующие характеристики модели:

$У = a_0 + a_1X_1 + a_2X_2$; $t_{a1} = 25$; $t_{a2} = 30$.

Варианты ответа:

1) фактор X_2 оказывает более достоверное влияние на $У$, чем фактор X_1 ;

2) фактор X_1 оказывает более достоверное влияние на $У$, чем фактор X_2 .

Вопрос 51. Определение мультиколлинеарности.

Мультиколлинеарностью называется свойство переменных, если между ними существует

Варианты ответа:

1) сильная корреляционная зависимость;

2) слабая корреляционная зависимость.

Вопрос 52. Влияние мультиколлинеарности на характеристики модели.

Известно, что коэффициенты модели считаются по формуле

$$\mathbf{A = (X'X)^{-1} X'У},$$

чтобы вычислить обратную матрицу $\mathbf{(X'X)^{-1}}$ надо знать определитель матрицы $\mathbf{(X'X)}$. Если столбцы матрицы $\mathbf{X}$ будут линейно зависимыми, то столбцы матрицы $\mathbf{X'X}$ тоже будут линейно зависимыми и определитель матрицы $\mathbf{X'X}$ будут стремиться к нулю. Тогда элементы обратной матрицы $\mathbf{(X'X)^{-1}}$ будут стремиться к бесконечности. Следовательно, ошибки коэффициентов будут стремиться к бесконечности и модель станет недостоверной.

Варианты ответа:

1) да;

2) нет.

Вопрос 53. Признаки наличия сильной мультиколлинеарности.

Если между объясняемыми переменными существует мультиколлинеарность, то модели зависимости У от каждой независимой переменной в отдельности могут быть достоверными. Модели, включающие мультиколлинеарные переменные, вероятнее всего, станут недостоверными.

Варианты ответа:

1) да;

2) нет.

Вопрос 54. Определение мультиколлинеарности.

Известна матрица парных коэффициентов корреляции

	X_1	X_2	X_3	У
X_1	1	0,10	0,85	0,95
X_2		1	0,25	0,2
X_3			1	0,80
У				1

Между переменными X_1 и X_3 существует:

1) сильная мультиколлинеарность;
2) слабая мультиколлинеарность.

Вопрос 55. Определение факторов для включения в модель.

Известна матрица коэффициентов корреляции

В модель может иметь следующий вид:

	X_1	X_2	X_3	У
X_1	1	0,10	0,85	0,95
X_2		1	0,25	0,2
X_3			1	0,80
У				1

Варианты ответа:

1) У = f(X_2);
2) У = f(X_1, X_3);
3) У = f(X_1);
4) У = f(X_3).

Пояснения. Правильное решение задачи заключается в том, чтобы модель содержала факторы, слабо связанные между собой, но сильно связанные с зависимой переменной. Анализ матрицы коэффициентов корреляции показывает, что между переменными X_1 и X_3 существует сильная связь, У сильно связана с переменными X_1 и X_3, следовательно эти переменные вместе в модель включать нельзя. Их можно включать в модель только по одному. У слабо зависит от X_2, поэтому переменную X_2 в модель включать нежелательно. Целесообразно в модель включить фактор X_1 или X_3, но переменная X_1 сильнее связана с У, чем X_3, следовательно, в модель необходимо включить фактор X_1.

Вопрос 56. Определение гомоскедастичности.

Гомоскедастичностью называется:

Варианты ответа:

1) неоднородность дисперсий признака;

2) однородность дисперсий признака.

Вопрос 57. Определение гетероскедастичности.

Гетероскедастичностью называется:

Варианты ответа:

1) неоднородность дисперсий признака;

2) однородность дисперсий признака.

Вопрос 58. Признаки появления гетероскедастичности остатков.

Гетероскедастичность остатков возникает в том случае, если с увеличением объясняемой переменной варьирование остатков:

Варианты ответа:

1) возрастает;

2) остается постоянной.

Вопрос 59. Последствия гетероскедастичности остатков.

Доверительные интервалы уравнения регрессии

Варианты ответа:

1) соответствуют действительности;

2) не соответствуют действительности.

Доверительные интервалы прогноза

Варианты ответа:

1) являются ошибочно меньше или больше действительности;

2) соответствуют действительности.

Вопрос 60. Способы устранения гетероскедастичности.

Для устранения гетероскедастичности остатков необходимо:

Варианты ответа:

1) все переменные регрессионного уравнения разделить на остатки;

2) все переменные регрессионного уравнения умножить на остатки.

Вопрос 61. Получение точечного и интервальных прогнозов для исходных данных при условии наличия гетероскедастичности остатков.

Для преобразованных переменных определяется точечный и интервальный прогноз, затем эти значения:

Варианты ответа:

1) умножаются на остатки, которые использовались для преобразования данных;

2) делятся на остатки, которые использовались для преобразования данных.

Вопрос 62. Определение автокорреляции.

Понятие автокорреляции относится только к временным рядам.

Автокорреляцией называется коэффициент корреляции, рассчитанный между двумя временными рядами, один из которых является исходным временным рядом, а вторым является:

Варианты ответа:

1) тем же временным рядом, только сдвинутым по времени на m дат;

2) тем же временным рядом, только не сдвинутым по времени на m дат.

Пояснения. Понятие, причины и последствия автокорреляции остатков.

Эконометрический анализ регрессионной модели должен сопровождаться изучением остатков модели. Если график зависимости остатков объясняемой переменной содержит регулярности, то появляется возможность улучшить модель, уменьшить ошибку модели, повысить точность прогноза.

Регулярности остатков могут быть трех видов: увеличение или уменьшение квадратов остатков в зависимости от значений объясняемой переменной, которое называется гетероскедастичностью; наличие нециклической и циклической тенденций.

Нециклические и циклические тенденции возникают вследствие плохой спецификации модели. Нециклическая тенденция в остатках может описываться следующими математическими функциями: параболической, гиперболической, логарифмической, экспоненциальной и др.

Циклическая тенденция в остатках динамических моделей проявляется наличием устойчивых чередований положительных и отрицательных остатков.

Циклическая тенденция в остатках аппроксимируется новыми переменными, которые включаются в основную модель динамического процесса. Возможны комбинации разных видов регулярностей.

Существуют два способа обнаружения регулярностей остатков: визуальный анализ графиков остатков и расчет коэффициентов автокорреляции остатков.

Визуальный анализ графика остатков является самым мощным методом обнаружения всех видов регулярностей и позволяет указать пути улучшения модели.

Наличие автокорреляции остатков означает, что в остатках существует регулярность, которую можно устранить улучшением спецификации модели.

Улучшение спецификации модели должно привести к уменьшению ошибки модели, увеличению достоверности модели и повышению точности прогнозов.

Вопрос 63. Причины появления автокорреляции остатков временных рядов.

Автокорреляция остатков временных рядов возникает вследствие:

Варианты ответа:

1) плохой спецификации модели и появления в остатках регулярностей в виде нециклической и циклической тенденций;

2) хорошей спецификации модели и отсутствия в остатках регулярностей.

Вопрос 64. Прогнозирование с учетом автокорреляции остатков.

Точечный прогноз, полученный обычным способом:

Варианты ответа:

1) корректируется на величину зависимости от предшествующего значения остатка модели

2) не корректируется в зависимости от автокорреляции остатков.

Пояснения. Модель, учитывающая автокорреляцию остатков.

Допустим, по критерию Дарбина-Уотсона установлено достоверное наличие в остатках автокорреляции первого порядка. Можно избрать два пути устранения обнаруженной регулярности. Первый путь заключается в улучшении спецификации модели с помощью включения в модель факторов, которые вызывают эту регулярность. Второй путь состоит в построении модели, учитывающей автокорреляцию остатков.

Продемонстрируем схему построения модели с учетом автокорреляции остатков. Допустим, коэффициент автокорреляции первого порядка является достоверным, тогда динамическую модель процесса можно представить в следующем виде:

$$У_t = a_0 + a_1X_t + e_t = a_0 + a_1X_t + pe_{t-1} + v_t,$$

где $e_t = pe_{t-1} + v_t$;

p — коэффициент показывает какая доля e_{t-1} входит в состав e_t;

v_t — случайная составляющая, не содержащая автокорреляцию первого порядка.

Расчеты коэффициентов полученной модели произведем в два шага.

Шаг 1. Вычислим коэффициенты модели $У_t = a_0 + a_1X_t + e_t$ и остатки e_t.

Шаг 2. Вычислим коэффициент p из модели

$$e_t = b_0 + pe_{t-1} + v_t,$$

и характеристики модели

$$У_t = a_0 + a_1x_t + pe_{t-1} + v_t.$$

Примечание. Коэффициент p можно вычислить по формуле

$$p = 1 - d/2,$$

где d — критерий Дарбина-Уотсона[1].

Прогнозирование по модели, учитывающей автокорреляцию остатков, производится на период $t = n+1$ по формуле

$$У_{п(t=n+1)} = a_0 + a_1X_{(t=n+1)} + pe_{(t=n)} + v_t,$$

которая состоит из двух частей. Первая часть содержит обычный прогноз, вторая часть содержит поправку на автокорреляцию остатков.

Примечание. Мы рассмотрели упрощенный вариант модели, учитывающей автокорреляцию остатков. Упрощение заключалось в том, что в модели

$$У_t = a_0 + a_1x_t + p \times e_{t-1} + v_t$$

пропущен коэффициент b_0, который входит в состав

$$e_t = b_0 + p \times e_{t-1} + v_t.$$

Включение в модель коэффициента b_0 приведет к смещению коэффициентов a_0 и a_1, что вызовет изменение остатков e_t и коэффи-

---

[1] *Грубер Й.* Эконометрика. Том 1. Введение в эконометрию. — Киев: Астарта, 1996.

циента ρ. Корректное построение модели, учитывающей автокорреляцию остатков производится двухшаговым методом Дарбина, который позволяет незначительно уменьшить ошибку модели по сравнению с упрощенной моделью.

Вопрос 65. Определение понятия лага.

Лагом называют:

Варианты ответа:

1) временную задержку между причиной и следствием;

2) отсутствие временной задержки между причиной и следствием.

Вопрос 66. Составляющие временного лага.

Временной лаг состоит из:

Варианты ответа:

1) времени передачи сигнала о возникновении причины, обработке информации, принятия решения, перевода системы из одного устойчивого состояния в другое;

2) времени только передачи сигнала от причины к объекту.

Вопрос 67. Примеры распределенных лагов.

Инвестиции, вложенные в производство, позволяют получить увеличение прибыли:

Варианты ответа:

1) в момент вложения инвестиций в производство;

2) на протяжении некоторого времени.

Вопрос 68. Прогнозирование с использованием лаговых факторов.

Использование лаговых факторов в регрессионной модели позволяет:

Варианты ответа:

1) уменьшить ошибку прогноза за счет того, что ожидаемые значения лагового фактора точно известны;

2) увеличить ошибку прогноза.

Вопрос 69. Получение прогнозных значений с учетом лагового фактора.

Известны коэффициенты уравнения

$У_t = a_0 + a_1X_t + a_2X_{t-1} + a_3X_{t-2}$,
где $a_0 = 2{,}3$; $a_1 = 3$; $a_2 = 4$; $a_3 = 3{,}2$; $X_{n+1} = 10$; $X_n = 9$; $X_{n-1} = 5$.

Получить прогнозные значения $У_{пр}$ на период n+1.

Варианты ответа:

1) $У_{пр} = 2{,}3 + 3 \times 10 + 49 + 3{,}25 = 84{,}55$;

2) $У_{пр} = 2{,}3 + 3 \times 10 = 32{,}3$.

Вопрос 70. Структура многофакторной модели динамического процесса.

Имеется модель динамического процесса

$У_t = a_0 + a_1t + a_2X_t + a_3X_{t-1} + a_4У_{t-1} + a_5У_{t-2} + a_6\sin(2\pi t/T) + a_7\cos(2\pi t/T) + a_8e_{t-1} + v_t$,
которая имеет следующую структуру:

Варианты ответа:

1) линейную тенденцию, модель распределенного лага, авторегрессию зависимой переменной, периодическую составляющую, модель автокорреляции остатков

2) линейную тенденцию, модель распределенного лага.

Вопрос 71. Необходимость использования системы одновременных уравнений.

Все показатели экономической системы между собой:

Варианты ответа:

1) взаимосвязаны;

2) не взаимосвязаны.

Вопрос 72. Определение эндогенных и факторов.

Эндогенные факторы в системе одновременных уравнений:

Варианты ответа:

1) являются зависимой переменной;

2) не являются зависимой переменной.

Вопрос 73. Определение эндогенных и факторов.

Экзогенные факторы в системе одновременных уравнений:

Варианты ответа:

1) являются только объясняемыми переменными;

2) являются зависимыми переменными.

Вопрос 74. Прогнозирование эндогенных переменных.

Прогноз эндогенных переменных можно произвести с использованием:

Варианты ответа:

1) системы структурных уравнений;

2) системы приведенных уравнений.

Вопрос 75. Качественные переменные.

Шкалой измерения качественной переменной:

Варианты ответа:

1) служит перечень классов, в которых может находиться объект.

2) служит численное значение переменной.

Пояснения. Если экономический объект характеризуется качественной переменной, то значением качественной переменной будет служить номер класса, к которому принадлежит объект.

Главный редактор — *А. Е. Илларионова*
Редактор — *В. Н. Рогожкин*
Художник — *В. А. Антипов*
Верстка — *Н. В. Байкова*
Корректор — *Н. А. Тимофеева*

Ответственный за выпуск — *С. А. Булатова*

Учебное издание

Валентинов Вячеслав Аркадьевич

Эконометрика

Санитарно-эпидемиологическое заключение
№ 77.99.02.953.Д.004609.07.04 от 13.07.2004 г.

Подписано в печать 10.09.2008. Формат 60×84 1/16.
Печать офсетная. Бумага офсет № 1. Печ. л. 28,0.
Тираж 2500 экз. (1-й завод 1–1250 экз.) Заказ №6368.

Издательско-торговая корпорация «Дашков и К°»
129347, Москва, Ярославское шоссе, д. 142, к. 732.
Для писем: 129347, Москва, п/о И-347
Тел./факс: (095) 182-01-58, 182-11-79, 183-93-01
E-mail: sales@dashkov.ru — отдел продаж
office@dashkov.ru — офис;
http://www.dashkov.ru

Отпечатано в соответствии с качеством предоставленных диапозитивов
в ФГУП «Производственно-издательский комбинат ВИНИТИ»,
140010, г. Люберцы Московской обл., Октябрьский пр-т, 403. Тел.: 554-21-86