Издательско-торговая корпорация «Дашков и К°»

В. А. Валентинов

ЭКОНОМЕТРИКА

Практикум

3-е издание

Москва
2010

УДК 330.115
ББК 65В6
В15

Рецензенты:

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

В. Ф. Тулинов — доктор физико-математических наук, профессор кафедры высшей математики и естественно-научных дисциплин.

Валентинов В. А.

В15 Эконометрика: Практикум / В. А. Валентинов. — 3-е изд. — М.: Издательско-торговая корпорация «Дашков и К°», 2010. — 436 с.

ISBN 978-5-394-00682-1

Практикум составлен на основании учебника (В. А. Валентинов) и электронного модульного мультимедийного обучающего комплекса Econ3. В нем 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 и косвенного метода наименьших квадратов.

По каждой лабораторной работе приведены целевые установки, вопросы для самостоятельной подготовки, даны задания четырех уровней сложности, приведены примеры решения задачи, даны варианты выполнения самостоятельной работы студентов, даны дополнительные задания, приведены биографии ведущих ученых в области эконометрики, приведены примеры выполнения расчетов с помощью пакетов прикладных программ, дан список рекомендованной литературы.

Для студентов специальностей «Прикладная информатика в экономике», «Бухгалтерский учет и аудит» и других экономических специальностей.

ISBN 978-5-394-00682-1

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Эконометрику можно рассматривать как:
- — науку;
- — алгоритмы решения экономических задач;
- — средство реализации алгоритмов;
- — примеры решения типичных экономических задач.

Эконометрика как наука должна содержать теоретическое обоснование эконометрических моделей, с помощью которых можно решать экономические задачи. Теоретическое обоснование моделей имеется в учебниках по эконометрике.

Эконометрика как алгоритмы решения экономических задач должна содержать формулы, позволяющие рассчитать все характеристики модели и решить поставленные задачи.

Эконометрика как средство реализации алгоритмов должна предложить различные средства проведения расчетов.

Средствами проведения расчетов могут быть такими:
- — визуальные;
- — средства калькулятора;
- — средства вычислительных процессоров Excel, Mat-Lab, Mat-Cad;
- — пакеты прикладных программ.

Каждое средство имеет свои сильные и слабые стороны, возможности и угрозы, описание которых можно предложить студентам в виде реферата по эконометрике.

Эконометрика позволяет получить навыки решения аналогичных задач последующем обучении и в производственной деятельности выпускника.

В практикуме рассматривается теоретическая основа эконометрики в виде ответов на вопросы самостоятельной работы при входном тестировании, алгоритмы решения типичных эконометрических задач, средства их реализации и примеры построения моделей экономических процессов.

Целями книги заключаются в том, чтобы студент смог быстро решить эконометрическую задачу, при этом он должен знать, какие требования предъявляются к выбранной модели, какие данные нужно собрать и каким требованиям они должны удовлетворять, куда ввести данные, с помощью каких средств и как можно быстро выполнить расчеты, как провести анализ полученных расчетов, какие выводы можно сделать.

Быстрота решения задачи обеспечивается двумя способами: визуальным и с использованием ПВМ.

Первый способ состоит в том, что студент должен визуально быстро приблизительно вычислить коэффициенты модели и ее характеристики, точечный и интервальный прогноз.

Второй способ заключается в быстром и точном проведении расчетов коэффициентов и характеристик модели с помощью вычислительных средств ПВМ.

Практикум составлен на основе учебника [3] и модульного электронного мультимедийного обучающего комплекса по эконометрике Econ3, который можно использовать для дистанционного обучения и имеющего нелинейную структуру, реализованного в Excel. Описание комплекса имеется в учебнике. Темы комплекса трансформировались в отдельные лабораторные работы. Нелинейная структура комплекса превращалась в линейную с потерей анимации. Были добавлены новые лабораторные работы в соответствии с учебником.

Для получения возможности работы с Econ3 можно обратиться по адресу http://www.kbfcenter.da.ru.

Практикум предназначен для изучения основных моделей эконометрики.

Для выполнения лабораторных работ студент должен иметь знания по следующим дисциплинам: информатика, теория вероятности, математическая статистика и уметь выполнять расчеты в среде Excel.

Практикум содержит задания разного уровня сложности в зависимости от целей, которые ставит перед собой студент.

Чем выше сложность задания, тем в большем объеме студент сможет изучить эконометрику.

Логическим продолжением практикума является работа [2], в которой приводятся примеры построения моделей экономических процессов таких как: модель ценообразования на основной капитал; анализ факторов, влияющих на заработную плату; прогнозирование совокупных инвестиционных расходов; модель спроса на электроэнергию; модель зависимости объем продаж от рекламы; моделирование взаимосвязанного спроса на факторы производства и др.

Для выполнения заданий по углубленному изучению теории эконометрики можно рекомендовать литературу: [6].

Использование в эконометрике средств и методов системы качества планируется изложить в отдельном издании.

Условные обозначения и расчетные формулы полностью соответствуют учебнику [3].

Условные обозначения

Обозначим выборочные наблюдения через

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

$У_1, У_2, ..., У_n$

и введем их арифметические средние

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

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

где 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

$У = a_0 + a_1 X + e$

для выборочной совокупности,

где $e = У - (a_0 + a_1 X) = У - У_р$ — отклонение или остаток (Гаусс называл его убытком, В эконометрической литературе принято e называть остатком), учитывающий влияние всех факторов, не включенных в модель;

У — фактические значения зависимой переменной;

$У_р = a_0 + a_1 X$ — 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

$У = a_0 + a_1 X + e$

имеет два коэффициента: a_0 и a_1, следовательно $k = 2$.

t — индекс времени в моделях временных рядов.

Например, $У_t = a_0 + a_1 X_t + e_t$, 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 α_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.

$$\begin{cases} У_1 = a_0+a_1X_1+a_2У_2+ e_1, \\ У_2 = b_0+b_1X_2 +b_2У_1 + e_2 \end{cases}$$

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, составляющей N семей. Численные значения переменных выборки мы будем записывать в таблице.

Таблица

Исходные значения выборочной совокупности

i	X_{1i}	X_{2i}	$У_i$
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_{1i} обозначает величину переменных X_d для i-го домашнего хозяйства. Гипотеза о линейной зависимости У от X_1 и X_2 может быть записана в виде

$У_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \varepsilon_{i,}\ i = 1, 2, ..., n.$

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.

E — 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.

M — 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

$\mathbf{A = (X'X)^{-1} X'У}$

A, X, У — матрицы.

Расчетные формулы. Расчет коэффициентов регрессионного уравнения

$У = a_0 + a_1 X + e$

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

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

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

где **А** — матрица коэффициентов модели;

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

Ошибка модели:

$$E = \sqrt{\frac{\sum (\text{У}_i - \text{У}_{pi})^2}{n-k}},$$

где $\text{У}_{pi} = a_0 + a_1 X_i$.

Основное вариационное уравнение

$$\sum \left(\text{У}_i - \overline{\text{У}}\right)^2 = \sum \left(\text{У}_i - \text{У}_{pi}\right)^2 + \sum \left(\text{У}_{pi} - \overline{\text{У}}\right)^2$$

где $= \sum \left(\text{У}_i - \overline{\text{У}}\right)^2 = C_{\text{общ}}$ — вариация общая;

$\sum \left(\text{У}_i - \text{У}_{pi}\right)^2 = C_{\text{ост}}$ — вариация остатков;

$\sum \left(\text{У}_{pi} - \overline{\text{У}}\right)^2 = C_{\text{рег}} = C_{\text{общ}} — C_{\text{ост}}$ — вариация регрессии.

$\dfrac{\sum (\text{У}_i - \overline{\text{У}})^2}{n-1}$ $\quad S_{\text{общ}}{}^2 =$ — дисперсия общая.

$\dfrac{\sum (\text{У}_i - \text{У}_{pi})^2}{n-k}$ $\quad S_{\text{ост}}{}^2 =$ — дисперсия остатков.

$\dfrac{\sum (\text{У}_{pi} - \overline{\text{У}})^2}{k-1}$ $\quad S_{\text{рег}}{}^2 =$ — дисперсия регрессии.

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

Множественный коэффициент детерминации:

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

Критерий Фишера:

$$F = \frac{S_{\text{рег}}^2}{S_{\text{ост}}^2}.$$

Ошибка коэффициента a_0:

$$S_{a_0} = E\sqrt{\frac{\sum X_i{}^2}{n\sum(X_i - \overline{X})^2}}.$$

Ошибка коэффициента a_1:

$$S_{a_1} = E\sqrt{\frac{1}{\sum(X_i - \overline{X})^2}}.$$

Критерий Стьюдента для коэффициента a_1:

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

Частный коэффициент детерминации для фактора X_1:

$$r^2_{X_1 . X_2 , . X_3} = \frac{t_1^2 R^2}{t_1^2 + t_2^2 + t_3^2},$$

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%-ный интервальный прогноз для математического ожидания У:

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

Расчет коэффициентов модели методом Эйткена:

$\mathbf{B = (X_н{}'X_н)^{-1} X_н{}'У_н = (X'V^{-1}X)^{-1}X'V^{-1}}$

Парный коэффициент корреляции:

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

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:

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

Критерий Дарбина — Уотсона (Дарбина−Ватсона):

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

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:

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

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:

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

модель периодических составляющих временного ряда:

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

модель экспоненциально взвешенного среднего:

$Z_t = \lambda У_t + \lambda (1-\lambda)У_{t-1} + \lambda (1-\lambda)^2 У_{t-2} + \lambda (1-\lambda)^3 У_{t-3} + \ldots =$

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

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:

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

1. ВИЗУАЛЬНЫЕ СРЕДСТВА РАСЧЕТА ХАРАКТЕРИСТИК ЛИНЕЙНОЙ МОДЕЛИ

Постановка задачи

Объект — пространственные данные экономического показателя.

Предмет — регулярности динамики экономического показателя.

Цель — научить быстрому построению графиков и их анализу.

Актуальность — от результатов анализа графиков зависит выбор модели экономического показателя.

Рабочая гипотеза — визуальный анализ графика динамики экономического показателя является самым мощным средством обнаружения всех регулярностей зависимости экономического показателя от фактора.

Метод — визуальный анализ зависимости экономического показателя от фактора, представленного в графическом виде.

Способ — использование мастера диаграмм, тип графика — точечный.

Задачи:

- повторить определение эконометрики;
- повторить основные понятия регрессионного анализа;
- научиться быстрому построению графиков и проведению расчетов коэффициентов и характеристик линейной модели;
- научиться проводить анализ регулярностей значений экономического показателя;
- научиться выполнять редактирование графика.

Ожидаемый результат — визуальный анализ графика зависимости значений экономического показателя от изучаемого фактора позволит быстро приблизительно вычислить коэффициенты и характеристики линейной модели, обнаружить все регулярности значений экономического показателя, которые можно будет воспроизвести с помощью регрессионной модели.

Метод для сравнения — сравнение результатов визуального анализ с точными значениями характеристик линейной модели, которые будут выполнены в следующей теме.

Задания первого уровня сложности

Задание 1. Подготовить устные сообщения по вопросам самостоятельной работы.

1. Дайте определение эконометрики.
2. Назовите объект, предмет, цели, задачи, методы, структуру и область использования эконометрики.
3. Укажите связь эконометрики с родственными науками.
4. Расскажите историю эконометрики.
5. Приведите примеры использования эконометрических методов для решения экономических задач.
6. Укажите последовательность построения графика в среде Excel. [3, с. 23–45]

Методические указания работы с литературой

Основная цель подготовки ответов на вопросы самостоятельной подготовки — ознакомление с основными понятиями эконометрики и пополнение словарного запаса студентов по новой дисциплине. Однако многие понятия уже использовались в следующих дисциплинах: теория вероятности, математическая статистика, статистика. Ответы на вопросы необходимо озвучивать и постоянно улучшать профессиональный эконометрический язык. Без знания теоретической части курса нельзя приступать к практическому решению задач.

По курсу эконометрики имеется несколько учебников разного уровня сложности.

Следует ознакомиться со всеми предложенными учебниками и выбрать такой, который соответствует вашему уровню подготовки.

Изучение эконометрики следует начинать с определения основных понятий по простым учебникам, не требующих специальной математической подготовки: [3, 10, 4, 7, 12] с обязательным рассмотрением предложенных решений сквозных примеров. Затем можно обратиться к учебникам, содержащим теоретические обоснования эконометрических методов: [1, 8].

Задание 2. Пройдите входное тестирование.

Выберите варианты правильных ответов:

1. Эконометрика:

а) это наука, которая использует методы математической статистики для описания теоретических моделей реальных хозяйственных экономических процессов;

б) это наука, которая использует методы статистики для описания теоретических моделей реальных хозяйственных экономических процессов.

2. Объектом эконометрики являются:

а) экономические процессы, происходящие в экономической системе общества;

б) биологические процессы, происходящие в биологической системе.

3. Предметом эконометрики является:

а) количественная оценка взаимосвязи между случайными событиями, признаками, показателями, факторами, переменными экономических объектов, проверка теоретических моделей реальных экономических процессов для получения прогнозов деятельности экономических систем;

б) оценка взаимосвязи между случайными событиями, признаками, показателями, факторами, переменными экономических объектов, проверка теоретических моделей реальных экономических процессов.

4. Целью эконометрики является:

а) оценка точечных и интервальных прогнозов деятельности генеральной совокупности объектов экономической системы на основании расчетов по данным выборочной совокупности или реализации случайного процесса;

б) расчет точечных прогнозов деятельности генеральной совокупности объектов экономической системы по данным генеральной совокупности.

5. Задачами эконометрики являются:

а) разработка эконометрических методов, повышающих точность прогноза;

б) разработка эконометрических методов, снижающих точность прогноза.

6. Методами эконометрики являются:

а) метод наименьших квадратов;

б) метод наибольших квадратов.

7. Построение графика в среде Excel производится в такой последовательности:

а) ввести данные Х и У, выделить данные У и Х, мастер диаграмм, точечная, готово;

б) ввести данные Х и У, выделить данные У и Х, мастер диаграмм, гистограмма, готово.

Задание 3. Построить график зависимости У от Х.

В табл. 1.1 представлены пространственные данные зависимости У — розничный товарооборот от Х — количество продавцов, при условии что в магазине имеется очередь.

Таблица 1.1

База данных зависимости У от Х

№ магазина	Х (чел.)	У (тыс. руб.)
1	1	4
2	3	6
3	2	7
4	4	10
5	5	?

У — зависимая переменная, которая изучается в соответствии с миссией, видением, стратегическими и оперативными целями и задачами деятельности предприятия, а также с учетом актуальности и важности проведения исследования.

Примечание. Одним из критериев правильно выбранной зависимой переменной является материальное вознаграждение руководством данного исследования.

Различия в розничном товарообороте зависит от количества продавцов и неучтенных факторов.

К неучтенным факторам можно отнести: наличие сертификатов соответствия на товар и систему качества, безопасность товара, насыщенность товаров по ассортименту, эффективность использования торгового оборудования, квалификацию продавцов, культуру обслуживания, стиль руководства и эффект взаимодействия продавцов между собой, мотивацию, рабочую среду, изменение спроса в зависимости от места расположения магазина.

Пояснения.

1. Скопируйте табл. 1.1 и вставьте ее в лист Excel.

2. В листе Excel выделите значения X и У, мастер диаграмм, точечная, готово. На экране появится график, представленный на рис. 1.1.

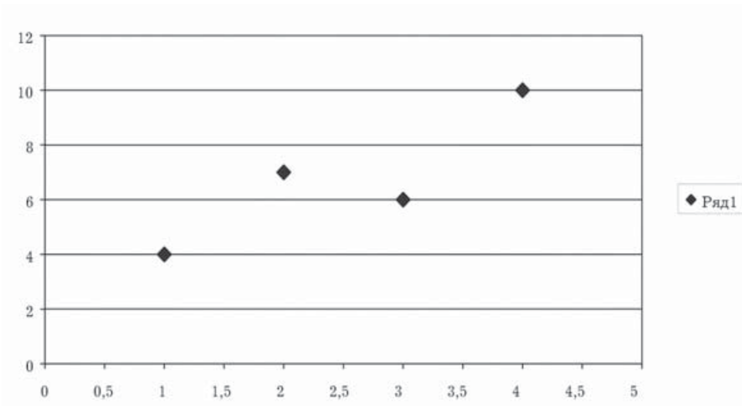


Рис. 1.1. График зависимости У от X

Задание 4. Визуально определите коэффициенты линейной модели $У = a_0 + a_1X + e$ и точечный прогноз при ожидаемом значении количества продавцов, равного 5 чел.,

где У — фактические значения зависимой величины;

a_0, a_1 — коэффициенты линейной модели;

X — фактор;

e — остатки модели, $e = У - (a_0 + a_1X) = У - У_р$;

$У_р = a_0 + a_1X$ — расчетные значения У.

Пояснения. Возьмите линию из набора инструментов “рисование” и проведите ее по рис. 1.1 так, чтобы:

— линия проходила через точку $\overline{X}$ и $\overline{У}$;

— линия прошла как можно ближе к исходным данным У (рис. 1.2).

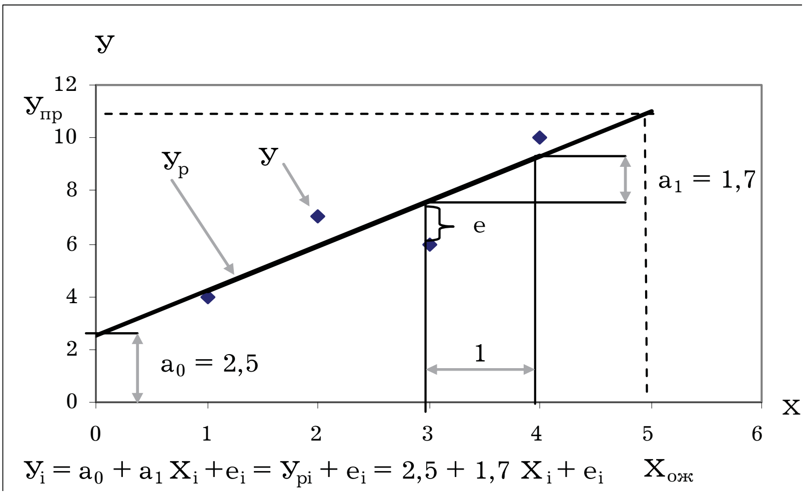


Рис. 1.2. Построение тенденции зависимости У от X

Изучите визуальное определение коэффициентов a_0 и a_1, получение точечного прогноза [3, с. 33–38].

Примечание: a_0 равно расчетному значению У в точке X = 0, a_1 равно приросту расчетного значения У при изменении X на единицу. Прогнозное значение $У_{пр}$ равно расчетному значению У при X = 5.

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

$$У_i = 2{,}5+1{,}7X_i+e_i.$$

Средняя производительность труда продавцов составляет $a_1 = 1{,}7$ тыс. руб.

Если в магазине не будет продавцов, то товарооборот составит $a_0 = 2{,}5$ тыс. руб. Этот вывод является абсурдным и объясняется тем, что:

— во-первых, расчетное значение У при X = 0, равное a_0, является прогнозным значением "назад" и не входит в нашу задачу;

— во-вторых, для исходных данных, изменяющихся от одного продавца и больше, модель хорошо отражает имеющуюся зависимость У от X.

— в-третьих, для получения прогноза У при X = 0 необходима другая модель, учитывающая особенность объекта моделирования, в частности необходимо вести ограничения $a_0 = 0$.

Примечание: при проведении расчетов в Excel с помощью функции "Линейн" имеется возможность ввести ограничение на $a_0 = 0$.

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

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

Задание 5. Визуально приблизительно определите ошибку модели $У = a_0 + a_1X + e$.

Пояснения. Приблизительно ошибку модели (E) можно рассчитать как среднее кваратическое отклонение остатков (5е)по формуле

$S_e = E = (e_{мах} — e_{min)}/6.$

Обоснование предложенной формулы состоит в том, что если остатки имеют нормальный закон распределения, то с вероятностью 0,99 можно утверждать, что между максимальным и минимальным значением случайной величины будет помещаться 6 среднеквадратических отклонений[1].

Для рис. 1.2

$$e_{мах} = 1 \text{ для } X = 2; \; e_{min} = -8 \text{ для } X = 3.$$

$$S_e = E = (e_{мах} — e_{min)}/6 = [2 — (-8)]/6 = 10/6 = 1{,}6 \text{ (тыс. руб.)}.$$

Задание 6. Визуально приблизительно определите индекс детерминации.

Пояснение. Индекс детерминации равен доли объясненной вариации или равен доли исходных данных, которые имеет выбранную тенденцию.

Для рис. 1.2 изменения У при X = 1, X = 2 и изменения У при X = 3, X = 4 совпадают с выбранной тенденцией. Следовательно, из трех изменений У в двух случаях изменения совпадают с выбранной тенденцией, тогда индекс детерминации определится по формуле

$$R^2 = 2/3 = 0{,}66.$$

Вывод: 66% исходных данных имеют линейную тенденцию.

Творческое задание. Как приблизительно определить индекс детерминации при условии, что зависимость У от X представлена как облако значений (рис. 1.3).

Пояснение. Одним из вариантов может быть следующее решение.

С помощью средних значений X и У разделим все облако на четыре квадрата. Определим, какая доля данных находится во втором и четвертом квадратах. Ответ: 9 из 14. Тогда индекс детерминации будет равен 9/14 = 0,64.

---

[1] См. свойства нормального закона распределения и “правило трех сигм” в учебниках по математической статистике

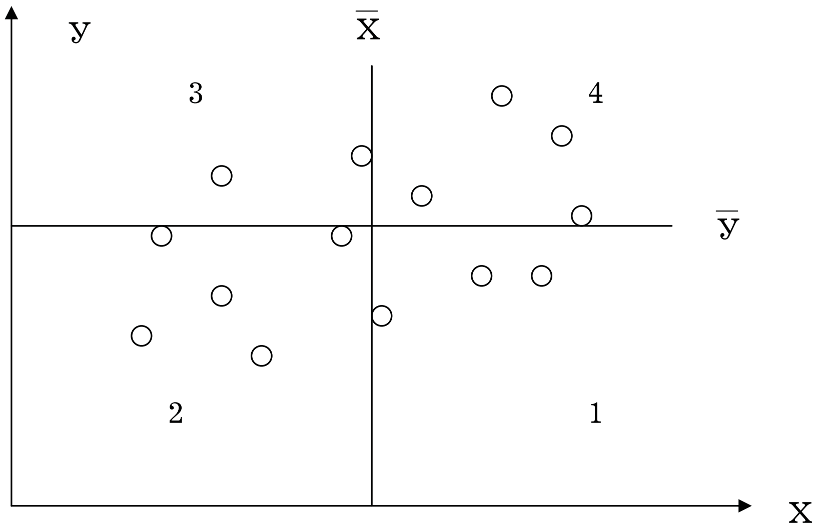


Рис. 1.3. Зависимость У от Х для пространственной выборки

Задание 7. Получите приблизительные 95%-ные доверительные прогнозные интервалы для прогнозного значения У.

Пояснения. При объеме выборки больше 20 95%-ный доверительный интервал приблизительно вычисляется по формуле $У_{пр} \pm 2Е$[1], (99%-ный доверительный интервал вычисляется по формуле $У_{пр} \pm 3Е$[2])

Выполним расчеты: $У_{пр} \pm 2Е = 11 \pm 2{\times}1{,}6 = 11 \pm 3{,}2$.

Вывод. С вероятность 0,95 можно утверждать, что прогнозное значение У будет находиться в интервале 11 ± 3,2.

Задание 8. Изучите основные виды тенденций, которые имеются во временных рядах и остатках экономических показателей.

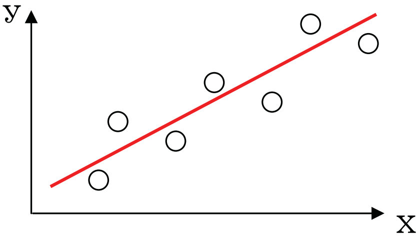
[1] Число 2 соответствует критическому значению критерия Стьюдента на уровне значимости 0,05.

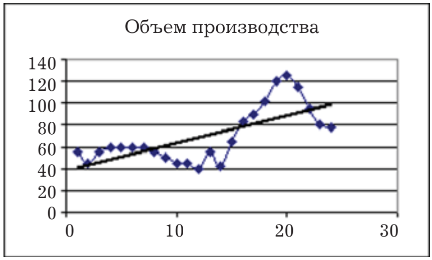
[2] Число 3 соответствует критическому значению критерия Стьюлента на уровне значимости 0,01.

Закономерности во временных рядах

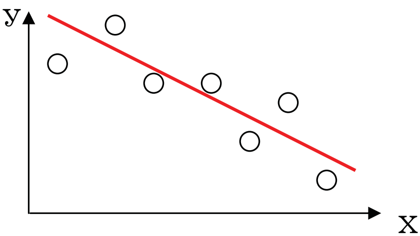
Условные данные	*Фактические данные временных рядов*

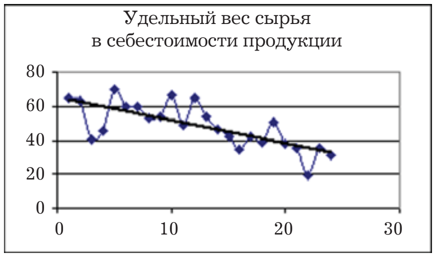
Положительная линейная тенденция



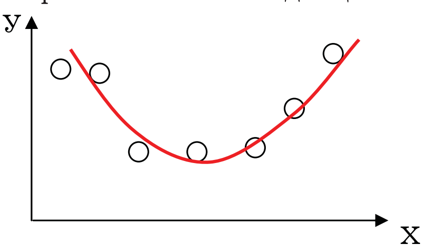


Отрицательная линейная тенденция

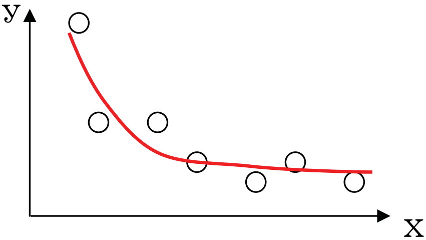




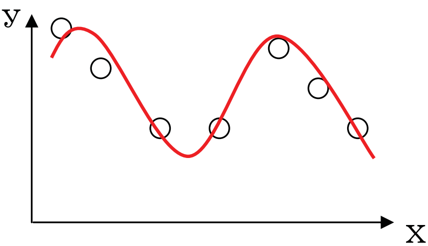
Параболическая тенденция

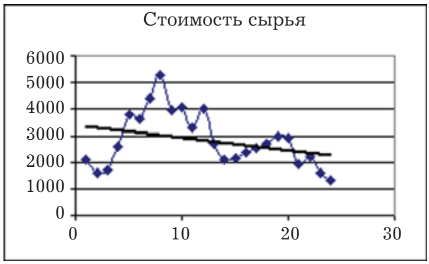


Гиперболическая тенденция

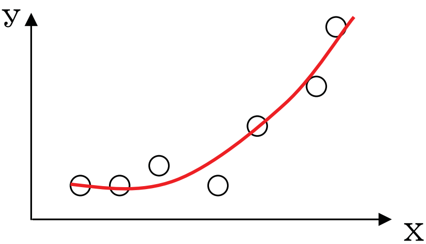


Периодическая тенденция

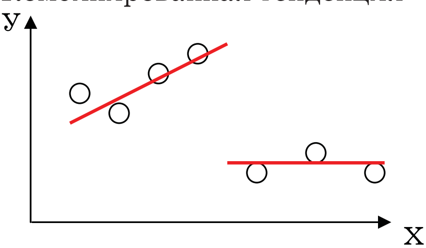


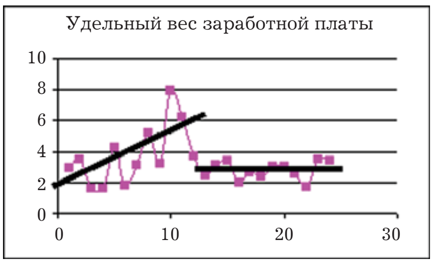


Экспоненциальная тенденция



Комбинированная тенденция

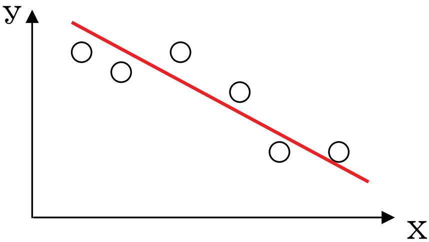




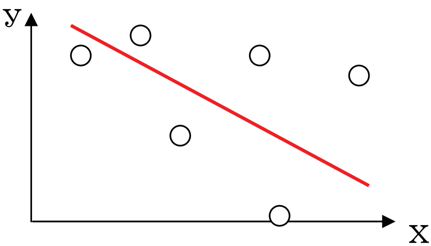
Закономерности остатков линейной модели

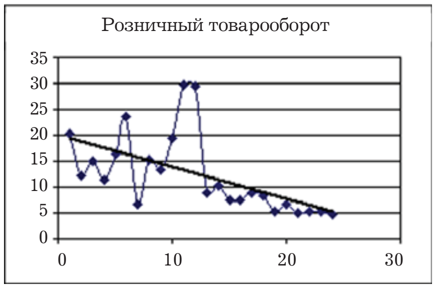
Условные данные	*Фактические данные временных рядов*

Остатки однородны (гомоскедастичны)

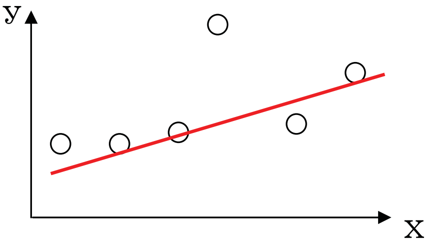


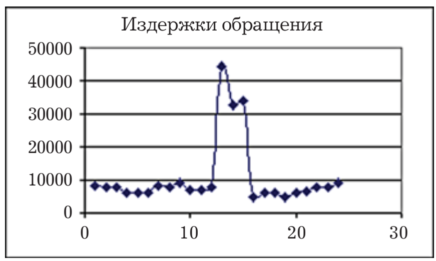
Остатки неоднородны (гетероскедастичны)





Остатки содержат выбросы





Задание 9. Изучите виды тенденций временных рядов на фактических данных условной области, представленных в таблицах и рисунках.

Таблица 1.2

Показатели деятельности коммунального предприятия (в действующих ценах)

t	$У_t$	t	$У_t$
1	908	13	-5963
2	4219	14	804
3	-8914	15	-1002
4	4770	16	-6950
5	-6674	17	-4596
6	-7798	18	-9671
7	3882	19	-2793
8	3409	20	-6734
9	7294	21	-2104
10	-9623	22	7857
11	-1032	23	3003
12	-6734	24	-8964

где t — время (месяцы) за два условных года;
У — прибыль (тыс. руб.).

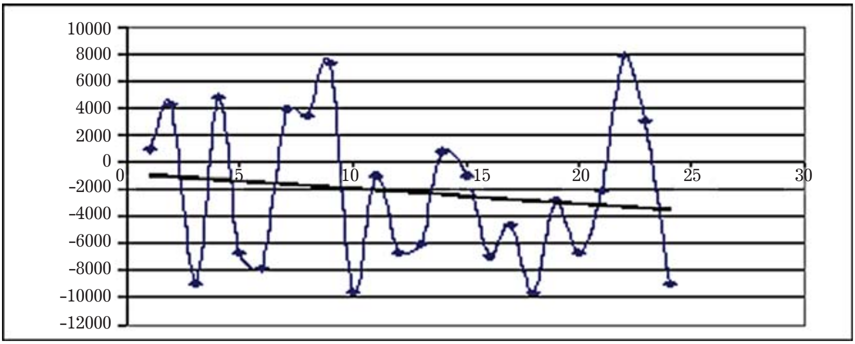


Рис. 1.4. Динамика прибыли (У) (тыс. руб.)

Анализ рис. 1.4 показывает, что данные содержат слабо выраженную отрицательную линейную тенденцию с большими однородными остатками (остатки гомоскедастичны)

Таблица 1.3

База данных показателей деятельности хлебокомбината (в действующих ценах)

t	X_{1t}	X_{2t}	X_{3t}	X_{4t}	$У_t$
1	65,2	2,9	34	68,26	24 549
2	63,02	3,47	34	70,17	25 045
3	40,67	1,62	35	81,51	36 090
4	45,75	1,58	35	90,96	36 478
5	69,69	4,22	35	64,1	25 386
6	59,85	1,83	36	68,65	35 992
7	59,9	3,15	36	70,93	35 867
8	52,72	5,17	36	77,88	36 143
9	53,52	3,24	36	78,66	35 227
10	66,23	7,89	36	90,22	23 248
11	48,98	6,23	36	75,89	27 346
12	64,99	3,69	34	65,16	19 444
13	53,63	2,41	34	67,03	18 640
14	45,93	3,14	34	68,01	17 661
15	41,82	3,43	34	69,51	19 032
16	34,72	1,97	34	88,35	25 164
17	42,28	2,63	35	81,68	18 998
18	38,69	2,32	36	73,24	17 035
19	50,74	3,02	36	65,12	18 166
20	38,12	3,01	36	81,52	23 204
21	35,75	2,55	36	78,66	21 117
22	19,16	1,74	36	80,81	20 995
23	35,08	3,51	35	80,06	17 863
24	30,99	3,37	35	77,95	17 525

где t — время (месяцы) за два условных года;

X_1 _ удельный вес сырья в общей себестоимости продукции (%);

X_2 _ удельный вес суммы заработной платы производственных рабочих в общей себестоимости продукции (%),

X_3 _ колличество рабочих, (чел.);

X_4 _ уровень затрат (%);

У — себестоимость продукции (руб).

Графики динамики показателей представлены на рис. 1.5−1.9

Анализ рис. 1.5 показывает, что данные содержат четко выраженную отрицательную линейную тенденцию с циклической составляющей

(Т = 6...8 м).

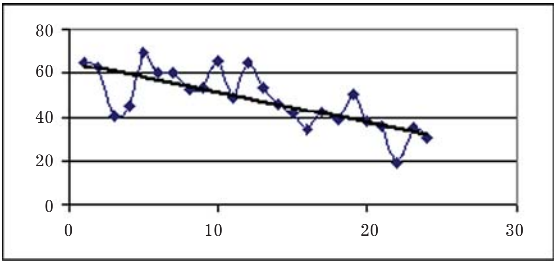


Рис. 1.5. Динамика удельного веса сырья (X1) в общей себестоимости продукции

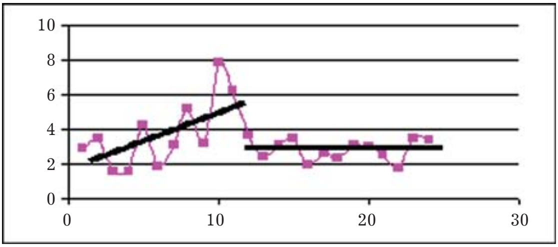


Рис. 1.6. Динамика удельного веса суммы заработной платы (X_2) производственных рабочих в общей себестоимости продукции

Анализ рис. 1.6 показывает, что данные имеют функциональную неоднородность: в предыдущем году была положительная линейная тенденция с большими колебаниями остатков, в последующем году тенденция не отличалась от среднего значения с малыми колебаниями остатков.

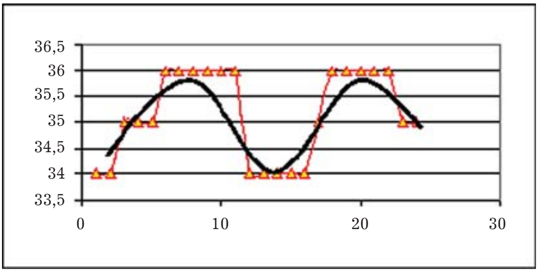


Рис. 1.7. Динамика колличества рабочих (X_3)

Анализ рис. 1.7 показывает, что данные имеют четко выраженную циклическую тенденцию с периодом 12 месяцев (сезонная волна).

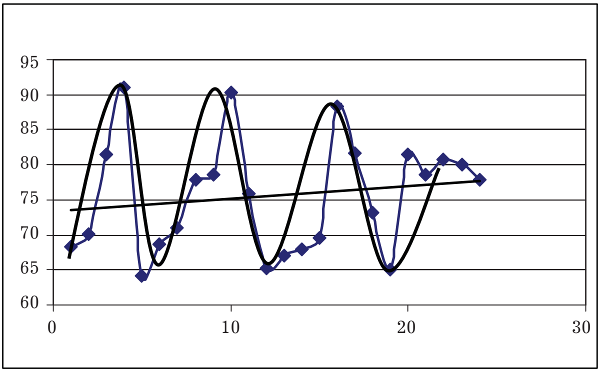


Рис. 1.8. Динамика уровня затрат (X_4)

Анализ рис. 1.8 показывает, что данные имеют четко выраженную циклическую тенденцию с периодом 6 месяцев.

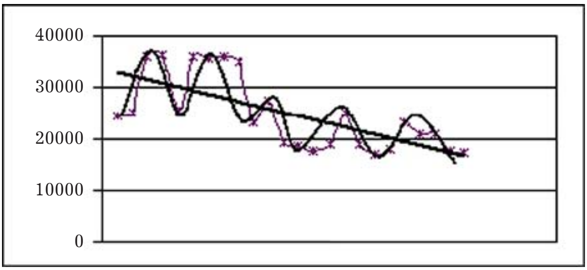


Рис. 1.9. Динамика себестоимости продукции (У)

Анализ рис. 1.9 показывает, что данные содержат четко выраженную отрицательную линейную тенденцию с циклической составляющей (Т = 6…8 м), с учетом всех регулярностей остатки являются однородными.

Таблица 1.4

Показатели деятельности ОАО комбината по производству сахара

t	$У_t$	t	$У_t$
1	189,5	13	-120
2	128,3	14	-80
3	-541,1	15	-113,4
4	-9,3	16	-75,9
5	-10,2	17	-27,3
6	190,8	18	-5
7	-21,4	19	-4,9
8	-31,6	20	-12,3
9	43,9	21	-200,4
10	-49,5	22	17,2
11	-189,5	23	-400,3
12	441,1	24	-448,9

где t — время (месяцы) за два условных года;

У — прибыль, тыс. руб. в действующих ценах.

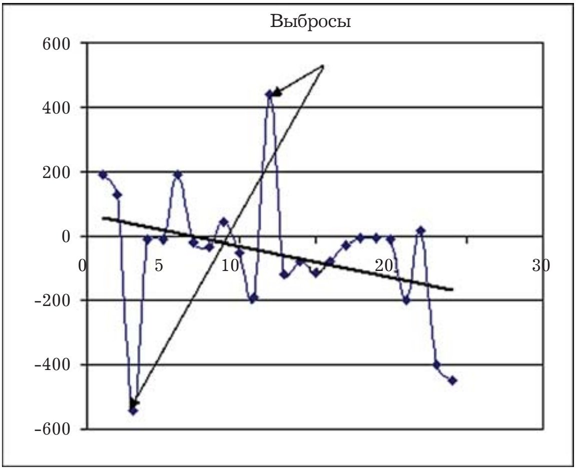


Рис. 1.10. Динамика прибыли (У)

Анализ рис. 1.10 показывает, что данные содержат слабо выраженную отрицательную линейную тенденцию с выбросами в 3-м и 12-м месяцах предыдущего года и 11-м, 12-м месяцах последующего года остатки с учетом всех регулярностей являются однородными.

Таблица 1.5

Показатели деятельности райпотребсоюза

t	$У_t$	t	$У_t$
1	8190	13	44500
2	7610	14	32450
3	8000	15	34050
4	5990	16	4630
5	6060	17	6039
6	6200	18	6231
7	8300	19	4820
8	7820	20	6170
9	8980	21	6510
10	6900	22	7930
11	6850	23	8035
12	8000	24	9035

где t — время (месяцы); за два условных года;
У — издержки обращения в действующих ценах (руб.)

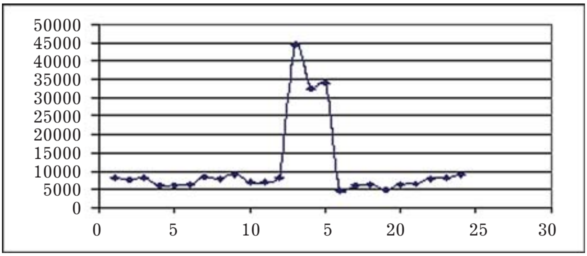


Рис. 1.11. Динамика издержек обращения (У)

Анализ рис. 1.11 показывает, что данные имеют тенденцию, которая не отличается от среднего значения, имеются четко выраженные выбросы в 13-м, 14-м, 15-м месяцах, остатки без учета выбросов являются однородными.

Задание второго уровня сложности

Задание. Используя базу условных данных, расположенных в табл. 1.6, постройте график с использованием разнообразных графических средств Excel (рис. 1.12).

Таблица 1.6

База условных данных зависимости У от t

t	$У_t$		t	$У_t$
1			10	
2			11	
3			15	
4			16	
5			30	
6			69	
7			60	
8			80	

где t — порядковый номер времени;

У — значение экономического показателя в условных единицах.

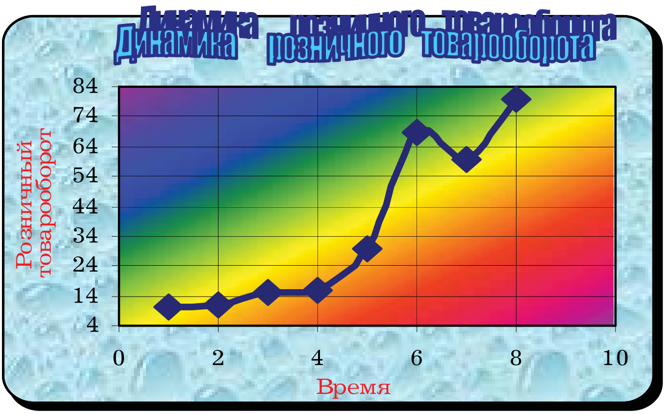


Рис. 1.12. Пример использования графических средств Excel

Задание третьего уровня сложности

Задание 1. Имеется матрица парных коэффициентов корреляции, представленная в табл. 1.7. Необходимо графически представить матрицу парных коэффициентов корреляции с использованием графических средств Excel трехмерной графики.

Таблица 1.7

Матрица парных коэффициентов корреляции

	X_1	X_2	X_3	У
X1	1	0,8	-0,7	0,2
X2	0,8	1	0,5	0,3
X3	-0,7	0,5	1	-0,9
У	0,2	0,3	-0,9	1

Пример выполнения задания см. рис. 1.13.

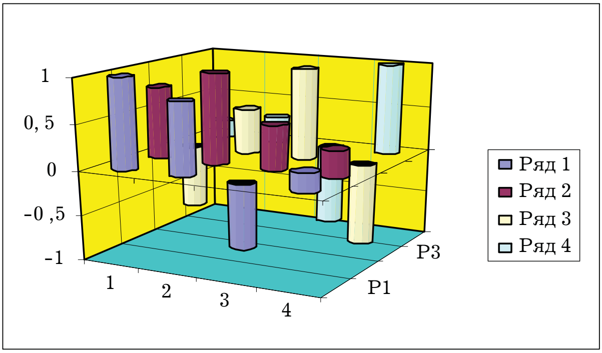


Рис. 1.13. Трехмерный график матрицы парных коэффициентов корреляции

Задание четвертого уровня сложности

Задание. Имеется зависимость У от двух факторов, численные значения которых приведены в табл. 1.8.

Таблица 1.8

База условных данных зависимости У от X_1 и X_2

i	X_1	X_2	У
1	12	25	126
2	13	23	123
3	16	26	124
4	12	29	158
5	14	31	169
6	21	35	171

Необходимо с помощью пакета программ Table Curve 3D построить трехмерный график зависимости У от X_1 и X_2 (или зависимость Z от X и Y).

Пояснение. С помощью данного пакета можно построить поверхность по 39 функциям с анализом свойств полученной поверхности. Однако, надо знать английский язык.

Приводим пример решения задачи (рис. 1.14).

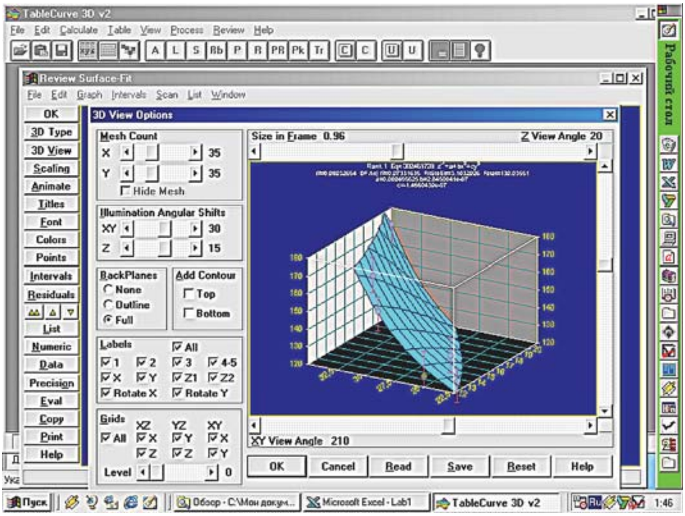


Рис. 1.14. Трехмерный график, построенный с помощью Table Curve 3D

Дополнительное задание

Задание. Составьте кроссворд из следующих ключевых слов по вертикали:

1. Наука, которая использует методы математической статистики для описания теоретических моделей реальных хозяйственных экономических процессов **(эконометрика)**.

2. Задачей эконометрики является разработка эконометрических методов, повышающих (**точность)** прогноза.

По горизонтали

1. Объектом эконометрики являются **(экономические)** процессы, происходящие в экономической системе общества.

2. Предметом эконометрики является количественная оценка взаимосвязи между случайными событиями, признаками, показателями, факторами, переменными экономических объектов, проверка теоретических моделей реальных экономических процессов для получения (**прогнозов)** деятельности экономических систем.

3. Построение графика в среде Excel производится в такой последовательности: ввести данные X и У, выделить данные У и X, мастер диаграмм, (**точечная)**, готово.

Пример кроссворда

		1Э										
		К		2Т								
1Э	К	О	Н	О	М	И	Ч	Е	С	К	И	Е
		Н		Ч								
2П	Р	О	Г	Н	О	З	О	В				
		М		О								
		Е		С								
		Т		3Т	О	Ч	Е	Ч	Н	А	Я	
		Р		Ь								
		И										
		К										
		А										

Выходное тестирование

Задание. Дана база данных (табл. 1.9), по которой в течение 10 с надо в Excel построить график зависимости У от i, без надписей и редактирования (рис. 1.15).

Таблица 1.9

База условных данных динамики экономического показателя

i	$У_i$	i	Уi
1	10	6	16
2	12	7	20
3	16	8	25
4	13	9	29
5	19	10	30

где i — порядковый номер месяца;

$У_i$ — численные значения экономического показателя в условных единицах.

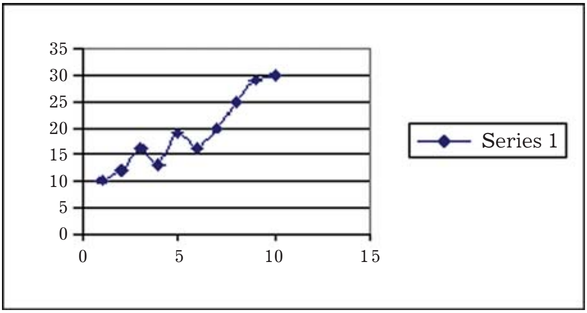


Рис. 1.15. Пример построения графика за 7 с

Нерешенные проблемы по теме визуального анализа тенденций зависимости экономического показателя от фактора: нет хорошей методики определения основных характеристик регрессионных моделей с использованием только визуального анализа тенденций экономического показателя.

2. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Эконометрика — это не только построение модели,
но и решение проблемы *С.А. Айвазян*

Постановка задачи

Объект — пространственные данные экономического показателя.

Предмет — регулярности экономического показателя.

Цель — научить вычислять коэффициенты линейного уравнения регрессии с помощью метода наименьших квадратов.

Актуальность — расчет коэффициентов большинства эконометрических моделей выполняется с помощью метода наименьших квадратов.

Рабочая гипотеза — метод наименьших квадратов позволит вычислить коэффициенты линейной модели.

Метод — метод наименьших квадратов.

Способ — использование Excel как калькулятора.

Задачи:

- — рассмотреть этапы эконометрического моделирования как реализацию цикла Деминга;
- — вычислить коэффициенты линейной модели с помощью средств Excel как калькулятора.

Ожидаемый результат — коэффициенты модели, вычисленные визуальным способом и методом наименьших квадратов, должны совпадать.

Метод для сравнения — сравнение результатов визуального анализ с точными значениями характеристик линейной модели.

Задания первого уровня сложности

Задание 1. Подготовьте устные сообщения по вопросам самостоятельной работы.

1. Дайте определение модели.

2. Приведите классификацию моделей.

3. Приведите виды абстрактных моделей.

[3, с. 46–51; 7, с. 9–21; 1, с. 20–24; 8, с. 21–25; 9, с. 40–43].

4. Приведите цикл Деминга улучшения процессов и этапы эконометрического моделирования.

[3, с. 51–53; 7, с. 21–23; 1, с. 31–36; 9, с. 43–48].

5. Приведите основные свойства экономической системы как объекта моделирования.

[3, с. 53–54; 10, с. 98–100].

6. Приведите классификацию переменных в эконометрических исследованиях.

[3, с. 55–61; 7, с. 13–17; 9, с. 25–31; 4, с. 25–33; 8, с. 25–26].

7. Приведите методы: выявление проблем, существующих на предприятии; выявления наиболее значимой проблемы; выявление причин появления проблемы; выявления наиболее значимых причин, влияющих на проблему. [3, с. 352–356; 16, с. 36–52].

8. Приведите общий вид и структуру множественной регрессии. [3, с. 70–72; 7, с. 9–11; 1, с. 42–494; 5, с. 34–40; 8, с. 60–64]..

9. В чем сущность спецификации модели?

10. Приведите условия идентифицируемости модели (ограничения, накладываемые на свойства переменных, их количества, вид модели).

11. Выведите формулы оценок параметров парной и множественной регрессии методом наименьших квадратов.

12. Укажите основные предпосылки метода наименьших квадратов.

13. Выясните свойства оценок параметров регрессионной модели (несмещенность, состоятельность, эффективность).

[3, с. 70–79; 7, с. 17–107; 1, с. 48–198; 4, с. 34–176; 8, с. 27–81; 10, с. 13–200; 6, с. 4–141; 12, с. 3–132, 165–196; 9, с. 231–241].

Задание 2. Пройдите входной тест.
Выберите вариант правильного ответа.

1. Модель это:
а) условный образ объекта исследования;
б) точное копирование объекта исследования.

2. Абстрактные модели подразделяются:
а) на графические, математические, словесно-описательные;
б) математические.

3. Цикл Деминга состоит из следующих этапов:
а) планируй, делай, анализируй, совершенствуй;
б) планируй, делай.

4. Этапы эконометрического моделирования:

а) Постановочный. Априорный. Информационный. Спецификация модели. Идентификация модели. Определение качества модели. Верификация модели. Выводы и предложения.

б) Постановочный. Априорный. Информационный. Спецификация модели. Идентификация модели. Выводы и предложения.

5. Основные свойства экономической системы как объекта моделирования:

а)

— все экономические процессы происходят в пространстве и во времени;
— экономическая система — самонастраиваемая система, которая может находиться в состоянии динамического равновесия;
— экономическая система обладает инерционными свойствами;
— текущее состояние экономической системы испытывает влияние прошлых, настоящих и будущих значений переменных этой системы;
— для всех явлений в природе между причиной и следствием существует временной лаг, или временная задержка;
— все временные экономические процессы происходят циклически;

— последние значения временного ряда оказывают бо́льшее влияние на прогнозное значение, чем первые значения временного ряда;
— прошлые значения показателя временного ряда оказывают влияние на его текущее значение, но не зависят от него.

б)
— все экономические процессы происходят не в пространстве и не во времени;
— экономическая система — не самонастраиваемая система, которая не может находиться в состоянии динамического равновесия;
— экономическая система не обладает инерционными свойствами;
— текущее состояние экономической системы не испытывает влияния прошлых, настоящих и будущих значений переменных этой системы;
— для всех явлений в природе между причиной и следствием не существует временного лага, или временной задержки;
— все временные экономические процессы происходят не циклически.
— последние значения временного ряда не оказывают большее влияние на прогнозное значение, чем первые значения временного ряда;
— прошлые значения показателя временного ряда не оказывают влияния на его текущее значение и зависят от него.

6. Классификация переменных

а) Входные. Выходные. Производные переменные. Зависимая переменная. Главный фактор. Второстепенный фактор. Числовая переменная. Качественная переменная. Исходные переменные. Преобразованные переменные. Доступные переменные. Скрытные, или латентные переменные. Фиктивная переменная. Инструментальная переменная. Текущие. Лаговые. Эндогенные. Экзогенные. Статические. Динамические. Средние динамические равновесные. Синхронизирующие.

б) Входные. Выходные.

7. Методы выявления проблем:

а) самооценка по методик, отраженной в премии РФ в области качества;

б) причинно–следственная диаграмма.

8. Метод выявления наиболее значимой проблемы:

а) диаграмма Парето по проблемам;

б) метод: почему? почему?

9. Метод выявления причин проблемы:

а) причинно-следственная диаграмма Исикавы, метод: почему? почему?, мозговой штурм, расслоение;

б) расслоение.

10. Метод выявления наиболее значимой причины:

а) диаграмма Парето по причинам, мозговой штурм;

б) мозговой штурм.

11. Спецификация модели — это:

а) определение вида математической функции, которая описывает влияние объясняемых переменных на зависимую переменную;

б) определение факторов, влияющих на зависимую переменную.

12. Идентифицируемость модели — это:

а) возможность вычисления коэффициентов регрессионной модели методом наименьших квадратов (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); возможность вычисления коэффициентов структурной системы одновременных уравнений по коэффициентам приведенной системы одновременных уравнений;

б) возможность вычисления коэффициентов регрессионной модели методом наименьших квадратов

13. Вывод формул оценок параметров парной и множественной регрессии методом наименьших квадратов

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 Х:

$У_i = a_0 + a_1 X_i + e_i$.

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

$У_i = a_0 + a_1 X_i + e_i$

будет минимальной.

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:

У = ХА + е,

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:

е'е $\Rightarrow$ min, где **е = У — ХА**.

е'е = (У — ХА)'(У — ХА) = (У' — А'Х')(У- ХА) = У'У — А'Х'У — У'ХА + А'Х'ХА = = У'У — 2А'Х'У +А'Х'ХА.

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

У'ХА = (У'ХА)' = А'Х'У.

Примечание. Свойства операций транспонирования.

А, В, С — матрицы;

(A)’ = A;

(A+B)’ = A’ + B’;

(AB)’ = B’A’;

(ABC)’ = C’B’A’.

Возьмем первую частную производную от **е’е** по переменной **А**, приравняем ее к нулю и определим оценки параметров модели, при которых **е’е** 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 **B, C** — вектор столбцы, D — симметрическая матрица.

Полагая, что **А** и **Х’У** векторы столбцы, найдем

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

Полагая, что **Х’Х** — симметрическая матрица, **А** — вектор столбец, найдем

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

После несложных преобразований можно получить несколько выражений.

Х’ХА = Х’У — система нормальных уравнений, представленная в матричном виде.

А = (Х’Х)$^{-1}$ Х’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.

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

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

14. Расчет коэффициентов линейной модели $У_i = a_0 + a_1X_i + e_i$ в алгебраическом виде

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

15. Предпосылки метода наименьших квадратов:

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

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

где M — математическое ожидание;

ε_i — возмущения модели;

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

Задание 3. Решить задачу.

Имеются численные значения двух показателей: количество продавцов и розничного товарооборота по четырем выборочным однородным магазинам одной фирмы, представленные в табл. 2.1.

Таблица 2.1

Показатели работы четырех магазинов одной фирмы

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

где i — номер филиала фирмы;

Х — количество продавцов;

У — величина розничного товарооборота;

$Х_i$ — количество продавцов (чел.);

$У_i$ — значение розничного товарооборота (тыс. руб.).

Примечание. Х и У могут означать различные переменные, например:

1) Х — время (порядковый номер: дня, месяца, квартала, года), У — временной ряд — экономический показатель предприятия (прибыль, количество работников, количество потерь от брака, затраты на качество, стоимость, качество, сервис, темпы);

2) Х — инвестиции, У — прибыль;

3) Х — затраты на рекламу, У — розничный товарооборот;

4) Х — спрос, У — предложение;

5) Х — спрос, У — цена;

6) Х — таможенный налог на ввоз товара, У — цена товара;

7) Х — доходы дебитора, У — размер кредита, представляемый банком;

8) Х — степень усталости оператора ЭВМ, У — количество совершенных ошибок оператором ЭВМ;

9) Х — время работы ЭВМ, У — риск появления сбоя ЭВМ;

10) Х — качество сырья, У — объем выработанной продукции;

11) Х — затраты на качество, У — цена продукции;

12) Х — риск, У — прибыль.

Необходимо:

— вычислить коэффициенты a_0 и a_1 выборочной парной линейной регрессии:

$$У = a_0 + a_1Х + e;$$

— получить прогнозное значение $У_{пр}$, при ожидаемом значении $Х_{ож} = 5$;

— вычислить расчетные значения $У_{pi}$ для каждого значения $Х_i$;

— построить график зависимости У и $У_р$ от Х.

— решение задачи выполнить в соответствии с этапами эконометрического моделирования.

Решение

Целью моделирования является решение приоритетной проблемы. Проблемы можно выявить с помощью самооценки по модели премии правительства РФ в области качества (рис. 2.1).

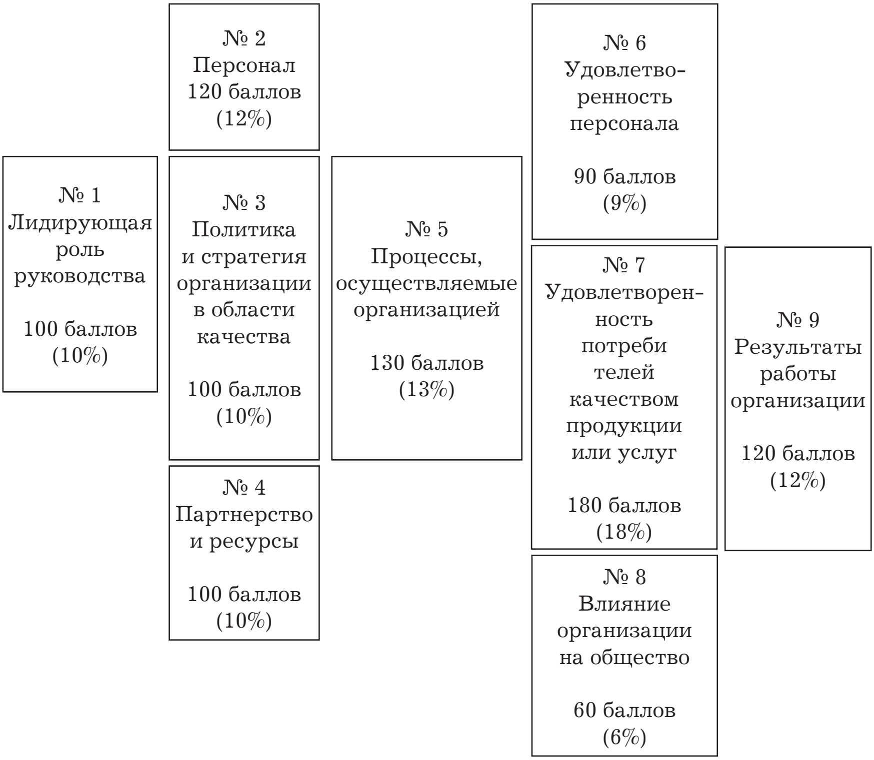


Рис. 2.1. Модель премии правительства РФ в области качества

Модель содержит 9 критериев, каждый из которых имеет несколько подкритериев, включающих вопросы, на которые надо дать ответы[1].

[1] http: //www.vniis.ru

Нас интересуют результаты деятельности фирмы, которые содержатся в девятом критерии.

Критерий 9: *Результаты работы организации*

Содержание критерия: Результаты, которых добилась организация в отношении запланированных целей в работе.

Составляющие критерия

Критерий включает в себя следующие две составляющие, по которым должна быть представлена информация:

9а) финансовые показатели работы организации

Оцениваются на основе результатов анализа, представленных в виде графиков, таблиц, диаграмм, в том числе:

- показателей прибыли и убытков, включая:
  - выручку от реализации;
  - прибыль от реализации;
  - чистую прибыль;
  - прибыль от финансово-хозяйственной деятельности,
  - отчисления из прибыли на благотворительные цели;
- показатели бухгалтерского баланса, включая:
  - валюту баланса;
  - внеоборотные активы;
  - оборотный капитал;
  - капитальные резервы;
  - долгосрочные и краткосрочные пассивы, запасы;
- показатели движения денег, включая:
  - платежи в бюджет;
  - капиталовложения (источники и структура);
  - управление движением наличности (кредиты банка, в том числе непогашенные, дебиторская и кредиторская задолженность);
  - эксплуатационные расходы;
- показателей финансового состояния, включая:
  - финансовую устойчивость;
  - ликвидность;
  - оборачиваемость оборотного капитала;

— коэффициент возможного банкротства (коэффициент восстановления платежеспособности);

- других показателей, включая:
  - — доход на акционерный капитал;
  - — доход на капитал;
  - — кредитный рейтинг (платежеспособность);
  - — общий доход акционеров, в том числе долю выплаченных доходов;
  - — доходы от акций;
  - — затраты на качество;
  - — рентабельность.

(Показатели могут быть выражены в абсолютных, условных или относительных величинах, в том числе, например, к капиталу или на одного работающего).

9б качество продукции (услуг) и другие результаты работы организации

Объектами оценки могут быть любые нефинансовые результаты. В частности, могут оцениваться следующие показатели:

- качество продукции (услуг):
  - — сопоставление с продукцией (услугами) лучших организаций;
  - — уровень дефектности;
  - — добровольная сертификация, в том числе сертификация систем менеджмента;
- процессы:
  - — внедрение передовых процессов и технологий;
  - — улучшения в процессах;
  - — время выполнения процессов;
  - — завершенность;
  - — производительность;
  - — уровень дефектности;
- информация и знания:
  - — средства распространения информации об организации и ее продукции (услугах);
  - — полнота;
  - — доступность;

— достоверность;
— уместность;
— своевременность;
— ценность интеллектуальной собственности;

- деятельность в целом:
  — доля, которую занимает продукция (услуги) организации на рынке;
  — экспорт;
  — освоение новых видов продукции (услуг);
  — время ввода новой продукции (услуг) на рынок;
  — объем продаж в натуральном выражении;
  — использование и состояние основных фондов, в том числе их среднегодовая стоимость, износ, загрузка, коэффициенты обновления, выбытия или интенсивности обновления, автоматизация и механизация производства;
- партнерство:
  — взаимодействие с партнерами, включая факты совместного создания большей ценности, совместных улучшений и инноваций, поощрение партнеров;
  — взаимодействие с поставщиками, включая цену закупаемой продукции (услуг), ее дефектность, отзывчивость поставщиков на обращения организации;
- имущество, в том числе здания и оборудование:
  — обесценивание;
  — стоимость обслуживания;
  — использование.

(Показатели могут быть выражены в абсолютных, условных или относительных величинах).

Из всех показателей нас будут интересовать прибыль и доходы, а также что является более важным потери от недополученных прибылей и доходов по сравнению с поставленными целями. Предположим, что все эти потери известны и обозначим их такими переменными:

$У_1$ — потери прибыли от реализации,

$У_2$ — потери прибыли от финансово-хозяйственной деятельности,

$У_3$ — потери доходов на акционерный капитал,

$У_4$ — потери доходов на капитал,

$У_5$ — потери доходов от акций.

Представим ранжированные численные значения переменных в таблице 2.2.

Таблица 2.2

Значения потерь (данные условные)

Переменные потерь	Значения (млн. руб.)	Накопленная сумма (млн. руб.)
У1	30	30
У2	15	45
У3	8	53
У4	5	58
У5	2	60

Построим диаграмму Парето по результатам с использованием данных таблицы 2.1 в такой последовательности:

— скопировать данные таблицы 2.2 и разместить их на листе Excel;

— выделить три колонки, мастер диаграмм, нестандартные, график|гистограмма, готово.

Результат построения диаграммы Парето изображен на рис. 2.2.

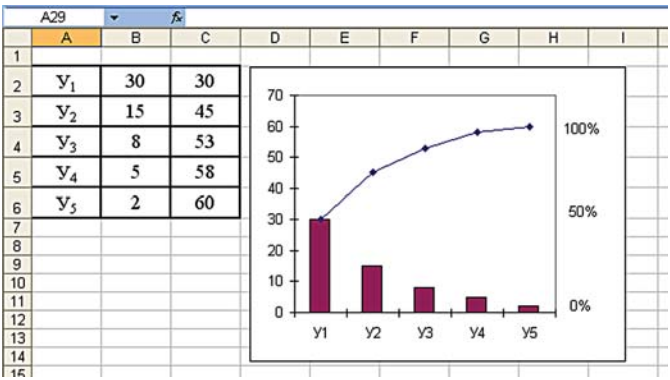


Рис. 2.2. Диаграмма Парето по результатам

Анализ рис. 2.2 показывает, что значения $У_1$ и $У_2$ составляют более 70% от всех потерь предприятия по анализируемым пяти показателям. При этом наибольшие потери (более 50%) составляет недополученная прибыль от реализации ($У_1$).

Следовательно, в первую очередь следует решить проблему более низкой прибыли от реализации по сравнению с поставленными целями. Необходимо определить причину низкой выручки от реализации или величины товарооборота. Анализ причин низкого товарооборота можно выполнить с помощью диаграммы Исикавы.

Примечание. Для выявления причин низкого товарооборота можно использовать другие средства, такие как: мозговой штурм: почему? почему?; корреляционный анализ и др.

Выбор переменных

Величина товарооборота зависит от шести основных факторов: *сырья и материалов* (ассортимент товаров и их качество); *машин* (оборудование и дизайн магазина); *методов* (форма обслуживания покупателей: форма самообслуживания, культура обслуживания); *людей* (количество продавцов, уровень квалификации, образование, возраст, деловые характеристики); *физической и абстрактной среды* (температура, влажность в магазине; традиции, психологический климат существующие в магазине; правовая среда существования магазина, информационное обеспечение магазина); *времени* (тенденции динамического развития магазина) (рис. 2.3).

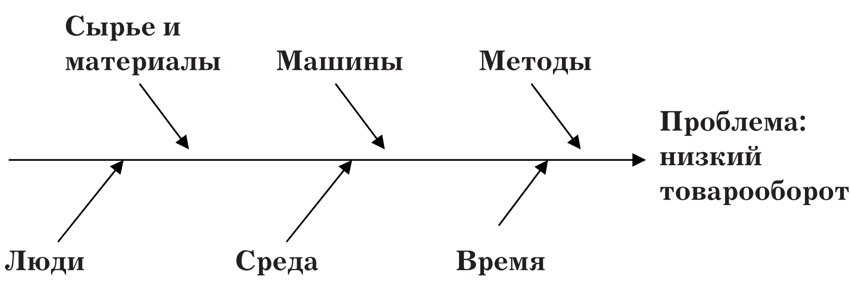


Рис. 2.3. Причинно-следственная диаграмма Исикавы

Выделение входных и выходных переменных

По условию задачи численность продавцов влияет на величину розничного товарооборота, следовательно, входной и переменной будет численность продавцов, выходной переменной-розничный товарооборот. Однако, в общем случае, на основании рис. 2.3, входными переменными могут быть 6 групп Факторов.

Сбор статистических данных

Была произведена выборка по четырем магазинам фирмы с учетом величины розничного товарооборота и количества продавцов (X). Исходные данные представлены в табл. 1.1. График зависимости У от X изображен на рис. 2.4.

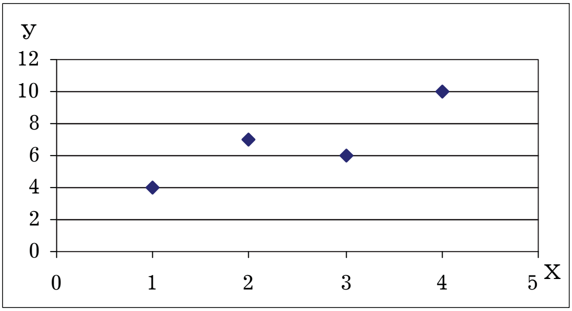


Рис. 2.4. График зависимости У от X

Анализ рис. 2.4 показывает, что с ростом числа продавцов увеличивается розничный товарооборот. Однако эта зависимость является стохастической (вероятностной), а не детерминированной (строго функциональной).

Выдвижение гипотез

В теоретическом плане увеличение продавцов должно привести к росту товарооборота до тех пор, пока существует очередь. Если очереди не будет, то товарооборот не станет зависеть от количества продавцов.

Примечание. Зависимость товарооборота от количества продавцов при наличии и отсутствии очереди можно выразить с помощью логистической функции (рис. 2.5).

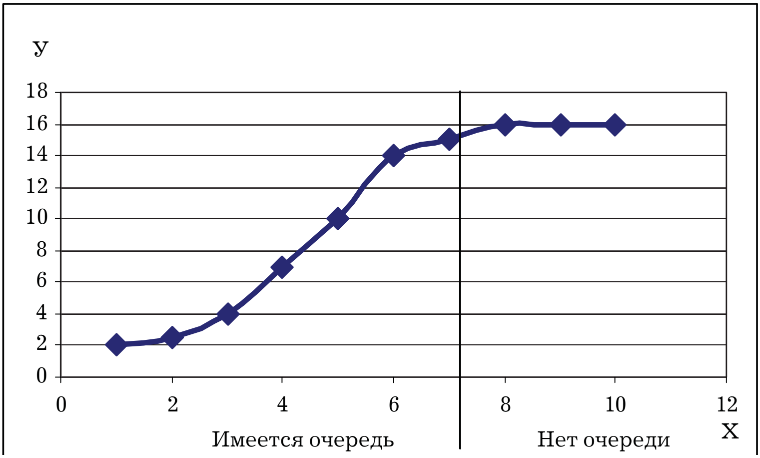


Рис. 2.5. График зависимости розничного товарооборота У от количества продавцов X при наличии и отсутствии очереди

Предположим, что розничный товарооборот зависит линейно от количества продавцов при условии наличия очереди.

Формулировка допущений

Розничный товарооборот будет линейно зависеть от количества продавцов при условии, что в магазине имеется очередь. Продавцы в магазине могут заменять друг друга.

Спецификация модели

Предположим, что фактическая зависимость У от X для всех магазинов 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. 2.6.

Где $M(У|X_i)$ — математическое ожидание значений У при фиксированном значении X_i;

Рис. 2.6. Зависимость У от Х для данных генеральной совокупности

α_0 — свободный параметр модели, равный $У_р$ при Х = 0 (численно равный розничному товарообороту самообслуживания при условии, что в магазине нет продавцов). На рис. 2.6 свободный параметр модели равен 3,33;

α_1 — параметр модели, равный $\partial У_р/\partial Х$ — первой производной от $У_р$ по Х, что соответствует приросту $У_р$ при изменении Х на 1 или математическое ожидание производительности труда продавцов. На рис. 2.6 параметр α_1 равен 1,66;

У — фактические значения зависимой переменной;

Х — фактор, причина, случайная или детерминированная величина, оказывающая влияние на У;

$У_р$ — расчетное значение У или математическое ожидание случайной величины У при фиксированном значении Х;

$У_{пр}$ — точечное прогнозное значение У при ожидаемом значении $Х_{ож} = 5$.

$У_{пр} = 3{,}33 + 1{,}66Х_{ож} = 3{,}33 + 1{,}66 \times 5 = 11{,}63$ тыс. руб.

$Х_i$ — фиксированные значения Х. Для каждого фиксированного значения Х можно получить большое (в идеальном случае бесконечное) множество значений У, которое имеет определенный закон распределения, или условную плотность распределения. Условная плотность распределения имеет несколько характеристик: математическое ожидание (или среднее значение), дисперсию,

ε_i — остатки (возмущения) или случайная составляющая, равна разности между $У_i$ и $У_{pi}$, учитывающая влияние факторов, не вошедших в модель.

Для каждого фиксированного значения X_i можно получить одно значение $У_{pi}$ и множество значений У, а также соответствующее количество остатков, которые имеют определенный закон распределения и характеристики этого распределения: математическое значение остатков $M(\varepsilon_i)$, дисперсию остатков $\sigma_{\varepsilon i}^2$.

Остатки при разных фиксированных значениях X_i (например остатки при X_1 и X_2: ε_1 и ε_2, остатки при X_1 и X_3: ε_1 и ε_2 и т. д.) могут быть связанными между собой или нет. Степень взаимосвязи остатков определяется с помощью коэффициента автокорреляции.

Если дисперсии остатков $\sigma_{\varepsilon i}^2$ при каждом фиксированном значении X_i равны между собой или однородны, то это свойство остатков называется гомоскедастичностью.

Если дисперсии остатков $\sigma_{\varepsilon i}^2$ при каждом фиксированном значении X_i не равны между собой или не однородны, то это свойство остатков называется гетероскедастичностью.

Классическая нормальная линейная модель парной регрессии (КНЛМПР)

Предполагаем, что зависимость У от X можно представить в виде классической нормальной линейной модели парной регрессии.

$$У_i = f(X_i) + \varepsilon_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i = У_{pi} + \varepsilon_i, \ i = 1, 2, ..., N,$$

где $У_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i$ — парное уравнение регрессии для генеральной совокупности;

i — номер фиксированного значения X;

$f(X_i) = M_x(У) = M(У|X = X_i) = У_{pi} = \alpha_0 + \alpha_1 X_i$ — математическая функция зависимости $У_i$ от X_i;

$M_x(У) = M(У|X = X_i)$ — две записи условного математического ожидания У при фиксированном значении X;

ε_i — возмущение при фиксированном значении X_i;

$У_i$ — случайная (стохастическая) переменная;

X_i — неслучайная (детерминированная) переменная;

α_0, α_1 — неизвестные параметры;

ε_i — неизвестное возмущение;

N — объем генеральной совокупности.

Относительно возмущения ε_i выдвигаем следующие гипотезы:

гипотеза 1

$M(\varepsilon_i) = 0$ при всех $i = 1, 2, ..., N$;

гипотеза 2

$M(\varepsilon_i \varepsilon_J) = 0$ при $i \neq j$; $i, j = 1, 2, ..., N$;

гипотеза 3

$M(\varepsilon_i \varepsilon_J) = \sigma_{\varepsilon i}^2$ при $i = j$; $i, j = 1, 2, ..., N$,

где M — знак математического ожидания;

ε_i — нормально распределенная случайная величина со средним $\varepsilon = 0$ и дисперсией равной $\sigma_{\varepsilon i}^2$,

$\sigma_{\varepsilon i}^2$ — неизвестная дисперсия возмущения ε_i;

первая гипотеза означает, что расчетные значения $У_{pi}$ должны пройти через условные математические ожидания;

вторая гипотеза означает, что остатки не связаны между собой для всех пар фиксированных значений X_i;

третья гипотеза означает, что дисперсии остатков являются однородными (гомоскедастичными) и равны константе $\sigma_{\varepsilon i}^2$ для всех фиксированных значений X_i.

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 розничного товарооборота от количества продавцов. Поэтому, можно с помощью исходных данных оценить параметры линейной модели, при условии наличия очереди в магазине.

Идентификация модели

(Предложить и реализовать процедуру оценивания параметров модели).

Нам желательно знать параметры модели α_0, и α_1, однако, если мы не можем проанализировать все магазины генеральной совокупности, то воспользуемся выборочной совокупностью магазинов, с помощью которой вычислим коэффициенты a_0 и a_1 выборочной парной линейной регрессии:

$$У = a_0 + a_1 X + e. \qquad (2.1)$$

Коэффициенты a_0 и a_1 являются оценкой параметров модели α_0, и α_1 и могут существенно отличаться от них. Очевидно, при увеличении объема выборки коэффициенты будут стремиться к их параметрам.

Необходимо определить такие значения коэффициентов a_0 и a_1, при которых сумма квадратов остатков регрессионной модели $У_i = a_0 + a_1 X_i + e_i$ будет минимальной.

Примечание. Вывод формулы расчета коэффициентов модели см. в вопросе 13 входного теста.

$\mathbf{A = (X'X)^{-1}X'У}$ — расчетная формула коэффициентов модели или формула оценки параметров модели для генеральной совокупности.

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

$\mathbf{X'}$ **—** транспонированная матрица, полученная от матрицы $\mathbf{X}$.

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

Представим систему нормальных уравнений

$\mathbf{X'XA = X'У}$

в развернутом виде:

$$\left.\begin{aligned} &n \\ &a_0 + a_1 \sum X_i = \sum У_i \\ &a_0 \sum X_i + a_1 \sum X_i^2 = \sum X_i У_i \end{aligned}\right\} \qquad (2.2)$$

Решим систему уравнений и получим формулы расчета коэффициентов, представленных в скалярном виде:

$$a_1 = \frac{\sum (X_i - \overline{X})(У_i - \overline{У})}{\sum (X_i - \overline{X})^2}, \qquad (2.3)$$

$$a_0 = \overline{У} - a_1 \overline{X}, \qquad (2.4)$$

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

Для оценки параметров и характеристик условного математического ожидания У, при фиксированных значениях X воспользуемся выборочным регрессионным уравнением, коэффициенты и характеристики которого определим по выборочной совокупности объемом n. Это уравнение называется выборочной регрессионной моделью зависимости У от X:

$$У_i = a_0 + a_1X_i + e_i = У_{pi} + e_{i,}$$

где $У_i$ — исходные значения зависимой переменной;

X_i — фактор, влияющий на У;

i — порядковый номер измерения,

$У_{pi}$- расчетные значения зависимой переменной, отражают существующую связь между У и X,

$У_{pi} = a_0 + a_1X_i$; (2.5)

$e_i = (У_i — У_{pi})$;

e_i — остатки модели, отражают влияние неучтенных факторов, a_0, a_1 — коэффициенты выборочной регрессии, которые определим методом наименьших квадратов.

Приведем самый простой вариант расчета коэффициентов модели, который часто использовался до появления ЭВМ и предназначен только для понимания алгоритма проведения расчетов.

Определим коэффициенты парной регрессии с помощью решения системы нормальных уравнений методом наименьших квадратов (2.2).

Для расчетов воспользуемся табл. 2.3.

Таблица 2.3

Расчет коэффициентов парной регрессии

i	X_i	$У_i$	X_{i2}	$X_iУ_i$	Урi
1	1	4	1	4	4,2
2	3	6	9	18	7,6
3	2	7	4	14	5,9
4	4	10	16	40	9,3
Сумма	10	27	30	76	

Составим систему нормальных уравнений:

$n = 4$

$$\left.\begin{array}{l} 4a_0 \times 4 + a_1 \times 10 = 27 \\ a_0 \times 10 + a_1 \times 30 = 76 \end{array}\right\}$$

Определим a_0 из первого уравнения системы уравнений (2.6) и подставим его во второе.

$a_0 = (27 — a_1 10)/4$

$((27 — a_1 10)/4)10 + a_1\ 30 = 76$

Определим коэффициент a_1

$a_1 = (76 — 27 \times 10/4)/(30 — 10 \times 10/4) = 1,7.$

Подставим значение коэффициента a_1 в первое уравнение системы уравнений (2.11) и определим значение a_0.

$a_0 = (27 — 1,7 \times 10)/4 = 2,5.$

Выборочное линейное парное уравнение регрессии будет иметь следующий вид:

$У_i = a_0 + a_1X_i + e_i = 2,5 + 1,7X_i + e_i.$

Примечание. Расчеты, выполненные визуальным способом в первой лабораторной работе, совпадают с точными расчетами, выполненными методом наименьших квадратов.

Определим прогнозное значение $У_{пр}$ при ожидаемом значении $X_{ож} = 5$.

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

Вычислим расчетные значения $У_{pi}$ по формуле (2,5):

$$У_{pi} = a_0 + a_1 X_i = 2{,}5 + 1{,}7 X_i.$$

Численные значения $У_{pi}$ занесем в табл. 2.3

Построим график зависимости $У_i$ и $У_{pi}$ от X_i, а также покажем на графике расчетное значение У для генеральной совокупности(рис.2.7).

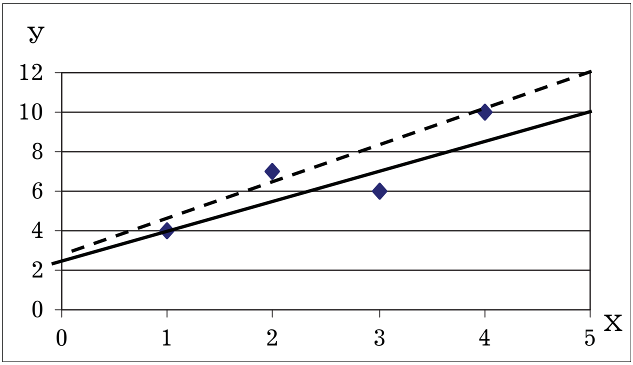


Рис. 2.7. График зависимости У, $У_р$ от Х для выборочной совокупности (сплошная линия) и генеральной совокупности (пунктирная линия)

Анализ рис. 2.7 показывает, что параметр α_0 и коэффициент a_0 модели соответственно для генеральной и выборочной совокупностей почти равны между собой.

Параметр α_1 и коэффициент a_1 модели соответственно для генеральной и выборочной совокупностей немного отличаются между собой.

Прогнозные значения товарооборота при количестве продавцов 5 чел. для генеральной и выборочной совокупностей, соответственно равные 12 и 11 тыс. руб., отличаются между собой.

Вывод. Для выборочной совокупности средний товарооборот, приходящийся на одного продавца равен 1,7 тыс. руб. При количестве продавцов, равном нулю (при работе магазина в режиме самообслуживания), товарооборот составит 2,5 тыс. руб. При ожидаемом значении количества продавцов, равном 5, точечное прогнозное значение товарооборота будет равно 11 тыс. руб.

Если увеличивать объем выборки, то можно ожидать, что выборочные характеристики будут приближаться к характеристикам генеральной совокупности.

Верификация модели

(Проверка достоверности коэффициентов модели и самой модели на обучающей совокупности. Проверка прогноза модели для контрольной совокупности.)

Этап верификации модели будет проводиться на следующем занятии.

Задание второго уровня сложности

Задание. Имеется база данных (табл. 2.4) зависимости величины розничного товарооборота У магазина от известных затратах на рекламу X.

Таблица 2.4

База данных

i	X_i	Уi
1	12	35
2	13	40
3	11	36
4	14	48
5	16	?

где X — расходы на рекламу (тыс. руб.);
У — величина розничного товарооборота (тыс. руб.);
i — номер магазина.

Необходимо:

— найти выборочные коэффициенты линейной парной регрессии визуальным способом и решением системы нормальных уравнений, определить прогнозное значение $У_{пр}$ при ожидаемом расходе на рекламу, равном $X_{ож} = 16$ тыс. руб.;
— получить расчетные значения $У_р$;
— построить график зависимости У, $У_р$ от X.

Пояснения

1. Скопируйте таблицу 2.4 и вставьте ее в лист Excel.

2. Постройте график зависимости У от Х.

3. Визуальным способом проведите линейную тенденцию (рис. 2.8).

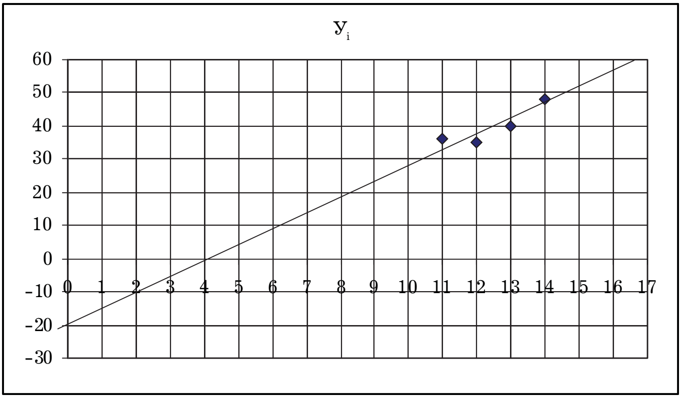


Рис. 2.8. Построение линейной тенденции визуальным способом

Примечание. На рис. 2.8 изменены форматы осей с помощью закладки “шкала” меню “формат оси”.

4. Визуальным способом определите коэффициенты линейной модели (2.1)

$$У_i = a_0 + a_1 X_i + e_i = -20 + 5X_i + e_i$$

5. Для составления системы нормальных уравнений (2.2) произведите необходимые расчеты и занесите их в табл. 2.5.

Таблица 2.5

Расчет элементов системы одновременных уравнений (2.2)

i	X_i	$У_i$	X_{i2}	$X_iУ_i$	$У_{pi}$
1	12	35	144	420	37,7
2	13	40	169	520	41,8
3	11	36	121	396	33,6
4	14	48	196	672	45,9
Сумма	50	159	630	2008	
5	16	?			54,1

6. Составим систему нормальных уравнений (2.2)

$$\left.\begin{aligned} &4a_0 + a_1 50 = 159 \\ &a_0\,50 + a_1 630 = 2008 \end{aligned}\right\}$$

7. Решим систему нормальных уравнений:

$a_0 = (159 — a_1 50)/4;$

$((159 — a_1 50)/4)50 + a_1\,630 = 2008;$

$a_1 = 4.1;$

$a_0 = (159 — 4.1{\times}50)/4 = — 11{,}5.$

$У_i = a_0 + a_1 X_i + e_i = -11{,}5 + 4{,}1X_i + e_i$

Сравнение значений коэффициентов a_0 и a_i, вычисленных визуальным способом и методом наименьших квадратов, показывает, что они отличаются. Однако для предварительного анализа точности визуального способа вполне достаточно.

8. Получим расчетные значения $У_{pi}$ по формуле (2.5)

$У_{pi} = a_0 + a_1 X_i = -11{,}5 + 4{,}1X_i$

$У_{p1} = -11{,}5 + 4{,}1{\times}12 = 37{,}7$

$У_{p2} = -11{,}5 + 4{,}1{\times}13 = 41{,}8$

$У_{p3} = -11{,}5 + 4{,}1{\times}11 = 33{,}6$

$У_{p4} = -11{,}5 + 4{,}\,1{\times}14 = 45{,}9$

9. Получим прогнозные значения У при $X_{ож} = 16$

$$У_{пр} = -11{,}5 + 4{,}1{\times}16 = 54{,}1$$

9. Построим график зависимости $У_р$ от X и прогнозное значение $У_{пр}$ при $X_{ож} = 16$ (рис. 2.9).

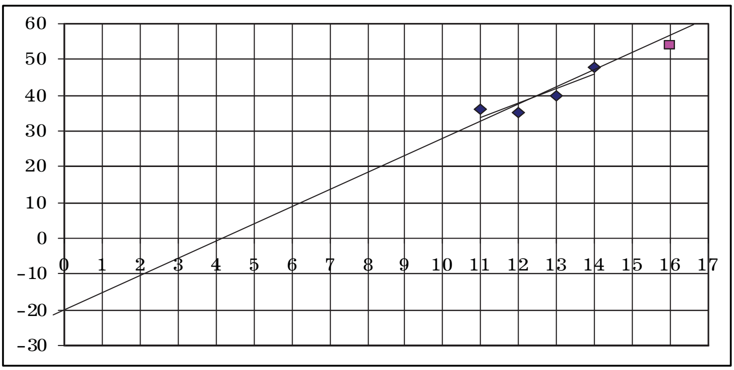


Рис. 2.9. График зависимости $У_р$ от X и прогнозное значение $У_{пр}$ при X = 16

Задание третьего уровня сложности

Задание. Произведите прогнозирование с помощью графических средств Excel.

Пояснения. Пример решения задания изучите по литературе [3, с. 171–177]. Пример решения задачи на рис. 2.10.

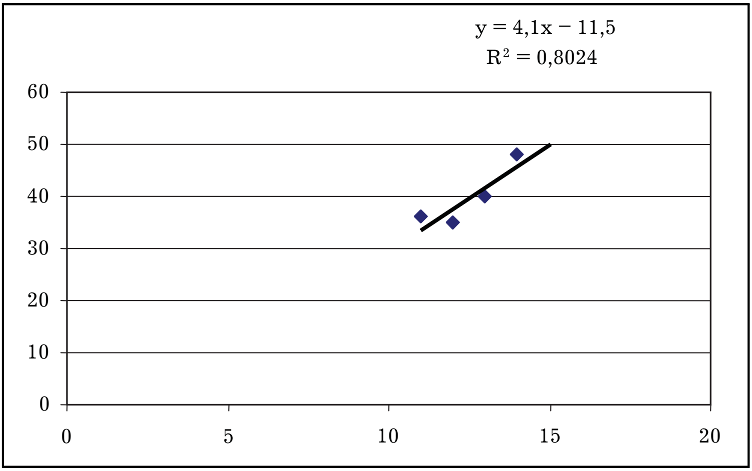


Рис. 2.10. Пример получения прогноза с помощью графических средств Excel.

Задание четвертого уровня сложности

Задание. Определите коэффициенты линейной модели с помощью программы Excel "Поиск решения".

Пояснения. Пример решения задачи изучите по литературе [3, с. 191–194].

Пройдите выходной тест

Исходные данные представлены в табл. 2.6. Необходимо в Excel построить график зависимости У от X и, не проводя линии тенденции, визуально определить коэффициенты линейной модели.

С помощью графических средств Excel проверьте правильность расчетов, выполненных визуальным способом.

Таблица 2.6

База данных выходного теста

i	$У_i$
1	12
2	15
3	18
4	19

3. РАСЧЕТ ХАРАКТЕРИСТИК ЛИНЕЙНОЙ РЕГРЕССИИ

Постановка задачи

Объект — экономический показатель.

Предмет — тенденция экономического показателя.

Цель — получить навыки проведения и анализа расчетов характеристик линейной тенденции экономического показателя.

Актуальность — полученные навыки проведения и анализа расчетов характеристик линейной тенденции экономического показателя можно будет использовать при исследовании любых тенденций.

Рабочая гипотеза — предполагается, что тенденция экономического показателя имеет линейный вид.

Метод — использование полуавтоматического метода проведения расчетов.

Способ — использование табличного процессора Excel.

Задачи:

- построить график временного ряда;
- произвести расчет характеристик линейной модели с использованием математических и статистических функций;
- проверить достоверность модели и ее коэффициентов;
- получить точечный и интервальный прогноз;
- представить результаты расчетов в графическом виде.

Ожидаемый результат — ожидается, что подтвердится гипотеза о том, что тенденция экономического показателя имеет линейный вид и можно будет получить точечный и интервальный прогноз, которые хорошо отражают свойства регулярности экономического показателя.

Расчеты для сравнения — автоматизированное проведение расчетов с помощью функции “Линейн”.

Задания первого уровня сложности

Задание 1. Подготовить сообщения по вопросам самостоятельной работы.

1. Изучите основные характеристики регрессионной модели: коэффициенты регрессионной модели, дисперсионный анализ регрессионной модели, ошибку модели, коэффициент множественной детерминации, ошибки коэффициентов модели, статистические критерии проверки достоверности модели и ее коэффициентов, доверительный интервал уравнения регрессии, точечный и интервальный прогноз.

2. Выясните возможности графического представления результатов эконометрического анализа.

3. Какие должны быть структура и состав отчета эконометрического анализа?

(3, с. 111–164; 7, с. 64–78; 1, с.17–107; 42–198; 5, с.34–176; 8, с. 27–81; 10, с. 13–200; 6, с.4–141; 12, с. 3–132, 165–196; 9, с. 231–241).

Задание 2. Пройдите входное тестирование

1. Подтвердите правильность обозначений на рис. 3.1.

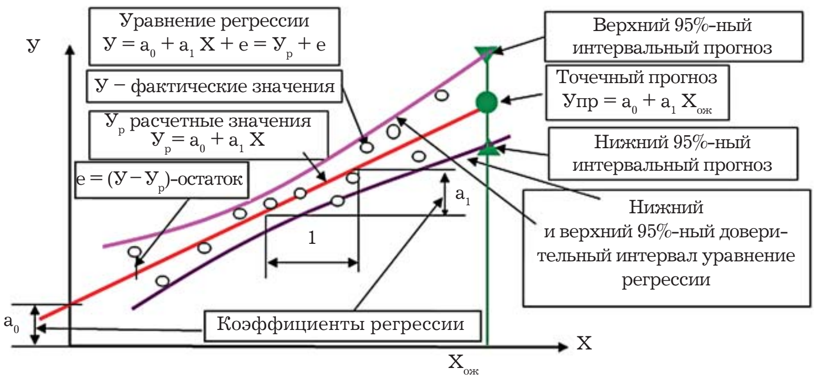


Рис. 3.1. Основные понятия и обозначения регрессионного анализа

а) да;
б) нет.

2. Подтвердите правильность расчетных формул.
Расчет коэффициентов регрессионного уравнения:
$У = a_0 + a_1X + e$

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

$a_0 = \overline{У} — a_1\overline{X}$
А = (Х'Х)-1 Х'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 $У_{pi} = a_0 + a_1X_i$.
Основное вариационное уравнение

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

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

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

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

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

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

$S_{рег}{}^2 = \dfrac{\sum(У_{pi} - \overline{У})^2}{k-1}$ — дисперсия регрессии.

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

Множественный коэффициент детерминации:

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

Критерий Фишера:

$$F = \frac{S^2_{рег}}{S^2_{ост}}.$$

Ошибка коэффициента a_0:

$$S_{a_0} = E\sqrt{\frac{\sum X_i^2}{n\sum(X_i - \overline{X})^2}}.$$

Ошибка коэффициента a_1:

$$S_{a_1} = E\sqrt{\frac{1}{\sum(X_i - \overline{X})^2}}.$$

Критерий Стьюдента для коэффициента a_1:

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

Частный коэффициент детерминации для фактора X_1:

$$r^2_{X_1 \bullet X_2, X_3} = \frac{t_1^2 R^2}{t_1^2 + t_2^2 + t_3^2},$$

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%-ный интервальный прогноз для математического ожидания У:

Задание 3. Решите задачу

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) (табл. 3.1). Визуальный анализ регулярностей этой зависимости показывает, что она имеет четко выраженную линейную тенденцию с однородными остатками.

Таблица 3.1

База данных

i	X_i	Уi
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 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.

3. Произвести эконометрический анализ линейной модели.

Решение задачи

Приводим графическое представление решения задачи на рис. 3.1.

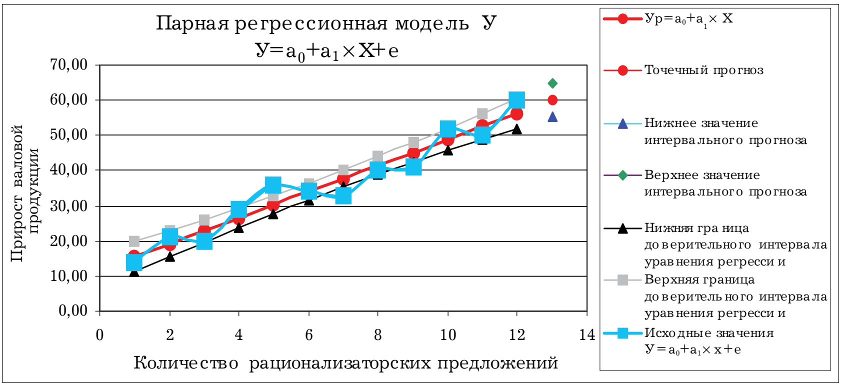


Рис. 3.1. Парная регрессионная модель прироста валовой продукции

На рис. 3.1 показаны результаты моделирования с получением точечных и интервальных значений уравнения регрессии и прогноза.

Постановка задачи

Цель моделирования

Целью моделирования является решение приоритетной проблемы, связанной с повышением валовой продукции предприятия по сравнению с лучшими аналогичными предприятиями. Методика выявления приоритетной проблемы изложена в "методе наименьших квадратов".

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) (рис. 3.2).

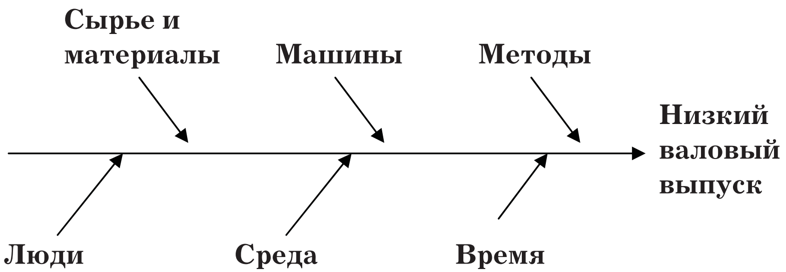


Рис. 3.2. Причинно-следственная диаграмма Исикавы.

Выделение входных и выходных переменных

Предприятие имеет множество характеристик, но в соответствии с возникшей проблемой, связанной с низким 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 подразделениях). Рационализаторские предложения направлены на совершенствование методов улучшение качества выпускаемой продукции за счет улучшения процессов. Премию работникам можно выдавать только за совершенствование процессов. Совершенствование процессов можно выполнить по ???????????????, изложенных в: Европейской модели идеального предприятия и международных стандартах системы менеджмента качества ИСО серии 9000: 2000.

Входными переменными будут количество поданных рационализаторских предложений, выходными показателями будет прирост валовой продукции.

Сбор статистических данных

Была произведена пространственная группировка по двум признакам: У — величина прироста валовой продукции и Х — количество поданных рационализаторских предложений. Исходные данные представлены в таблице 3.1. График зависимости У от Х изображен на рис. 3.3.

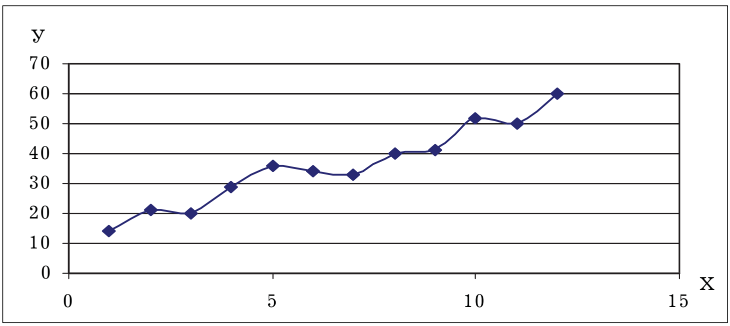


Рис. 3.3. График зависимости У от Х

Анализ рис.3.3 показывает, что с ростом числа поданных рационализаторских предложений происходит прирост валовой продукции предприятия. Однако эта зависимость является стохастической.

Выдвижение гипотез

Теоретически при увеличении числа рационализаторских предложений должна наблюдаться тенденция роста валовой продукции предприятия. Однако, прирост валовой продукции может происходить и в результате совершенствования процессов управления производством. Поэтому зависимость У от Х является стохастической.

Предположим, что прирост валовой продукции линейно зависит от количества поданных рационализаторских предложений.

Формулировка допущений

Прирост валовой продукции в большей степени зависит от количества поданных рационализаторских предложений, чем от других мер совершенствования процессов.

Спецификация модели

Предположим, что фактическая зависимость У от Х для всех предприятий генеральной совокупности можно представить в виде классической нормальной линейной модели парной регрессии, спецификация которой представлена соотношениями

$$У_i = f(X_i) + \varepsilon_i = \alpha_0 + \alpha_1 X_i + \varepsilon_i, \text{ при } i = 1, 2, ..., n,$$

где $f(X_i) = M_x(У) = M(У|X = X_i) = \alpha_0 + \alpha_1 X_i$ — функция зависимости У от Х.

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.

Идентификация модели

(Предложить и реализовать процедуру оценивания параметров модели).

Для оценки параметров и характеристик истинной зависимости У от Х воспользуемся регрессионным уравнением коэффициенты и характеристики которого определим по выборочной совокупности объемом n. Это уровнение называется выборочной регрессионной моделью зависимости У от Х:

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

где Уi — исходные значения зависимой переменной;

Xi — фактор, влияющий на У;

i — порядковый номер измерения.

$У_{pi} = a_0 + a_1 X_i$ — расчетные значения зависимой переменной, отражают существующую связь между У и Х;

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

a_0 — коэффициент модели, определенный методом наименьших квадратов по формуле

$$a_0 = \overline{У} + a_1 \overline{X},$$

численно равный значению $У_p$ при значении Х равного нулю, что следует из соотношения (2.1).

a_1 — коэффициент модели, определенный методом наименьших квадратов по формуле численно равный приросту значения **У** при изменении **Х** на единицу:

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

$\sum (X_i — \overline{X}) (У_i — \overline{У})$ — ковариация или сумма произведений отклонений значений **Х** и **У** от своих средних значений, отражает синхронность изменений У и Х;

$\overline{X}$ — среднее значение Х;

$\overline{У}$ — среднее значение У;

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

Модуль 1. Расчет коэффициентов a_1, a_0. Воспользуемся табл. 3.2.

Таблица 3.2

Расчет коэффициентов a_1, a_0

i	X_i	$У_i$	$X_i - \overline{X}$	$У_i - \overline{X}_у$	$(X_i - \overline{X})(Уi - \overline{У})$	$(Xi - \overline{X})^2$
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

$$a_1 = \frac{\sum (X_i - \overline{X})(У_i - \overline{У})}{\sum \left(X_i - \overline{X}\right)^2} = \boxed{3{,}7133},$$

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

$a_0 = \overline{У} - a_1 \overline{X}$ $\boxed{11{,}697}$,

где a_0 — коэффициент модели $У_i = a_0 + a_1X_i + e_i$.

$\overline{X} = \boxed{6{,}5}$ — среднее значение X,

$a_1 = \boxed{3{,}7133}$ — коэффициент модели $У_i = a_0 + a_1X_i + e_i$.

$\overline{У}$ $\boxed{35{,}833}$ — среднее значение У.

Верификация модели

(Проверка достоверности коэффициентов модели и самой модели на обучающей совокупности. Проверка прогноза модели для контрольной совокупности.).

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 ошибка модели 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. Воспользуемся табл. 3.3.

Таблица 3.3

Расчет ошибки модели

i	X_i	$У_i$	$У_{pi}$	$e_i = У_i — У_{pi}$	$(Уi — У_{pi})^2$
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

i	X_i	$У_i$	$У_{pi}$	$e_i = У_i — У_{pi}$	$(Уi — У_{pi})^2$
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

$$E = \sqrt{\frac{\sum (У_i - У_{pi})^2}{n - k}} = \boxed{3{,}5201},$$

где E — стандартная ошибка модели;

$У_{pi} = a_0 + a_1X_i$ — расчетные значения У.

$a_0 =$	11,697	— коэффициент a_0 модели $У_i = a_0 + a_1X_i + e_i$.
$a_1 =$	3,7133	— коэффициент a_1 модели $У_i = a_0 + a_1X_i + e_i$.
$\sum (У_i — У_{pi})^2 =$	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: $\alpha_1 = 0$.

Если $|t_{a1}| > t_{\alpha/2}(\alpha = 0{,}05;\ m = n — k)$, то нулевая гипотеза H_0 отвергается с вероятностью $1 — \alpha = 1 — 0{,}05 = 0{,}95$ 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 S_{a1};

$$t_{a_0} = \frac{a_0}{S_{a_0}}; \quad t_{a_1} = \frac{a_1}{S_{a_1}},$$

где: S_{a0} — ошибка коэффициента a_0 или стандартное отклонение коэффициента a_0, равное среднеквадратическому отклонению коэффициента a_0 от своего математического ожидания α_0

$$S_{a_0} = E\sqrt{\frac{\sum X_i^2}{n\sum(X_i - \overline{X})^2}};$$

S_{a1} — ошибка коэффициента a_1, или стандартное отклонение коэффициента a_1, равное среднеквадратическому отклонению коэффициента a_1 от своего математического ожидания α_1.

$$S_{a_1} = E\sqrt{\frac{1}{\sum(X_i - \overline{X})^2}}$$

$t_{\alpha/2}(\alpha = 0,05;\ m = n - k)$ — вычисляется по функции СТЬЮДРАСПОБР.

Модуль 3. Расчет ошибок коэффициентов a_1 и a_0.

Таблица 3.4

Расчет ошибок коэффициентов a_1 и a_0.

i	X_i	X_{i2}	$(X_i - \overline{X})^2$
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

i	X_i	X_{i2}	$(X_i - \overline{X})^2$
10	10	100	12,25
11	11	121	20,25
12	12	144	30,25
Сумма		650	143
Среднее	6,5		

$$S_{a_1} = E\sqrt{\frac{1}{\sum\left(X_i - \overline{X}\right)^2}} = \boxed{0{,}2944},$$

где S_{a1} — ошибка коэффициента a_1 или среднее квадратическое отклонение коэффициента a_1 от параметра $\alpha 1$, рассчитанного для всех объектов генеральной совокупности.

$E = 3{,}5201$ — ошибка модели;

$\sum (X_i - \overline{X})^2 = 143$ — вариация переменной X.

$$S_{a0} = E\sqrt{\frac{\sum\left(X_i\right)^2}{n\sum\left(X_i - \overline{X}\right)^2}} = \boxed{2{,}1665},$$

где S_{a0} — ошибка коэффициента a_0 или среднее квадратическое отклонение коэффициента a_0 от параметра, рассчитанного для всех объектов генеральной совокупности.

$E = 3{,}5201$ — ошибка модели;

$\sum (X_i - \overline{X})^2 = 143$ — вариация переменной X;

$\sum (X_i)^2 = 650$ — сумма квадратов переменной X;

$n = 12$ — объем выборки.

Модуль 4. Вычисление критерия Стьюдента для коэффициентов a_1 и a_0.

$$t_{a_1} = \frac{a_1}{S_{a_1}} = \boxed{12{,}615},$$

Где t_{a1} — критерий Стьюдента для коэффициента a_1.

a_1 =	3,7133	— коэффициент модели $У_i = a_0 + a_1X_i + e_i$.
S_{a1} =	0,2944	— среднее квадратическое отклонение коэффициента a_1

$$t_{a_0} = \frac{a_0}{S_{a_0}} = \boxed{5{,}3991},$$

Где t_{a0}- критерий Стьюдента для коэффициента a_0.

a_0 =	11,697	— коэффициент модели $У_i = a_0 + a_1X_i + e_i$.
S_{a0} =	2,1665	— среднее квадратическое отклонение коэффициента a_0

Модуль 5. Определение критического значения критерия Стьюдента на уровне значимости α = 0,05 с помощью статистической функции СТЬЮДРАСПОБР.

$t_{\alpha/2}(\alpha = 0.05,\ m = n\text{-}k) =$		2,2281	— критическое значение критерия Стьюдента;
где	n =	12	— объем выборки,
	k =	2	— количество коэффициентов в модели;
	α =	0,05	— уровень значимости критерия, или вероятность совершить ошибку при отклонении нулевой гипотезы.

Модуль 6. Проверка достоверности коэффициента a_1.

Проверим нулевую гипотезу H_0: “$\alpha_1 = 0$”.

Так как $|t_{a1}| = 12{,}615 > t_{\alpha/2}(\alpha = 0.05;\ m = n\text{-}k) = 2{,}2281$

то нулевая гипотеза H_0: “ $\alpha_1 = 0$ “ отвергается с вероятностью 1 — α = 0,95

Модуль 7. Вычисление доли объясненной вариации переменной У. Воспользуемся табл. 3.5.

Таблица 3.5

Расчет коэффициента детерминации и критерия Фишера

i	$У_i$	$У_{pi}$	$(У_i — \overline{У})^2$	$(У_{pi} — \overline{У})^2$	$(У_i — У_{pi})^2$
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			

Определение доли объясненной вариации.

R^2- коэффициент детерминации равен доли общего варьирования, которое объясняется за счет уравнения регрессии.

$$R^2 = \frac{\sum(У_{pi} - \overline{У})^2}{\sum(У_i - \overline{У})^2}.$$

Расчет коэффициент детерминации

$$R^2 = \frac{\sum\left(У_{pi} - \overline{У}\right)^2}{\sum\left(У_i - \overline{У}\right)^2} = \boxed{0,9409},$$

Где R^2 — коэффициент детерминации.

$\sum(У_{pi} — \overline{У})^2 = \boxed{1971,8}$ — вариация регрессии;

$\sum(У_i — \overline{У})^2 = \boxed{2095,7}$ — вариация переменной У.

Проверка достоверности модели.

Выдвигается нулевая гипотеза H_0: “Модель недостоверна”, или “$У_{pi} = \bar{У}$”, или "$S^2_{рег} = S^2_{ост}$"

Если $F > F_{кр}(\alpha = 0{,}05;\ m_1 = k\text{-}1;\ m_2 = n\text{-}k)$, то нулевая гипотеза H_0 отвергается с вероятностью совершить ошибку, равную $\alpha = 0{,}05$, или нулевая гипотеза H_0 отвергается с вероятностью $1\text{-}\ \alpha = 1\text{-}\ 0{,}05 = 0{,}95$ и утверждается, что модель является достоверной с вероятностью 0,95.

Если $F < F_{кр}(\alpha = 0{,}05;\ m_1 = k\text{-}1;\ m_2 = n\text{-}k)$, то нулевая гипотеза H_0 принимается и утверждается, что модель является недостоверной.

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 E%.

$$F = \frac{S^2_{рег}}{S^2_{ост}}.$$

$S_{рег}{}^2$ — дисперсия, обусловленная регрессией или сумма квадратов отклонений расчетных значений У от своего среднего значения, деленная на число степеней свободы,

$$S^2_{рег} = \frac{\sum (У_{pi} - \bar{У})^2}{k - 1}.$$

$S_{ост}{}^2$ — дисперсия остатков или сумма квадратов отклонения фактических от расчетных значений переменной У, деленная на число степеней свободы,

$$S^2_{ост} = \frac{\sum (У_i - У_{pi})^2}{n - k}.$$

$F_{кр}(\alpha = 0{,}05;\ m_1 = k\text{-}1;\ m_2 = n\text{-}k)$ — критическое значение критерия Фишера, определенное на уровне значимости α = 0.05, числе степеней свободы для большей дисперсии m_1 и числе степеней свободы для меньшей дисперсии m_2. $F_{кр}$ определяется по функции Excel FРАСПОБР.

α — уровень значимости критерия, или вероятность совершить ошибку при отклонении нулевой гипотезы.

Модуль 8. Вычисление критерия Фишера.

$$F = \frac{S_{рег}^2}{S_{ост}^2} = \frac{(n-k)\sum\left(У_{pi} - \overline{У}\right)^2}{(k-1)\sum(У_i - У_{pi})^2} = \boxed{159{,}13}$$

Где F- критерий Фишера.

$\sum(У_{pi} — \overline{У})^2 = \boxed{1971{,}8}$ — вариация регрессии,

$\sum(У_i — \overline{У}_{pt})^2 = \boxed{123{,}91}$ — вариация остатков,

$n = \boxed{12}$ — объем выборки,

$k = \boxed{2}$ — количество коэффициентов в модели,

$S_{рег}^2$ — дисперсия регрессии,

$S_{ост}^2$ — дисперсия остатков.

Модуль 9. Расчет табличного критерия Фишера.

Вычислим критическое 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

$F_{кр}(\alpha = 0{,}05,\ m_1 = k\text{-}1;\ m_2 = n\text{-}k) = 4{,}9646$

$n = \boxed{12}$ — объем выборки,

$k = \boxed{2}$ — количество коэффициентов в уравнении.

Модуль 10. Проверка достоверности модели.

Проверим нулевую гипотезу H_0: “Модель недостоверна”, или “$У_p = \overline{У}$”, или “$S_{рег}^{\ 2} = S_{ост}^{\ 2}$”,

где $S_{рег}^{\ 2}$- дисперсия регрессии,

$S_{ост}^{\ 2}$- дисперсия остатков.

Так как $F = 159{,}13 > F_{кр}(\alpha = 0.05,\ m_1 = k\text{-}1;\ m_2 = n\text{-}k) = \boxed{4{,}9646}$

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

$$У_{мi} = У_{пр} \pm t_{\alpha/2}\, S_{УМ},$$

где $У_{мi}$ — математическое ожидание У для значения $X_{ож}$;

$S_{Ум}$ — стандартная ошибка прогноза математического ожидания У, равная среднеквадратическому отклонению между точечным прогнозом и математическим ожиданием У:

$$S_{Ум} = E\sqrt{\frac{1}{n} + \frac{(X_{ож} - \overline{X})^2}{\sum (X_i - \overline{X})^2}};$$

$t_{\alpha/2}(\alpha = 0.05,\ m = n — k)$- критическое значение двухстороннего критерия Стьюдента.

95%-ный прогнозный доверительный интервал нахождения прогнозного индивидуального значения У определяется по формуле

$$У_{и} = У_{пр} \pm t_{\alpha/2}\, S_{УИ},$$

где $У_{и}$ — индивидуальное значение У для значения $X_{ож}$;

$S_{Уи}$- стандартная ошибка прогноза индивидуального значения У, равная среднеквадратическому отклонению между точечным прогнозом и индивидуальным значением У:

$$S_{Уи} = E\sqrt{1 + \frac{1}{n} + \frac{(X_{ож} - \overline{X})^2}{\sum (X_i - \overline{X})^2}}.$$

Получение доверительных интервалов уравнения регрессии.

95%-ные доверительные интервалы математических ожиданий $У_i$ для значений X_i вычисляются по формулам

$У_{min(м)i} = У_{pi} — t_{\alpha/2}\, S_{УМi}$ — нижний 95%-ный доверительный интервал для математического ожидания У,

$У_{min(м)i} = У_{pi} — t_{\alpha/2}\, S_{УМi}$ — верхний 95%-ный доверительный интервал для математического ожидания У,

где $S_{Умi}$ — стандартная ошибка расчетного значения $У_{pi}$ для значения X_i по отношению к математическому ожиданию значения $У_i$:

$$S_{Уиi} = E\sqrt{\frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum (X_i - \overline{X})^2}}.$$

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, будет находиться в интервале от $У_{min(м)i}$ до $У_{max(м)i}$.

95%-ные доверительные интервалы индивидуальных значений $У_i$ для значений X_i вычисляются по формулам

$У_{min(и)i} = У_{pi} — t_{\alpha/2}\, S_{УИi}$ — нижний 95%-ный доверительный интервал для индивидуального значения У,

$У_{max(и)i} = У_{pi} — t_{\alpha/2}\, S_{УИi}$ — верхний 95%-ный доверительный интервал для индивидуального значения У.

где: $S_{Уиi}$ — стандартная ошибка расчетного значения $Ур_i$ для значения X_i по отношению к индивидуальному значению $У_i$,

$$S_{Уиi} = E\sqrt{1 + \frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum(X_i - \overline{X})^2}}.$$

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, $X_{ож} = 13$.

Получим точечное прогнозное значение $У_{прв}$ для выборочной совокупности при ожидаемом значении $X_{ож}$.

$У_{прв} = a_0 + a_1X_{ож} =$	59,97	
где $a_O =$	11,697	— коэффициент a_0 модели $У_i = a_0 + a_1X_i + e_i$,
$a_1 =$	3,7133	— коэффициент a_1 модели $У_i = a_0 + a_1X_i + e_i$,
$X_{ож} =$	13	— ожидаемое значение X.

Модуль 11. Рассчитаем 95%-ный доверительный интервальный прогноз нахождения прогнозного математического ожидания (среднего значения).

$У_{прг}$ по всем предприятиям отрасли (для всей генеральной совокупности) при *ожидаемом* значении $X_{ож}$, *который вычисляется по формуле*

$$У_{прг} = У_{пр} \pm t_{таб} E\sqrt{\frac{1}{n} + \frac{\left(X_{ож} - \overline{X}\right)^2}{\sum(X_i - \overline{X})^2}}.$$

$У_{прг} =$ 59,97 ± 4,8272 — (1-α = 0,05) 100 = 95%-ный доверительный интервал прогноза математического ожидания У.

где $У_{прг}$ — прогнозное значение У для генеральной совокупности, или математическое ожидание У,

$У_{прв} = a_0 + a_1 X_{ож} =$	59,97	— прогнозное значение У для выборочной совокупности;
$X_{ож} =$	13	— ожидаемое значение X
$t_{\alpha/2}(\alpha = 0{,}05, m = n - k) =$	2,2281	— критическое значение критерия Стьюдента;
$n =$	12	— объем выборки,
$k =$	2	— количество коэффициентов в модели;
$E =$	3,5201	— ошибка модели,
$\sum (X_i - \overline{X})^2 =$	143	— вариация переменной X,
$\overline{X} =$	6,5	— среднее значение X,

$$t_{\alpha/2} E \sqrt{\frac{1}{n} + \frac{\left(X_{ож} - \overline{X}\right)^2}{\sum (X_i - \overline{X})^2}} = \boxed{4{,}8272} \text{ — ошибка прогноза}$$

Модуль 12. Вывод по определению интервального прогноза.

С вероятностью 95% можно утверждать, что среднее значение объема производства для генеральной совокупности ($У_{прг}$) при ожидаемом значении $X_{ож} = 13$ будет находиться в интервале от 55,142 до 64,797.

Модуль 13. Рассчитаем 95%-ный 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

$$У_{мi} = У_{pi} \pm t_{таб} E \sqrt{\frac{1}{n} + \frac{\left(X_i - \overline{X}\right)^2}{\sum (X_i - \overline{X})^2}}$$

где $У_{мi}$ — математическое ожидание У для i- ого измерения,
$У_{pi} = a_0 + a_1 X_i$ — расчетное значение У для i -ого измерения;

X_i		— значения X при i -м измерении;
$t_{\alpha/2}(\alpha = 0.05,\ m = n\text{-}k) =$	2,2281	— критическое значение критерия Стьюдента;
n =	12	— объем выборки;
K =	2	— количество коэффициентов в модели:
E =	3,5201	— ошибка модели;
$\sum (X_i - \overline{X})^2 =$	143	— вариация переменной X;
$\overline{X} =$	6,5	— среднее значение X.

Таблица 3.6

Расчет 95%-го доверительного интервала математического ожидания У

i	X_i	$У_{pi}$	$У_{мmin(i)}$	$У_{мmax(i)}$
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 $У_{мmin(i)}$ — 95% нижний доверительный интервал математических ожиданий У,

$У_{мmax(i)}$ — 95% верхний доверительный интервал математических ожиданий У.

Модуль 14. Произведем эконометрический анализ линейной модели

Линейная модель экономического процесса имеет следующий вид:

$У_i = a_0 + a_1 X_i + e_i$

$У_i =$	11,697	+	3,7133	$X_i + e_i$		
$S_{a0} =$	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 $У_i = a_0 + a_1X_i + e_i$,

S_{a0} — ошибка коэффициента a_0,

S_{a1} — 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,

$t_{\alpha/2}$ — критическое значение критерия Стьюдента,

$F_{кр}$ — критическое значение критерия Фишера.

Проверка достоверности модели.

Проверим нулевую гипотезу H_0: “модель недостоверна”, или “$У_р = \bar{У}$”, или “$S_{рег}^2 = S_{ост}^2$”,

где $S_{рег}^2$- дисперсия регрессии;

$S_{ост}^2$- дисперсия остаточная.

Так как F =	159,13	>	$F_{кр}$(α = 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

$У_i\ a_0 + a_1X + e_i =$ $\boxed{11{,}697}$ + $\boxed{3{,}7133}$ $X_i + e_i$

построенного по выборочной совокупности равной 12 месяцам, является достоверной с вероятностью 0,95

Проверка достоверности коэффициента a_1.

Проверим нулевую гипотезу H_0: "$\alpha_1 = 0$".

Так как $|t_{a1}| =$ $\boxed{12{,}615}$ $\boxed{>}$ $t_{\alpha/2}(\alpha = 0{,}05; m = n - k) =$ $\boxed{2{,}2281}$

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 $X_i + e_i$

и является достоверной с вероятностью 0,95

Коэффициент $a_1 =$ 3,7133 достоверно отличается от нуля

В прогнозном периоде в январе следующего года при $X_{ож} = 13$ точечный прогноз, или наиболее вероятное прогнозное значение прироста объема производства, будет равно $У_{прв} = 59{,}97$

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 (рис 3.4).

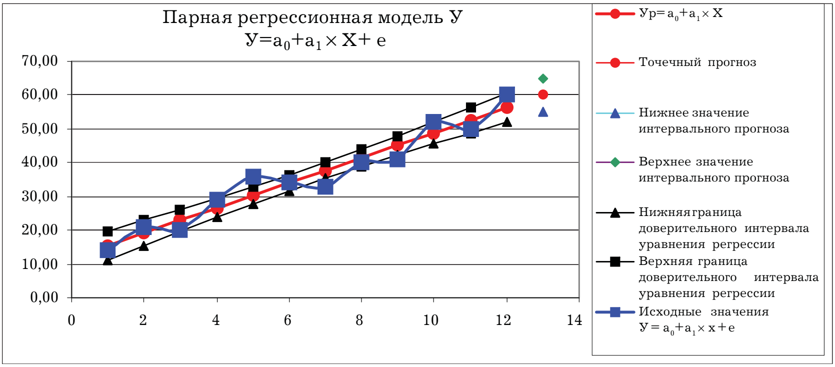


Рис. 3.4. График полученных расчетов

Модуль 16. Проверка расчетов (функция "Линейн").

Произведем проверку всех расчетов с использованием статистической функции "Линейн" в среде Excel 97, которая выполняется в следующей последовательности:

— в рабочую ячейку введите функцию Линейн с указанием диапазона ячеек для зависимой переменной У и объясняемой переменной X, введите условие вывода всех характеристик функции "Линейн";

— выделить два столбца и пять строчек, F2, Ctrl+Shift+ +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

— результаты расчетов коэффициентов и всех характеристик линейной модели позволяют произвести эконометрический анализ регрессионной модели.

$У = 11{,}697 + 3{,}7133\,X$
$S_{a0} = 2{,}1665$, $S_{a1} = 0{,}2944$
$t_{a0} = 5{,}3991$ $t_{a1} = 12{,}615$
$E = 3{,}5201$
$R^2 = 0{,}9409$
$F = 159{,}13$

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 следующей последовательности.

Шаг 1. Разделим всю выборочную совокупностиь на две части — обучающую (табл. 3.7), которая составит 2/3 объема выборки (табл. 3.8)и контрольную, которая составит 1/3 объема выборки.

Таблица 3.7

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 3.8

Контрольная выборка

i	X_i	$У_i$
1	9	41
2	10	52
3	11	50
4	12	60

Шаг 2. Вычислим коэффициенты ошибки 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

$У_р = 12,78 + 3,464X$
$E = 3,507$

Шаг 3. Вычислим ошибку модели на контрольной выборке.
$E = 5,58$

Шаг 4. Вычислим критерий Фишера для контрольной выборки.
$F = 2,2$; $F_{кр}(0,05;\ m_1 = k — 1 = 2-1;\ m_2 = n — k = 4-2) = 18,51$

Шаг 5. Проверим достоверность модели по контрольной выборке.

Обучающая выборка — Расчеты по функции “Линейн”.

i	Xi	Уобуч	Уробуч	3,464285714	12,78571429
1	1	14	16,25	0,54124133	2,733134741
2	2	21	19,71429	0,872253639	3,507644712
3	3	20	23,17857	40,96806967	6
4	4	29	26,64286	504,0535714	73,82142857
5	5	36	30,10714		
6	6	34	33,57143		
7	7	33	37,03571		
8	8	40	40,5		
F =	40,96807		Fкр(0,05; к-1 = 2-1; n-k = 8-2) =		5,987374152
E =	3,507645				

Контрольная выборка

i	Xi	Укон	УрКон	(Укон-УрКон)2	(УрКон-$\overline{У}$)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			
Sост2 = ($\sum$(Укон-УрКон)2)/ (n-k) =				31,1619898	
Sper2 = ($\sum$(УрКон-$\overline{У}$)2)/(k-1) =					70,10969388
F = Sper2/Sост2 =		2,2498	Fкр(0,05; m1 = k-1 = 2-1; m2 = n-k = 4-2) =		18,51276465
E = корень(Sост2) =		5,5822			

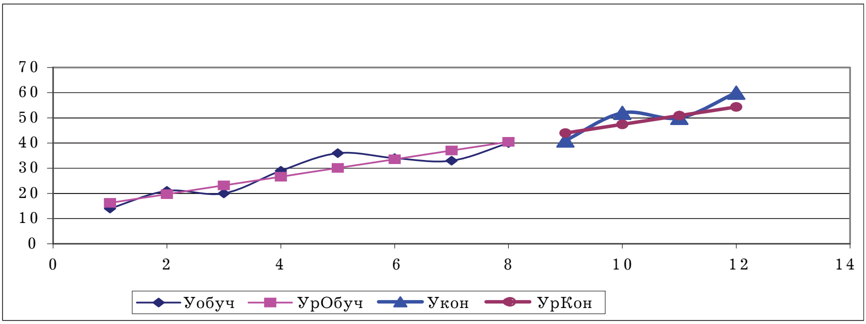


Рис. 3.5. График проверки достоверности модели на контрольной совокупности

Анализ рис. 3.5 показывает, что остатки модели на обучающей и контрольной совокупности имеют одинаковые значения. Для предварительного анализа этого достаточно, чтобы утверждать о хорошем воспроизведении исходных данных контрольной совокупности с помощью модели, построенной по обучающей 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 $F = 40.9 > F_{кр}(0{,}05; m_1 = k — 1 = 2-1; m_2 = n — k = 8\text{-}2) = 5{,}98$. 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 объясняется малым объемом 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.

Выходное тестирование

1. Подтвердите правильность обозначений линейной модели, изображенной на рис. 3.6.

$У_i = a_0 + a_1 X_i + e_i$

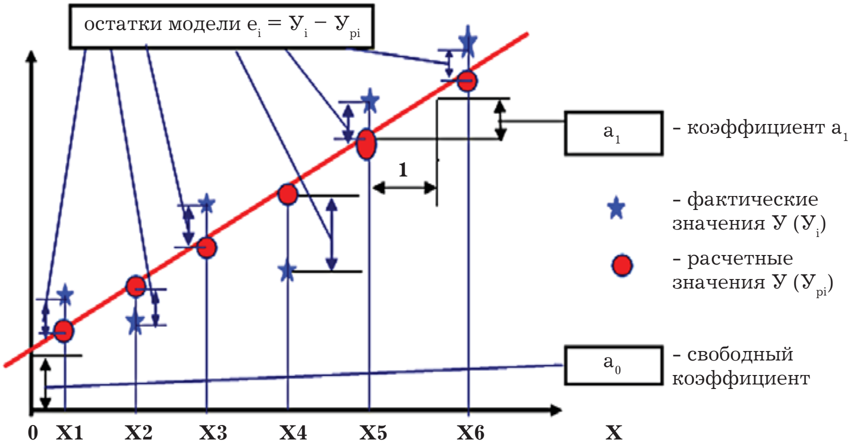


Рис. 3.6. Условные обозначения регрессионной модели

а) да;
б) нет.

4. НЕЛИНЕЙНЫЕ МОДЕЛИ РЕГРЕССИИ И ИХ ЛИНЕАРИЗАЦИЯ. СПЕЦИФИКАЦИЯ МОДЕЛИ

Постановка задачи

Объект — зависимости между экономическими показателями.

Предмет — тенденции зависимости между двумя переменными.

Цель — получить навыки определения математической функции, которая хорошо отражает тенденцию зависимости между двумя переменными.

Актуальность — полученные навыки проведения и анализа расчетов характеристик по любым тенденциям зависимости между двумя переменными можно будет использовать при исследовании любых переменных.

Метод — перебор основных математических функций.

Способ — сравнение характеристик моделей связи между двумя переменными.

Задачи — построить график зависимости между двумя переменными,

- произвести анализ регулярностей связи между двумя переменными,
- произвести расчет характеристик моделей с использованием различных тенденций,
- проверить достоверность модели и ее коэффициентов,
- сравнить характеристики моделей для определения лучшей.

Рабочая гипотеза — предполагается, что зависимость между двумя переменными имеет выбранную тенденцию.

Ожидаемый результат — ожидается, что подтвердится гипотеза о том, что

зависимость между двумя переменными будет иметь выбранную тенденцию.

Метод для сравнения — автоматизированное проведение расчетов с помощью пакетов прикладных программ.

Задания первого уровня сложности

Задание 1. Подготовьте краткие ответы на вопросы самостоятельной подготовки.

1. Приведите примеры моделей линейных по переменным и коэффициентам.
2. Приведите примеры нелинейных моделей по переменным и коэффициентам.
3. Расскажите правила линеаризации модели по переменным и коэффициентам.
4. Приведите примеры линеаризации модели по переменным и коэффициентам. [3, с. 240–243, 7, с.124–132, 1 с. 172–186, 5, с. 62–86, 10, с.200–229]

Задание 2. Пройдите входное тестирование.

Выберите вариант правильного ответа

1. Регрессионная модель является линейной относительно коэффициентов, если выполняются четыре условия:

— коэффициенты модели находятся в первой степени;
— переменные могут быть в любой степени и иметь любые математические преобразования;
— коэффициенты модели не являются степенью по отношению к другим коэффициентам или переменной;
— модель должна быть аддитивной.

Если условия линейности нарушаются, то она называется нелинейной относительно коэффициентов

а) да;
б) нет.

2. Регрессионная модель является линейной относительно переменных, если выполняются условия:

— переменные находятся в первой степени;

— модель должна быть аддитивной.

а) да;

б) нет.

3. Регрессионная модель является нелинейной относительно переменных, если выполняются условия:

— переменные возводятся в степень отличной от первой. В качестве степени может быть число, или коэффициент, или другая переменная;

— модель должна быть аддитивной.

а) да;

б) нет.

4. Регрессионная модель является нелинейной относительно связи между факторами, если выполняются условие: связь между факторами является мультипликативной.

а) да;

б) нет.

5. Метод наименьших квадратов применим к линейным относительно коэффициентов аддитивным регрессионным уравнениям следующего вида:

$$У = a_0 + a_1X_1 + a_2X_2 + ...+ a_кX_к + e,$$

где — коэффициенты регрессии находятся в первой степени,

— переменные могут быть в любой степени и иметь любые математические преобразования,

— коэффициенты модели не являются степенью по отношению к другим коэффициентам или переменной.

а) да;

б) нет.

6. Если нельзя рассчитать коэффициенты модели с помощью метода наименьших квадратов, то есть три направления решения этой проблемы:

— выводятся специальные формулы расчетов коэффициентов уравнения регрессии;

— используются приближенные итеративные методы расчетов коэффициентов;
— нелинейную модель преобразовывают к линейному аддитивному виду.

а) да;
б) нет.

7. Первое направление требует много времени и изобретательности.

Примечание. Особенно много хлопот доставила логистическая функция относительно оценки предела, к которому она стремится. Логистическая функция активно применяется для описания процессов, стремящихся к насыщению.

Второе направление предполагает использование итеративных программ:
— "Эврика" — программа, которая с помощью итеративных методов может рассчитать коэффициенты любых функций;
— программа "Поиск решения", имеющаяся в Excel, которая позволяет вычислить коэффициенты любых функций;
— специализированная программа TableCurve2D, которая позволяют рассчитать коэффициенты более 2500 различных функций.

Третье направление предполагает приведение нелинейной модели к линейному виду с помощью логарифмирования и замены переменной.

а) да;
б) нет.

8. Линеаризация — процесс преобразования функции к линейному аддитивную виду.

а) да;
б) нет.

9. Известны два метода линеаризации: логарифмирование функции и замена преобразованной переменной. Оба эти метода представлены в табл. 4.1.

Таблица 4.1

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	$У_p = a_0e^{(a1*X)}$	Экспоненциальная 2	$Z_1 = \ln У$ $Z_2 = \ln a0$	$Z_1 = Z_2+a_1X$
8	$У_p = e^{(a0+a1X)}$	Экспоненциальная 3	$Z_1 = \ln У$	$Z_1 = a_0+a_1X$
9	$У_p = a_0X^{a1}$	Экспоненциальная 4. Кривая Энгеля	$Z_1 = \ln У$ $Z_2 = \ln X$	$Z_1 = a_0+a_1Z_2$
10	$У_p = e^{(a0+a1/X)}$	Экспоненциальная 5. S образная	$Z_1 = \ln У$ $Z_2 = 1/X$	$Z_1 = a_0+a_1Z_2$
11	$У_p = a_0 + a_1e^X$	Экспоненциальная 6	$Z = e^X$	$У_p = a_0+a_1Z$
12	$У_p = a_0e^{(a1/X)}$	Экспоненциальная 7	$Z_1 = \ln У$ $Z_2 = \ln a_0$ $Z_3 = 1/X$	$Z_1 = Z_2+a_1Z_3$
13	$У_p = 1/(a_0+a_1e^{-X})$	Обратная экспоненциальная 8	$Z_1 = 1/У$ $Z_2 = e^{-X}$	$Z_1 = a_0+a_1Z_2$
14	$У_p = a_0X^{\,a1X}$	Степенная 1	$Z_1 = \ln У$ $Z_2 = \ln a_0$ $Z_3 = X\ln X$	$Z_1 = Z_2+a_1Z_3$
15	$У_p = a_0+a_1X^{(1/2)}$	Степенная 2	$Z = X^{1/2}$	$У_p = a_0+a_1Z$
16	$У_p = 1/(a_0+a_1X)$	Гипербола 2	$Z = 1/У$	$Z = a_0+a_1X$
17	$У_p = X/(a_0+a_1X)$	Гипербола 3	$Z = X/У$	$Z = a_0+a_1X$
18	$У_p = a_0X_1^{a1}X_2^{a2}$	Кобба — Дугласа	$Z_1 = \ln У$ $Z_2 = \ln a_0$ $Z_3 = \ln X_1$ $Z_4 = \ln X_2$	$Z_1 = Z_2+a_1Z_3+a_2Z_4$

а) да;

б) нет.

10. Расчет коэффициентов степенной функции следующего вида: $У_p = a_0X^{a_1X}$ с помощью функции Excel "Линейн".

Шаг 1. Линеаризуем степенную функцию к виду

$Z_1 = Z_2 + a_1Z_3$,

где $Z_1 = \ln У$, $Z_2 = \ln a_0$, $Z_3 = X\ln X$.

Шаг 2. Создадим базу данных значений X и У, в среде Excel введем значения X, У в таблицу, добавим к исходной базе данных две колонки lnУ и XlnX, в которые введем преобразованные значения.

Шаг 3. Выполним функцию "Линейн", в которой зависимой переменной будут значения колонки lnX, объясняемой переменной — значения колонки XlnX. Функция "Линейн" позволяет получить коэффициенты Z_2, a_1 модели

$Z_1 = Z_2+a_1Z_3$.

Шаг 4. Вычислим коэффициенты 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 $a_0 = e^{Z2}$.

Шаг 5. Вычислим характеристики модели $У_p = a_0X^{a1X}$ 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.

а) да;

б) нет.

Задание 3. Изучите решение задачи.

Была произведена группировка магазинов пространственной выборки за летний месяц по двум показателям: количество продавцов и величина месячного товарооборота. Данные группировки представлены в табл. 4.2.

Таблица 4.2

База данных нескольких магазинов

i	X_i	$У_i$	$У_i/X_i$
1	1	0,95	0,95
2	2	1,8	0,9
3	3	2,1	0,7
4	4	2,9	0,725
5	5	4,2	0,84
6	6	6	1
7	7	6,7	0,9571
8	8	6,5	0,8125
9	9	6,8	0,7556
10	10	6,75	0,675
11	11	6,9	0,6273
Ожидаем	12	?	

где X — количество продавцов (чел.),
У — розничный товарооборот (тыс. руб.)
Необходимо:
— получить прогноз розничного товарооборота при ожидаемом значении количества продавцов.
— определить оптимальное значение количества продавцов, при котором их производительность труда будет максимальной.

Решение

Целью моделирования является решение приоритетной проблемы, связанной с низкой производительностью труда продавцов по сравнению с лучшими аналогичными магазинами. Методика выявления приоритетной проблемы изложена во второй теме.

Выбор переменных

Величина товарооборота зависит от шести основных факторов: *сырье и материалы* (ассортимент товаров и их качество); *машины* (оборудование и дизайн магазина); *методы* (форма обслуживания покупателей: форма самообслуживания, культура обслуживания); *люди* (количество продавцов, уровень квалификации, образование, возраст, деловые характеристики); *физическая и абстрактная среда* (температура, влажность в магазине; традиции, психологический климат существующие в магазине; правовая среда существования магазина, информационное обеспечение магазина); *время* (нет влияния, так как получена пространственная выборка) (рис. 4.1).

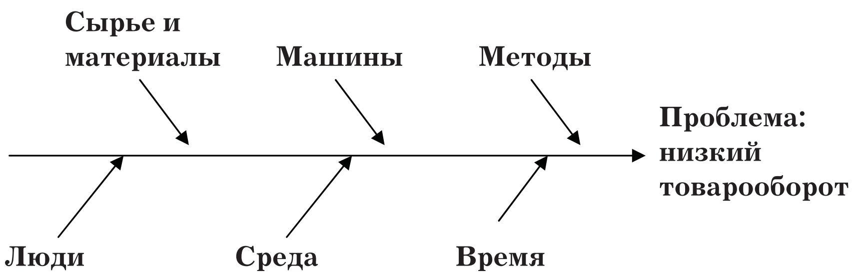


Рис. 4.1. Причинно-следственная диаграмма Исикавы

Выделение входных и выходных переменных

Предположим, что в данных магазинах существенным является влияние количества продавцов на величину розничного товарооборота.

Сбор статистических данных

Исходные данные представлены в табл. 4.2. График зависимости У от Х показан на рис. 4.2.

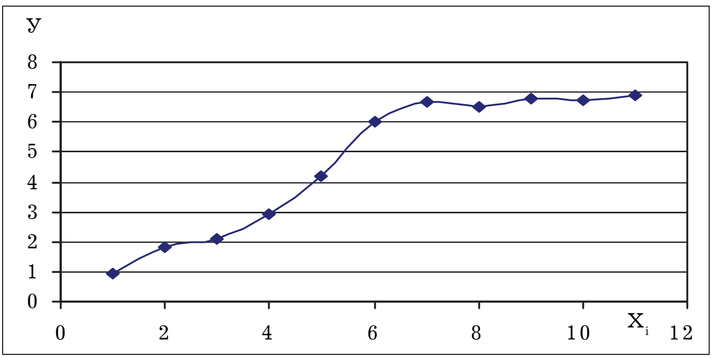


Рис. 4.2. График исходных данных

Выдвижение гипотез

Визуальный анализ рис. 4.2 показывает, что с увеличением количества продавцов розничный товарооборот сначала возрастает, а затем стремится к постоянной величине. Эта закономерность объясняется, очевидно, тем, что при определенном ассортименте товаров существует оптимальное количество продавцов, которое обеспечивает максимальную производительность труда. Критерием оптимальности количестве продавцов является наличие небольшой очереди. При этом длина очереди не должна превышать критического значения, при котором покупатель не уходит из канала обслуживания.

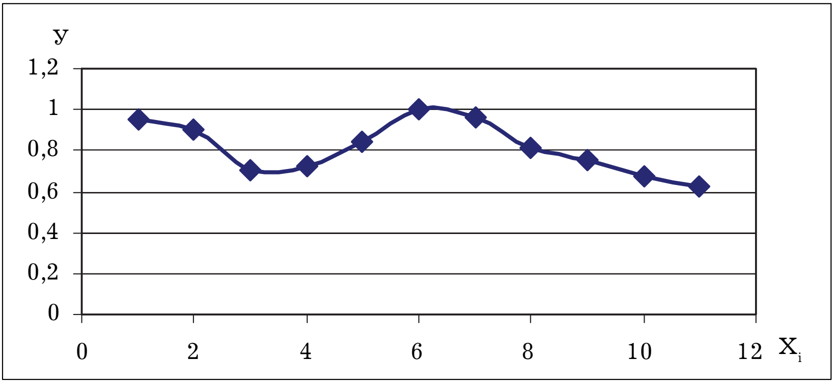


Рис. 4.3. Зависимость производительности труда продавцов от их численности

Визуальный анализ рис. 4.3 показывает, что с увеличением количества продавцов до четырех человек производительность труда падает, затем она возрастает и при шести продавцах после чего опять уменьшается. Таким образом оптимальное количество продавцов равно шести.

Зависимость розничного товарооборота от количества продавцов не является линейной и, очевидно, будет логистической с пределом, равным максимальному розничному товарообороту, который можно получить в магазине за один месяц.

Формулировка допущений

Розничный товарооборот будет нелинейно зависеть от количества продавцов при условии, что в магазине начиная с определенного количестве продавцов пропадает очередь.

Спецификация

Воспроизведение нелинейной зависимости У от Х будем производить с использованием усложняющих следующих основных функций: линейной, параболической, логарифмической, степенной, гиперболической, логистической.

Примечание. В восьмидесятые годы XIX столетия, практически во всех вычислительных центрах наблюдалось увлечение использования различных функций для описания экономических процессов. Оказалось, что одну и ту же тенденцию можно описать различными функциями, которые по критерию ошибки модели мало отличаются между собой. Возникла проблема выбора наиболее подходящей функции по определенному критерию. В качестве критериев были выбраны: простота математического выражения и возможность экономической интерпретации коэффициентов модели. Было обнаружено, что с помощью 20% известных функций можно описать 80% всех возможных тенденций.

В практике научных исследований с помощью около десяти основных функций можно описать практически все виды тенденций. Студенту важно знать свойства этих основных функций.

Спецификация модели зависимости У от Х с помощью линейной функции

$\mathbf{У = XA} + \varepsilon$ или $У_i = \alpha_0 + \alpha_1 X_{1i} + \varepsilon_i$,

при условии, что

$M(\varepsilon) = 0$, $M(\varepsilon\varepsilon') = \sigma_\varepsilon^{\,2} I$

Идентифицируемость модели

(Можно ли в принципе однозначно восстановить значения неизвестных параметров по выбранной модели).

Исходные данные соответствуют модели, поэтому можно оценить ее параметры.

Идентификация модели

(Предложить и реализовать процедуру оценивания параметров модели).

В табл. 4.3 представлена база данных для вычисления.

Таблица 4.3

База данных для вычисления коэффициентов функции

$$У_p = a_0 + a_1 X$$

i	X_i	$У_i$	$У_{pi}$	$(X_i - \overline{X})^2$	$У_{min\,м}$	$У_{max\,м}$
1	1	0,95	1,3568	25	0,2983	2,4153
2	2	1,8	2,0236	16	1,1113	2,9359
3	3	2,1	2,6905	9	1,9106	3,4703
4	4	2,9	3,3573	4	2,6878	4,0267
5	5	4,2	4,0241	1	3,4307	4,6175
6	6	6	4,6909	0	4,1251	5,2567
7	7	6,7	5,3577	1	4,7643	5,9511
8	8	6,5	6,0245	4	5,3551	6,694
9	9	6,8	6,6914	9	5,9115	7,4713
10	10	6,75	7,3582	16	6,4459	8,2705
11	11	6,9	8,025	25	6,9665	9,0835
Ожидаем	12		8,6918		7,4783	9,9053
Сумма	66			110		
Среднее	6					

где $У_{min.м}$ — 95%-ный нижний доверительный интервал математического ожидания У,

$У_{max.м}$ — 95%-ный верхний доверительный интервал математического ожидания У.

Примечание. Расчетные формулы $У_{min\,м}$ и $У_{max\,м}$ и примеры их расчета имеются в третьей теме.

$У_{min(i)} = У_{pi} - t_{\alpha/2}\, S_{УМi}$,

$У_{max(i)} = У_{pi} - t_{\alpha/2}\, S_{УМi}$.

$$S_{УМ_i} = E\sqrt{\frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum (X_i - \overline{X})^2}}.$$

Расчет коэффициентов и характеристик модели с помощью функции "Линейн":

0,6668	0,69
0,0791	0,5364
0,8876	0,8295
71,081	9
48,911	6,193

Представим результаты расчетов в общепринятом обозначении.

У =	0,69	+	0,6668	X
S_a =	0,5364		0,0791	
t_a =	1,2863		8,4309	
E =	0,8295			
R^2 =	0,8876		$t_{\alpha/2}$ =	2,2622
F =	71,081		$F_{кр}$ =	5,1174

где S_a — ошибки коэффициентов модели, расположенные под соответствующим коэффициентом.

$$S_{a0} = 0{,}5/364\text{: } S_{a1} = 0{,}0791$$

ta — критерий Стьюдента, расположенный под соответствующим коэффициентом

$$t_{a_o} = \frac{a_0}{S_{a_0}} = \frac{0{,}69}{0{,}5364} = 1{,}2863; \; t_{a_1} = \frac{a_1}{S_{a_1}} = \frac{0{,}6668}{0{,}0791} = 8{,}4309;$$

$t_{x/2}$ — табличные значения критерия Стьюдента на уровне значимости X = 0,05 и числом степеней свободы m = n-k.

Верификация

Представляем результаты эконометрического анализа.

1. 88,76% исходных данных имеют линейную тенденцию.

2. По критерию Фишера достоверность модели статистически доказана.

Так как $F>F_{кр}(\alpha = 0{,}05;\ m_1 = k\text{-}1;\ m_2 = n\text{-}k)$, то модель является достоверной с вероятность 0,95.

3. По критерию Стьюдента проверим достоверность коэффициентов модели.

Так как $|t_{a0}| < t_{\alpha/2}$, то коэффициент a_0 не отличается от нуля.

Так как $|t_{a1}| > t_{\alpha/2}$, то коэффициент a_1 с вероятностью 0,95 отличается от нуля.

4. При ожидаемом значении объясняемой переменной $X_{ож} = 12$ среднее значение зависимой переменной будет составлять $У_{пр} = 8{,}6918$.

Прогнозное значение зависимой переменной с вероятность 0,95 будет находиться в интервале от 7,4783 до 9,9053

5. Графическое представление результатов расчета имеет следующий вид (рис. 4.4).

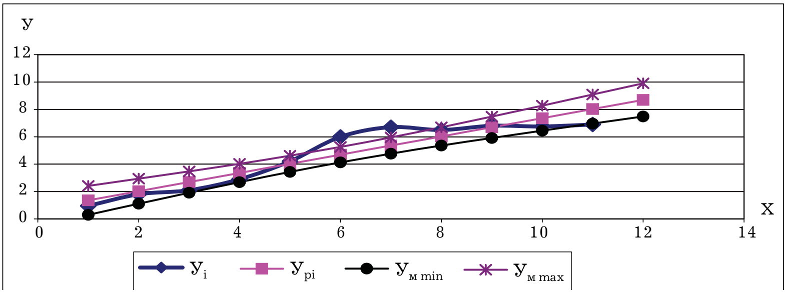


Рис. 4.4. Характеристики линейной модели

Анализ рис. 4.4 показывает, что линейная модель является достоверной, но прогноз плохо воспроизводит тенденцию исходных данных.

Спецификация модели зависимости У от Х с помощью параболической функции

$У = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + \varepsilon$

$M(\varepsilon) = 0,\ M(\varepsilon\varepsilon') = \sigma_\varepsilon^2 I.$

Идентифицируемость модели

Анализ параболической функции показывает, что она не является линейной относительно второй переменной X^2, так как она находится во второй степени.

Если произвести линеаризацию параболической функции, то можно оценить ее параметры.

Идентификация модели

Линеаризация параболической функции заключается в замене переменной X^2 на переменную Z, которая заменяет X^2. В дальнейшем переменная Z используется для расчета коэффициентов модели как линейная переменная.

Примечание. С помощью данного приема замены нелинейных переменных на линейные удается вычислять коэффициенты разнообразных функций с нелинейными переменными.

Например, $У_р = a_0+a_1\ln X + a_2X^3 + a_3X_1X_2$.

В табл. 4.4 представлена база данных для вычесления.

Таблица 4.4

База данных для оценки коэффициентов функции
$$\mathbf{У_р = a_0 + a_1X + a_2X_2}$$

i	X_i	X_{i2}	$У_i$	$У_{pi}$	$(Xi - \overline{X})^2$	$У_{min\,м}$	$У_{max\,м}$
1	1	1	0,95	0,4066	25	-0,356	1,1693
2	2	4	1,8	1,6436	16	0,9862	2,3009
3	3	9	2,1	2,7538	9	2,1919	3,3157
4	4	16	2,9	3,7373	4	3,255	4,2197
5	5	25	4,2	4,5942	1	4,1666	5,0218
6	6	36	6	5,3244	0	4,9167	5,732
7	7	49	6,7	5,9278	1	5,5003	6,3554
8	8	64	6,5	6,4046	4	5,9223	6,887
9	9	81	6,8	6,7547	9	6,1928	7,3166
10	10	100	6,75	6,9781	16	6,3208	7,6354
11	11	121	6,9	7,0748	25	6,3122	7,8375
Ожидаем	12	144		7,0448		6,1705	7,9192
Сумма	66				110		
Среднее	6						

-0,0633	1,427	-0,957
0,02	0,2466	0,6439
0,9501	0,5863	#Н/Д
76,146	8	#Н/Д
52,354	2,7502	#Н/Д

Представим результаты расчетов в общепринятом обозначении.

У =	-0,957	+	1,427	X	-0,063	X^2
S_a =	0,6439		0,2466		0,02	
t_a =	-1,486		5,7861		-3,165	
E =	0,5863					
R^2 =	0,9501		$t_{\alpha/2}$ =	2,306		
F =	76,146		$F_{кр}$ =	4,459		

Верификация

Представляем результаты эконометрического анализа.

1. 95,009% исходных данных имеют линейную тенденцию.

2. По критерию Фишера достоверность модели статистически доказана.

Так как $F > F_{кр}(\alpha = 0{,}05;\ m_1 = k — 1;\ m_2 = n — k)$,

то модель является достоверной с вероятность 0,95.

3. По критерию Стьюдента проверим достоверность коэффициентов модели.

Так как $|t_{a0}| < t_{\alpha/2}$, то коэффициент a_0 не отличается от нуля.

Так как $|t_{a1}| > t_{\alpha/2}$, то коэффициент a_1 с вероятностью 0,95 отличается от нуля.

Так как $|t_{a2}| > t_{\alpha/2}$, то коэффициент a_2 с вероятностью 0,95 отличается от нуля.

4. При ожидаемом значении объясняемой переменной $X_{ож} = 12$

среднее значение зависимой переменной будет составлять

$У_{пр} = 7{,}0448$.

Прогнозное значение зависимой переменной с вероятность 0,95 будет находиться в интервале от 6,1705 до 7,9192

5. Графическое представление результатов расчета имеет следующий вид (рис. 4.5).

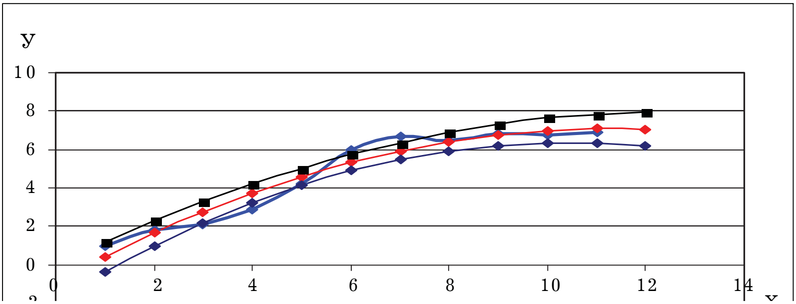


Рис. 4.5. Характеристики параболической модели

Анализ рис. 4.5 показывает, что параболическая модель хорошо воспроизводит тенденцию зависимости У от Х, прогноз соответствует исходным данным. Однако прогноз по параболической функции нельзя производить на длительный интервал.

Спецификация модели зависимости У от Х с помощью логарифмической функции

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

$M(\varepsilon) = 0,\ M(\varepsilon\varepsilon') = \sigma_\varepsilon^2 I.$

Идентифицируемость модели

Анализ логарифмической функции показывает, что она не является линейной относительно переменной Х. Если произвести линеаризацию логарифмической функции, то можно оценить ее параметры.

Идентификация модели

Линеаризацию логарифмической функции произведем методом замены lnX на переменную Z. В табл. 4.5 представлена база данных для вычисления.

Таблица 4.4

База данных для оценки коэффициентов функции

$$У_р = a_0 + a_1 \ln X$$

i	X_i	$У_i$	$\ln X_i$	$У_{pi}$	$(X_i - \bar{X})^2$	$У_{м.min(i)}$	$У_{м.max(i)}$
1	1	0,95	0	-0,057	25	-1,07	0,956
2	2	1,8	0,6931	2,0114	16	1,1525	2,8703
3	3	2,1	1,0986	3,2213	9	2,4916	3,951
4	4	2,9	1,3863	4,0797	4	3,4391	4,7203
5	5	4,2	1,6094	4,7456	1	4,1525	5,3387
6	6	6	1,7918	5,2896	0	4,7218	5,8574
7	7	6,7	1,9459	5,7496	1	5,1565	6,3427
8	8	6,5	2,0794	6,148	4	5,4849	6,8111
9	9	6,8	2,1972	6,4995	9	5,7338	7,2652
10	10	6,75	2,3026	6,8139	16	5,9243	7,7035
11	11	6,9	2,3979	7,0983	25	6,0537	8,1428
Ожидаем	12		2,4849	7,3579		6,1503	8,5655
Сумма	66				110		
Среднее	6						

2,9839	-0,057
0,3369	0,587
0,8971	0,7938
78,455	9
49,433	5,6708

Представим результаты расчетов в общепринятом обозначении.

У =	-0,057	+	2,9839	ln(X),
S_a =	0,587		0,3369	
t_a =	-0,097		8,8575	
E =	0,7938			
R^2 =	0,8971		$t_{\alpha/2}$ =	2,2622
F =	78,455		$F_{кр}$ =	5,1174

Верификация

Представляем результаты эконометрического анализа.

1. 89,709% исходных данных имеют линейную тенденцию;

2. По критерию Фишера достоверность модели статистически доказана.

Так как $F > F_{кр}(\alpha = 0{,}05;\ m_1 = k\text{-}1;\ m_2 = n — k)$,

то модель является достоверной с вероятность 0,95.

3. По критерию Стьюдента проверим достоверность коэффициентов модели.

Так как $|t_{a0}| < t_{\alpha/2}$, то коэффициент a_0 не отличается от нуля.

Так как $|t_{a1}| > t_{\alpha/2}$, то коэффициент a_1 с вероятностью 0,95 отличается от нуля.

4. При ожидаемом значении объясняемой переменной $X_{ож} = 12$ среднее значение зависимой переменной будет составлять $У_{пр} = 7{,}3579$.

Прогнозное значение зависимой переменной с вероятность 0,95 будет находиться в интервале от 6,1503 до 8,565.

5. Графическое представление результатов расчета имеет следующий вид (рис. 4.6).

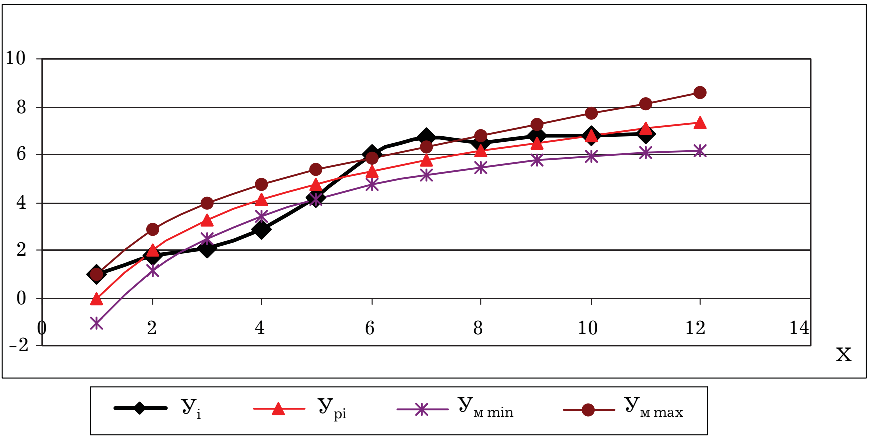


Рис. 4.6. Логарифмическая модель зависимости У от Х

Анализ рис. 4.6 показывает, что логарифмическая функция плохо воспроизводит начальные значения У и хорошо воспроизводит последние значения У.

Спецификация модели зависмости У от Х с помощью степенной функции

$У = \alpha_0 X^{\alpha 1} \times \varepsilon$

$M(\varepsilon) = 0, M(\varepsilon\varepsilon') = \sigma_\varepsilon{}^2 I.$

Идентифицируемость модели

Анализ степенной функции показывает, что она является нелинейной относительно коэффициента a_1, так как этот коэффициент является степенью для переменной Х.

Вторая особенность степенной функции заключается в том, что свободный коэффициент a_0 не прибавляется к переменной, а умножается на нее.

Третья особенность степенной функции заключается в том, что для нее в исходном виде нельзя применить метод наименьших квадратов расчета ее коэффициентов.

Следовательно, параметры степенной функции нельзя оценить, если она представлена в исходном виде.

Если степенную функцию привести к аддитивному виду, то в принципе можно оценить ее параметры.

Идентификация модели

Для линеаризации степенной функции используется метод логарифмирования.

Приведем мультипликативную функцию к аддитивному виду:

$$\ln У = \ln a_0 + a_1 \ln X$$

Затем к полученной функции применяем метод замены:

$$У_н = \ln У,\ a_н = \ln a_0,\ X_н = \ln X$$

и получаем функцию

$$У_н = a_н + a_1 X_н,$$

которая является линейной и ее коэффициенты можно вычислить методом наименьших квадратов.

Коэффициент a_0 вычисляется по формуле:

$$a_0 = \exp(a_н).$$

Если нелинейную функцию нельзя привести к линейному виду методом замены или логарифмирования, то для расчета ее коэффициентов используются приближенные итерационные методы В табл. 4.6 дана база данных для вычисления.

Таблица 4.6

База данных для оценки коэффициентов функции $У_р = a_0 X^{a1}$

i	Xi	Уi	lnX_i	$lnУ_i$	$У_{pi}$	$(X_i — \bar{X})^2$	$У_{м\,min(i)}$	$У_{м\,max(i)}$	$(У_i — У_{pi})^2$
1	1	0,95	0	-0,051	0,9353	25	-0,461	2,3315	0,00022
2	2	1,8	0,6931	0,5878	1,7513	16	0,586	2,9166	0,00237
3	3	2,1	1,0986	0,7419	2,5275	9	1,5799	3,4752	0,18279
4	4	2,9	1,3863	1,0647	3,2791	4	2,5244	4,0337	0,14369
5	5	4,2	1,6094	1,4351	4,0127	1	3,4025	4,623	0,03507
6	6	6	1,7918	1,7918	4,7325	0	4,1787	5,2863	1,60658
7	7	6,7	1,9459	1,9021	5,4409	1	4,8306	6,0511	1,58543
8	8	6,5	2,0794	1,8718	6,1396	4	5,385	6,8943	0,12986
9	9	6,8	2,1972	1,9169	6,8301	9	5,8825	7,7778	0,00091
10	10	6,75	2,3026	1,9095	7,5134	16	6,348	8,6787	0,5827
11	11	6,9	2,3979	1,9315	8,1901	25	6,7939	9,5863	1,66433
Ожидаем	12		2,4849		8,861		7,2264	10,496	
Сумма	66				51,353	110			5,93394
Среднее	6								

Рассчитаем коэффициенты уравнения $lnУ = a_н + a_1 lnX$ с помощью функции Excel “Линейн”

0,9049	-0,067
0,0627	0,1092
0,9586	0,1477
208,43	9
4,5458	0,1963

В результате расчетов коэффициентов получаем уравнение

$$lnУ = a_н + a_1 lnX = -0,067 + 0,9049 lnX,$$

где $a_н = lna_0 = -0,067$; $a_1 = 0,9049$; $S_{a1} = 0.0627$.

Определим коэффициент a_0 по формуле

$$a_0 = \exp(a_н) = \exp(-0,067) = 0,9353$$

Степенная функция имеет следующий вид:

$$У_р = a_0 X^{a1} = 0,9353X^{0,9049}$$

Вычислим ошибку E модели для степенной функции

$$E = 0,812$$

Представим результаты расчетов в общепринятом обозначении.

$У_p =$	0,9353	$X^{0,9049}$	
$S_a =$		0,0627	
$t_a =$		14,437	
E=	0,812		
		$t_{\alpha/2} =$	2,2622

где ошибка коэффициента a_1 $S_{a1} = 0.0627$, ошибка коэффициента a_0 неизвестна, так для ее расчета пока не нашли формул.

Верификация.

Представляем результаты эконометрического анализа.

1. По критерию Стьюдента достоверность коэффициентов модели статистически доказана.

Так как $|t_{a1}| > t_{\alpha/2}$., то коэффициент a_1 с вероятностью 0,95 отличается от нуля.

2. При ожидаемом значении объясняемой переменной $X_{ож} = 12$

среднее значение зависимой переменной будет составлять

$$У_{пр} = 8,861.$$

Прогнозное значение зависимой переменной с вероятность 0,95 будет находиться в интервале от 7,0241 до 10,698

3. Графическое представление результатов расчета имеет следующий вид.

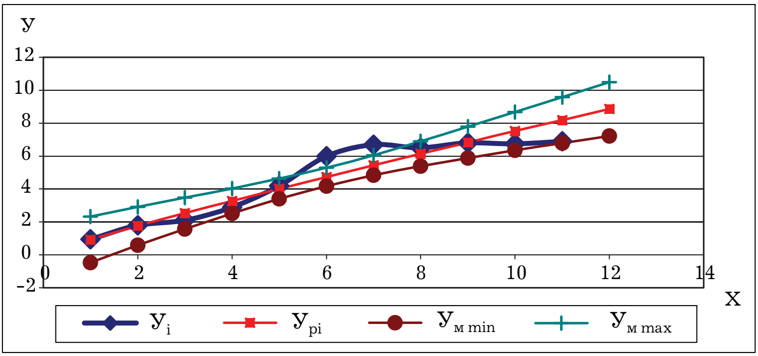


Рис. 4.7. Степенная модель зависимости У от X

Анализ рис. 4.7 показывает, что степенная модель плохо воспроизводит прогнозные значения У.

Спецификация модели зависимости У от Х гиперболической функцией

$У = \alpha_0 + \alpha_1/X + \varepsilon.$

$M(\varepsilon) = 0,\ M(\varepsilon\varepsilon') = \sigma_\varepsilon^{\ 2}I.$

Идентифицируемость модели

Анализ гиперболической функции показывает, что она не является линейной относительно переменной X. Если произвести линеаризацию гиперболической функции, то можно оценить ее параметры.

Идентификация модели

Линеаризацию гиперболической функции произведем методом замены 1/X на переменную Z.

В табл. 4.7 представлена база данных для вычисления

Таблица 4.7

База данных для оценки коэффициентов гиперболической функции $У_p = a_0 + a_1/X$

i	X_i	$У_i$	$1/X_i$	$У_{pi}$	$(X_i — \bar{X})^2$	$У_{м\ min\ (i)}$	$У_{м\ max(i)}$
1	1	0,95	1	-0,507	25	-2,296	1,2807
2	2	1,8	0,5	3,0753	16	1,5341	4,6166
3	3	2,1	0,3333	4,2696	9	2,9521	5,5871
4	4	2,9	0,25	4,8667	4	3,7358	5,9977
5	5	4,2	0,2	5,225	1	4,2225	6,2275
6	6	6	0,1667	5,4638	0	4,508	6,4197
7	7	6,7	0,1429	5,6345	1	4,632	6,6369
8	8	6,5	0,125	5,7624	4	4,6314	6,8934
9	9	6,8	0,1111	5,8619	9	4,5444	7,1795
10	10	6,75	0,1	5,9415	16	4,4003	7,4828
11	11	6,9	0,0909	6,0067	25	4,2185	7,7949
Ожидаем	12		0,0833	6,061		4,0109	8,111
Сумма	66				110		
Среднее	6						

Вычислим коэффициенты модели
$У = a_0 + a_1/X$:

-7,1656	6,6581
1,6413	0,6177
0,6792	1,4014
19,059	9
37,429	17,675

Представим результаты расчетов в общепринятом обозначении.

У =	6,6581	+	-7,166	/ X
S_a =	0,6177		1,6413	
t_a =	10,779		-4,366	
E =	1,4014			
R^2 =	0,6792		$t_{\alpha/2}$ =	2,2622
F =	19,059		$F_{кр}$ =	5,1174

Верификация

Представляем результаты эконометрического анализа.

1. 67,925% исходных данных имеют линейную тенденцию;

2. По критерию Фишера достоверность модели статистически доказана.

Так как $F > F_{кр}(\alpha = 0,05; m_1 = k\text{-}1; m_2 = n\text{-}k)$,

то модель является достоверной с вероятность 0,95.

3. По критерию Стьюдента проверим достоверность коэффициентов модели.

Так как $|t_{a0}| > t_{\alpha/2}$, то коэффициент a_0 с вероятностью 0,95 отличается от нуля.

Так как $|t_{a1}| > t_{\alpha/2}$, то коэффициент a_1 с вероятностью 0,95 отличается от нуля.

4. При ожидаемом значении объясняемой переменной $X_{ож} = 12$

среднее значение зависимой переменной будет составлять

$У_{пр} = 6,061$.

Прогнозное значение зависимой переменной с вероятность 0,95 будет находиться в интервале от 4,0109 до 8,111.

5. Графическое представление результатов расчета имеет следующий вид (рис. 4.8).

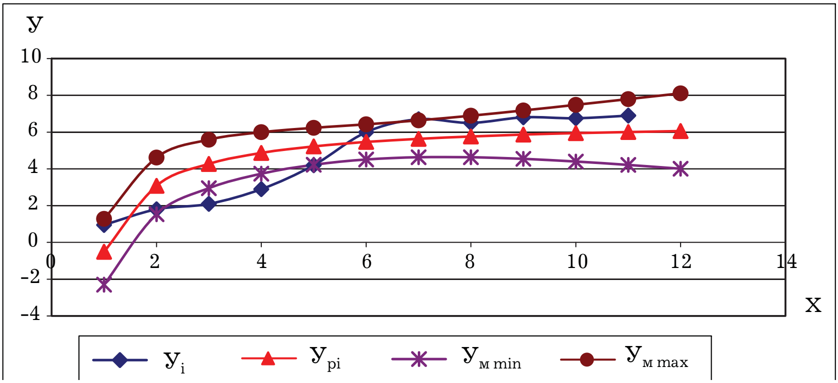


Рис. 4.8. Гиперболическая зависимость У от Х

Анализ рис. 4.8 показывает, что гиперболическая зависимость учитывает асимптотическую тенденцию У, однако прогнозные значения ниже ожидаемых.

Спецификация модели зависимости У от Х логистической функцией

$$У = \frac{\alpha}{1 + \beta e^{-\gamma t}} + \varepsilon,$$

где α, β, γ — неизвестные параметры,

ε — случайная составляющая,

t — порядковый номер измерения (время).

$M(\varepsilon) = 0,\ M(\varepsilon\varepsilon') = \sigma_\varepsilon^{\,2} I.$

Заменим параметры логистической функции ее выборочными оценками

$$У = \frac{\alpha}{1 + be^{-ct}} + v.$$

Анализ логистической функции показывает, что она является нелинейной относительно переменной и коэффициентов. Эту функцию нельзя привести в линейному виду ни методом замены, ни методом логарифмирования, так как логарифм знаменателя, содержащий сумму ($1 + be^{-ct}$) нельзя разложить на составляющие. Следовательно, для оценки параметров модели необходимо произвести специальные исследования.

Особый интерес для экономистов вызывает коэффициент а, который равен пределу этой функции при стремлении t к бесконечности. Для нашей задачи коэффициент а будет равен максимальному розничному товарообороту, который можно получить в магазине за месяц торговли.

Идентификация модели

Известно несколько видов логистической функции и несколько методов расчетов ее коэффициентов, большинство из которых не являются строго формализированными, а также имеются итерационные методы, входящие в состав статистических пакетов.

Приводим один из строго формализованных алгоритмов вычисления коэффициентов логистической модели

$$У = \frac{\alpha}{1 + be^{-ct}} + v,$$

где t — значение фактора с равными интервалами изменения его значений (обычно это время).

Расчетные формулы коэффициентов логистической модели выводятся на основе метода наименьших квадратов[1].

1. Вычислим методом наименьших квадратов коэффициенты c и d модели

$$У_{t+1} — У_t = cУ_t — dУ_t^2 + e_t,$$

2. Определим коэффициент а по формуле

$a = c/d$,

где коэффициент а равен пределу логистической функции при увеличении t.

3. Определим коэффициент b по формуле

$$b = \exp(\frac{1}{n}\sum \ln\left(\frac{a}{У_t} - 1\right) + \frac{c}{n}\sum t),$$

где n — объем выборки.

[1] Толбатов Ю.А. Економетрика. — К.: Четверта хвиля, 1997, с.130−132.

Предложенный алгоритм имеет одно ограничение на значения $У_t$: $У_t < a$, так как при расчете коэффициента b необходимо вычислять $\ln(a/У_t. — 1)$, где логарифм отрицательного числа не существует.

Создадим базу данных с использованием преобразованных переменных (табл. 4.8).

Таблица 4.8

База данных для оценки коэффициентов функции

$$\mathbf{У_{t+1} — У_t = cУ_t — dУ_t^2 + e_t}$$

i	t	$У_t$	$-У_t^2$	$У_{t+1}-У_t$	$У_p$	Z_{1t}	$(t — \bar{t})^2$	$У_{м\,min\,(i)}$	$У_{м\,max(i)}$
1	1	0,95	-0,903	0,85	0,5494	1,827	25	-0,12	1,21926
2	2	1,8	-3,24	0,3	1,0624	1,0325	16	0,4896	1,63525
3	3	2,1	-4,41	0,8	1,9093	0,8172	9	1,4252	2,39346
4	4	2,9	-8,41	1,3	3,0732	0,3102	4	2,664	3,4823
5	5	4,2	-17,64	1,8	4,3258	-0,459	1	3,9691	4,6825
6	6	6	-36	0,7	5,3647	-1,949	0	5,0273	5,70207
7	7	6,7	-44,89	-0,2	6,0556	-3,769	1	5,6989	6,41227
8	8	6,5	-42,25	0,3	6,4502	-2,909	4	6,0411	6,85938
9	9	6,8	-46,24	-0,05	6,6563	-4,826	9	6,1722	7,14044
10	10	6,75	-45,56	0,15	6,7589	-4,168	16	6,1861	7,33173
11	11	6,9			6,8088			6,139	7,47859
Ожидаем	12				6,8327			6,0607	7,60471
Сумма	66				49,015	-14,09	85		
Среднее	6								

Продолжение таблицы 4.8.

i	t	Z2t
1	1	0,1604
2	2	0,544
3	3	0,0364
4	4	0,03
5	5	0,0158
6	6	0,4036
7	7	0,4153
8	8	0,0025
9	9	0,0206
10	10	8E-05
11	11	
Ожидаем	12	
Сумма	66	1,6287

где $Z_{1t} = \ln[(a/У_t) — 1]$

$$Z_{2t} = (У — У_p)^2$$

Вычислим коэффициенты c, d модели

$$У_{t+1} — У_t = cУ_t — dУ_t^2 + e_t,$$

0,1086	0,7443
0,0193	0,1203
0,6856	0,3702
8,7209	8
2,3907	1,0965

Коэффициент c = 0,7443

Коэффициент d = 0,1086

a = c/d = 6,8545

$$b = \exp(\frac{1}{n}\sum \ln\left(\frac{a}{У_t} - 1\right) + \frac{c}{n}\sum t)$$

b = 24.155

n = 10 — объем выборки,

к = 3 — количество коэффициентов в модели.

Представим результаты расчетов в общепринятом обозначении.

$$У = \frac{6,8545}{1 + 24,155e^{-0,7443t}}.$$

E = 0,4512 $t_{\alpha/2}$ = 2,3646

F = 112,05 $F_{кр}$ = 4,7374

Верификация

Представляем результаты эконометрического анализа.

1. По критерию Фишера достоверность модели статистически доказана.

Так как $F > F_{кр}(\alpha = 0,05; m_1 = k\text{-}1; m_2 = n — k)$,

то модель является достоверной с вероятность 0,95.

2. При ожидаемом значении объясняемой переменной $X_{ож}$ = 12 среднее значение зависимой переменной будет составлять $У_{пр}$ = 6,8327.

Прогнозное значение зависимой переменной с вероятность 0,95 будет находиться в интервале от 6,0607 до 7,6047.

3. Графическое представление результатов расчета имеет следующий вид (рис. 4.9).

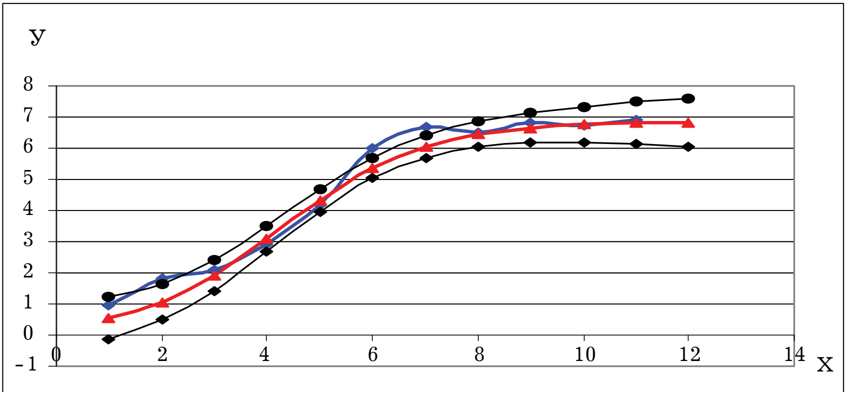


Рис. 4.9. Логистическая зависимость У от Х

Анализ рис. 4.9 показывает, что логистическая функция хорошо воспроизводит предел стремления розничного товарооборота при увеличении количества продавцов.

Проведем сравнительный анализ всех изученных моделей по их ошибке Е.

Таблица 4.9

Сравнительный анализ полученных моделей

№	Название функции	Ошибка модели Е
1	Линейная	0,8295
2	Параболическая	0,5863
3	Логарифмическая	0,7938
4	Степенная	0,812
5	Гиперболическая	1,4014
6	Логистическая	0,4512

Анализ табл. 4.9 показывает, что по ошибке модели лучшей функцией, отражающей специфическую зависимость розничного товарооборота от численности продавцов, является логистическая функция.

Задание 4. Повторите расчеты задания 3 для любой из шести функций.

Пояснение. Откройте файл Excel “Лаборат4Часть2” и повторите расчеты для любой из шести функций.

Задание 5. С помощью программы “Поиск решения” найдите коэффициенты любой нелинейной функции, например, логистической:

$$У_p = \frac{\alpha}{1 + be^{-ct}}.$$

Пояснение. Скопируйте исходные данные табл. 4.2 и вставьте в лист Excel, выделите три ячейки для коэффициентов и введите в них любые числа, кроме нуля, в таблице исходных данных добавьте колонку с $У_p$, вычисленных по логистической функции с использованием коэффициентов, определите квадрат остатков, найдите сумму квадратов остатков — она будет целевой функцией, обратитесь к программе Excel “Поиск решения”, укажите ячейки, которые надо изменять — это ячейки, в которых находятся значения коэффициентов, укажите ячейку, где находится целевая функция, устремите целевую функцию к минимуму, ОК, получите значения коэффициентов, сравните полученные коэффициенты с вычисленными по точным формулам задания 3.

Протокол расчетов представлен на рис. 4.10 и 4.11.

a = 7,0783;

b = 17,296;

c = 0,6868.

Сравним эти коэффициенты с найденными ранее:

a = 6,8545;

b = 24.155;

c = 0,7443,

которые не существенно отличаются между собой.

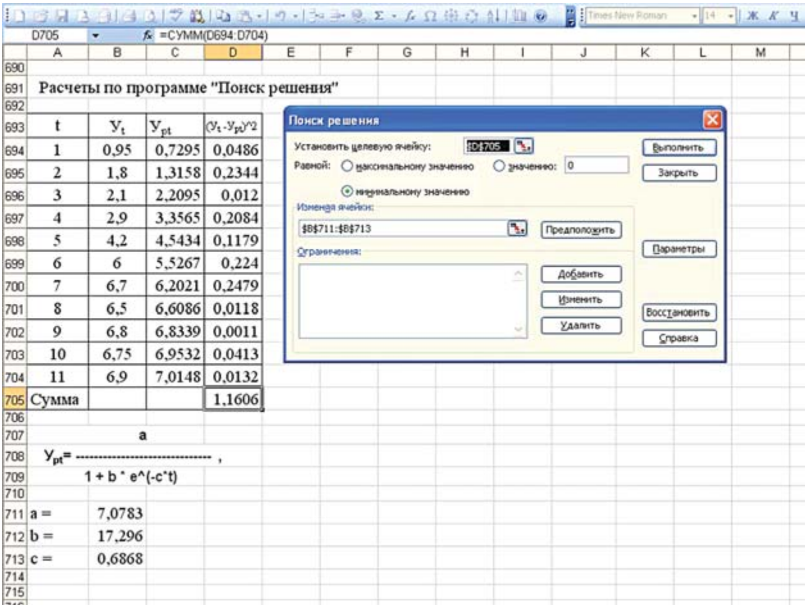


Рис. 4.10. Протокол решения задания 5 с помощью программы “Поиск решения”

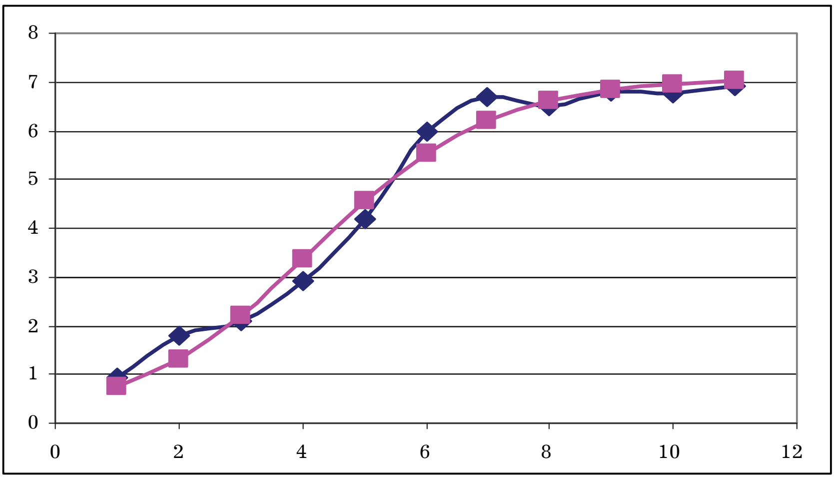


Рис. 4.11. Фактические и расчетные значения логистической функции, определенные с помощью программы “Поиск решения”

Анализ рис. 4.11 показывает, что расчетные значения логистической функции очень хорошо описывают тенденцию изменений данных.

Вывод. Расчеты по программе “Поиск решения” по времени значительно эффективнее, чем по специальным формулам. Следует отметить, что программа “Поиск решения” позволяет рассчитать коэффициенты любой функции.

Задания второго уровня сложности

Задание 1. Имеются различные виды математических функций, которые часто используются в экономическом анализе.

Необходимо:

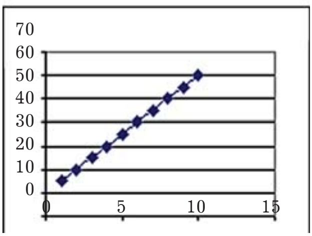
— изучить вид функции при фиксированных значениях коэффициентов и изменениях X = 1, 2, 3, ..., 10;

— сделать выводы о свойствах изучаемых функций.

Пояснение. Знание видов кривых или тенденций, которые может воспроизвести каждая математическая функция, позволит быстро подобрать необходимую функцию для хорошей спецификации модели экономического процесса.

Таблица 4.10

Набор линейных (относительно коэффициентов) функций

№ п/п	Вид функции	График функции
1	Линейная $У_р = a_0 + a_1X$ $a_0 = 10$ $a_1 = 5$	

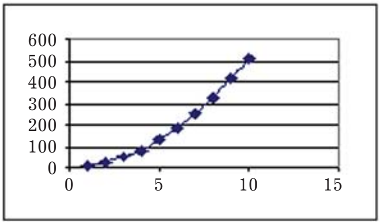
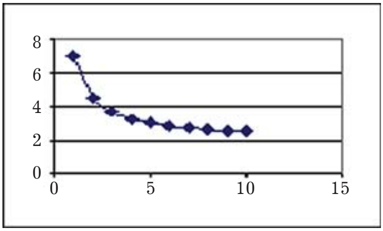
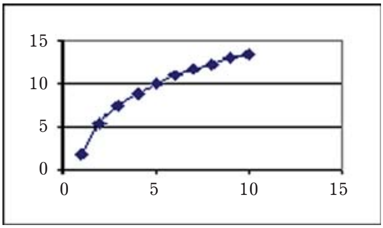
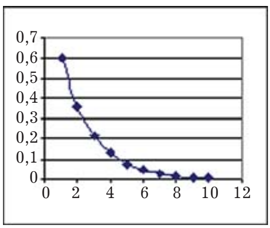
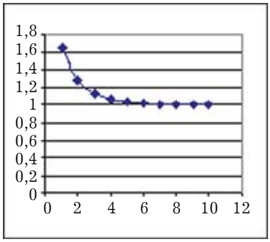
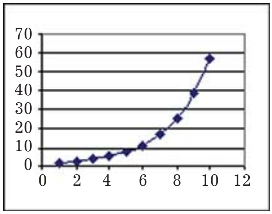
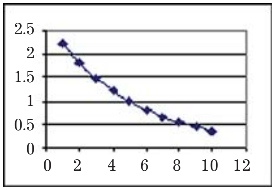
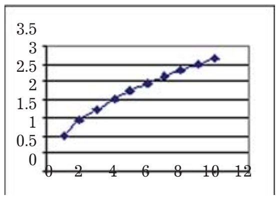
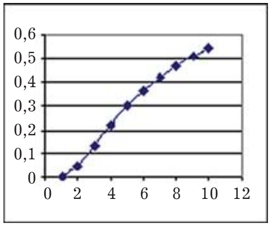
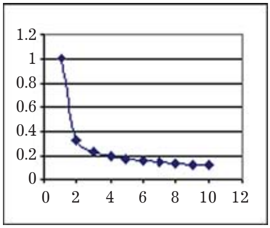
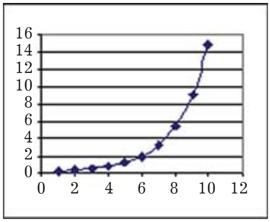
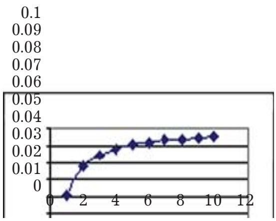
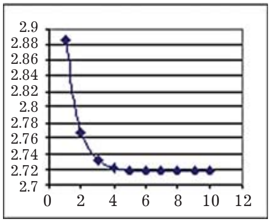
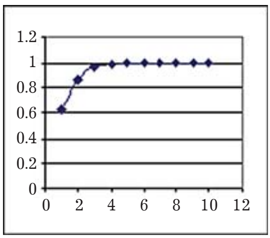
<table>
<tr>
<td>2</td>
<td>Парабола
$У_р = a_0 + a_1X +$
$+ a_2X^2$
$a_0 = 1$
$a_1 = 1$
$a_2 = 5$</td>
<td>
</td>
</tr>
<tr>
<td>3</td>
<td>Гипербола
$У_р = a_0 + a_1/X$
$a_0 = 2$
$a_1 = 5$</td>
<td>
</td>
</tr>
<tr>
<td>4</td>
<td>Логарифмическая
$У_р = a_0 + a_1 \ln X$
$a_0 = 2$
$a_1 = 5$</td>
<td>
</td>
</tr>
</table>

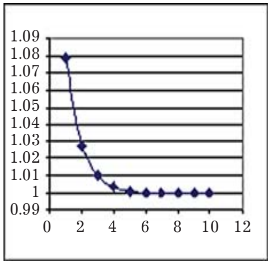
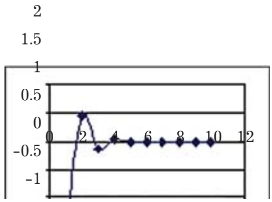
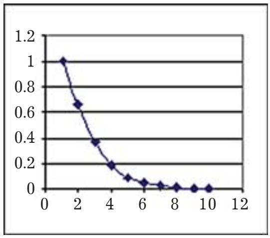
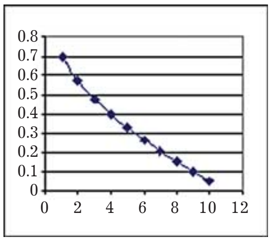
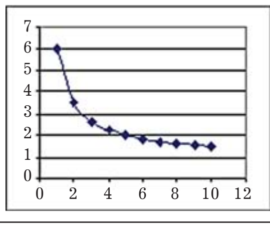
Таблица 4.11

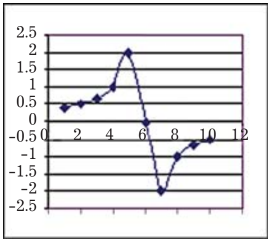
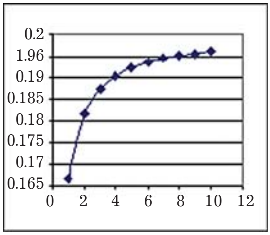
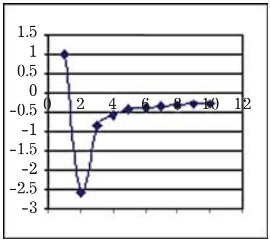
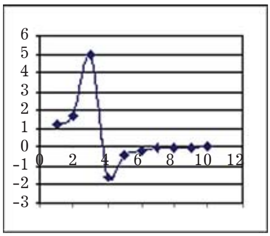
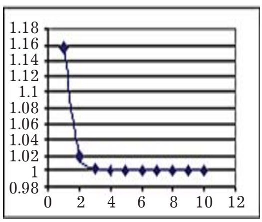
Набор нелинейных (относительно коэффициентов) функций, которые приводятся к линейному виду с помощью логарифмирования уравнения

<table>
<tr>
<th>№ п/п</th>
<th>Вид нелинейной функции</th>
<th>Преобразование</th>
<th>График функции</th>
</tr>
<tr>
<td>1</td>
<td>Экспоненциальная 1 $У_р = a_0 a_1^X$
$a_0 = 1$
$a_1 = 0,6$</td>
<td>$\ln У = \ln a_0 + X \ln a_1$</td>
<td>
</td>
</tr>
</table>

<table>
<tr><td>2</td><td>Экспоненциальная 2
$У_p = a_0 e^{a1X}$
$a_0 = 1$
$a_1 = 0,5$
(Доугерти с.122,
Льюис с. 88)</td><td>$\ln У = \ln a_0 + a_1 X$</td><td>
</td></tr>
<tr><td>3</td><td>Экспоненциальная 3
$У_p = a_0(1\text{-}a_1)^X$
$a_0 = 1$
$a_1 = -0,5$</td><td>$\ln У = \ln a_0 +$
$+ (\ln(1\text{-}a_1))X$</td><td>
</td></tr>
<tr><td>4</td><td>Экспоненциальная 4
$У_p = e^{a0 + a1X}$
$a_0 = 1$
$a_1 = -0,2$</td><td>$\ln У = a_0 + a_1 X$</td><td>
</td></tr>
<tr><td>5</td><td>Степенная
Кривая Энгеля
$У_p = a_0 X^{a1}$
$a_0 = 1$
$a_1 = 0,5$</td><td>$\ln У = a_0 + a_1 \ln X$</td><td>
</td></tr>
<tr><td>6</td><td>S образная
$У_p = e^{a0+a1/X}$
$a_0 = -0,01$
$a_1 = -6$</td><td>$\ln У = a_0 + a_1/X$
S образная, при
$a_1 < -a_0$</td><td>
</td></tr>
</table>

7	Обратная логарифмическая $У_p = 1/(a_0+a_1 \ln X)$ $a_0 = 1$ $a_1 = 3$	$1/У = a_0+a_1 \ln X$	
8	Экспоненциальная 5 $У_p = a_0 + a_1 e^X$ $a_0 = 0{,}1$ $a_1 = 0{,}5$	$У_p = a_0 + a_1 e^X$	
9	Экспоненциальная 6 $У_p = a_0 e^{a1/X}$ $a_0 = 0{,}1$ $a_1 = — 0{,}5$	$\ln У = \ln a_0 + a_1/X$	
10	Кривая Гомперца 1 $У_p = e^{a_0 + a_1 a_2^X}$ $a_0 = 1$ $a_1 = 0{,}2$ $a_2 = 0{,}3$	$\ln У = a_0+a_1 a_2^X$	
11	Кривая Гомперца 2 $У_p = a_0 a_1^{a_2^X}$ $a_0 = 1$ $a_1 = 0{,}2$ $a_2 = 0{,}3$	$\ln У = \ln a_0+\ln a_1 (a_2)^X$	

12	Обратная экспоненциальная $У_р = 1/(a_0 + a_1e^{-X})$ $a_0 = 1$ $a_1 = -0,2$	$1/У_р = a_0 + a_1e^{-X}$	
13	Экспоненциальная 7 $У_р = a_0 + a_1a_2e^X$ $a_0 = 1$ $a_1 = 5$ $a_2 = -0,3$ (Льюис с. 106)		
14	Степенная 1 $У_р = a_0X^{a1X}$ $a_0 = 1$ $a_1 = -0,3$	$\ln У = \ln a_0 +$ $+ a_1 X \ln X$	
15	Степенная 2 $У_р = a_0 + a_1X^{1/2}$ $a_0 = 1$ $a_1 = -0,3$	$У_р = a_0 + a_1X^{1/2}$	
16	Гипербола 1 $У_р = a_0 + a_1/X$ $a_0 = 1$ $a_1 = 5$	$У = a_0 + a_1/X$	

<table>
<tr><td>17</td><td>Гипербола 2
$У_р = 1/(a_0 + a_1X)$
$a_0 = 3$
$a_1 = -0,5$</td><td>$1/У = a_0 + a_1X$</td><td>
</td></tr>
<tr><td>18</td><td>Гипербола 3
$У_р = X/(a_0 + a_1X)$
$a_0 = 1$
$a_1 = 5$</td><td>$X/У = a_0 + a_1X$</td><td>
</td></tr>
<tr><td>19</td><td>Обратная логарифмическая
$У_р = 1/(a_0 + a_1 \ln X)$
$a_0 = 1$
$a_1 = -2$</td><td>$1/У = a_0 + a_1 \ln X$</td><td>
</td></tr>
<tr><td>20</td><td>Логистическая 1
$У_р = 1/(a_0 + a_1a_2^X)$
$a_0 = 1$
$a_1 = -0,1$
$a_2 = 2$</td><td></td><td>
</td></tr>
<tr><td>21</td><td>Логистическая 2
$У_р = a_0/(1 + a_1e^{a2X})$
$a_0 = 1$
$a_1 = -1$
$a_2 = 2$</td><td></td><td>
</td></tr>
</table>

Задание 2 Постройте график любой функции с условием изменения ее коэффициентов

Пояснение. Откройте лист Excel, выделите ячейки для коэффициентов функции и введите в них произвольные числа, введите в таблицу значения X и расчетные значения $У_р$ по выбранной функции, постройте график зависимости $У_р$ от X, изменяйте значения коэффициентов и наблюдайте за меняющимся графиком, определите зависимость графика от коэффициентов функции.

Пример решения задания (см. рис. 4.12).

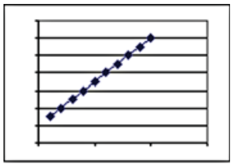
Вид функции	Примечание	Вид кривой
Линейная $У_р = a_0 + a_1X$ $a_0 =$ **10** $a_1 =$ **5**	Введите значения a_1 в пределах от -5 до 5	

Рис. 4. 12 Пример решения задания, изменяя значения a_1 можно наблюдать изменение поведения графика.

Задание третьего уровня сложности

Задание 1. Изучите вывод оценки параметров логистической регрессии[1]

Для прогнозирования спроса на товары длительного пользования часто используют логистическую регрессию

$У = A/(1+ce^{-at})$.

Спрос на товары длительного использования имеет такую тенденцию к изменениям: сначала некоторое время спрос растет, а затем колеблется вокруг постоянного числа.

Логистическая регрессия является нелинейной по оцениваемым параметрам.

Рассмотрим метод оценки параметров, который объединяет метод конечных разниц и метод наименьших квадратов.

[1] Толбатов Ю. А. Економетрика. — К.: Четверта хвиля, 1997, с.130–132.

Если через Y обозначить обеспеченность товаром, t — время, A — насыщенность рынка товаром, k-коэффициент пропорциональности, тогда изменение обеспеченности товаром можно записать в виде дифференциального уравнения Бернулли :

$$\frac{\partial У}{\partial t} = kУ\left(A - Y\right).$$

Пусть динамический ряд имеет вид:

y_i	y_1	y_2	...	y_n
t_i	1	2	...	n

Тогда дифференциальное уравнение можно записать в виде такого выражения:

$$\Delta Y = k\, Y\, (A - Y)\, \Delta t.$$

Если наблюдения за показателями проводилось через единицу времени ($\Delta t = 1$), то

$\Delta Y = k\, Y\, (A\text{-}Y)$ или $y_{i+1} - y_i = k\, Ay_i - ky_i^2$, $i = 1, ..., n\text{-}1$.

Обозначим через а произведение kA и используем МНК для оценки параметров:

$$Q(a,k) = \sum_{i=1}^{n-1}(\Delta y_i - ay_i + ky_i^2) \Rightarrow min$$

Если функционал Q (a,k) в точке (a, k) достигает экстремума, то в этой точке частные производные по параметрам равны нулю:

$$\frac{\partial Q(a,k)}{\partial a} = -2\sum_{i=1}^{n}(\Delta y_i - ay_i + ky_i^2)y_i,$$

$$\frac{\partial Q(a,k)}{\partial k} = -2\sum_{i=1}^{n}(\Delta y_i - ay_i + ky_i^2)y_i.$$

После преобразований получим систему нормальных уравнений для оценки параметров a и k:

$$\left.\begin{aligned} a\sum_{i=1}^{n-1} y_i^2 - k\sum_{i=1}^{n-1} y_i^3 = \sum_{i=1}^{n-1} y_i \Delta y_i \\ a\sum_{i=1}^{n-1} y_i^3 - k\sum_{i=1}^{n-1} y_i^4 = \sum_{i=1}^{n-1} y_i^2 \Delta y_i \end{aligned}\right\}$$

После оценки параметров a и k получим A = a/k. После интегрирования дифференциального уравнения Бернулли получим логистическое уравнение

$$Y_p = A/(1+ce^{-at}),$$

где постоянная c — неизвестна.

Для получения оценки параметра c преобразуем уравнение логистической регрессии.

$$\ln c = \ln(A/Y - 1) + at.$$

После подстановки всех значений динамического ряда в полученное выражение и суммирования этих значений получим:

$$n \ln c = \sum_{i=1}^{n} \ln(\frac{A}{y_i} - 1) + a\sum_{i=1}^{n} t_i,$$

$$c = \exp(\frac{1}{n}\sum_{i=1}^{n} \ln(\frac{A}{y_i} - 1) + \frac{a}{n}\sum_{i=1}^{n} t_i).$$

Рассмотрим характер изменения регрессии при $t \to \infty$

$$\lim_{t\to\infty} \frac{A}{1+ce^{-at}} = A.$$

То есть $У_p = A$ — прямая, которая является горизонтальной асимптотой.

С точки зрения экономики после некоторого времени уровень обеспеченности товарами будет стабильно ассимптотически стремится к значению A.

Предложенные формулы расчета коэффициентов дают несмещенные оценки параметров при малых значениях прироста Δy_i, Δt_i. С увеличением значений прироста увеличивается ошибка оценки параметров.

Задание 2. Вычислите характеристики следующей модели:

$$У_p = a_0 + a_1/X + a_2(\ln X)^4$$

Пояснение. Пример решения задачи представлен на рис. 4.13.

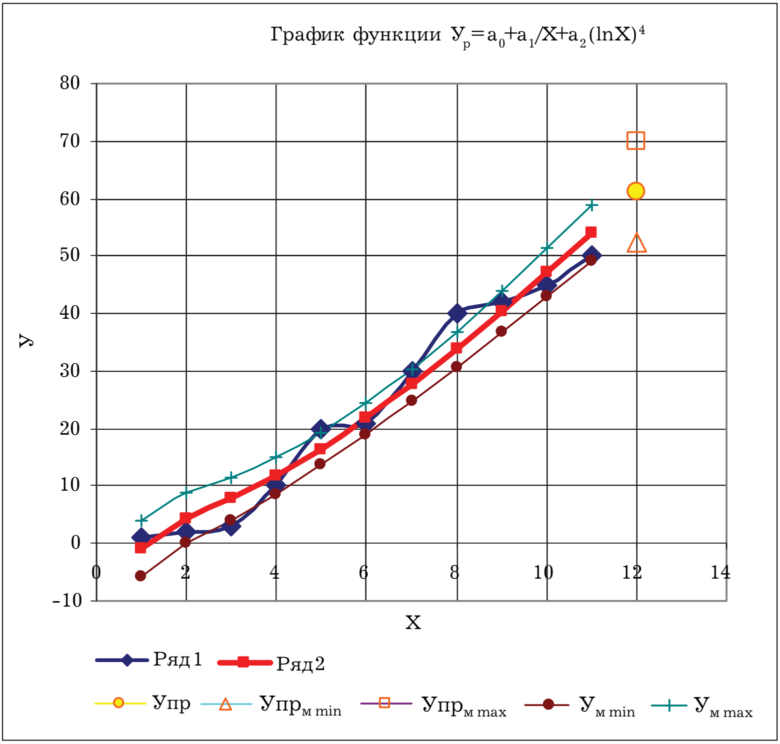


Рис. 4.13. Графическое представление результатов расчетов

Задание четвертого уровня сложности

Задание. Необходимо вычислить коэффициенты различных функций с использованием пакетов прикладных программ: Curve Exp (рис. 4.14) и Table Curve 2D (рис. 4.15).

Нерешенные проблемы по теме спецификация модели:

— необходимо найти такую модель, которая описывала бы только детерминированную составляющую зависимой переменной,

— нет удовлетворительных методик, с помощью которых можно выделить детерминированную и случайную составляющую зависимой переменной.

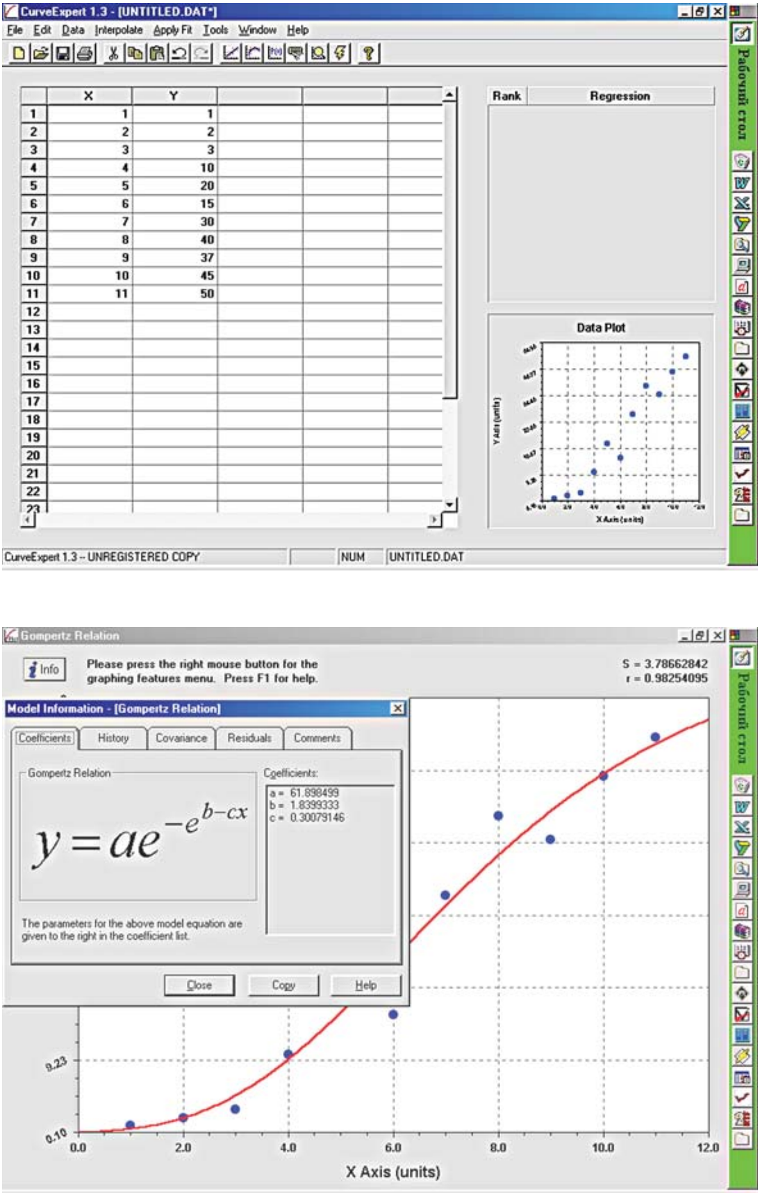


Рис. 4. 14. Протоколы расчетов по программе Curve Exp

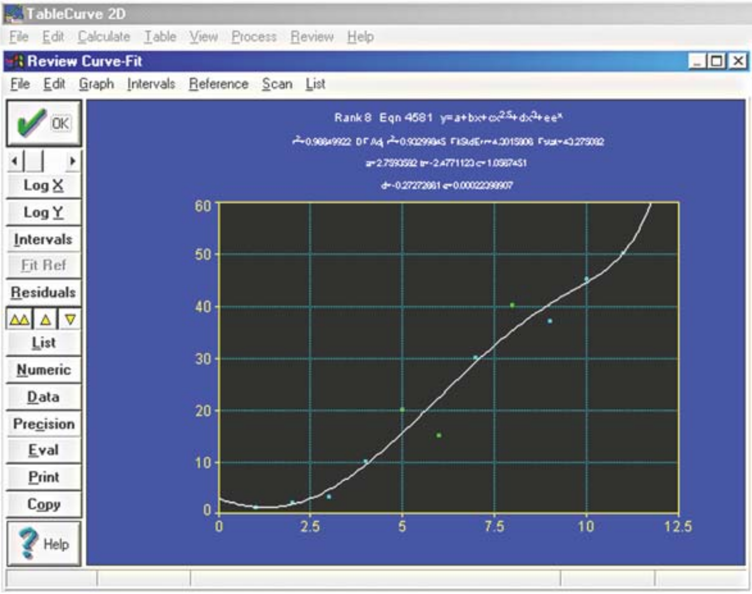


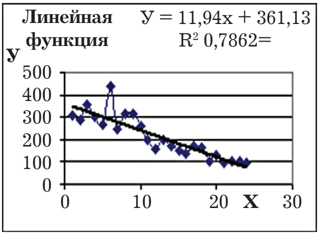
Рис. 4. 15. Протокол расчетов по программе Table Curve 2D

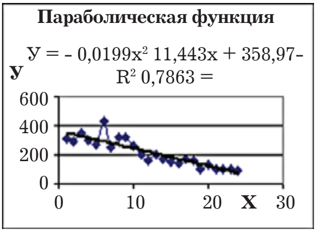
Дискуссия. Является спорным следующее утверждение — чем меньше ошибка модели, тем лучше модель. Зависимая переменная состоит из детерминированной и случайной составляющей. Если модель описывает детерминированную и часть случайной составляющей, то ошибка модели будет маленькой, но прогноз по этой модели будет хуже чем прогноз, полученный по модели, которая описывает только детерминированную составляющую, но с большей ошибкой модели.

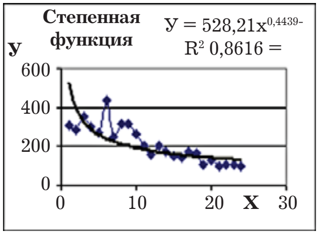
Выходной тест

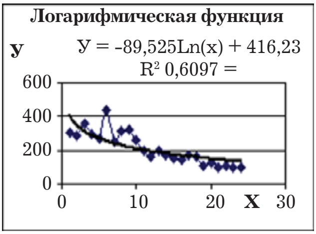
Имеются данные для Козельщенского Потребительского общества, Козельщенского Райсоюза, Полтавской области и их графики.

Х	У
1	306
2	288
3	355
4	300
5	270
6	435
7	247
8	316
9	318
10	262
11	200
12	160
13	201
14	172
15	148
16	140
17	173
18	162
19	104
20	129
21	99
22	103
23	102
24	95









где Х — порядковый номер месяца за два условных года;
У — налог в фонд Чернобыля (грн.),
На графиках представлены пять видов функций:
— линейная;
— логарифмическая;
— параболическая;
— степенная;
— экспоненциальная;

Определите лучшую функцию, по критерию ?детерминации к^2?, которая хорошо воспроизводит тенденцию динамики налогов в фонд Чернобыля.

5. ОБОБЩЕННАЯ ЛИНЕЙНАЯ МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ. МУЛЬТИКОЛЛИНЕАРНОСТЬ

Постановка задачи

Объект — показатели деятельности экономического объекта.

Предмет — зависимость результирующего признака от нескольких факторов.

Цель — получить навыки построения многофакторной модели при наличии мультиколлинеарности между факторами, определить факторы для включения в модель.

Актуальность — сильная мультиколлинеарность не позволяет определить коэффициенты и характеристики регрессионной модели.

Метод — анализ парных и частных коэффициентов корреляции.

Способ — построение корреляционных плеяд.

Задачи:

- построить графики зависимости между У и факторами;
- произвести анализ регулярностей связи между переменными;
- построить корреляционные плеяды;
- определить факторы для включения в модель;
- проверить достоверность модели и ее коэффициентов;
- получить точечный и интервальный прогноз;
- представить результаты расчетов в графическом виде.

Рабочая гипотеза — предполагается, что в экономической системе все переменные взаимосвязаны между собой.

Ожидаемый результат — ожидается, что можно определить факторы для включения в модель при соблюдении условия построения модели, — факторы, включаемые в модель должны быть сильно связаны с У и слабо связаны между собой.

Метод для сравнения — построение многофакторной модели, которая включает все изучаемые факторы.

Задания первого уровня сложности

Задание 1. Подготовьте устные сообщения по вопросам самостоятельной работы.

1. Приведите общий вид обобщенной линейной множественной регрессии.

2. Дайте определение мультиколлинеарности и приведите методы ее устранения.

3. Приведите метод корреляционных плеяд построения множественной модели.

4. Опишите суть шаговой регрессии, используемой для построения множественной регрессии.

5. Приведите пакет прикладных программ, в котором имеется построение модели с помощью шаговой регрессии.

[3, с. 71–72, 207–223, 7, с.108–115, 9, т. 2, с. 74–88, 8, с.92–94, 10, с.271–284, 12, с. 134–165, 1 т. 1, с.577–580].

Задание 2. Пройдите входной тест.

Выберите вариант правильного ответа.

1. Общий вид множественного регрессионного уравнения для выборочных данных можно представить следующим математическим выражением:

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

где $У_i$ — случайная зависимая переменная;

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

$f(X_{1i}, X_{2i}, X_{3i}, \dots)$ — математическая функция, отражающая детермини-рованную 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 вошедших в модель;

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

n — объем выборки.

а) да;

б) нет.

2. Мультиколлинеарность — это свойство переменных объектов изучения, которое заключается в том, что:

а) существует хотя бы одна пара переменных, между которыми имеется существенная связь, определяемая по коэффициенту корреляции;

б) не существует хотя бы одной пары переменных, между которыми имеется существенная связь, определяемая по коэффициенту корреляции.

3. Последствия мультиколлинеарности заключаются в том, что при построении многофакторной модели возникают следующие проблемы в определении коэффициентов и качества модели.

а) Нельзя определить коэффициенты модели по формуле

$$\mathbf{A = (X'X)^{-1} X'У}$$

так как нельзя найти обратную матрицу

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

в которой матрица факторов X содержит неортогональные переменные;

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 взаимосвязанных факторов может быть признаком обнаружения мультиколлинеарности.

а) да;

б) нет.

4. Известно два направления устранения мультиколлинарности:

— методы, которые используют преобразования данных;
— методы, которые используют разные приемы построения моделей.

а) да;

б) нет.

5. Известны следующие часто используемые методы построения моделей при наличии мультиколлинеарных факторов:

— шаговый метод построения модели;
— метод корреляционных плеяд.

а) да;

б) нет.

6. Шаговый метод включения построения модели заключается в том, что на первом шаге в модель вводится фактор, который по коэффициенту корреляции сильнее всего влияет на У, ищутся остатки и характеристики модели, если модель является достоверной, то выполняется следующий шаг — в модель включается фактор, который сильнее всего по коэффициенту корреляции связан с остатками модели, ищутся остатки модели, при условии, что в модели имеется два фактора, проверя-

ется достоверность модели, если модель является достоверной, то выполняется следующий шаг и так далее до тех пор, пока модель станет недостоверной. Недостоверность модели означает то, что в модель был включен фактор сильно связанный с теми факторами, которые уже находятся в модели.

а) да;

б) нет.

7. Метод корреляционных плеяд заключается в том, что на круговой диаграмме размещают обозначения факторов, линией соединятся те факторы, между которыми существует достоверный коэффициент корреляции, по отношению к которым применяют правило построения многофакторной модели: “факторы, включенные в модель должны быть сильно связаны с У и слабо связаны между собой”.

а) да;

б) нет.

8. Шаговая регрессия имеется в пакете прикладных программ STADIA 6

а) да;

б) нет.

Задание 3. Решите задачу.

Имеются данные по консервному заводу Райсоюза южной области РФ за каждый месяц условного года. Численность работников завода составила 1135 человек, в том числе производственных работников 843 человека (табл 5.1).

Таблица 5.1

База данных консервного завода

i	X_{1i}	X_{2i}	X_{3i}	X_{4i}	$У_i$
1	1	328	0,054	0,3	397
2	2	329	0,101	0,6	670
3	3	329	0,099	1,2	1209
4	4	347	0,019	0,1	138

i	X_{1i}	X_{2i}	X_{3i}	X_{4i}	$У_i$
5	5	352	0,065	0,3	373
6	6	370	0,053	0,1	79
7	7	378	0,178	2,3	1883
8	8	385	0,174	2,6	2124
9	9	396	0,298	5,5	5069
10	10	399	0,195	2,4	2618
11	11	390	0,102	1,6	1265
12	12	373	0,138	0,6	562
Ожидаем	13	392	0,142	0,72	?
Сумма	78	4376	1,476	17,6	16387
Среднее	6,5	364,67	0,123	1,46667	1365,6

где X_1 — время, номера месяцев,

X_2 — фондовооруженность (тыс. руб./чел);

X_3 — фондоотдача (тыс. руб. объема товар. прод./тыс. руб. осн.фондов);

X_4 — производительность труда (туб/чел.);

У — валовая продукция (туб.);

i — порядковый номер измерения;

туб.- тысяча условных банок.

Необходимо построить многофакторную модель, в которой будет соблюдаться основное правило построения модели “Факторы, включенные в модель, должны быть сильно связаны с зависимой переменной и слабо связаны между собой”.

Решение задачи необходимо произвести с использованием следующих методов:

- метода анализа парных и частных коэффициентов корреляции с использованием метода корреляционных плеяд.
- метода шагового регрессионного анализа.

Решение

Завод на протяжении нескольких лет работает стабильно. Однако по сравнению с лучшими консервными заводами имеются следующие проблемы: мал ассортимент продукции, высокая цена на продукцию, низкая производительность труда. Целью моделирования является определить за счет каких факторов можно увеличить объемы валовой продукции.

Выбор переменных

Величина валовой продукции зависит от шести основных факторов: *сырье и материалы* (цена и качество перерабатываемого сырья); *машины* (оборудование завода); *методы* (технология переработки сырья, степень использования современных технологий); *люди* (уровень квалификации, образование, возраст, деловые характеристики); *физическая и абстрактная среда* (температура, влажность, уровень шума, освещенность, наличие комнаты отдыха на заводе; традиции, психологический климат, существующие на заводе; правовая среда существования завода, информационное обеспечение завода); *время* (тенденции динамического развития завода, сезонность) (рис. 5.1).

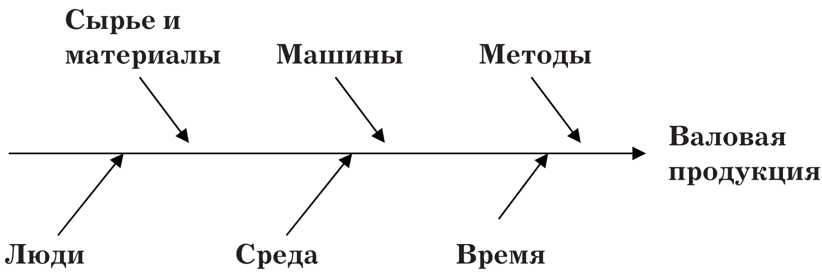


Рис. 5.1. Причинно-следственная диаграмма Исикавы

Выделение входных и выходных переменных

Консервный завод может иметь три вида показателей: входные, выходные и производные.

Входными показателями предприятия могут быть характеристики: машин, материалов, методов, персонала, абстрактной и физической среды, в которой действует предприятие.

Выходными показателями предприятия могут быть объемы конечной продукции, выраженные в натуральных или денежных единицах.

Производные показатели предприятия получаются делением выходных показателей на входные.

Обычно для целей управления производством определяется зависимость выходных показателей от входных. Однако в данной задаче определяется зависимость выходного показателя от производных: фондовооруженности, фондоотдачи и производительности труда, которые являются одними из основных показателей деятельности завода, и важно знать, как они влияют на конечный продукт.

Сбор статистических данных

Деятельность завода анализировалась на протяжении одного года. Основные показатели завода представлены в таблице 6.1. Временные ряды показателей изображены на рис. 5.2.

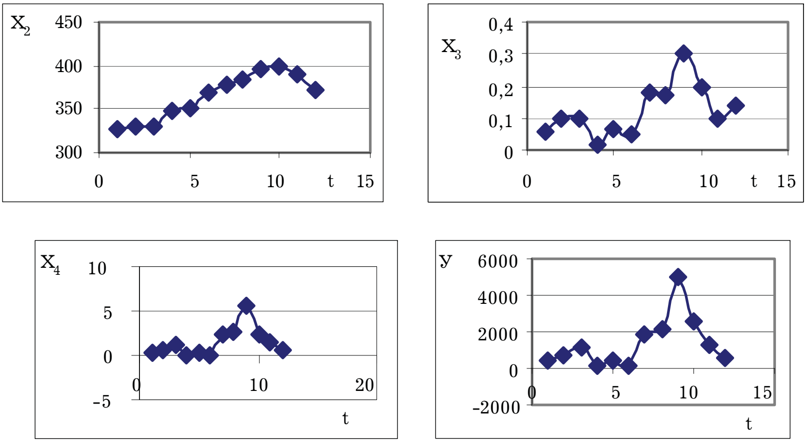


Рис. 5.2. Временные ряды показателей завода

Выдвижение гипотез

Объем валовой продукции должен зависеть от основных показателей деятельности завода. Между показателями завода должна присутствовать мультиколлинеарность, вызванная ложной корреляцией.

Между объясняемыми переменными возникает два вида ложной корреляции ложная корреляция Пирсона и ложная корреляция временных рядов.

Ложная корреляция Пирсона появляется между производными переменными, при расчетах которых используется общая переменная. Например, между фондовооруженностью и фондоотдачей существует ложная корреляция Пирсона, так как в расчетах производных переменных используется общая переменная — размер основных фондов. Между фондовооруженностью и производительностью труда существует ложна корреляция Пирсона, так как в расчетах производных переменных используется общая переменная — количество работников. Между производительностью труда и валовой продукцией существует ложная корреляция Пирсона, так как в расчетах производительности труда используется размер валовой продукции.

Все переменные измеряются во времени, поэтому между ними может возникнуть ложная корреляция временных рядов, так как время является общим фактором, влияющим на переменные.

Формулировка допущений

Если между факторами, включенными в модель существует небольшая мультиколлинеарность, то она не окажет существенного влияния на качество модели.

Спецификация

Для устранения ложной корреляции временных рядов фактор времени нужно обязательно включать в множественную модель.

Так как между фондовооруженностью и фондоотдачей, между фондовооруженностью и производительностью труда имеется ложная корреляция Пирсона, то это должно привести к появлению мультиколлинарности между объясняемыми переменными. Следовательно, возникнет проблема выбора факторов для включения в модель.

Обобщенная линейная модель множественной регрессии

Предполагаем, что зависимость У от X можно представить в виде обобщенной линейной модели множественной регрессии. Приводим спецификацию данной модели:

$\mathbf{У} = \mathbf{XA} + \varepsilon$, при условии, что

$M(\varepsilon) = 0$, $M(\varepsilon\varepsilon') = \sigma_\varepsilon{}^2 I$,

где $\mathbf{X}$ — матрица имеет размерность X_{nd}, n — число наблюдений,

d — количество всех объясняемых переменных, включая столбец, состоящий из единиц.

rang(**X**) = d<n — ранг матрицы **X** должен равняться d, количество переменных должно быть меньше числа измерений.

Условие — ранг матрицы **X** должен равняться количеству ее столбцов означает, что между объясняющими переменными не должно быть мультиколлинарности.

Идентифицируемость модели

(Можно ли в принципе однозначно восстановить значения неизвестных параметров по выбранной модели).

Если среди всех факторов можно выбрать хотя бы один, достоверно влиящий на зависимый признак, то модель будет построена.

Идентификация модели

(Предложить и реализовать процедуру оценивания параметров модели).

Метод анализа парных и частных коэффициентов корреляции с использованием метода корреляционных плеяд

Вычислим матрицу парных коэффициентов корреляции. по которым определим факторы для совместного включения в модель.

Таблица 5.2

Матрица парных коэффициентов корреляции

R =

	X_1	X_2	X_3	X_4	У
X_1	1,00	0,89	0,56	0,46	0,43
X_2	0,89	1,00	0,68	0,66	0,63
X_3	0,56	0,68	1,00	0,94	0,94
X_4	0,46	0,66	0,94	1,00	0,99
У	0,43	0,63	0,94	0,99	1,00

Пояснение. Матрицу парных коэффициентов корреляции можно вычислить в такой последовательности:

1) скопируйте данные табл. 5.1. и вставьте их в лист Excel;

2) выполните действия в среде Excel: сервис, анализ данных, корреляция, ОК;

3) укажите входной диапазон данных табл. 5.1, укажите выходной диапазон — куда вывести матрицу коэффициентов корреляции, ОК;

4) получите результат, представленный в табл. 5.2.

Вычислим критическое значение коэффициента корреляции $r_{кр}$.

$$r_{кр} = \frac{t_{\alpha/2}}{\left(t_{\alpha/2}{}^2 + n - 2\right)^{1/2}} = 0,57598,$$

где $t_{\alpha/2} = t_{\alpha/2}(\alpha = 0,05;\ m = n - 2) = 2,22814$ — табличное значение критерия Стьюдента;

$n = 12$.

Проведем анализ взаимосвязи между переменными по парным коэффициентам корреляции.

$|r\ (X_1,X_2)| = 0,89 > r_{кр} = 0,58$,переменные X_1 и X_2 связаны между собой

$|r\ (X_1,X_3)| = 0,56 < r_{кр} = 0,58$,переменные X_1 и X_3 не связаны между собой

$|r\ (X_1,X_4)| = 0,46 < r_{кр} = 0,58$,переменные X_1 и X_4 не связаны между собой

$|r\ (X_1,У)| = 0,43 < r_{кр} = 0,58$,переменные X_1 и У не связаны между собой

$|r\ (X_2,X_3)| = 0,68 > r_{кр} = 0,58$,переменные X_2 и X_3 связаны между собой

$|r\ (X_2,X_4)| = 0,66 > r_{кр} = 0,58$,переменные X_2 и X_4 связаны между собой

$|r\ (X_2,У)| = 0,63 > r_{кр} = 0,58$,переменные X_2 и У связаны между собой

$|r\ (X_3,X_4)| = 0,94 > r_{кр} = 0,58$,переменные X_3 и X_4 связаны между собой

$|r\ (X_3,У)| = 0,94 > r_{кр} = 0,58$,переменные X_3 и У связаны между собой

$|r\ (X_4,У)| = 0,99 > r_{кр} = 0,58$,переменные X_4 и У связаны между собой

Представим в графическом виде достоверную взаимосвязь между переменными, определенную по парным коэффициентам корреляции. Данный прием получил название метода корреляционных плеяд (рис. 5.3).

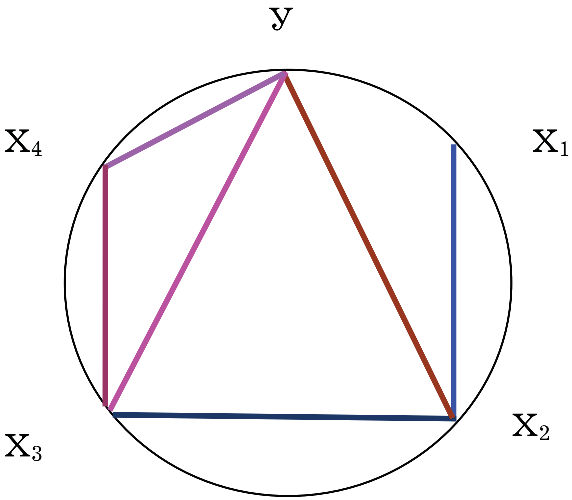


Рис.5.3. Корреляционные плеяды, определенные по парным коэффициентам корреляции

Вывод. Анализ парных коэффициентов корреляции показывает, что факторы X_2, X_3, X_4 достоверно связаны с У. Факторы X_1 не оказывают достоверного влияния ????. Факторы X_2 и X_3, X_3 и X_4 достоверно связаны между собой. Множественная модель может включать факторы X_2 или X_3, X_3 или X_4. Очевидно, модель должна включать факторы X_2 и X_4.

Вычислим матрицу частных коэффициентов корреляции, по которым определим факторы для совместного включения в модель.

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:

$$r_ч(X_i, X_j) = \frac{-Cij}{\sqrt{Cii \times Cjj}},$$

где $r_ч(X_i,X_j)$ — частный коэффициент корреляции,

C = R^{-1} — обратная матрица от матрицы **R**.

Пояснение. Обратную матрицу **C = R.**$^{-1}$ можно вычислить в Excel с помощью математической функции МОБР.

R — матрица парных коэффициентов корреляции, которая состоит из коэффициентов корреляции между всеми факторами.

$$\mathbf{C} = \mathbf{R}^{-1} = \begin{bmatrix} 7{,}2744 & -6{,}884 & -4{,}7803 & 1{,}5443 & 4{,}2089 \\ -6{,}884 & 8{,}5368 & 3{,}08901 & -4{,}469 & -0{,}9262 \\ -4{,}7803 & 3{,}089 & 13{,}6591 & -2{,}477 & -10{,}342 \\ 1{,}5443 & -4{,}469 & -2{,}4775 & 65{,}923 & -60{,}865 \\ 4{,}2089 & -0{,}926 & -10{,}342 & -60{,}87 & 69{,}897 \end{bmatrix}$$

Матрица частных коэффициентов корреляции.

	X_1	X_2	X_3	X_4	У
X_1	-1,00	0,87	0,48	-0,07	-0,19
X_2	0,87	-1,00	-0,29	0,19	0,04
X_3	0,48	-0,29	-1,00	0,08	0,33
X_4	-0,07	0,19	0,08	-1,00	0,90
У	-0,19	0,04	0,33	0,90	-1,00

Проведем анализ взаимосвязи между переменными по частным коэффициентам корреляции.

$|r(X_1,X_2)| = 0{,}87 > r_{кр} = 0{,}576$, переменные X_1 и X_2 связаны между собой

$|r(X_1,X_3)| = 0{,}48 < r_{кр} = 0{,}576$, переменные X_1 и X_3 не связаны между собой

$|r(X_1,X_4)| = 0{,}07 < r_{кр} = 0{,}576$, переменные X_1 и X_4 не связаны между собой

$|r(X_1,У)| = 0{,}19 < r_{кр} = 0{,}576$, переменные X_1 и У не связаны между собой

$|r(X_2,X_3)| = 0{,}29 < r_{кр} = 0{,}576$, переменные X_2 и X_3 не связаны между собой

$|r(X_2,X_4)| = 0{,}19 < r_{кр} = 0{,}576$, переменные X_2 и X_4 не связаны между собой

$|r(X_2,У)| = 0{,}04 < r_{кр} = 0{,}576$, переменные X_2 и У не связаны между собой

$|r(X_3,X_4)| = 0{,}08 < r_{кр} = 0{,}576$, переменные X_3 и X_4 не связаны между собой

$|r(X_3,У)| = 0{,}33 < r_{кр} = 0{,}576$, переменные X_3 и У не связаны между собой

$|r(X_4,У)| = 0{,}90 > r_{кр} = 0{,}576$, переменные X_4 и У связаны между собой

Представим в графическом виде достоверную взаимосвязь между переменными, определенную по частным коэффициентам корреляции (рис. 5.4).

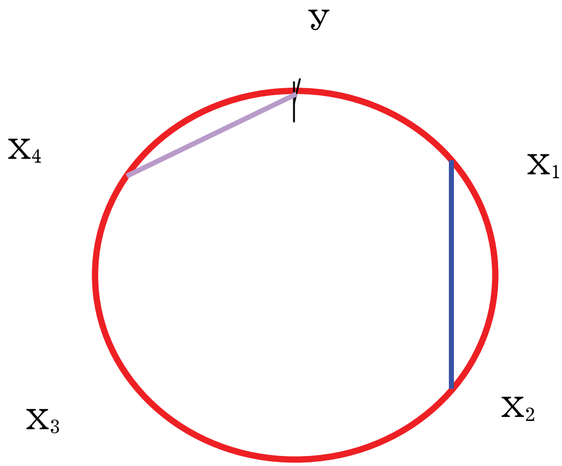


Рис. 5.4. Корреляционные плеяды, определенные по частным коэффициентам корреляции.

Вывод. Анализ частных коэффициентов корреляции показывает, что фактор X_4 достоверно связаны с У. Факторы X_1 и X_2 достоверно связаны между собой.

Множественная модель может включать фактор X_4.

Сравним характеристики модели, которые включают факторы X_2, X_4, а также все факторы X_1, X_2, X_3, X_4 (табл. 5.3).

Таблица 5.3

База данных для модели, которая включает факторы: X_2, X_4

i	X_{2i}	X_{4i}	$У_i$	$У_{pi}$
1	328	0,3	397	356,814
2	329	0,6	670	630,894
3	329	1,2	1209	1182,54
4	347	0,1	138	139,835
5	352	0,3	373	315,006
6	370	0,1	79	99,7691
7	378	2,3	1883	2108,53
8	385	2,6	2124	2372,16
9	396	5,5	5069	5019,28
10	399	2,4	2618	2163,89
11	390	1,6	1265	1444,04
12	373	0,6	562	554,247
Ожидаем	392	0,72		631,478
Сумма	4376	17,6	16387	
Среднее	364,7	1,4667	1365,6	

где X_{2i} — фондовооруженность (тыс. руб./чел);

X_{4i} — производительность труда (туб/чел.);

$У_i$ — валовая продукция (туб.), туб.- тысяча условных банок;

i — порядковый номер измерения.

$У_{pi} = a_0 + a_2 \bullet X_2 + a_4 \bullet X_4$

Пояснение. Расчеты многофакторной модели производятся с помощью функции “Линейн”.

919,4076	-1,742	652,36
50,97589	2,9983	1047,36
0,983872	200,25	#Н/Д
274,5152	9	#Н/Д
22015716	360893	#Н/Д

Представим результаты расчетов в общепринятом обозначении.

$У_{pi}$ =	652,4	-1,742	X_2 +	919,41	X_4
S_a =	1047	2,9983		50,976	
t_a =	0,623	-0,581		18,036	
E =	200,2				
R^2 =	0,984			$t_{\alpha/2}$=	2,26216
F =	274,5			$F_{кр}$=	4,25649

Верификация модели

Представляем результаты эконометрического анализа.

1. Определение доли объясненной вариации.

98,39% исходных данных имеют линейную тенденцию.

2. Проверка достоверности модели.

По критерию Фишера достоверность модели статистически доказана.

Так как $F > F_{кр}(\alpha = 0{,}05;\ m_1 = k-1,\ m_2 = n-k)$,

то модель является достоверной с вероятность 0,95.

3. Проверка достоверности коэффициентов.

По критерию Стьюдента проверим достоверность коэффициентов модели.

Так как $|t_{a0}| > t_{\alpha/2}$, то коэффициент a_0 с вероятностью 0,95 отличается от нуля.

Так как $|t_{a1}| < t_{\alpha/2}$, то коэффициент a_1 не отличается от нуля.

Так как $|t_{a2}| < t_{\alpha/2}$, то коэффициент a_2 не отличается от нуля.

Определение доли влияния факторов

В эконометрическом анализе важно знать долю влияния факторов, которые включены в модель. Это можно определить по формуле

$$r_{X_2}^{\ 2} = \frac{t_{X_2}^2 \quad R^2}{t_{X_2}^{\ 2} + t_{X_4}^{\ 2}},$$

где $r_{X2}^{\ 2}$ — коэффициент детерминации для фактора X_2,

t_{X4} — критерий Стьюдента для фактора X_4,

t_{X2} — критерий Стьюдента для фактора X_2,

R^2 — общий коэффициент детерминации,

$r_{X2}^{\ 2} = 0{,}001$ — доля влияния фактора X_2,

$r_{X4}^{\ 2} = 0{,}9829$ — доля влияния фактора X_4,

Сумма 0,9839 — доля влияния всех факторов.

4. Получение точечного и интервального прогноза

При ожидаемых значениях объясняемых переменных на период t = 13

$X_{2ож} = 392$, $X_{4ож} = 0{,}72$

Точечный прогноз будет равен

$У_{пр} = 631{,}48$

Прогнозное значение зависимой переменной с вероятностью 0,95 будет находиться в интервале от 178,49 до 1084,5.

5. Графическое представление результатов расчета (рис. 5.5).

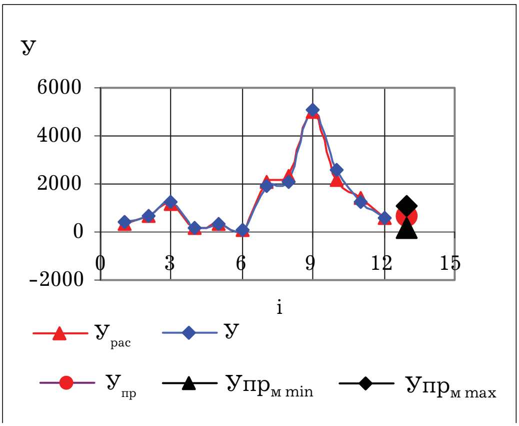


Рис. 5.5. Зависимость У от X_2 и X_4

Таблица 5.4

База данных для модели, которая включает факторы:X_1, X_2, X_3, X_4

i	X_{1i}	X_{2i}	X_{3i}	X_{4i}	$У_i$	$У_p$
1	1	328	0,054	0,3	397	362,06
2	2	329	0,101	0,6	670	703,13
3	3	329	0,099	1,2	1209	1148,1
4	4	347	0,019	0,1	138	51,336
5	5	352	0,065	0,3	373	313,5
6	6	370	0,053	0,1	79	111,9
7	7	378	0,178	2,3	1883	2170,4
8	8	385	0,174	2,6	2124	2377,8
9	9	396	0,298	5,5	5069	4989
10	10	399	0,195	2,4	2618	2238,8
11	11	390	0,102	1,6	1265	1324,9
12	12	373	0,138	0,6	562	596,1
Ожидаем	13	392	0,142	0,72	?	691,38
Сумма	78	4376	1,476	17,6	16387	
Среднее	6,5	364,67	0,123	1,46667	1365,6	

где X_1 — время (месяцы).

X_2 — фондовооруженность (тыс. руб./чел);

X_3 — фондоотдача (тыс. руб. объема товар. прод./тыс. руб. осн.фондов);

X_4 — производительность труда (туб/чел.);

У — валовая продукция (туб.), туб.- тысяча условных банок;

i — порядковый номер измерения.

Протокол расчетов по функции Excel “Линейн” имеет ??????????щий вид:

790,298	2704	0,7073	-23,82	-229,24
147,497	2877	7,0461	47,386	2297,67
0,98569	213,9	#Н/Д	#Н/Д	#Н/Д
120,569	7	#Н/Д	#Н/Д	#Н/Д
2,2E+07	3E+05	#Н/Д	#Н/Д	#Н/Д

Представим результаты расчетов в общепринятом обозначении.

У =	-229,2	-23,82	X_1 +	0,70734	X_2 +	2704,1	X_3 +	790,3	X_4
S_a=	2298	47,386		7,04607		2877,3		147,5	
t_a=	0,1	0,5027		0,10039		0,9398		5,3581	
E=	213,9								
R^2=	0,9857		$t_{\alpha/2}$=	2,36462					
F=	120,6		$F_{кр}$=	4,12031					

Представляем результаты эконометрического анализа.

1. Определение доли объясненной вариации.

98,57% данных имеют линейную тенденцию.

2. Проверка достоверности модели.

По критерию Фишера достоверность модели статистически доказана.

Так как $F > F_{кр}(\alpha = 0{,}05;\ m_1 = k — 1 = 5\text{-}1;\ m_2 = n — k = 12-5)$,

то модель является достоверной с вероятность 0,95.

3. Проверка достоверности коэффициентов.

По критерию **Стьюдента** проверим достоверность коэффициентов модели

Так как $|t_{a0}| < t\alpha$, то коэффициент a_0 не отличается от нуля.

Так как $|t_{a1}| < t\alpha$, то коэффициент a_1 не отличается от нуля.

Так как $|t_{a2}| < t\alpha$, то коэффициент a_2 не отличается от нуля.

Так как $|t_{a3}| < t\alpha$, то коэффициент a_3 не отличается от нуля.

Так как $|t_{a4}| > t\alpha$, то коэффициент a_4 с вероятностью 0,95 отличается от нуля.

Определим долю влияния факторов, включенных в модель.

$r(X_1)^2 = 0{,}0083$ — доля влияния фактора X_1,

$r(X_2)^2 = 0{,}0003$ — доля влияния фактора X_2,
$r(X_3)^2 = 0{,}0292$ — доля влияния фактора X_3,
$r(X_4)^2 = 0{,}9479$ — доля влияния фактора X_4.
Сумма 0,9857 — доля влияния всех факторов.

4. Получение точечного и интервального прогноза.

При ожидаемых значениях объясняемых переменных на период $t = 13$

$$X_{1ож} = 13,\ X_{2ож} = 392,\ X_{3ож} = 0{,}142,\ X_{4ож} = 0{,}72$$

Точечный прогноз или среднее значение зависимой переменной будет составлять $У_{пр} = 691{,}38$

Прогнозное значение зависимой переменной с вероятностью 0,95 будет находиться в интервале от 185,69 до 1197,1.

5. Графическое представление результатов расчета (рис. 5.6).

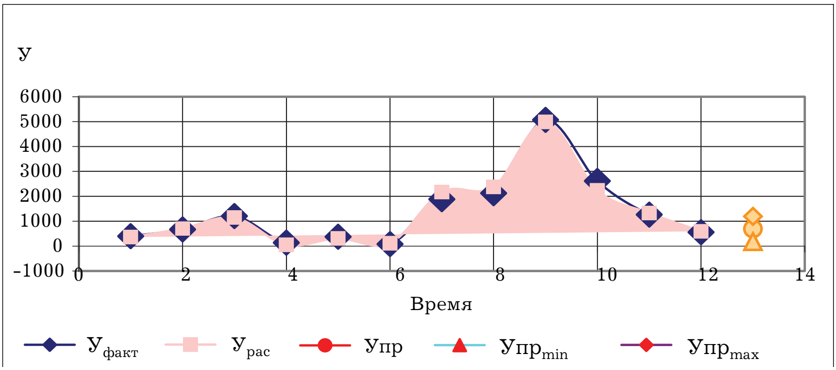


Рис. 5.6. График зависимости У от факторов X_1, X_2, X_3, X_4

Вывод. Эконометрический анализ полученной модели показывает, что модель является достоверной, однако коэффициенты a_1, a_2 и a_3 не отличаются от нуля.

Коэффициент a_2 имеет отрицательный знак, что противоречит экономическому содержанию, так как снижение фондовооруженности не должно приводить к увеличению валовой продукции. Обнаруженные свойства модели являются следствием наличия мультиколлинеарности между переменными, возникшей за счет взаимосвязи между X_1 и X_2, а также между X_3 и X_4.

Вывод. Модель, которая включает все факторы, имеет характеристики хуже, чем модель, включающая только два фактора.

Построение модели шаговой регрессии с помощью пакета прикладных программ Stadia 6

Примечание. До появления шаговой регрессии использовались разные схемы включения факторов в модель в зависимости от субъективного мнения исследователя, при этом получались разные модели с разными характеристиками, возникали проблемы выбора лучшей модели и исключения субъективизма исследователя[1]. Шаговая регрессия позволяет однозначно построить достоверную модель при наличии мультиколлинеарности между изучаемыми факторами[2].

Построение многофакторной модели с помощью шаговой регрессии проводится в такой последовательности.

1) запускаем программу Stadia 6 с помощью файла Stadia.exe;

2) вводим данные (рис. 5.7).

Примечание. Разделителем дробной части числа служит точка;

3) F9 — входим в статистические методы, выбираем пошаговую регрессию (рис. 5.8), выделяем факторы и зависимую переменную (рис. 5.9), утвердить;

4) устанавливаем селекцию переменных таким образом, чтобы последовательно включать все переменные (рис 5.10);

5) получаем протоколы расчетов (рис. 5.11) (А, Б, В, Г).

Анализ протоколов расчетов рис. 5.11 показывает, что на первом шаге была включена переменная X_4. Полученная модель является достоверной по критерию Фишера $F = 587{,}9$. При

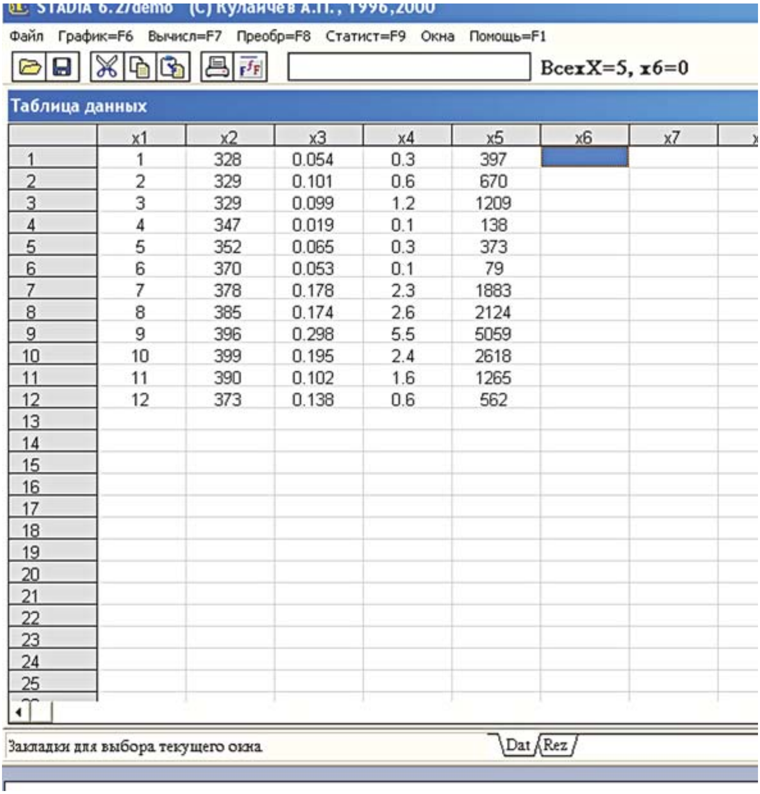
---

[1] Дукарский О. М. Закурдаев А.Г. Статистический анализ и обработка наблюдений на ЭВМ “Минск 32”. — М.: Статистика, 1971.

[2] Дрейпер Н., Смит Г. Прикладной регрессионный анализ. — М.: Статистика, 1973. (Имеется алгоритм Эфроимсона шаговой регрессии с контрольным примером).

последовательном включении факторов X_3 и X_1 характеристики модели ухудшаются, критерий Фишера уменьшается, что является следствием наличия мультиколлинеарности между факторами.

Если изменить селекцию переменных по критерию Фишера (F = уровень), как показано на рис. 5.10, то после включения в модель фактора X_4 построение модели заканчивается (рис. 5.12) (А, Б).



	x1	x2	x3	x4	x5	x6	x7
1	1	328	0.054	0.3	397		
2	2	329	0.101	0.6	670		
3	3	329	0.099	1.2	1209		
4	4	347	0.019	0.1	138		
5	5	352	0.065	0.3	373		
6	6	370	0.053	0.1	79		
7	7	378	0.178	2.3	1883		
8	8	385	0.174	2.6	2124		
9	9	396	0.298	5.5	5059		
10	10	399	0.195	2.4	2618		
11	11	390	0.102	1.6	1265		
12	12	373	0.138	0.6	562		

Рис. 5.7. База данных

Рис. 5.8. Меню статистических методов

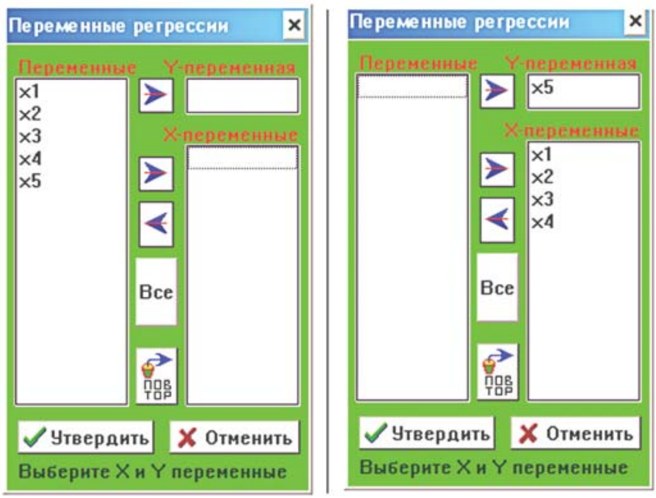
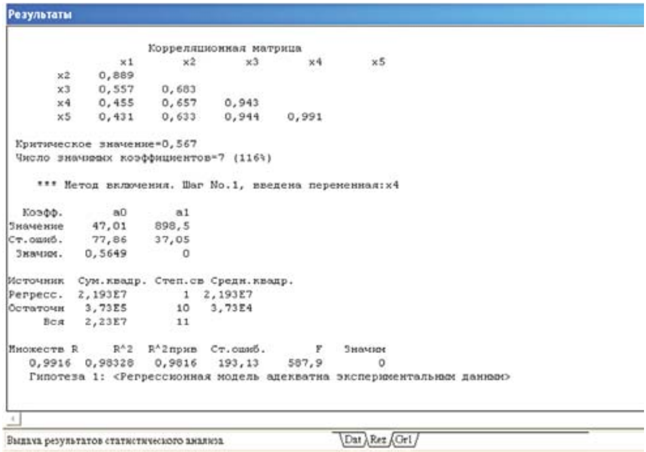


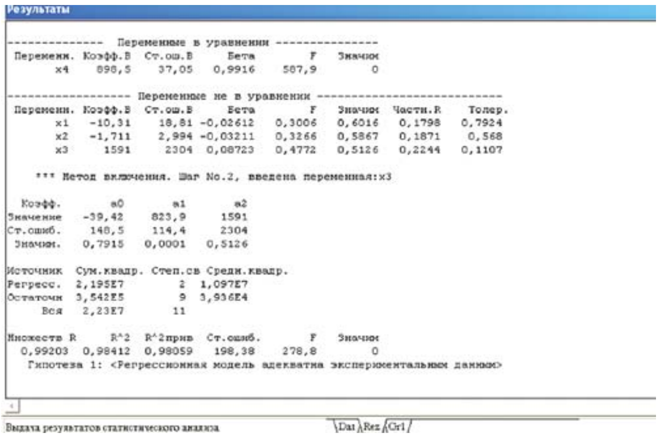
Рис. 5.9. Выбор переменных модели

Рис. 5.10. Ввод параметров селекции переменных



```
Результаты

                    Корреляционная матрица
              x1         x2         x3         x4         x5
      x2    0,889
      x3    0,557      0,683
      x4    0,455      0,657      0,943
      x5    0,431      0,633      0,944      0,991

 Критическое значение=0,567
 Число значимых коэффициентов=7 (116%)

    *** Метод включения. Шаг No.1, введена переменная:x4

  Коэфф.        a0        a1
Значение     47,01     898,5
Ст.ошиб.     77,86     37,05
 Значим.    0,5649         0

Источник  Сум.квадр. Степ.св Средн.квадр.
Регресс.  2,193E7         1  2,193E7
Остаточн   3,73E5        10   3,73E4
     Вся   2,23E7        11

Множеств R      R^2  R^2прив  Ст.ошиб.        F   Значим
   0,9916  0,98328   0,9816   193,13    587,9        0
   Гипотеза 1: <Регрессионная модель адекватна экспериментальным данным>

Выдача результатов статистического анализа        Dat  Rez  Grl
```

A)



```
Результаты

--------------  Переменные в уравнении ---------------
 Переменн. Коэфф.B  Ст.ош.B     Бета         F   Значим
       x4    898,5    37,05   0,9916     587,9        0

------------------ Переменные не в уравнении ---------------------------
 Переменн. Коэфф.B  Ст.ош.B     Бета         F   Значим  Части.R    Толер.
       x1   -10,31    18,81 -0,02612    0,3006   0,6016   0,1798   0,7924
       x2   -1,711    2,994 -0,03211    0,3266   0,5867   0,1871    0,568
       x3     1591     2304  0,08723    0,4772   0,5126   0,2244   0,1107

    *** Метод включения. Шаг No.2, введена переменная:x3

  Коэфф.        a0        a1        a2
Значение    -39,42     823,9      1591
Ст.ошиб.     148,5     114,4      2304
 Значим.    0,7915    0,0001    0,5126

Источник  Сум.квадр. Степ.св Средн.квадр.
Регресс.  2,195E7         2  1,097E7
Остаточн  3,542E5         9  3,936E4
     Вся   2,23E7        11

Множеств R      R^2  R^2прив  Ст.ошиб.        F   Значим
  0,99203  0,98412  0,98059   198,38    278,8        0
   Гипотеза 1: <Регрессионная модель адекватна экспериментальным данным>

Выдача результатов статистического анализа        Dat  Rez  Grl
```

Б)

```
Результаты
-------------  Переменные в уравнении ---------------
 Перемени. Коэфф.B  Ст.ош.B      Бета         F   Значим
       x4   823,9    114,4   0,9093     51,87   0,0001
       x3    1591     2304  0,08723    0,4772   0,5126

------------------ Переменные не в уравнении --------------------------
 Перемени. Коэфф.B  Ст.ош.B      Бета         F   Значим  Частн.R    Толер.
       x1  -19,52     20,8 -0,04942    0,8801   0,6215   0,3148    0,6446
       x2  -2,393    3,147 -0,04489    0,5779   0,5257   0,2596    0,5309

   *** Метод включения. Шаг No.3, введена переменная:x1

  Коэфф.        a0        a1        a2        a3
Значение    0,9217     795,5      2633    -19,52
Ст.ошиб.     155,6     119,1      2571      20,8
 Значим.    0,9911    0,0003    0,3373    0,6215

Источник  Сум.квадр. Степ.св Средн.квадр.
Регресс.  2,198E7          3  7,328E6
Остаточн  3,191E5          8  3,989E4
     Вся   2,23E7         11

Множеств R      R^2  R^2прив  Ст.ошиб.        F   Значим
  0,99282  0,98569  0,98033    199,72     183,7        0
   Гипотеза 1: <Регрессионная модель адекватна экспериментальным данным>

Измен.R^2         F    Значим

Выдача результатов статистического анализа        Dat  Rez  Grl
```

В)

```
Результаты
-------------  Переменные в уравнении ---------------
 Перемени. Коэфф.B  Ст.ош.B      Бета         F   Значим
       x4   795,5    119,1    0,878     44,63   0,0003
       x3    2633     2571   0,1443     1,049   0,3373
       x1  -19,52     20,8 -0,04942    0,8801   0,6215

------------------ Переменные не в уравнении --------------------------
 Перемени. Коэфф.B  Ст.ош.B      Бета         F   Значим  Частн.R    Толер.
       x2  0,7637    7,029  0,01433   0,01181   0,9129  0,04103    0,1173

    Хэксп     Yэксп     Yрегр   остаток Ст.остат  Ст.ошиб Довер.инт
      0,3       397     362,3     34,68   0,2038      190     445,4
      0,6       670     703,3    -33,31  -0,1957    188,1       441
      1,2      1209      1146     62,69   0,3684      186       436
      0,1       138     51,85     86,15   0,5062    191,6     449,1
      0,3       373     314,4     58,58   0,3442      190     445,4
      0,1        79     113,9    -34,93  -0,2053    191,6     449,1
      2,3      1883      2169      -286    -1,68      188     440,6
      2,6      2124      2376    -251,6   -1,479    189,8     444,9
      5,5      5059      4981     77,53   0,4556    231,5     542,7
      2,4      2618      2238     380,1    2,233    188,5     441,9
      1,6      1265      1324    -58,68  -0,3448    185,8     435,6
      0,6       562     597,2    -35,21  -0,2069    188,1       441

Выдача результатов статистического анализа        Dat  Rez  Grl
```

Г)

Рис. 5.11. Протокол расчетов шаговой регрессии

```
Результаты
    *** Метод включения. Шаг No.1, введена переменная:x4

  Коэфф.        a0        a1
Значение     47,01     898,5
Ст.ошиб.     77,86     37,05
 Значим.    0,5649         0

Источник  Сум.квадр. Степ.св Средн.квадр.
Регресс.   2,193E7         1  2,193E7
Остаточн    3,73E5        10   3,73E4
     Вся    2,23E7        11

Множеств R      R^2  R^2прив  Ст.ошиб.        F    Значим
   0,9916  0,98328   0,9816    193,13    587,9         0
   Гипотеза 1: <Регрессионная модель адекватна экспериментальным данным>

Измен.R^2          F    Значим
   0,9833      587,9         0

-------------- Переменные в уравнении ---------------
 Перемени. Коэфф.В  Ст.ош.В     Бета        F    Значим
       x4    898,5    37,05   0,9916    587,9         0
```

А)

```
Результаты
----------------- Переменные не в уравнении --------------------------
 Перемени. Коэфф.В  Ст.ош.В     Бета         F   Значим  Частн.R   Толер.
       x1   -10,31    18,81 -0,02612    0,3006   0,6016   0,1798   0,7924
       x2   -1,711    2,994 -0,03211    0,3266   0,5867   0,1871    0,568
       x3     1591     2304  0,08723    0,4772   0,5126   0,2244   0,1107

    Хэксп    Yэксп    Yрегр  остаток Ст.остат  Ст.ошиб Довер.инт
      0,3      397    316,5    80,45   0,4369    205,6    453,2
      0,6      670    586,1    83,91   0,4557    203,6    448,7
      1,2     1209     1125    83,84   0,4553    201,3    443,6
      0,1      138    136,9    1,142 0,006203    207,3    456,9
      0,3      373    316,5    56,45   0,3066    205,6    453,2
      0,1       79    136,9   -57,86  -0,3142    207,3    456,9
      2,3     1883     2113   -230,5   -1,252    203,4    448,2
      2,6     2124     2383     -259   -1,407    205,4    452,6
      5,5     5059     4989    70,47   0,3827    250,5    552,1
      2,4     2618     2203    414,7    2,252      204    449,6
      1,6     1265     1485   -219,5   -1,192    201,1    443,2
      0,6      562    586,1   -24,09  -0,1308    203,6    448,7
```

Б)

Рис. 5.12. Протокол расчетов при включении в модель фактора X_4

Вывод. Метод корреляционных плеяд и шаговая регрессия позволили построить одинаковую модель, включающую фактор X_4.

Задание второго уровня сложности

Задание. Есть данные хлебокомбината РПС (расположенного в южной области РФ) за каждый месяц в течение двух лет (табл. 5.4).

Таблица 5.4

База данных хлебокомбината

i	X_{1i}	X_{2i}	X_{3i}	X_{4i}	$У_i$
1	1	65,2	2,9	35964	68,26
2	2	63,02	3,47	35693	70,17
3	3	40,67	1,62	44275	81,51
4	4	45,75	1,58	40102	90,96
5	5	69,69	4,22	39603	64,1
6	6	59,85	1,83	52431	68,65
7	7	59,9	3,15	50564	70,93
8	8	52,72	5,17	46406	77,88
9	9	53,52	3,24	42328	78,66
10	10	66,23	7,89	44785	90,22
11	11	48,98	6,23	30310	75,89
12	12	64,99	3,69	25619	65,16
13	13	53,63	2,41	28607	67,03
14	14	45,93	3,14	25968	68,01
15	15	41,82	3,43	27379	69,51
16	16	34,72	1,97	28482	88,35
17	17	42,28	2,63	25610	81,68
18	18	38,69	2,32	23259	73,24
19	19	50,74	3,02	27898	65,12
20	20	38,12	3,01	28465	81,52
21	21	35,75	2,55	26844	78,66
22	22	19,16	1,74	25978	80,81
23	23	35,08	3,51	22311	80,06
24	24	30,99	3,37	22483	77,95
Ожидаем	25	31,35	3,45	22500	
Сумма	300	1157,43	78,09	801364	1814,33
Среднее	12,5	48,22625	3,25375	33390,17	75,59708

где i — порядковый номер измерения;

X_1 _ время (месяцы);

X_2 — удельный вес стоимости сырья в общей себестоимости продукции (%);

X_3 — удельный вес заработной платы в общей себестоимости продукции (%);

X_4 _ объем производства (руб);

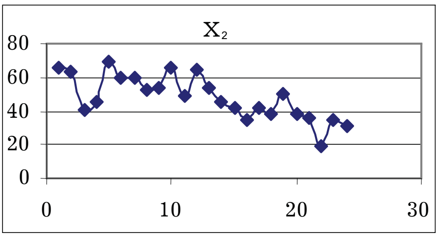
$У_i$ — уровень затрат (%).

Необходимо:

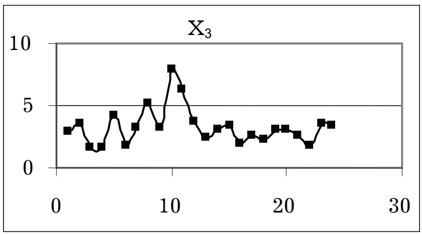
— произвести визуальный анализ динамики всех переменных и определить факторы, которые можно включить в модель;
— произвести корреляционный анализ и определить факторы для включения в модель;
— определить лучшую модель при различных сочетаниях факторов;
— произвести эконометрический анализ для лучшей модели.

Визуальный анализ динамики переменных

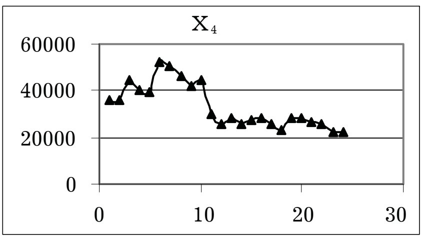
Визуальный анализ динамики переменных (рис. 5.13) показывает, что временной ряд У не содержит четко выраженной линейной тенденции, но содержит четко выраженную периодическую составляющую с периодом 5–6 месяцев. Переменные У и X_2 имеют наибольшую степень синхронизации изменений своих значений.



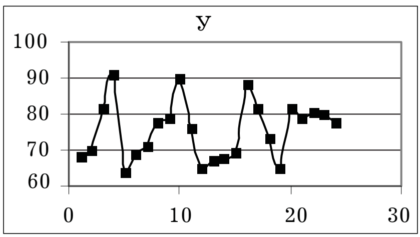
А)



Б)



В)



Г)

Рис. 5.13 Динамика переменных

Вывод. Визуальный анализ динамики переменных показал, что модель зависимой переменной должна содержать — периодическую составляющую с периодом 5–6 месяцев.

Корреляционный анализ факторов для определения включения их в модель

Вычислим матрицу парных коэффициентов корреляции и разместим ее в табл. 5.5.

Таблица 5.5

Матрица парных коэффициентов корреляции r**(X_i,X_j)**

$r(X_i,X_j)$.	X_1	X_2	X_3	X_4	У
X_1	1,00	-0,74	-0,10	-0,77	0,16
X_2	-0,74	1,00	0,44	0,58	-0,46
X_3	-0,10	0,44	1,00	0,16	0,06
X_4	-0,77	-0,46	0,16	1,00	0,06
У	0,16	-0,46	0,06	0,06	1,00

Примечание. Коэффициент корреляции рассчитывается с помощью статистической функции Excel “корреляция” или с помощью программы Excel — пакет анализа, корреляция.

$$r_{кр} = \frac{t_{\alpha/2}}{\left(t_{\alpha/2^2} + n - 2\right)^{1/2}} = 0,404386,$$

где $t_{\alpha/2} = t_{\alpha/2}(\alpha = 0,05;\ m = n — 2\ ) = 2,073873$ — двухстороннее табличное значение критерия Стьюдента;

$n = 24$.

Построим корреляционные плеяды с использованием табл. 5.5 (рис. 5.14).

Вывод. Анализ парных коэффициентов корреляции показывает, что следующие переменные достоверно связаны между собой: X_1 и X_2, X_1 и X_4, X_2 и X_4, X_2 и X_3, У и X_2. Три фактора X_1, X_2, X_4 тесно связаны между собой и, очевидно испытывают влияние со стороны общего фактора.

В модель следует включить фактор X_2.

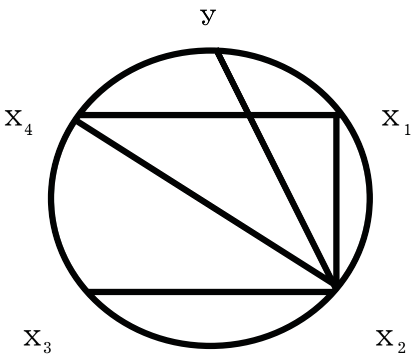


Рис. 5.14. Корреляционные плеяды, в которых переменные соединяются линией при наличии достоверной связи ($r > r_{кр} = 0{,}404$)

Определение лучшей модель при различных сочетаниях факторов

Вычислим характеристики модели при различных сочетаниях факторов, результаты расчетов представим в табл. 5.6.

Таблица 5.6

Характеристики моделей при различных комбинациях факторов, включенных в модель

Номер модели	Состав модели	E	F	Fкр
1	$У_р = f(X_1)$	8,075	0,604	4,300949
2	$У_р = f(X_2)$	7,268	5,903	
3	$У_р = f(X_3)$	8,1724	0,070	
4	$У_р = f(X_4)$	8,1720	0,072	
5	$У_р = f(X_1,X_2)$	7,103	4,108	3,4668
6	$У_р = f(X_1,X_3)$	8,243	0,347	
7	$У_р = f(X_1,X_4)$	7,917	1,259	
8	$У_р = f(X_2,X_3)$	7,044	4,352	
9	$У_р = f(X_2,X_4)$	6,668	6,076	
10	$У_р = f(X_3,X_4)$	8,355	0,058	
11	$У_р = f(X_1,X_2,X_3)$	6,348	5,524	3,098391
12	$У_р = f(X_1,X_2,X_4)$	6,829	3,870	
13	$У_р = f(X_1,X_3,X_4)$	8,107	0,809	
14	$У_р = f(X_2,X_3,X_4)$	6,198	6,125	
15	$У_р = f(X_1,X_2,X_3,X_4)$	6,183	4,889	2,895107

$F_{кр}$ — критические или табличные значения критерия Фишера.

Примечание. Если факторы не связаны между собой, то при вводе в модель нового фактора ошибка модели должна уменьшаться. Если факторы мультиколлинеарны (связаны между собой), то при вводе в модель нового фактора ошибка модели может возрастать.

Анализ табл. 5.6

1. Среди однофакторных моделей лучшей является модель $У_р = f(X_2)$

Среди двухфакторных моделей лучшей является модель $У_р = f(X_2,X_4)$

Среди трехфакторных моделей лучшей является модель $У_р = f(X_2,X_3,X_4$

Среди всех моделей лучшей является модель $У_р = f(X_2,X_3,X_4)$.

2. Все достоверные модели содержат фактор $X_{2.}$

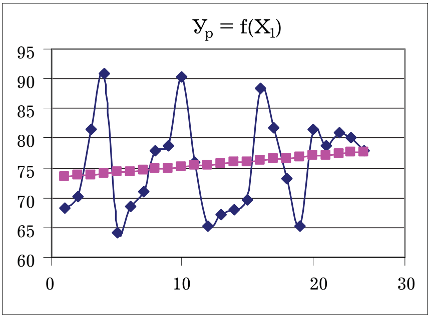
3. Сравнение моделей $У_р = f(X_1)$ и $У_р = f(X_1,X_3)$, $У_р = f(X_4)$ и $У_р = f(X_3,X_4)$ показывает, что введение в модель нового фактора приводит к увеличению ошибки модели, следовательно факторы X_1 и X_3, X_3 и X_4 должны быть мультиколлинеарны, т. е. связанными между собой. Однако парные коэффициенты корреляции между этими факторами не являются достоверными.

Этот феномен можно объяснить тем, что модели с участием только факторов X_1, X_3, X_4 не являются достоверными и если в модели встречаются два недостоверных фактора, то модель становится еще хуже.

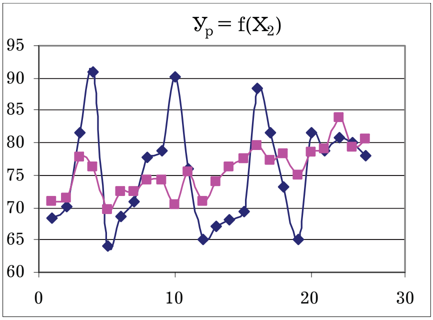
4. Факторы X_1, X_2, X_4 достоверно связаны между собой (по парным коэффициентам корреляции), однако совместное их присутствие в модели не приводит к ее ухудшению, следовательно мультиколлинеарность между факторами X_1, X_2, X_3 не является достоверной.

Вывод. Среди всех моделей лучшей является модель $У_р = f(X_2,X_3,X_4)$.

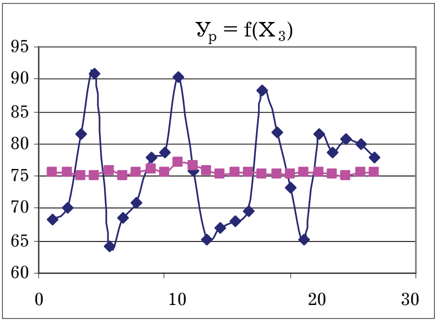
Произведем визуальный анализ аппроксимации зависимой переменной У (рис. 5.15) с помощью различных комбинаций факторов.



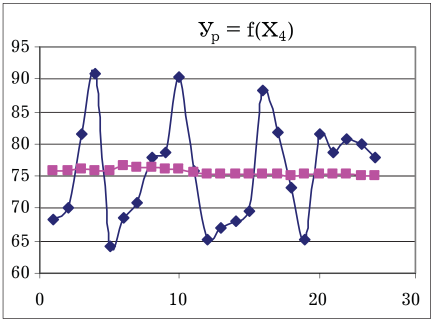
А)



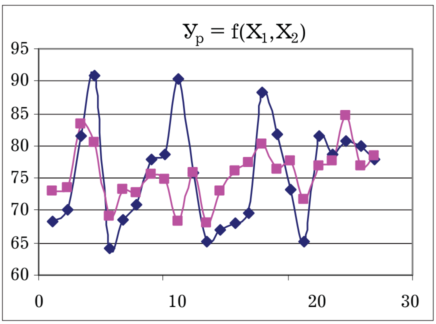
Б)



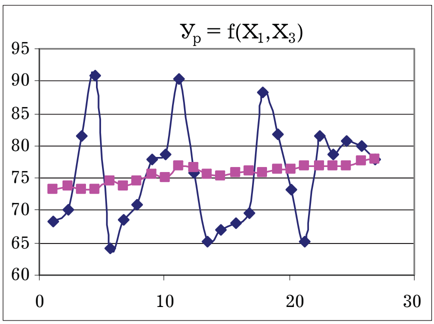
В)



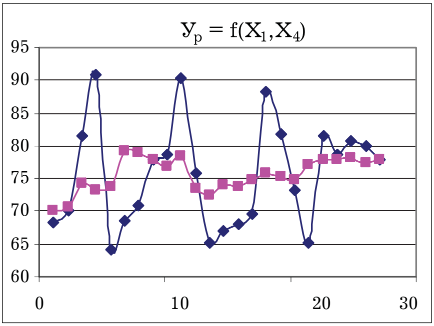
Г)



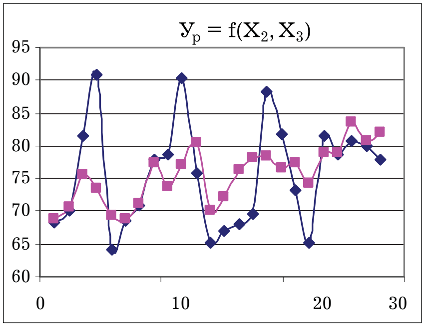
Д)



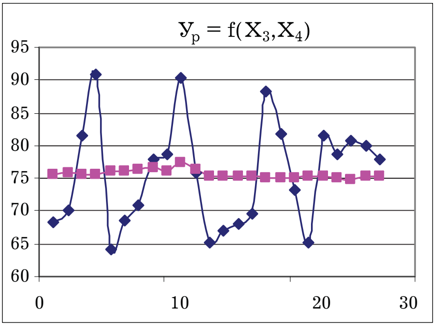
Е)


$У_р = f(X_1, X_4)$
95
90
85
80
75
70
65
60
0
10
20
30

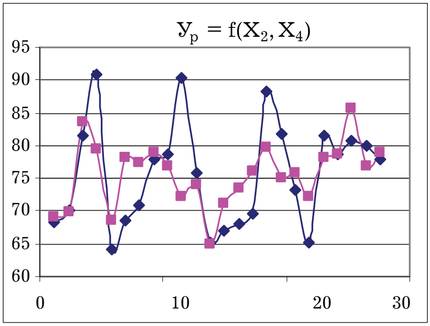
Ж)


$У_р = f(X_2, X_3)$
95
90
85
80
75
70
65
60
0
10
20
30

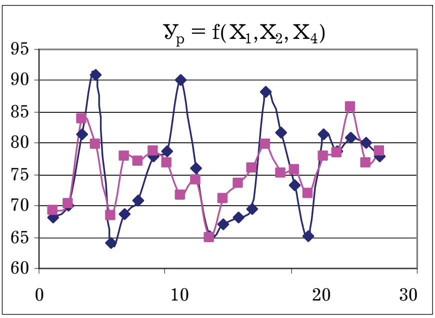
З)


$У_р = f(X_3, X_4)$
95
90
85
80
75
70
65
60
0
10
20
30

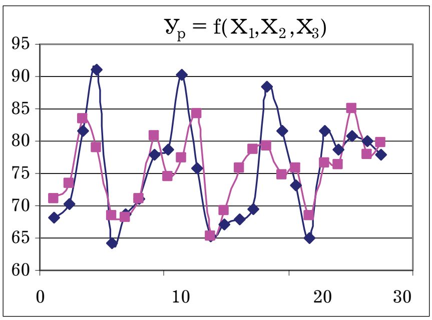
И)


$У_р = f(X_2, X_4)$
95
90
85
80
75
70
65
60
0
10
20
30

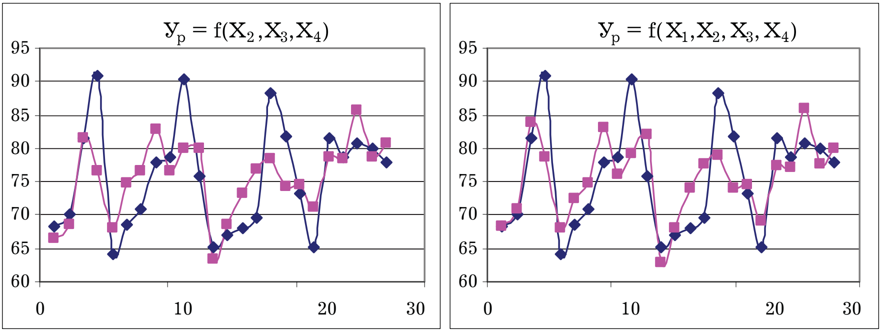
К)


$У_р = f(X_1, X_2, X_4)$
95
90
85
80
75
70
65
60
0
10
20
30

Л)


$У_р = f(X_1, X_2, X_3)$
95
90
85
80
75
70
65
60
0
10
20
30

М)



Н) О)

Рис. 5.15. Графики моделей при различных комбинациях факторов

Вывод. Визуальный анализ графиков рис. 5.15 (А — О) зависимости У от разного сочетания факторов показывает, что модели, которые содержат фактор X_2, хорошо отражают тенденцию зависимой переменной У. Однако, модель будет хорошо специфицирована, если в нее включить периодическую составляющую с периодом 5 — 6 месяцев.

Приводим график зависимости У от фактора X_1 и периодической составляющей с периодом 6 месяцев (E = 5,9689; F = 7,1242) (рис. 5.16).

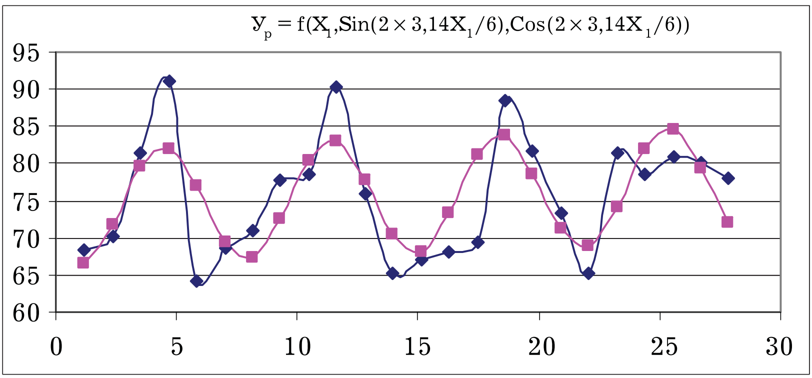


Рис. 5.16. График модели, содержащей периодическую составляющую

Полученная модель является лучшей из тех, которые были изучены. Если в эту модель добавить фактор X_2, то можно будет значительно уменьшить ошибку модели.

Если в модель включить фактор X_1, периодическую составляющую с периодом 6 месяцев и фактор X_2, то: E = 5,8558, F = 5,9966.

Включение фактора X_2 в модель позволило незначительно уменьшить ошибку модели. Последний вариант модели будем считать окончательным. На этом построение модели можно закончить.

Произведем эконометрический анализ окончательной многофакторной модели.

-0,2172	-3,98	-5,213	-0,166	88,14
0,1628	1,9445	1,7631	0,2861	10,959
0,558	5,8558	#Н/Д	#Н/Д	#Н/Д
5,9966	19	#Н/Д	#Н/Д	#Н/Д
822,49	651,51	#Н/Д	#Н/Д	#Н/Д

$У_р$ =	88,14	-0,166	X_1	-5,213	Z_1	-3,98	Z_2	-0,217	X_2
t_a =	8,043	-0,579		-2,957		-2,047		-1,334	
R_2 =	0,558								
E =	5,8558		$t_{таб}$ =	2,093					
F =	5,9966		$F_{таб}$ =	2,8951					

где $Z_1 = Sin(2 \times 3{,}14X_1/6)$ — периодическая составляющая регрессионной модели;

$Z_2 = Cos(2 \times 3{,}14X_1/6)$ — периодическая составляющая регрессионной модели;

X_1 — время (месяцы);

X_2 — удельный вес стоимости сырья в общей себестоимости продукции (%),;

У — уровень затрат (%).

1. Определение доли объясненной вариации.

55,8% исходных данных имеют выбранную тенденцию;

2. Проверка достоверности модели.

Достоверность модели проверяется по критерию Фишера.

Так как $F = 5{,}9966 > F_{таб}(\alpha = 0{,}05,\ m_1 = k — 1,\ m_2 = n — k) = 2{,}8951$, то модель

является достоверной с вероятность 0,95.

3. Проверка достоверности коэффициентов модели.

По критерию Стьюдента проверяется достоверность коэффициентов модели

Так как $t_{a0} = 8{,}043 > t_{таб} = 2{,}093$, то коэффициент a_0 является достоверным с вероятностью 0,95.

Так как $t_{a1} = 0{,}5788 < t_{таб} = 2{,}093$, то коэффициент a_1 не отличается от нуля.

Так как $t_{a2} = 2{,}9568 > t_{таб} = 2{,}093$, то коэффициент a_2; является достоверным с вероятностью 0,95.

Так как $t_{a3} = 2{,}047 < t_{таб} = 2{,}093$, то коэффициент a_3 не отличается от нуля.

Так как $t_{a4} = 1{,}3342 < t_{таб} = 2{,}093$, то коэффициент a_4 не отличается от нуля.

4. Доля влияния каждого фактора.

$r^2(X_1) = 0{,}0124$ — доля влияния фактора X_1;

$r^2(Z_1) = 0{,}3242$ — доля влияния фактора Z_1;

$r^2(Z_2) = 0{,}1554$ — доля влияния фактора Z_2;

$r^2(X_2) = 0{,}066$ — доля влияния фактора X_2;

Всего = 0,558 — доля влияния всех факторов.

Примечание. Определение доли влияния факторов.

В эконометрическом анализе важно знать долю влияния факторов, которые включены

в модель. Ее можно определить по формуле

$$r(X_i)^2 = \frac{t(X_i)^2 R^2}{t(X_1)^2 + t(X_2)^2 + t(X_3)^2 + \ldots + t(X_i)^2},$$

где $r(X_j)^2$ — коэффициент детерминации для фактора $X_{j;}$

$t(X_j)$ — критерий Стьюдента для фактора X_j;

R^2 — общий коэффициент детерминации;

X_i — факторы, включенные в модель.

5. Точечный и интервальный прогноз.

При ожидаемых значениях переменных на период t = 25 $X_{1ож} = 25$; $Z_1 = 0{,}8593$; $Z_2 = 0{,}5114$; $X_2 = 31{,}35$,

точечный прогноз зависимой переменной будет равен $У_{пр} = 70{,}675$.

Фактическое значение зависимой переменной с вероятностью 0,95 будет находиться в интервале от 58,418 до 82,931.

6. Графическое представление эконометрического анализа.

На рис. 5.17 показаны результаты расчетов по лучшей модели.

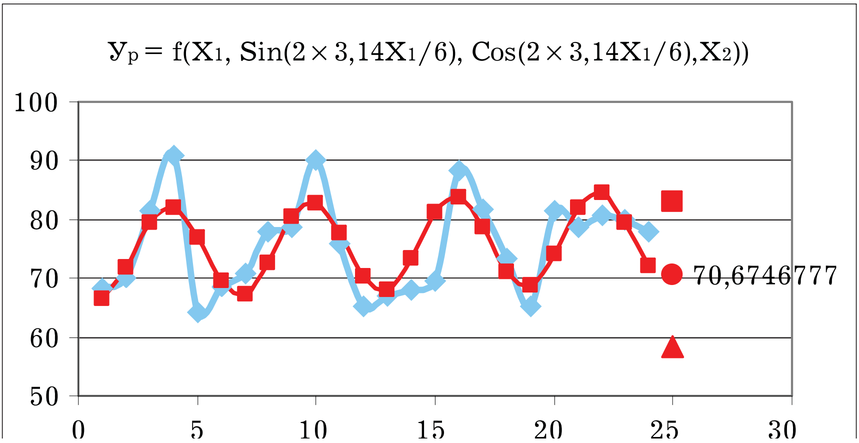


Рис. 5.17. Результаты расчетов по лучшей модели

Вывод. Многофакторный анализ зависимости У от четырех факторов показал, что лучшей является линейная модель следующего вида:

$У = a_0 + a_1X_1 + a_2\mathrm{Sin}(2\times3{,}14X_1/6) + a_3\mathrm{Cos}(2\times3{,}14X_1/6) + a_4X_2$.

На прогнозный период t = 25 точечный прогноз равен $У_{пр} = 70{,}675\%$.

Задание третьего уровня сложности

Задание. Реализуйте в Excel алгоритм шаговой регрессии включения.

Пояснение. Шаговая регрессия включения реализуется в такой последовательности:

1) ищется матрица парных коэффициентов корреляции между У и всеми X;

2) с помощью функции "Линейн" вычисляются характеристики модели, содержащей фактор, который сильнее всего (по коэффициенту корреляции) связан с У;

3) проводится проверка достоверности модели. Если модель достоверная, то построение модели продолжается;

4) ищутся остатки модели и матрица коэффициентов корреляции между остатками модели и всеми факторами, которые не вошли в модель. Ищется фактор, который сильнее всего связан с остатками модели;

5) с помощью функции "Линейн" вычисляются характеристики модели, содержащей дополнительный фактор, определенный в пункте 4. Затем выполняется пункт 3, 4, 5 до тех пор, пока модель станет недостоверной. Модель считается окончательной до момента включения в нее фактора, делающего ее недостоверной;

6) проводится полный эконометрический анализ полученной достоверной модели.

Задание четвертого уровня сложности

Задание. Реализуйте в Excel с помощью Visual Basic алгоритм Эфроимсона шаговой регрессии включения[1].

Выходное тестирование

Имеется матрица коэффициентов корреляции, вычисленная при $n = 24$.

[1] Дрейпер Н., Смит Г. Прикладной регрессионный анализ. — М.: Статистика, 1973. (Имеется алгоритм Эфроимсона шаговой регрессии с котрольным примером).

	X_1	X_2	X_3	У
X_1	1	0,1	0,85	0,95
X_2		1	0,25	0,2
X_3			1	0,8
У				1

Какой фактор можно включить в модель зависимости У от X:

а) X_1;

б) X_3;

в) X_1, X_3.

6. ОБОБЩЕННЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ. ВЫПОЛНЕНИЕ МАТРИЧНЫХ ОПЕРАЦИЙ

Введение

Объект — показатели деятельности экономического объекта.

Предмет — нарушения предпосылок метода наименьших квадратов.

Цель — изучить обобщенный метод наименьших квадратов (метод Эйткена). Получить навыки проведения расчетов с использованием матричных операций.

Актуальность — неоднородность остатков приводит к увеличению ошибок коэффициентов модели.

Метод Эйткена для расчета коэффициентов модели.

Способ — проведение расчетов с помощью матричных операций.

Задача — произвести расчеты коэффициентов и всех харктеристик модели с помощью матричных операций,

Рабочая гипотеза — предполагается, что остатки модели являются неоднородными.

Ожидаемый результат — ожидается, что можно определить коэффициенты и характеристики модели с помощью матричных операций.

Метод для сравнения — расчеты коэффициентов с помощью матричных операций и в развернутом виде.

Результаты расчетов

A = **X'X)$^{-1}$X'У =**

2,7
0,8

Коэффициенты модели

$У = a_0 + a_1X + e$

имеют следующие значения:

$a_0 = 2{,}7$;

$a_1 = 0{,}8$.

Задания первого уровня сложности

Задание 1. Подготовьте устные сообщения по вопросам самостоятельной работы.

1. Повторите правила выполнения основных матричных операций: транспонирование, умножение, определитель, обратная матрица.

2. Повторите, что означают собственные значения и собственные вектора матрицы.

3. Укажите название и место нахождения в Excel основных матричных операций.

4. Изучите последовательность вычисления коэффициентов модели с помощью матричной формулы

A = (X'X)$^{-1}$X'У.

5. Опишите обобщенный метод наименьших квадратов (метод Эйткена) для оценки параметров линейной модели

[3, с. 164–167, 7, с.150–155, 9, т. 2, с. 93–102].

Задание 2. Пройдите входной тест

1. Существует два вида записи расчетных формул:

— в развернутом по ее элементам виде;

— в матричном виде.

а) да;

б) нет.

2. 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 матричных формул не зависит от количества факторов, включенных в модель, и является очень удобным при расчетах характеристик многофакторных моделей. Имеются доступные инструментальные средства выполнения матричных операций: Mat-Lab, Mat-Cad, Excel, которые позволяют выполнять все необходимые для лабораторных работ матричные операции. Однако Excel не содержит операции вычисления собственных значений и собственных векторов матрицы, которые имеются в Mat-Lab и Mat-Cad.

Примечание. В эконометрической литературе формулы записываются в развернутом по элементам виде или в матричном виде. Доля формул в матричном виде достаточно велика. До появления ЭВМ использование матричных формул сдерживало отсутствие средств выполнения их расчетов. Например, можно с помощью калькулятора вычислить обратную матрицу третьего порядка, но неэффективно это делать для матрицы выше пятого порядка. При этом возникают проблемы, связанные с точностью расчетов. С появлением средств выполнения матричных операций использование матричных формул стало актуальным.

Недостатком использования в расчетах матричных формул является необходимость хорошего знания матричной алгебры и наличия программных средств проведения матричных операций.

а) да;

б) нет.

3. В Excel имеются такие матричные операции и координаты их нахождения.

Транспонирование — Вставка функции, Категория: Ссылки и массивы, Функции: ТРАНСП.

Вычисление обратной матрицы — Вставка функции, Категория: Математическая, матрицы, Функции: МОБР.

Умножение матриц — Вставка функции, Категория: Математическая, Функции: МУМНОЖ.

Регрессионный анализ — Вставка функции, Категория: Статистические, Функции: “Линейн”.

а) да;

б) нет.

4. Матричные операции в Excel выполняются в следующей последовательности:

— в одну ячейку вводится и выполняется матричная операция;
— для распространения матричной операции на все элементы матрицы необходимо выполнить такие действия:
— выделить ячейку с матричной операцией и все ячейки, где должны быть ее элементы:
— нажать клавишу F2 — вызов формулы для редактирования;
— последовательно не отпуская нажать три клавиши Ctrl, Shift, Enter, затем сразу отпустить все клавиши.

а) да;

б) нет.

5. Функция вычисления среднего значения находится по адресу:

Вставка функ, Категория: Статистические, Функции: СР-ЗНАЧ.

а)да;

б) нет.

6. Транспонирование.

Имеется вектор столбец **A**

$$\mathbf{A} = \begin{vmatrix} 2 \\ 1 \\ 3 \end{vmatrix}$$

Необходимо транспонировать матрицу **A**.

Предлагаются следующие варианты транспонирования:

а) $\mathbf{A'} = \begin{vmatrix} 3 \\ 1 \\ 2 \end{vmatrix}$

б) $\mathbf{A'} = \begin{vmatrix} 2 & 1 & 3 \end{vmatrix}$

Пояснение. При транспонировании матрицы столбцы исходной матрицы становятся строками транспонированной матрицы.

7. Произведение матриц.

Даны две матрицы **A** и **B**:

$\mathbf{A} = \begin{vmatrix} 2 \\ 1 \\ 3 \end{vmatrix}$, $\mathbf{A'} = \begin{vmatrix} 2 & 1 & 3 \end{vmatrix}$, $\mathbf{B} = \begin{vmatrix} 1 \\ 2 \\ 3 \end{vmatrix}$

Необходимо получить произведение матриц **A'B**.

а) **A'B** = 17

б) **A'B** = 13

Пояснения. При произведении вектора строки на вектор столбец элементы вектора строки умножаются на соответствующие значения вектора столбца, результаты умножения складываются.

8. Свойства обратных матриц.

Получение обратной матрицы от произведения матриц.

а) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

б) $(\mathbf{AB})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}$

Пояснение

$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

9. Свойства обратных матриц.

$\mathbf{A}^{-1}\mathbf{A}$

а) $\mathbf{A}^{-1}\mathbf{A} = \mathrm{I}$

б) $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}$

Пояснение

$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$

10. Вычисление коэффициентов линейной модели в матричном виде.

а) $\mathbf{А = (Х'Х)^{-1}\, Х'У}$

б) $\mathbf{А = (Х'Х)\, Х'У}$

Пояснение

$\mathbf{А = (Х'Х)^{-1}\, Х'У}$

Задание 3. Решите задачу 1.

Имеется пространственная база данных для определения зависимости товарооборота от затрат на рекламу.

Таблица 6.1

База данных по четырем однородным фирмам за один месяц

i	$Х_i$	$У_i$
1	2	4
2	3	6
3	4	5
4	5	7
Ожидаем	6	?
Сумма		

где $Х_i$ — затраты на рекламу (тыс. руб.);

$У_i$– товарооборот (тыс. руб.);

i — порядковый номер измерения.

Необходимо: произвести расчеты коэффициентов и характеристик линейной модели матричным способом.

Решение задачи

Целью моделирования является определение зависимости между товарооборотом и затратами на рекламу.

Выбор переменных

Величина товарооборота зависит от шести основных факторов: *сырье и материалы* (ассортимент товаров и их качество); *машины* (оборудование и дизайн магазина); *методы* (форма обслуживания покупателей: форма самообслуживания, культура обслуживания, рекламная деятельность); *люди* (количество продавцов, уровень квалификации, образование, возраст, деловые характеристики); *физическая и абстрактная среда* (температура, влажность в магазине; традиции, психологический климат существующие в магазине; правовая среда существования магазина, информационное обеспечение магазина); *время* (нет влияния, так как имеется пространственная выборка.).

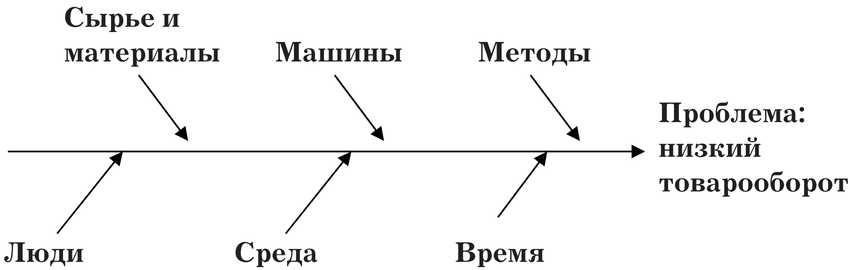


Рис. 6.1. Причинно-следственная диаграмма Исикавы

Выделение входных и выходных переменных

Магазин имеет множество характеристик, но в соответствии с возникшей проблемой, связанной с низким розничным товарооборотом по сравнению с конкурентами, появилось предложение увеличить товарооборот за счет активизации рекламы. При этом важно знать прогнозное значение розничного товарооборота при ожидаемом значении величины затрат на рекламу.

Примечание. Реклама доводит до населения характеристики товара, но не создает ее новые потребительские свойства. Повысить конкурентоспособность товара можно только за счет увеличения качества продукции и снижения себестоимости с помощью совершенствования процессов. Совершенствование процессов можно произвести с привлечением Европейской модели идеального предприятия и международных стандартов системы менеджмента качества ИСО серии 9000: 2000.

Объясняющей переменной X будут затраты на рекламу, зависимой переменной У — товарооборот.

Сбор статистических данных

Была произведена пространственная выборка по четырем магазинам фирмы с учетом У — величины товарооборота и

X — затрат на рекламу. Исходные данные представлены в таблице 6.1. График зависимости У от X изображен на рис. 6.2.

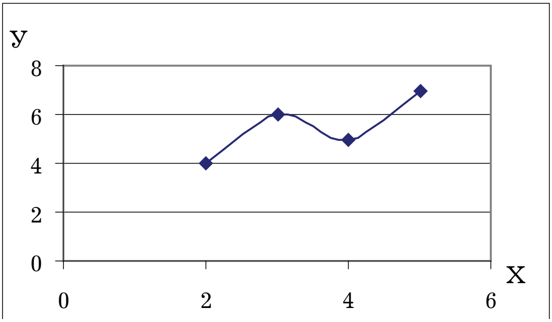


Рис. 6.2. График зависимости У от X

Анализ рис. 6.2 показывает, что с ростом затрат на рекламу увеличивается товарооборот. Однако эта зависимость является стохастической.

Предмодельный анализ

Выдвижение гипотез

В теоретическом плане при увеличении затрат на рекламу товарооборт должен возрастать до определенного значения, а затем его рост замедлится и будет стремиться к пределу, величина которого определяется охватом всех покупателей данной продукцией.

Данная зависимость описывается логистической функцией. Однако в первом приближении предположим, что товарооборот зависит линейно от затрат на рекламу.

Формулировка допущений

Товарооборот будет линейно зависеть от затрат на реклму при условии, что товарооборот не достиг максимального значения и наши данные соответствуют линейной части зависимости У от X.

Спецификация

Предполагаем, что спецификация зависмости У от X cоотетствует (КНЛМПР) классической нормальной линейной модели парной регрессии, описание которой смотрите в первой теме.

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 товарооборота от затрат на рекламу. Поэтому можно с помощью исходных данных оценить параметры линейной модели.

Идентификация модели

(Предложить и реализовать процедуру оценивания параметров модели).

Формулы расчета характеристик выборочной регрессионной модели

$$У_i = a_0 + a_1 X_i + e_i.$$

Характеристики модели рассчитывают по формулам, представленным в поэлементном и матричном видах.

Расчет коэффициентов модели a_0, a_1 производится по формулам:

$$a_1 = \frac{\sum (X_i - \overline{X})(У_i - \overline{У})}{\sum (X_i - \overline{X})^2}, \quad a_0 = \overline{У} - a_1 \overline{X}.$$

Расчет коэффициентов модели a_0, a_1 в матричном виде производится по формуле

$$\mathbf{A = (X'X)^{-1} X'У},$$

где **X** — матрица исходных данных, включающая вектор столбец переменной для свободного коэффициента, векторы столбцы объясняемых факторов;

A — вектор столбец коэффициентов модели;

X' — 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

$$\mathbf{A = (X'X)^{-1}X'У}$$

в следующей последовательности.

Шаг 1. Сформируем исходные матрицы.

Х =

1	2
1	3
1	4
1	5

У =

4
6
5
7

Шаг 2. Получим транспонированную матрицу **Х'**.

Х' =

1	1	1	1
2	3	4	5

Шаг 3. Получим произведение матриц **Х'Х.**

Х'Х =

4	14
14	54

Шаг 4. Вычислим обратную матрицу $\mathbf{C = (X'X)^{-1}}$.

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

2,7	-0,7
-0,7	0,2

Шаг 5. Вычислим матрицу $\mathbf{(X'X)^{-1}X'}$

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

1,3	0,6	-0,1	-0,8
-0,3	-0,1	0,1	0,3

Шаг 6. Вычислим коэффициенты модели по формуле

$\mathbf{A = (X'X)^{-1}X'У}$

А = $\mathbf{X'X)^{-1}X'У}$ **=**

2,7
0,8

Вывод. Коэффициенты модели

$У_{pi} = a_0 + a_1 X + e$

имеют следующие значения:

$a_0 = 2,7$;

$a_1 = 0,8$.

Вычисление стандартной ошибки модели

Расчет стандартной ошибки модели E в векторном виде производится по формуле:

$$E = \sqrt{\frac{\sum e_i^{\,2}}{n-k}},$$

$$e_i = У_i — (a_0 + a_1 X_i) = У_i — У_{pi}.$$

Расчет стандартной ошибки модели E в матричном виде производится по формуле

$$E = \sqrt{\frac{e'e}{n-k}}$$

в следующей последовательности.

Шаг 1. Определим $\mathbf{У_p = XA}$.

$\mathbf{У_p = XA =}$

4,3
5,1
5,9
6,7

Шаг 2. Определим остатки модели по формуле $e_i = У_i — У_p$.

Примечание. Матичной операции вычитания не существует, поэтому расчеты необходимо произвести поэлементно.

$e_i = У_i — У_p =$

-0,3
0,9
-0,9
0,3

Шаг 3. Определим **e'.**

e' =

-0,3	0,9	-0,9	0,3

Шаг 4. Определим **e'e**.

e'e =

1,8

Шаг 5. Определим E = КОРЕНЬ((e'e)/(n-k)), при n = 4, k = 2
E = КОРЕНЬ((e'e)/(n-k)) = 0,9487

Вычисление стандартных отклонений коэффициентов модели a_0 и a_1.

Расчет стандартных отклонений коэффициентов a_0, a_1

$$S_{a_0} = E\sqrt{\frac{\sum X_i^2}{n\sum (X_i - X_{ср})^2}},\ S_{a_1} = E\sqrt{\frac{1}{\sum (X_i - X_{ср})^2}}$$

можно выполнить по эквивалентным формулам, представленным в матричном виде:

S_{a0} = E×КОРЕНЬ(c_{11});

S_{a1} = E×КОРЕНЬ(c_{22}),

где c_{11} и c_{22}- диагональные элементы матрицы **C = (X’X)^-1^**.

S_{a0} = E×КОРЕНЬ(c_{11}) = 1,5588

S_{a1} = E×КОРЕНь(c_{22}) = 0,4243

Расчет критерия Фишера

Расчет критерия Фишера производится по формулам:

$$F = \frac{S_{рег}^2}{S_{ост}^2},\ S_{ост}^2 = \frac{\sum (У_i - У_{p_i})^2}{n-k},\ S_{рег}^2 = \frac{\sum (У_{p_i} - \overline{У})^2}{k-1},$$

которые можно выполнить по эквивалентным формулам, представленных в матричном виде:

$$F = \frac{S_{рег}^2}{S_{ост}^2},\ S_{ост}^2 = \frac{\mathbf{e'e}}{n\text{ - }k},\ S_{рег}^2 = \frac{\mathbf{A'X'У} - n\left(\overline{У}\right)}{k\text{ - }1},$$

где n — объем выборки,

k — количество коэффициентов в модели[1]

Вычислим сумму квадратов отклонений, обусловленную регрессией, по формуле

A’X’У’ — n $(\overline{У})^2$.

A’ =	2,7	0,8		
A’X’ =	4,3	5,1	5,9	6,7
A’X’У =	124,2			

[1] Вучков И., Бояджиева Л., Солаков Е. Прикладной линейный регрессионный анализ. — М.: Финансы и статистика, 1987, с. 42

Вычислим $n\,(\overline{\text{У}})^2$ с использованием функции вычисления среднего значения СРЗНАЧ.

$\overline{\text{У}} =$	5,5		
$n\,(\overline{\text{У}})^2 =$		121	
$\mathbf{А'Х'У} - n\,(\overline{\text{У}})^2 =$			3,2

Вычислим значение дисперсии, обусловленное регрессией по формуле:

$$S_{\text{рег}}{}^2 = \frac{\text{А'Х'У'} - n\,(\overline{\text{У}})^2}{k-1} = 3{,}2.$$

Вычислим значение дисперсии остатков по формуле:

$$S_{\text{ост}}{}^2 = \frac{e'e}{n-k} = 0{,}9.$$

Вычислим значение критерия Фишера по формуле:

$$F = \frac{S_{\text{рег}}{}^2}{S_{\text{ост}}{}^2} = 3{,}556.$$

Расчет коэффициента детерминации

Расчет коэффициента детерминации производится по формуле:

$$R^2 = \frac{\sum(\text{У}_{\text{р}i} - \overline{\text{У}})^2}{\sum(\text{У}_i - \overline{\text{У}})^2}.$$

Расчет коэффициента детерминации в матричном виде производится по формуле[1]:

$$R^2 = \frac{\mathbf{А'Х'У} - n(\overline{\text{у}})^2}{\mathbf{У'У} - n(\overline{\text{у}})^2}.$$

Вычислим значение общей суммы квадратов отклонений фактических значений от своей средней величины по формуле:

$\mathbf{У'У} - n\,(\overline{\text{У}})^2$

[1] Джонстон Д. Эконометрические методы. — М.: Статистика, 1980, с. 130.

У' =	**4**	**6**	**5**	**7**
У'У =	**126**			

У'У — n $(\overline{У})^2$ = **5**

Вычислим значение коэффициента детерминации

$$R^2 = \frac{A'X'У' - n\left(\overline{У}\right)^2}{У'У' - n\left(\overline{У}\right)^2} = \frac{3{,}2}{5} = 0{,}64.$$

Вычисление коэффициентов и характеристик модели с помощью функции "Линейн".

Пояснение.

В качестве независимой переменной укажите вектор столбец значений X (без фиктивной переменной), в качестве зависимой переменной укажите вектор столбец У.

Представим результаты расчетов по функции "Линейн".

0,8	2,7
0,4243	1,5588
0,64	0,9487
3,5556	2
3,2	1,8

Представим результаты расчетов в общепринятом обозначении.

$У_{pi}$ =	2,7	+	0,8	X
S_a =	1,5588		0,4243	
t_a =	1,7321		1,8856	
E =	0,9487			
R_2 =	0,64		$t\,\alpha/2$ =	4,3027
F =	3,5556		Fкр =	18,513

Вывод. Анализ результатов расчета характеристик модели с использованием функции "Линейн" и матричных операций показывает, что матричные операции выполнены правильно.

Построим графики фактических и расчетных значений зависимой переменной от объясняемой переменной (рис. 6.3).

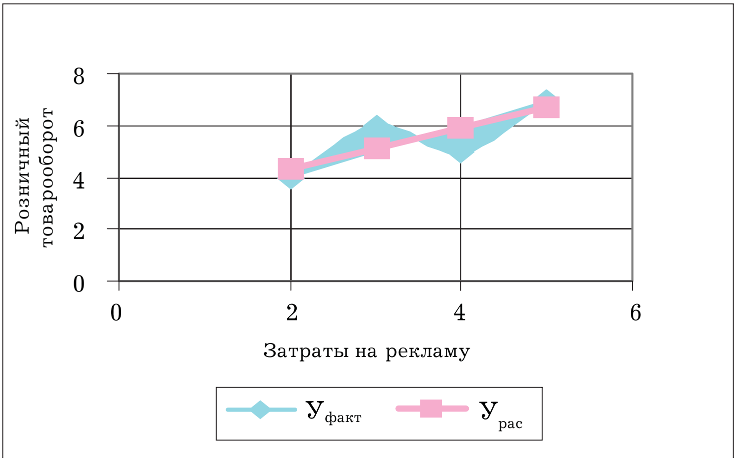


Рис. 6.3. График фактических и расчетных значений У.

Задание второго уровня сложности

Задание 1. Решите задачу 2.

Имеется решение задачи 1.

Необходимо:

— дополнить задачу 1 вычислением коэффициентов Эйткена при условии, что матрицу **V** можно оценить с помощью остатков вектора столбца е выборочного уравнения регрессии;

— рассчитать значения $У_р$ при использовании коэффициентов Эйткена и исходных данных;

— произвести анализ полученных расчетов, сделать выводы.

Решение задачи

В задаче 1 остаются неизменными: постановка задачи, сбор статистических данных, предмодельный анализ, отличие начинается со спецификации модели.

Спецификация

Обобщенная линейная модель парной регрессии

Предполагаем, что зависимость У от Х можно представить в виде обобщенной линейной модели парной регрессии. Приводим спецификацию данной модели:

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

где **У** — вектор столбец случайной величины;

XA + ε — уравнение регрессионной модели;

XA — уравнение регрессии;

X -неслучайная матрица факторов;

A — вектор столбец неизвестных параметров;

ε — вектор столбец неизвестного возмущения.

Относительно возмущения ε выдвигаем следующие гипотезы:

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

$M(\varepsilon\varepsilon') = \mathbf{V} \neq \sigma^2_\varepsilon \mathbf{I}$,

где **V** — симметрическая и положительно определенная ковариационная матрица.

При этом возмущения могут быть автокоррелированными и неоднородными (гетероскедастичными).

Идентифицируемость модели

(Можно ли в принципе однозначно восстановить значения неизвестных параметров по выбранной модели).

Если вычислить коэффициенты модели методом наименьших квадратов по формуле

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

то полученные коэффициенты будут несмещенными, состоятельными, но неэффективными.

Идентификация модели

(Предложить и реализовать процедуру оценивания параметров модели).

Если вычислить коэффициенты модели с помощью обобщенного метода наименьших квадратов (при известной матрице V), по формуле Эйткена:

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

то полученные коэффициенты будут несмещенными, состоятельными и эффективными по отношению к определенным образом преобразованным исходным значениям У и Х.

Если исходные значения Х 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.

Примечание. После преобразования исходных данных получается фактически новая модель с новыми переменными, которые часто не имеют экономического смысла.

Если матрица **V** является симметрической и положительно определенной, то возможно следующее соотношение:

$\mathbf{V = PP'}$,

где матрица Р является невырожденной.

$\mathbf{P = ((hD)^{-1})}$,

$$\mathbf{D} = \mathbf{I}\frac{1}{\sqrt{\mathbf{h'Vh}}} = \mathbf{I}\frac{1}{\sqrt{\mathbf{k}}},$$

I — единичная матрица.

h — собственный (характеристический) вектор матрицы **V**;

k — собственное (характеристическое) значение матрицы **V**.

Собственные значения **g** матрицы **V** можно определить из уравнения.

$\mathbf{Vh = kh}$.

Умножим уравнение модели $\mathbf{У = XA} + \varepsilon$ слева на матрицу $\mathbf{P}^{-1}$, получим

$\mathbf{У_н = X_н B} + \varepsilon_н$,

где $\mathbf{У_н = P^{-1}У}$; $\mathbf{X_н = P^{-1}X}$ и $\varepsilon_н = \mathbf{P}^{-1}\varepsilon$.

Коэффициенты В можно вычислить с помощью обобщенного метода наименьших квадратов (метод Эйткена) по формуле

$$\mathbf{B = (X_н'X_н)^{-1}X_н У_н = (X'V^{-1}X)^{-1}X'V^{-1}У}$$

Коэффициенты В будут несмещенными, состоятельными и эффективными по отношению к преобразованным данным $\mathbf{У_н}$ и $\mathbf{X_н}$.

Если условие

$\mathbf{M(\varepsilon\varepsilon') = V = \sigma_\varepsilon^2 I}$,

выполняется, то коэффициенты **A** и **B**, вычисленные соответственно обычным и обобщенным методом наименьших квадратов, будут равны между собой.

Матрица **V** является симметрической, если выполняется условие:

$\mathbf{V = V'}$

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.

Главные миноры матрицы — определители различных подматриц, образованные в результате вычеркивания строк и столбцов исходной матрицы, имеющих одинаковые номера.

Матрица **V** может иметь собственные значения и собственные вектора, которые находятся в следующем соотношении:

$\mathbf{Vh = kh}$,

где **V** — исходная матрица размером $n \times n$;
h — собственный вектор матрицы **V**;
k- собственное значение матрицы **V**.

Оценим матрицу **V** по величине остатков выборочного уравнения регрессии.

Имеется вектор столбец остатков:

$e_i = У_i — У_p =$

-0,3
0,9
-0,9
0,3

Определим матрицу **V = ee’**

e’ =

-0,3	0,9	-0,9	0,3

V = ee’ =

0,09	-0,27	0,27	-0,09
-0,27	0,81	-0,81	0,27
0,27	-0,81	0,81	-0,27
-0,09	0,27	-0,27	0,09

Проверим матрицу **V** на вырожденность. Если определитель матрицы **V** равен нулю, то она является вырожденной.

Определитель **V** = 0

Корреляционная матрица для матрицы **V** =

r(1,1) =	1
r(1,2) =	-1
r(1,3) =	1
r(1,4) =	-1

Определитель матрицы **V** равен нулю, так как ее столбцы линейно зависят друг от друга (коэффициенты корреляции между столбцами показывают на их линейную зависимость).

Примечание. Метод Эйткена в эконометрике является одним из самых сложных, поэтому необходимо повторить свойства матриц по литературе, посвященной линейной алгебре.

Вывод. Определитель матрицы **V** равен нулю, следовательно, она является вырожденной и мы не сможем вычислить коэффициенты Эйткена. Однако при определенных значениях матрицы **V** она может быть невырожденной и можно будет вычислить коэффициенты Эйткена, что будет сделано на следующем занятии.

Дискуссия. Можно ли с помощью выборочных матриц **ee’** определить математическое ожидание матрицы **V**, которая была бы невырожденной? Предположим, по выборочным данным нам удалось вычислить несколько матриц **ee’**, по которым вычислим средние значения для каждого ее элемента, получим среднюю матрицу $\bar{\mathbf{V}}$.

Очевидно, что диагональные элементы средней матрицы будут положительные числа, как правило, отличные от нуля.

Недиагональные элементы средней матрицы могут быть как положительными, так и отрицательными, равными или неравными нулю.

Если недиагональные элементы средней матрицы $\overline{\mathbf{V}}$ не равны нулю (т. е. между остатками существует автокорреляция), то она остается вырожденной и нельзя будет вычислить коэффициенты Эйткена. Однако из теории эконометрики известно, что при определенных свойствах автокорреляций остатков, недиагональные элементы средней матрицы не равны нулю, но матрица становится невырожденной. Следовательно, от нее можно найти обратную матрицу и определить коэффициенты Эйткена.

Если автокорреляция остатков отсутствует, то недиагональные элементы средней матрицы $\overline{\mathbf{V}}$ будут равны нулю, матрица станет невырожденной, то можно определить от нее обратную матрицу и следовательно вычислить коэффициенты Эйткена.

Задание третьего уровня сложности

Задание. Произведите расчеты коэффициенты линейной модели по матричной формуле

$$\mathbf{A = (X'X)^{-1}X'У}$$

с помощью одного математического выражения с использованием вложенных матричных операций.

Пояснение. Эту задачу можно решить двумя способами: прямым и обратным.

Прямой способ состоит в том, что в математическое выражение последовательно добавляются матричные операции в том порядке, как они выполняются (см. решение задачи 1). Например, определяется транспонированная матрица **X'**, затем она умножается на исходную матрицу **X** и т. д. Фрагмент этого выражения будет иметь следующий вид при условии, что исходная матрица **X** находится в ячейках A1:B4:

= МУМНОЖ((ТРАНСП(A1:B4));(A1:B4)).

Обратный способ состоит в том, что математическое выражение формируется с последних операций. Например, производится матричное умножение выражения **(X'X)$^{-1}$X'** на **У**. При этом матрица **У** известна, а матрица **(X'X)$^{-1}$X'** не известна и в дальнейшем раскрывается. Фрагмент этого выражения будет иметь следующий вид при условии, что матрица **У** находится в ячейках C1:C4.

= МУМНОЖ((ТРАНСП(:));(C1:C4)).

Если формула введена правильно, то расчеты по ней должны совпадать с решением задачи 1.

Задание четвертого уровня сложности

Задание. Реализуйте в Excel алгоритм расчета собственных значений и собственных векторов матрицы.

Пояснение. Имеется много разных алгоритмов (около двадцати) расчета собственных значений и собственных векторов матрицы, описанных в книгах по линейной алгебре, в программах, написанных на разных языках. Алгоритм расчета собственных значений и собственных векторов матрицы хорошо и доступным языком описан, с решением с решением контрольного примера, в литературе[1]. Реализация этого алгоритма в Excel не составила особого труда.

Выходной тест

Имеется две матрицы **A**, **B**:

$$\mathbf{A} = \begin{vmatrix} 1 & 2 & 3 & 1 \end{vmatrix},$$

$$\mathbf{B} = \begin{vmatrix} 1 & 2 \\ 2 & 5 \\ 4 & 1 \end{vmatrix}.$$

Необходимо получить произведение матриц **AB**

а) нет решения;

б) **AB** = |17 15|.

[1] Иберла К. Факторный анализ. — М.: Статистика, 1980.

7. ЛИНЕЙНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ С ГЕТЕРОСКЕДАСТИЧНЫМИ ОСТАТКАМИ

Постановка задачи

Объект — показатели деятельности экономического объекта.

Предмет — нарушения предпосылок метода наименьших квадратов.

Цель — Изучить обобщенный метод наименьших квадратов (метод Эйткена)

Актуальность — неоднородность остатков приводит к увеличению ошибок коэффициентов модели, доверительные интервалы регрессии и прогноза не соответствуют действительности.

Метод — Эйткена.

Способ — проведение расчетов с помощью матричных операций.

Задачи — получить доверительные интервалы уравнения регрессии и прогноза с учетом гетероскедастичности остатков.

Рабочая гипотеза — предполагается, что известна ковариационная матрица остатков.

Ожидаемый результат — ожидается, что доверительные интервалы уравнения регрессии и прогноза будут учитывать неоднородности остатков.

Метод для сравнения — построение доверительных интервалов уравнения регрессии и прогноза без учета и с учетом неоднородности остатков.

Задания первого уровня сложности

Задание 1. Подготовьте устные сообщения по вопросам самостоятельной работы.

1. Дайте определение гетероскедастичности остатков.
2. Перечислите возможные причины возникновения гетероскедастичности остатков.
3. Перечислите последствия наличия в модели гетероскедастичности остатков.
4. Укажите основные направления устранения гетероскедастичности остатков.

1.5.Опишите критерии обнаружения гетероскедастичности остатков с помощью графического анализа.

6. Опишите суть метода Эйткена устранения гетероскедастичности остатков.

[3, с. 306–322, 7, с. 150–167, 8, с. 138–152, 12, с. 200–217, 1, т. 2 с. 102–111).

Задание 2. Пройдите входной тест.

Выберите вариант правильного ответа

1. Гомоскедастичность происходит от слова гомо... (гр. homos — равный, одинаковый) и означает однородность.

Гомоскедастичность остатков означает равенство дисперсий остатков для каждого значения X_i.

а) да;

б) нет.

2. Гетероскедастичность происходит от слова гетеро (гр. heteros — другой, неодинаковый), означает неоднородность.

Гетероскедастичность остатков означает различие дисперсий остатков при всех фиксированных значениях объясняемой переменной.

а) да;

б) нет.

3. Если в регрессионной модели допускается наличие гетероскедастичности остатков, то оценки параметров модели, вычисленные методом наименьших квадратов, остаются несмещенными, но становятся неэффективными.

Неэффективность оценок параметров модели означает, что возрастают ошибки: коэффициентов, модели; снижается достоверность модели. Доверительные интервалы уравнения регрессии и прогноза не соответствуют действительным данным.

а) да;

б) нет.

4. На рис. 7.1 приведен пример гетероскедастичных дисперсий остатков модели.

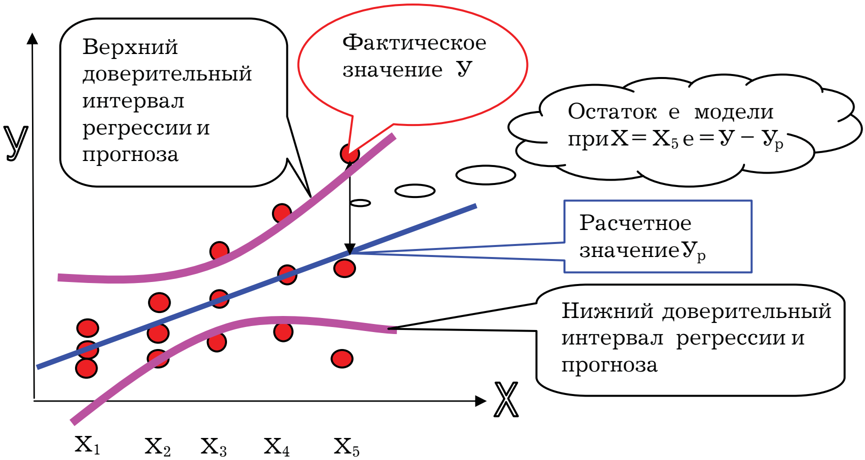


Рис. 7.1. Пример гетероскедастичных дисперсий остатков модели для каждого значения Х при наличии повторных значений У

Анализ рис. 7.1 показывает, что при малых значениях Х доверительный интервал уравнения регрессии превышает ошибку модели, а при больших значениях Х — меньше ошибки модели.

а) да;

б) нет.

5. Если в модели:

$\mathbf{У = XA} + \varepsilon$

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

$\mathbf{M\ (\varepsilon\varepsilon') = V} \neq \sigma_\varepsilon{}^2\mathbf{I}$,

где **V** — симметрическая и положительно определенная ковариационная матрица, то возмущения могут быть автокоррелированными и неоднородными (гетероскедастичными). Коэффициенты модели, вычисленные по формуле

$$\mathbf{A = (X'X)^{-1}\ X'У}$$

позволят получить оценки параметров модели несмещенными, состоятельными, но неэффективными.

Если расчеты коэффициентов модели произвести по формуле

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

то они будут несмещенными, состоятельными и эффективными для новых преобразованных данных.

а) да;

б) нет.

Задание 3. Решите задачу.

Было проведено пространственное бюджетное обследование семей с разными уровнями дохода за интервал времени, равный одному году. Изучалась зависимость размера потребления предметов роскоши от уровня дохода.

Выборочные данные одной повторности представлены в табл. 7.1.

Таблица 7.1

Данные бюджета семьи (данные условные)

i	X_i	$У_i$
1	1	4
2	2	2
3	3	7
4	4	3
5	5	12
6	6	4
Ожидаем	7	?

где X_i — среднегодовой уровень дохода на одного члена семьи (усл. д.е .),

$У_i$ — среднегодовое потребление семьей предметов роскоши (усл. д. е.),

i — порядковый номер измерения.

Необходимо решить следующие задачи:

1. Получить характеристики модели с гетероскедастичными остатками.

2. Устранить гетероскедастичность обобщенным методом наименьших квадратов.

3. Получить прогнозные значения У для $X_{ож} = 7$, $|e_{ож}| = 6$

4. Сравнить характеристики полученных моделей.

Решение

У предприятия, реализующего предметы роскоши, возникла проблема, связанная с тем, что спрос на них имеет различное значение в зависимости от района города, где располагаются его магазины. Неоднородность спроса по районам города приводит к дополнительным издержкам на хранение и реализацию неходового товара.

Целью моделирования является определение однородных групп районов, в которых спрос на предметы роскоши имеет одинаковую дисперсию. Распределение товаров по однородным районам с учетом спроса на предметы роскоши будет способствовать снижению издержек хранения неходового товара.

Выбор переменных

Спрос на предметы роскоши зависит от шести основных факторов: *сырье и материалы* (ассортимент товаров и их качество, цена, уровень дохода семьи); *машины* (оборудование и дизайн магазина, наличие автотранспорта для доставки товара); *методы* (форма обслуживания покупателей: самообслуживание, культура обслуживания, способ доставки товара, формы перечисления денег за покупку); *люди* (количество продавцов, уровень квалификации, образование, возраст, деловые характеристики); *физическая и абстрактная среда* (растояние от магазина до станций метро, до центра города, до центра района; температура и влажность в магазине; традиции, психологический климат; правовая среда существования магазина, информационное обеспечение); *время* (нет влияния, так как имеется пространственная выборка) (рис. 7.2).

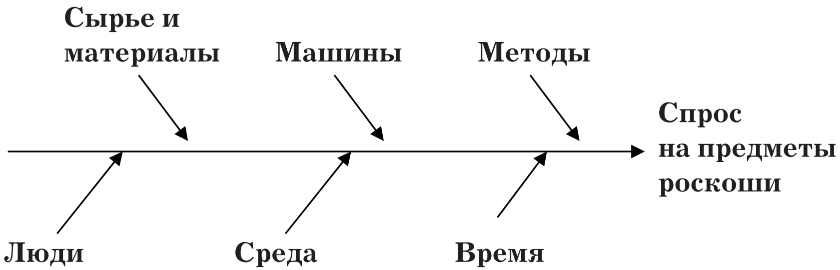


Рис. 7.2. Причинно-следственная диаграмма Исикавы.

Социологическое исследование районов города показало их неоднородность по среднему благосостоянию семей. Возникла гипотеза о том, что спрос на предметы роскоши зависит от уровня благосостояния семьи и если перераспределить поставку предметов роскоши в магазины с учетом на них спроса, то можно получить дополнительную прибыль. Отдел маркетинга предприятия поставил задачу перед отделом эконометрических исследований: определить зависимость потребление предметов роскоши от уровня благосостояния семьи.

Выделение входных и выходных переменных

Семья, как экономический объект имеет множество характеристик, но в соответствии с возникшей проблемой, связанной с неоднородностью (стратификацией) потребления предметов роскоши, было решено определить зависимость среднегодового потребления предметов роскоши от среднегодового уровня дохода на одного члена семьи.

Сбор статистических данных

Для определения зависимости потребления предметов роскоши (У) от уровня дохода (Х) были выбраны 6 семей с различными среднегодовыми уровнями дохода на одного человека и определены соответствующие им величины среднегодового потребления предметов роскоши. Исходные данные представлены в табл. 7.1. График зависимости У от Х изображен на рис. 7.3

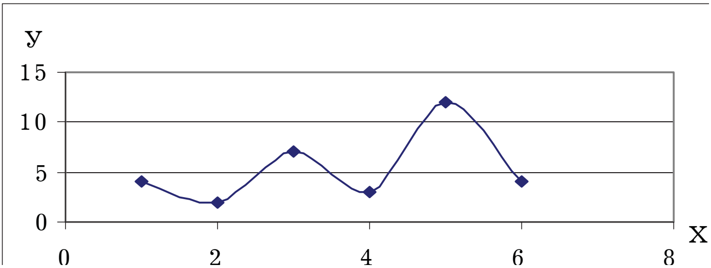


Рис. 7.3. График зависимости У от Х

Анализ рис. 7.3 показывает, что с увеличением значений уровня доходов размах значений У увеличивается.

Потребления предметов роскоши

Выдвижение гипотез

Социологические исследования показали, что с увеличением дохода семьи возрастают затраты на предметы роскоши. Однако замечено, что семьи по признаку затрат на предметы роскоши подразделяются на две группы: “скромные” — с ростом доходов у них слабо возрастают затраты на предметы роскоши, не демонстрируя своего богатства; “вызывающие” — демонстрируют всем свое богатство и стремятся к тому, чтобы у них было все самое лучшее, при этом чем больше доход, тем больше расходы на предметы роскоши, и только при определенных больших доходах расходы на предметы роскоши перестанут увеличиваться.

Формулировка допущений

Предполагаем, что расходы на предметы роскоши линейно зависят от дохода как для “скромных”, так и для “вызывающих” семей, и не содержат участка насыщения расходов. Предполагаем также, что автокорреляция остатков отсутствует.

Спецификация модели

Визуальный анализ рис. 7.3 показал, что амплитуда остатков возрастает с увеличением значений X. Следовательно, необходимо произвести спецификацию модели с учетом неоднородности (гетероскедастичности) остатков.

Обобщенная линейная модель парной регрессии

Предполагаем, что зависимость У от X можно представить в виде обобщенной линейной модели парной регрессии. Приводим спецификацию данной модели:

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

где **У** — вектор столбец случайной величины;

XA + ε — уравнение регрессионной модели;

XA — уравнение регрессии;

X — неслучайная матрица факторов;

A — вектор столбец неизвестных параметров;

ε — вектор столбец неизвестного возмущения.

Относительно возмущения ε выдвигаем следующие гипотезы:

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

$\mathbf{M}\ (\varepsilon\varepsilon') = \mathbf{V} \neq \sigma_\varepsilon{}^2\mathbf{I},$

где **V** — симметрическая и положительно определенная ковариационная матрица. При этом возмущения могут быть автокоррелированными и неоднородными (гетероскедастичными).

Идентифицируемость модели

(Можно ли в принципе однозначно восстановить значения неизвестных параметров по выбранной модели).

Если вычислить коэффициенты модели методом наименьших квадратов по формуле:

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

то полученные коэффициенты будут несмещенными, состоятельными, но неэффективными.

Идентификация модели

Получим характеристики модели 1 с гетероскедастичными остатками. Необходимо произвести эконометрический анализ линейной модели

$$У_i = a_0 + a_1 X_i + e_i\,.$$

Представим результаты расчетов с помощью Функции Excel “Линейн” в следующем виде:

0,7429	2,7333
0,9077	3,535
0,1434	3,7972
0,6697	4
9,6571	57,676

Представим результаты расчетов в общепринятом обозначении.

$У_р =$	2,7333	0,7429	X
$S_a =$	3,535	0,9077	
$t_a =$	0,7732	0,8184	
E =	3,7972		
$R^2 =$	0,1434	$t_{\alpha/2} =$	2,7765
F =	0,6697	$F_{кр} =$	7,7086

Определим доверительные интервалы математических ожиданий У (табл. 7.2).

Таблица 7.2

Расчет доверительных интервалов математических ожиданий У

i	X_i	$(X_i — \bar{X})^2$	$У_i$	$У_{pi}$	e_i	$У_{м\,min(i)}$	$У_{м\,max(i)}$
1	1	6,25	4	3,4762	0,5238	-4,1542	11,107
2	2	2,25	2	4,219	-2,219	-1,5095	9,9476
3	3	0,25	7	4,9619	2,0381	0,4771	9,4467
4	4	0,25	3	5,7048	-2,7048	1,22	10,19
5	5	2,25	12	6,4476	5,5524	0,7191	12,176
6	6	6,25	4	7,1905	-3,1905	-0,4399	14,821
Ожидаем	7			7,9333		-1,8815	17,748
Сумма		17,5					
$\bar{X} =$	3,5						

где $У_{pi} = a_0 + a_1\, x_i$, $e_i = У_i — У_{pi}$,

$У_{м\,min(i)} = У_р - t_{\alpha/2} S_{Уm}$

$У_{м\,max(i)} = У_р - t_{\alpha/2} S_{Уmi}$

$$S_{Ymi} = E\sqrt{\frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum (X_i - \overline{X})^2}};$$

$t_{\alpha/2} = 2{,}7765$, $E = 3{,}7972$, $n = 6$.

Построим график основных характеристик регрессионной модели (рис. 7.4).

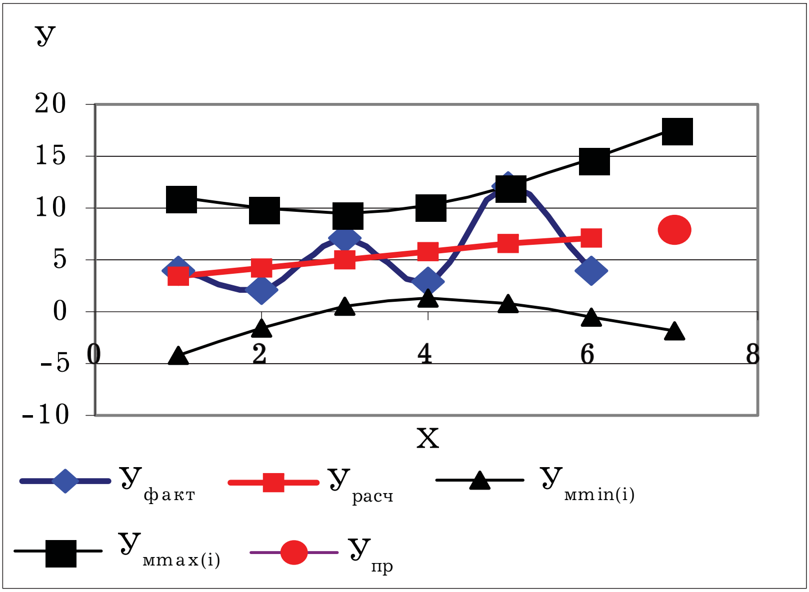


Рис. 7.4. График основных характеристик регрессионной модели 1, вычисленных методом наименьших квадратов (с гетероскедастичными остатками)

Верификация

Анализ рис. 7.4 показывает, что модель 1 с гетероскедастичными остатками имеет доверительные интервалы уравнения регрессии, которые не учитывают реальных изменений зависимой переменной. При малых значениях Х доверительные интервалы уравнения регрессии больше размаха значений У, при больших значениях Х доверительный интервала регрессии больше размаха значений У.

Для коэффициентов модели a_0 и a_1, вычисленных методом наименьших квадратов, имеются критерии Стьюдента, соответственно равные 0,7732 и 0,818382, которые являются недостоверными, так как имеется неоднородность остатков.

Вывод. При наличии гетероскедастичности остатков модель имеет большие ошибки как модели, так и коэффициентов.

Идентификация модели

Обобщенный метод наименьших квадратов

Если вычислить коэффициенты модели с помощью обобщенного метода наименьших квадратов (при известной матрице **V**) по формуле Эйткена

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

то полученные коэффициенты будут несмещенными, состоятельными и эффективными по отношению к определенным образом преобразованным исходным значениям У и Х.

Метод Эйткена позволяет получить преобразованные данные, для которых гетероскедастичность будет отсутствовать.

Примечание. Однако очень часто новым переменным нельзя дать экономическую интерпретацию.

Приводим алгоритмы обобщенного метода наименьших квадратов, или метода Эйткена.

Метод преобразования данных

Метод Эйткена лучше понять, если использовать его эквивалент в виде метода преобразования данных.

Метод преобразований данных для учета и устранения гетероскедастичности остатков эквивалентен методу Эйткена и заключается в том, что переменные уравнения регрессии

$$У_i = a_0 X_{1i} + a_1 X_{2i} + e_i$$

преобразуются в новые переменные с ????????????? “Н” по формулам:

$\mathbf{У_н = P^{-1} У}$ или $У_{нi} = У_i / |e_i|$,

$\mathbf{X_{н1} = P^{-1} X_1}$ или $X_{н1i} = 1 / |e_i|$,

$\mathbf{X_{н2} = P^{-1} X_2}$ или $X_{н2i} = X_{2i} / |e_i|$, где X_1 - переменная, содержащая единицы, служащие для учета свободного коэффициента;

X_2 — объясняемая переменная;

$|e_i|$ — значения остатков, взятые по модулю;

e_i — остатки модели определяются из уравнения

$У_i = a_0 + a_1 X_{1i} + e_i$;

a_0, a_1 _ коэффициенты модели, определенные обычным методом наименьших квадратов по исходным непреобразованным данным.

Если матрица **V** является симметрической и положительно определенной, то возможно следующее соотношение:

$$\mathbf{V = PP'},$$

где матрица P является невырожденной.

Матрицу **P** можно вычислить по формуле:

$$\mathbf{P = ((hD)^{-1}},$$

$$\mathbf{D} = \mathbf{I}\frac{1}{\sqrt{\mathbf{h'Vh}}} = \mathbf{I}\frac{1}{\sqrt{\mathbf{k}}}$$

где **I** — единичная матрица,

h — собственный (характеристический) вектор матрицы **V,**

k — собственное (характеристическое) значение матрицы **V**.

Собственные значения **k** матрицы **V** можно определить из уравнения

$$\mathbf{Vh = kh}.$$

обычно значение **h** и **k** вычисляются с помощью статистических пакетов и прикладных программ. В Excel нет расчетов **h** и **k**.

Предполагается, что

$$\mathbf{V = M}\ (\varepsilon\varepsilon'),$$

V — положительно определенная матрица;

ε — остатки модели, рассчитанные по уравнению регрессии, учитывающие параметры модели для генеральной совокупности;

M — знак математического ожидания.

Симметрическую, положительно определенную матрицу **V** можно представить как произведение двух матриц по формуле

$$\mathbf{V = PP'},$$

$$\mathbf{D} = \begin{pmatrix} 1/\sqrt{k_1} & 0 & \ldots & 0 \\ 0 & 1/\sqrt{k_2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ & 0 & \ldots & 1/\sqrt{k_m} \end{pmatrix},$$

k_i — собственные значения матрицы **V**,
m — размерность матрицы **V.**

Примечание. Если матрица **V** — диагональная, то расчет матрицы **P** упрощается

V = PP' = PP, P = $\sqrt{\mathbf{V}}$, так как, если матрица **V** диагональная, то матрица **P** тоже диагональна, тогда **P = P'**

$$\mathbf{P} = \begin{pmatrix} \sqrt{V_{11}} & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \sqrt{V_{22}} & \ldots & 0 \\ 0 & 0 & \ldots & \sqrt{V_{mm}} \end{pmatrix},$$

$$\mathbf{P}^{-1} = \begin{pmatrix} 1/\sqrt{V_{11}} & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 1/\sqrt{V_{22}} & \ldots & 0 \\ 0 & 0 & \ldots & 1/\sqrt{V_{mm}} \end{pmatrix}$$

Предположим, что матрица **V** является диагональной, элементы которой равны квадратам остатков линейной модели, определенным по выборочным данным. Тогда матрица $\mathbf{P}^{-1}$ будет иметь вид:

$$\mathbf{P}^{-1} = \begin{pmatrix} 1/|e_1| & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 1/\,|e_2| & \ldots & 0 \\ 0 & 0 & \ldots & 1/\,|e_m| \end{pmatrix}$$

Следовательно, соблюдается эквивалентность преобразований

$\mathbf{У_н = P^{-1} У}$ или $У_{нi} = У_i / |e_i|$,

$\mathbf{X_{н1} = P^{-1} X_1}$ или $X_{н1i} = 1 / |e_i|$,

$\mathbf{X_{н2} = P^{-1} X_2}$ или $X_{н2i} = X_{2i} / |e_i|$.

После преобразований анализируется следующая модель

$У_{нi} = b_0 X_{н1i} + b_1 X_{н2i} + z_i$.

Примечание. Следует заметить, что в этой модели нет свободного коэффициента. Однако с помощью функции Линейн можно вычислить коэффициенты модели, при условии, что свободный коэффициент равен нулю.

Расчет коэффициентов Эйткена можно произвести по формуле, представленной в матричном виде:

$$\mathbf{B = (X_н' X_н)^{-1} X'_н У_н = (X'V^{-1}X)^{-1} X'V^{-1}У.}$$

В дальнейшем производится расчет всех характеристик модели и эконометрический анализ преобразованных данных по формулам регрессионного анализа с использованием традиционного метода наименьших квадратов.

Примечание. Характеристики модели, вычисленные обобщенным методом наименьших квадратов для исходных данных, и характеристики модели, вычисленные традиционным методом наименьших квадратов для преобразованных данных, должны совпадать между собой.

Для перехода к исходным непреобразованным данным необходимо полученные точечные и интервальные прогнозы, а также расчетные значения зависимой преобразованной переменной и доверительные интервалы уравнения регрессии для преобразованных данных умножить на соответствующие значения ожидаемых или текущих остатков, взятых по модулю;

$$(|e_{ож}| \text{ или } |e_i|),$$

где $|e_i|$- значение остатка, взятое по модулю;

e_i — остатки модели, определяются из уравнения

$$У_i = a_0 + a_1 X_{1i} + e_i,$$

a_0, a_1 — коэффициенты модели, определенные обычным методом наименьших квадратов по исходным непреобразованным данным.

Примечания

1. Данная процедура перехода к исходным данным не вызывает особых возражений, так как переход к преобразованным значениям осуществлялся делением переменных на абсолютное значение остатков, то обратный переход предлагается производить умножением на абсолютное значение этих же остатков. Доказательство данной процедуры отсутствует, но пока в научной литературе мы не смогли обнаружить метода данного перехода.

2. Расчеты показывают, что данная процедура позволяет получить более обоснованные доверительные интервалы прогноза и уравнения регрессии.

Устраним гетероскедастичность обобщенным методом наименьших квадратов.

Преобразуем исходные переменные с помощью остатков модели по формулам:

$$У_{н\cdot i} = У_i/|e_i|,\ X_{1н\cdot i} = 1/|e_i|,\ X_{2н\cdot i} = X_i/|e_i|,$$

где $У_i$, $У_{нi}$ — фактические и преобразованные значения зависимой переменной;

X_i, $X_{н\cdot i}$ — фактические и преобразованные значения объясняемой переменной;

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

a_0, a_1 — коэффициенты, определенные методом наименьших квадратов,

$|e_i|$ — модуль остатка.

Таблица 7.3

Исходные и преобразованные данные

i	X_0	X_i	$У_i$	e_i	$X_{1нi}$	$X_{2нi}$	$У_{нi}$
1	1	1	4	0,5238	1,9091	1,9091	7,6364
2	1	2	2	-2,219	0,4506	0,9013	0,9013
3	1	3	7	2,0381	0,4907	1,472	3,4346
4	1	4	3	-2,7048	0,3697	1,4789	1,1092
5	1	5	12	5,5524	0,1801	0,9005	2,1612
6	1	6	4	-3,1905	0,3134	1,8806	1,2537

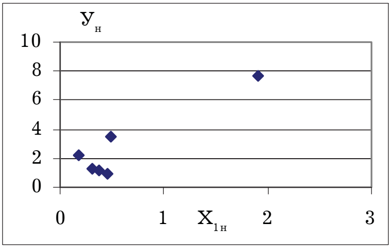


Рис. 7.5. Зависимость $У_н$ от $X_{1н}$

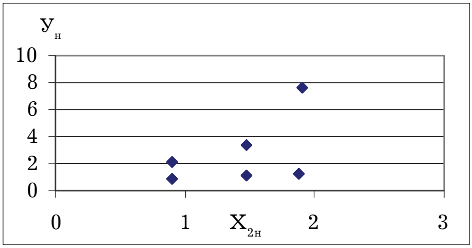


Рис. 7.6. Зависимость $У_н$ от $X_{2н}$

Анализ графиков, изображенных на рис. 7.5 и рис. 7.6, показывает, что чем меньше остаток модели для конкретного измерения, тем дальше оно удаляется от остальных измерений и этому измерению будет “уделено больше внимания” при расчетах коэффициентов модели.

Определим коэффициенты Эйткена $в_1$ и $в_2$ линейной модели 2 по преобразованным данным с помощью функции Линейн без свободного коэффициента:

$У_{нi} = в_1 X_{1нi} + в_2 X_{2нi} + z_i$ — модель 2.

0,318	3,6575
0,4988	0,8669
0,8586	1,0808
12,143	4
28,368	4,6725

Представим результаты расчетов в общепринятом обозначении.

$У_{нр}$ =	3,6575	$X_{1н}$ +	0,318	$X_{2н}$
S_a =	0,8669		0,4988	
t_a =	4,219		0,6376	
E =	1,0808			
R^2 =	0,8586		$t_{\alpha/2}$ =	2,7765
F =	12,143		$F_{кр}$ =	7,7086

Построим график основных характеристик модели 2 (рис. 7.7).

Анализ расчетов показывает, что для преобразованных данных модель 2 стала достоверной, коэффициент $в_1$ является достоверным, коэффициент $в_2$ является недостоверным.

Определим доверительные интервалы уравнения регрессии для модели 2.

Таблица 7.4

Расчет доверительных интервалов уравнения регрессии

i	$X_{1нi}$	$X_{2нi}$	$У_{нi}$	$У_{рнi}$	z_i	$У_{mini}$	$У_{maxi}$
1	1,9091	1,9091	7,6364	7,5896	0,0468	4,5888	10,59
2	0,4506	0,9013	0,9013	1,9349	-1,0336	-1,0659	4,9356
3	0,4907	1,472	3,4346	2,2627	1,1719	-0,7381	5,2635
4	0,3697	1,4789	1,1092	1,8226	-0,7134	-1,1782	4,8233
5	0,1801	0,9005	2,1612	0,9451	1,2161	-2,0557	3,9459
6	0,3134	1,8806	1,2537	1,7445	-0,4907	-1,2563	4,7452
Ожидаем	0,1667	1,1667		0,9806		-2,0202	3,9814

где $У_{рнi} = в_1X_{1нi} + в_2X_{2нi}$, $z_i = У_{нi} — У_{рнi}$,

$У_{mini} = У_{рнi} — t_{\alpha/2}\,E$, $У_{maxi} = У_{рнi} + t_{\alpha/2}\,E$,

$t_{\alpha/2} = 2{,}7765$, $E = 1{,}0808$.

Примечание. Доверительные интервалы уравнения регрессии рассчитаны по упрощенной формуле потому, что модель 2 является множественной регрессией, так как содержит два фактора. Доверительные интервалы математических ожиданий М(У|X) множественной регрессии вычисляются по формуле

$$\mathbf{c'A} \pm t_{\alpha/2}E\sqrt{\mathbf{c'(X'X)^{-1}c}},$$

где $\mathbf{c} = \{1\ X_{1i}\ X_{2i}\ \ldots\ X_{mi}\}$ — вектор столбец значений **X** для i-го измерения;

X — матрица всех факторов, включенных в модель, с учетом столбца единиц для свободного коэффициента;

A — вектор столбец коэффициентов модели;

E — ошибка модели;

m — количество факторов, включенных в модель;

$t_{\alpha/2}$ — двухсторонний критерий Стьюдента, определяемый на уровне значимости α и числе степеней свободы n — k, где n — объем выборки, k — число всех коэффициентов в модели. [13, с. 154].

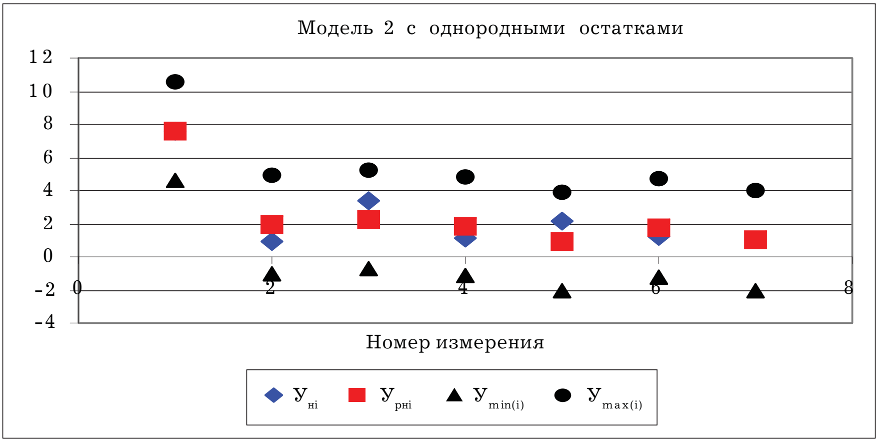


Рис. 7.7. Характеристики модели 2 с преобразованными значениями и однородными остатками

Анализ рис. 7.7 показывает, что остатки модели 2 стали гомоскедастичными или однородными, т. е. гетероскедастичность устранена. Однако не ясно экономическое содержание процедуры получения новых переменных делением исходных значений на модуль остатков и что теперь означают новые переменные $Х_{1н}$, $Х_{2н}$, $У_{н}$

Примечание. Если квадраты амплитуд остатков были бы пропорциональны значениям Х, то можно преобразование исходных данных произвести их делением на Х, тогда новые переменные приобретут экономическое содержание. Например, для нашей задачи новая переменная У/Х будет означать величину среднегодового потребления предметов роскоши, приходящихся на одну усл, д. е. среднегодого уровня дохода на одного члена семьи; 1/Х будет означать обратную величину среднегодового уровня дохода на одного члена семьи.

Вывод. Полученная модель 2 является новой моделью с новыми переменными, которые не имеют экономического смысла, но нам удалось в этой модели исключить гетероскедастичность остатков и получить достоверную модель. Однако остается проблема учета гетероскедастичности для исходных, а не для преобразованных данных.

Верификация

Метод учета гетероскедастичности для исходных данных

В учебной и научной литературе по эконометрике нам не удалось обнаружить метода устранения гетероскедастичности для исходных данных. Поэтому выдвигается следующая гипотеза: если преобразование исходных данных производилось делением их на модуль остатков, то доверительные интервалы уравнения регрессии, построенного по преобразованным значениям следует умножить на модуль тех же остатков, тогда, очевидно, полученные доверительные интервалы регрессии будут учитывать гетероскедастичность исходных данных. Обоснования данного метода пока нет, так как он основан на интуиции и простой логике. Реализуем предложенный метод учета гетероскедастичности для исходных данных (табл. 7.5).

Таблица 7.5

Доверительные интервалы уравнения регрессии для исходных данных

i	$Х_i$	$У_i$	$У_{рнi}$	$У_{mini}$	$У_{maxi}$	e_i	$У_{р.Э.i}$	$У_{min.Э.i}$	$У_{max.Э.i}$
1	1	4	7,5896	4,5888	10,59	0,5238	3,9755	2,4037	5,5473
2	2	2	1,9349	-1,0659	4,9356	-2,219	4,2935	-2,3654	10,952
3	3	7	2,2627	-0,7381	5,2635	2,0381	4,6116	-1,5043	10,727
4	4	3	1,8226	-1,1782	4,8233	-2,7048	4,9296	-3,1868	13,046
5	5	12	0,9451	-2,0557	3,9459	5,5524	5,2476	-11,414	21,909
6	6	4	1,7445	-1,2563	4,7452	-3,1905	5,5657	-4,0083	15,14
Прогноз	7	0	0,9806	-2,0202	3,9814	6	5,8837	-12,121	23,888

где $У_{р.Э\,i} = У_{рнi}\,|e_i|$, $У_{min.Э.i} = У_{mini}\,|e_i|$, $У_{max.Э.i} = У_{maxi}\,|e_i|$;

$У_{р.Э\,i}$ — расчетные значения У, полученные с использованием коэффициентов Эйткена и исходных значений Х.

Примечание. Экспериментальным методом получено, что $У_{р.Э.i} = У_{рнi} |e_i| = в_1 + в_2 Х_i$, где $в_1$, $в_2$ — коэффициенты Эйткена. Значения $У_{р.Э.i}$ проходят ближе к тем значениям $У_i$, у которых остатки e_i имеют меньшую величину. В этом заключается метод взвешенной регрессии.

$У_{min.Э.i}$, $У_{max.Э.i}$ — нижний и верхний доверительные интервалы Эйткена уравнения регрессии с учетом гетроскедастичности остатков,

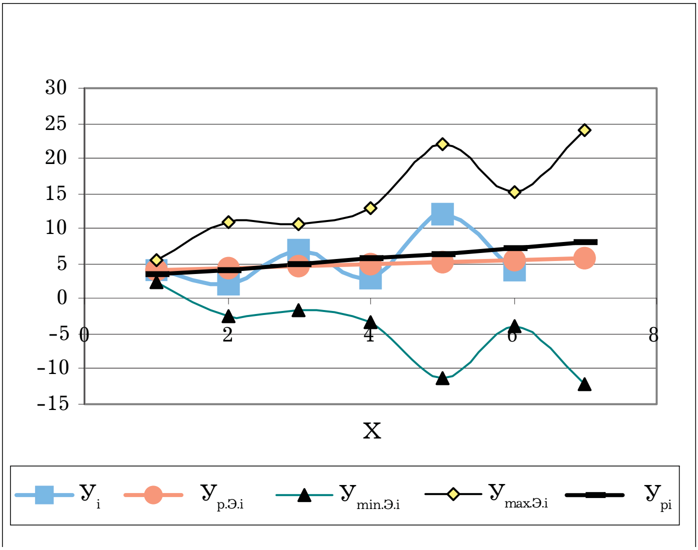


Рис. 7.8. График доверительных интервалов регрессии с учетом гетероскедастичности остатков для исходных данных

$У_{pi}$ — расчетные значения, найденные обычным методом наименьших квадратов

Анализ рис. 7.8 показывает, что доверительные интервалы Эйткена уравнения регрессии учитывают гетероскедастичность остатков.

Вывод. Метод Эйткена позволил получить более обоснованные доверительные интервалы уравнения регрессии при наличии гетероскедастичности остатков.

Задание второго уровня сложности

Задание. Получите модель взвешенной регрессии

Коэффициенты Эйткена используются для получения взвешенной регрессии.

Если коэффициенты Эйткена использовать для исходных непреобразованных данных, то расчетные значения $У_{р.Э.i} = в_1 + в_2 X_i$ будут проходить ближе к тем значениям У, которые имеют меньшую ошибку модели (весовой коэффициент). Для иллюстрации расчетов воспользуемся данными задачи 1.

Таблица 7.6

Расчет взвешенной регрессии

i	X_i	$У_i$	$У_{pi}$	$У_{р.Эi}$
1	1	4	3,4762	3,9755
2	2	2	4,219	4,2935
3	3	7	4,9619	4,6116
4	4	3	5,7048	4,9296
5	5	12	6,4476	5,2476
6	6	4	7,1905	5,5657
Прогноз	7	0	7,9333	5,8837

где $У_{pi} = a_0 + a_1 x_i$, $У_{р.Э.i} = в_1 + в_2 x_i$;

$a_0 = 2,7333$;

$a_1 = 0,7429$;

$в_1 = 3,6575$;

$в_2 = 0,318$,

a_0, a_1 — коэффициенты модели $У_i = a_0 + a_1 X_i + e_i$, определенные методом наименьших квадратов;

$в_1$, $в_2$ — коэффициенты Эйткена модели $У_{нi} = в_1 X_{1нi} + в_2 X_{2нi} + z_i$, определенные обобщенным методом наименьших квадратов.

Анализ графиков, изображенных на рис. 7.9 показывает, что уравнение регрессии Эйткена (взвешенная регрессия) проходит ближе к тем значениям У, ошибка модели которой e_i является наименьшей.

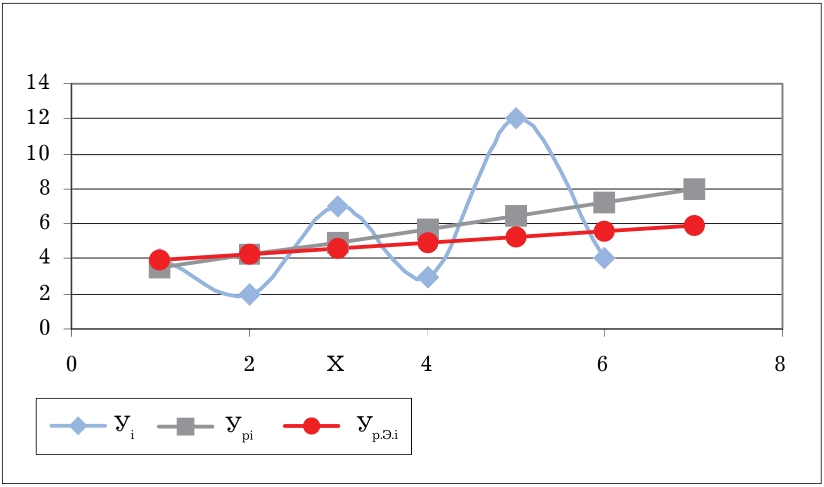


Рис. 7.9. Графики расчетных значений У, полученных методом наименьших квадратов и обобщенным методом наименьших квадратов (метод Эйткена)

Задние третьего уровня сложности

Задание. Вычислить коэффициенты Эйткена матричным способом.

Для иллюстрации расчетов воспользуемся данными задачи 1. Коэффициенты Эйткена можно вычислить с помощью обобщенного метода наименьших квадратов по матричным формулам:

$$\mathbf{B = (X_н'X_н)^{-1}X'_нУ_н = (X'V^{-1}X)^{-1}\,X'V^{-1}У,}$$

где

$$\mathbf{X} = \begin{array}{|cc|} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{array} \quad \mathbf{У} = \begin{array}{|c|} 4 \\ 2 \\ 7 \\ 3 \\ 12 \\ 4 \end{array} \quad |\mathbf{e}| = \begin{array}{|c|} 0{,}5238 \\ 2{,}219 \\ 2{,}0381 \\ 2{,}7048 \\ 5{,}5524 \\ 3{,}1905 \end{array}$$

$$|\mathbf{e}|' = \begin{array}{|c|c|c|c|c|c|} 0{,}5238 & 2{,}219 & 2{,}0381 & 2{,}7048 & 5{,}5524 & 3{,}1905 \end{array}$$

$\mathbf{V_1 = ee' =}$

0,2744	1,1624	1,0676	1,4168	2,9084	1,6712
1,1624	4,9242	4,5226	6,002	12,321	7,0798
1,0676	4,5226	4,1538	5,5126	11,316	6,5025
1,4168	6,002	5,5126	7,3157	15,018	8,6295
2,9084	12,321	11,316	15,018	30,829	17,715
1,6712	7,0798	6,5025	8,6295	17,715	10,179

В матрице $\mathbf{V_1}$ оставим только диагональные элементы, так как согласно модели автокорреляция остатков отсутствует.

$\mathbf{V =}$

0,2744	0	0	0	0	0
0	4,9242	0	0	0	0
0	0	4,1538	0	0	0
0	0	0	7,3157	0	0
0	0	0	0	30,829	0
0	0	0	0	0	10,179

$\mathbf{V^{-1} =}$

3,6446	0	0	0	0	0
0	0,2031	0	0	0	0
0	0	0,2407	0	0	0
0	0	0	0,1367	0	0
0	0	0	0	0,0324	0
0	0	0	0	0	0,0982

Нетрудно проверить, что диагональные элементы матрицы $\mathbf{V^{-1}}$ равны

$\mathbf{V^{-1}}_{\mathbf{ii}} = 1/|e_i|$.

Так как матрица $\mathbf{V}$ диагональная, симметрическая и положительно определенная, то ее можно представить в виде

$\mathbf{V = PP'}$.

Так как матрица $\mathbf{P}$ тоже должна быть диагональной, то $\mathbf{P = P'}$, тогда $\mathbf{V = PP}$, $\mathbf{P} = \sqrt{\mathbf{V}}$

$$\mathbf{P} = \sqrt{\mathbf{V}} = \begin{bmatrix} 0{,}5238 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2{,}219 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2{,}0381 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2{,}7048 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5{,}5524 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3{,}1905 \end{bmatrix}$$

$$\mathbf{P}^{-1} = \begin{bmatrix} 1{,}9091 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0{,}4506 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0{,}4907 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0{,}3697 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0{,}1801 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0{,}3134 \end{bmatrix}$$

Нетрудно проверить, что диагональные элементы матрицы $\mathbf{P}^{-1}$ равны

$$\mathbf{P}^{-1}_{ii} = 1/|e_i|.$$

$$\mathbf{X}_н = \mathbf{P}^{-1}\,\mathbf{X} = \begin{bmatrix} 1{,}9091 & 1{,}9091 \\ 0{,}4506 & 0{,}9013 \\ 0{,}4907 & 1{,}472 \\ 0{,}3697 & 1{,}4789 \\ 0{,}1801 & 0{,}9005 \\ 0{,}3134 & 1{,}8806 \end{bmatrix} \qquad \mathbf{У}_н = \mathbf{P}^{-1}\,\mathbf{У} = \begin{bmatrix} 7{,}6364 \\ 0{,}9013 \\ 3{,}4346 \\ 1{,}1092 \\ 2{,}1612 \\ 1{,}2537 \end{bmatrix}$$

Вычислим коэффициенты Эйткена по формуле

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

$\mathbf{X}_н{}' =$

1,9091	0,4506	0,4907	0,3697	0,1801	0,3134
1,9091	0,9013	1,472	1,4789	0,9005	1,8806

$$\mathbf{X}_н{}'\mathbf{X}_н = \begin{bmatrix} 4{,}3558 & 6{,}0714 \\ 6{,}0714 & 13{,}158 \end{bmatrix}$$

$(X_н'X_н)^{-1} =$

0,6433	-0,2968
-0,2968	0,213

$(X_н'X)^{-1} X_н' =$

0,6615	0,0224	-0,1213	-0,2011	-0,1514	-0,3566
-0,1601	0,0582	0,1678	0,2052	0,1383	0,3075

$(X_н'X)^{-1} X_н' У_н =$

3,6575	$= в_1$
0,318	$= в_2$

Вычислим коэффициенты Эйткена по формуле

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

$X' =$

1	1	1	1	1	1
1	2	3	4	5	6

$X'V^{-1} =$

3,6446	0,2031	0,2407	0,1367	0,0324	0,0982
3,6446	0,4062	0,7222	0,5468	0,1622	0,5894

$X'V^{-1}X =$

4,3558	6,0714
6,0714	13,158

$(X'V^{-1}X)^{-1} =$

0,6433	-0,2968
-0,2968	0,213

$(X'V^{-1}X)^{-1}X' =$

0,3465	0,0496	-0,2472	-0,544	-0,8409	-1,1377
-0,0839	0,1291	0,3421	0,555	0,768	0,981

$(X'V^{-1}X)^{-1} X'V^{-1} =$

1,2628	0,0101	-0,0595	-0,0744	-0,0273	-0,1118
-0,3057	0,0262	0,0823	0,0759	0,0249	0,0964

$(X'V^{-1}X)^{-1}X'V^{-1}У =$

3,6575	$= в_1$
0,318	$= в_2$

Вычислим коэффициенты Эйткена с помощью функции "Линейн" при использовании матрицы $\mathbf{X}_н$ и $\mathbf{У}_н$ без свободного коэффициента.

0,318	3,6575
0,4988	0,8669
0,8586	1,0808
12,143	4

$в_1 =$	3,6575
$в_2 =$	0,318

Сравнение коэффициентов Эйткена, определенных разными способами, показывает, что они равны между собой.

Если матрица **V** имеет только автокорреляцию остатков, то имеется возможность вычисления коэффициентов Эйткена, которые будут несмещенными, состоятельными и эффективными для специальным образом преобразованных данных.

Модели с автокоррелированными остатками мы будем изучать на следующем занятии.

Задание четвертого уровня сложности

Задание. Проверьте наличие гетероскедастичность в исходных данных задачи 1 по статистическим критериям и тестам:

— критерий Гольдфельда- Квандтома,

— тест ранговой корреляции Спирмена,

— тест Глейзера,

— тест Бреуша-Пагана.

Выходной тест

Расчет коэффициентов обобщенной линейной модели производится по формуле

$\mathbf{(X'V^{-1}X)^{-1}X'\,VУ}$

а) да;

б) нет.

Нерешенная проблема по теме гетероскедастичность:

Нет примеров одновременного устранения гетероскедастичности и автокорреляции остатков.

Пояснение. Если автокорреляция остатков будет иметь такой вид:

$$e_t = pe_{t-1} + v_t,$$

то матрица **V** будет положительно определенной и невырожденной, от которой можно получить обратную матрицу и матрицу **P**.

8. РЕГРЕССИОННЫЕ МОДЕЛИ С ПЕРЕМЕННОЙ СТРУКТУРОЙ

Постановка задачи

Объект — пространственные и временные данные экономического показателя.

Предмет — регулярности экономического показателя.

Цель — воспроизведение структурной и функциональной неоднородности с помощью фиктивных переменных.

Актуальность — структурная и функциональная неоднородность трудно воспроизводится с помощью обычных переменных.

Рабочая гипотеза — структурная и функциональная неоднородность объясняется факторами, численные значения которых трудно определить.

Метод — включение в модели фиктивных переменных.

Способ — для расчета характеристик модели можно использовать функцию Excel “Линейн”.

Задача — определить численные значения фиктивной переменной, которые объясняли бы структурную и функциональную неоднородность.

Ожидаемый результат — фиктивная переменная должна воспроизвести структурную и функциональную неоднородность.

Метод для сравнения — сравнение результатов без учета и с учетом фиктивной переменной.

Задания первого уровня сложности

Задание 1. Подготовить устные сообщения по вопросам самостоятельной работы.

1. Дайте определение регрессионной модели с переменной структурой.

2. Приведите примеры структурной и функциональной неоднородности.

2. Дайте определение фиктивной переменной.

3. Перечислите ограничения на численные значения нескольких фиктивных переменных.

4. Объясните, почему сумма нескольких фиктивных переменных для каждого измерения не должна равняться постоянному числу? [3, с. 224–226,7, с. 117–118, 115–123 , 1, т. 2 с. 155–171, 8, с. 95–100, 10, с. 285–309].

Задание 2. Пройдите входное тестирование.

Выберите вариант правильного ответа.

1. Регрессионная модель с переменной структурой — это такая модель, которая учитывает неоднородность количественной зависимой переменной с помощью качественного фактора (фиктивной переменной).

а) да;

б) нет.

2. Различают два типа неоднородностей зависимой переменной: функциональную и структурную.

а) да;

б) нет.

3. Функциональной неоднородностью называют такую тенденцию зависимой переменной, которая имеет четко выраженное изменение вида тенденции.

а) да;

б) нет.

4. Структурной неоднородностью называют такую тенденцию зависимой переменной, которая имеет четко выраженные выбросы для временных рядов или расслоения для пространственных данных.

а) да;

б) нет.

5. **Предупреждение.** В модели могут находиться несколько фиктивных переменных, на которые накладывается одно ограничение: сумма значений фиктивных переменных по каждой строчке не должна быть равна одному и тому же числу. Если это произойдет, то нельзя будет определить коэффициенты многофакторной модели.

а) да;

б) нет.

6. Достоинство фиктивной переменной состоит в том, что только с ее помощью можно воспроизвести выброс у зависимой переменной и повысить точность прогноза, если известно значение фиктивной переменной на ожидаемый период.

а) да;

б) нет.

Задание 3. Решите задачу.

Имеются данные временного ряда объема продажи мороженного (табл. 8.1).

Таблица 8.1

Данные временного ряда

t	$У_t$
1	1
2	2
3	12
4	3
5	5
6	4
7	?

где t — порядковый номер дней в июле месяце,

$У_t$ — объем продажи мороженного, произведенной одной торговой точкой (тыс. руб.).

Необходимо ввести в модель новую переменную, которая учитывала бы влияние фиктивной переменной, и произвести эконометрический анализ модели.

Решение

Цель моделирования

Необходимо улучшить точность прогноза за счет учета влияния фиктивной переменной.

Выбор переменных

Величина товарооборота зависит от шести основных факторов: *сырье* и *материалы* (ассортимент товаров и их качество); *машины* (холодильное оборудование торговой точки); *методы* (технология продажи); *персонал* (уровень квалификации, образование, возраст, деловые характеристики); *физическая* и *абстрактная среда* (температура и влажность воздуха, интенсивность людского потока около торговой точки; правовая среда существования и информационное обеспечение торговой точки); *время* (сезонность) (рис. 8.1).

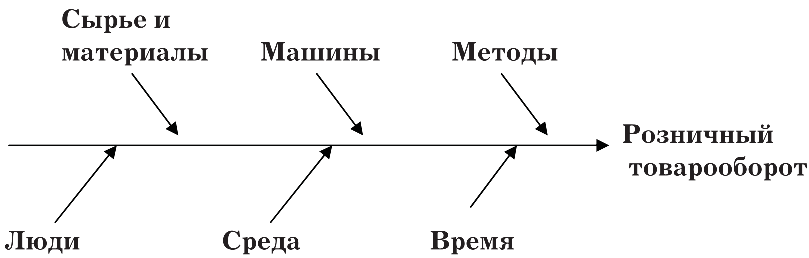


Рис. 8.1. Причинно-следственная диаграмма Исикавы

Выделение входных и выходных переменных

В качестве объясняемых факторов могут быть время и температура воздуха, зависимой переменной является розничный товарооборот мороженного.

Сбор статистических данных

Исходные данные представлены в табл. 8.1. График динамики У изображен на рис. 8.2.

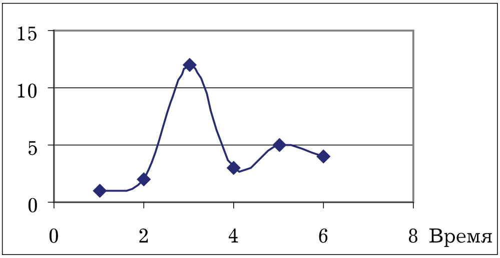


Рис. 8.2. График исходных данных

Анализ рис. 8.2 показывает, что в третий день наблюдался выброс товарооборота мороженного, который может быть вызван изменением одного или одновременно нескольких влияющих факторов с возможным временным лагом. Допустим, в этот день наблюдалась повышенная температура воздуха при отсутствии временного лага.

Можно выразить повышение температуры в виде кода, рассматривая температуру как фиктивную переменную.

Предмодельный анализ

Выдвижение гипотез

Предполагаем, что выброс в реализации мороженного был вызван выбросом температуры.

Формулировка допущений

Предполагаем, что временной лаг отсутствует между выбросами товарооборота мороженного и температурой.

Спецификация

Обобщенная линейная модель множественной регрессии.

Предполагаем, что зависимость У от X можно представить в виде обобщенной линейной модели множественной регрессии. Приводим спецификацию данной модели:

$$\mathbf{У = XA} + \varepsilon \text{ или } У_t = \alpha_0 + \alpha X_{1t} + \alpha_2 X_{2t} + \varepsilon_t$$

при условии, что

$M(\varepsilon) = 0$,

$M(\varepsilon\varepsilon') = \sigma_\varepsilon^2 I$, где **X** — матрица имеет размерность X_{nd},

n — число наблюдений,

d — количество всех объясняемых переменных, включая столбец, состоящий из единиц.

$$\mathrm{rang}(\mathbf{X}) = d < n.$$

Первое условие rang(**X**) = d — ранг матрицы **X** должен равняться d. это означает следующее: между объясняющими переменными не должно быть мультиколлинарности. Следствием из этого условия является утверждение того, что сумма значений фиктивных факторов не должна равняться константе.

Второе условие (d < n) означает следующее: количество переменных должно быть меньше числа измерений.

В множественную модель включены два фактора: первый учитывает время, второй — фиктивная переменная — температура воздуха. В модель можно включать несколько фиктивных переменных при условии, что после их кодирования сумма кодов по каждому измерению не будет одинаковым числом. При кодировании фиктивной переменной можно пользоваться следующими правилами: бо´льшим выбросам зависимой переменной можно дать бо´льшим код фиктивной переменной; при получении прогноза фиктивной переменной можно дать ожидаемое значение, которое может составлять определенную долю от его максимального значения. Допустим, в третий день дадим фиктивной переменной значение 1, в остальные дни — 0. В ожидаемый седьмой день предполагается повышение температуры, составляющей 0,6 от его максимального значения, равного 1.

Идентифицируемость модели

(Можно ли в принципе однозначно восстановить значения неизвестных параметров по выбранной модели).

Если сумма кодов фиктивных переменных для каждого измерения не будет равна постоянному числу, то можно определить коэффициенты модели.

Идентификация модели

(Предложить и реализовать процедуру оценивания параметров модели).

Построим множественную модель с фиктивной переменной в следующей последовательности.

Шаг 1.

Составим базу данных с двумя объясняемыми переменными (табл. 8.2).

Таблица 8.2

Данные временных рядов

t	X_{1t}	X_{2t}	$У_t$	$У_{pt}$
1	1	0	1	1,186
2	2	0	2	1,8837
3	3	1	12	12
4	4	0	3	3,2791
5	5	0	5	3,9767
6	6	0	4	4,6744
Ожидаем	7	0,6	?	11,023

где t — порядковый номер времени в июле месяце (дни);

X_{1t} — фактор времени;

X_{2t} — фиктивная переменная, учитывающая влияние температуры воздуха;

$У_t$ — объем продажи мороженного, произведенной одной торговой точкой (тыс.руб.);

$У_{pt}$ — расчетные значения У.

Шаг 2.

Вычислим коэффициенты многофакторной модели

$$У_{pt} = a_0 + a_1X_{1t} + a_2X_{2t}$$

методом наименьших квадратов (МНК) с помощью функции “Линейн”.

9,419	0,698	0,488
0,814	0,178	0,719
0,979	0,737	#Н/Д
69,911	3,000	#Н/Д
75,872	1,628	#Н/Д

где $a_o =$ 0,488

$a_1 =$ 0,698

$a_2 =$ 9,419

Шаг 3.

Определим расчетные значения зависимой переменной $У_{pt}$ по формуле

$$У_{pt} = a_0 + a_1X_{1t} + a_2X_{2t}$$

и занесем их в табл. 8.2.

Шаг 4.

Построим график фактических и расчетных значений У (рис. 8.3).

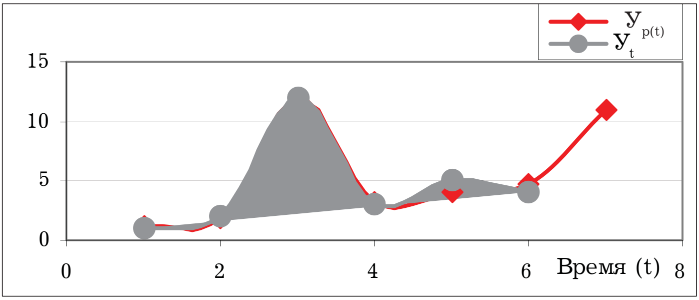


Рис. 8.3. Зависимость У, $У_p$ от времени t

Анализ рис. 8.3 показывает, что наблюдаемый выброс товарооборота мороженного хорошо воспроизводится с помощью фиктивной переменной, учитывающей температуру воздуха. Прогноз на седьмой день учитывает ожидаемое значение фиктивной переменной.

Основной проблемой при вводе фиктивной переменной является экономическое обоснование фиктивной переменной и ее кодовых значений.

Задание второго уровня сложности

Задание. Измените в табл. 8.2 численные значения У так, чтобы в них появилось два выброса, подберите такие значения фиктивной переменной X_2 в табл. 8.2, чтобы ошибка модели стала минимальной.

Пояснение. При подборе значений фиктивной переменной надо вначале всем значениям дать значения 0, затем проранжировать по значению все выбросы У и последовательно этим измерениям присваивать убывающие значения фиктивной переменной, минимизируя ошибку модели.

Задание третьего уровня сложности

Задание. С помощью фиктивной переменной воспроизвести функциональную неоднородность.

Пояснение. Динамика повышения производительности труда за предыдущий и отчетный год по каждому месяцу имеет такую закономерность: в предыдущем году производительность труда незначительно изменялась около среднего значения; в отчетном году было принято решение о внедрении системы качества на предприятии. Это привело к тому, что оживилась работа по устранению несоответствий и повышению производительности труда.

Исходные данные и решения представлены в табл. 8.3, 8.4, рис. 8.4, 8.5 при различных значениях фиктивной переменной.

Таблица 8.3

Исходные данные У и расчетные значения $У_р$, при значениях фактора времени t и постоянной фиктивной переменной $X_ф$

t	$X_{фt}$	$У_t$	$У_{pt}$	t	$X_{фt}$	$У_t$	$У_{pt}$
1	1	12	3,9933	13	1	16	24,435
2	1	13	5,6968	14	1	18	26,139
3	1	14	7,4003	15	1	20	27,842
4	1	12	9,1038	16	1	25	29,546
5	1	15	10,807	17	1	30	31,249
6	1	16	12,511	18	1	32	32,952
7	1	12	14,214	19	1	35	34,656
8	1	13	15,918	20	1	41	36,359
9	1	13	17,621	21	1	42	38,063
10	1	12	19,325	22	1	45	39,766
11	1	15	21,028	23	1	50	41,47
12	1	14	22,732	24	1	51	43,173

Протокол расчетов по функции “Линейн”.

0	1,7035	2,2899
0	0,1801	2,5735
0,8026	6,1078	#Н/Д
89,454	22	#Н/Д
3337,1	820,72	#Н/Д

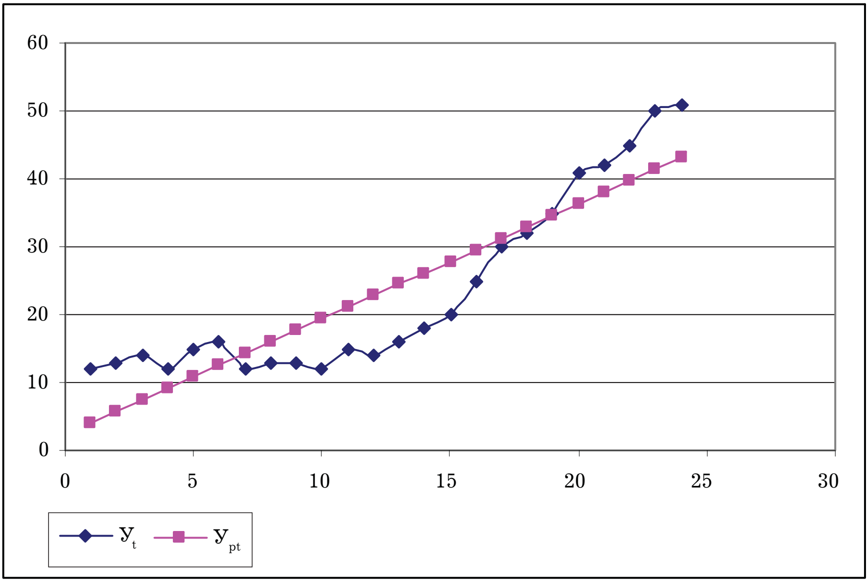


Рис. 8.4. Фактические и расчетные значения У при отсутствии влияния фиктивной переменной

Вывод. В табл. 8.3 фиктивная переменная является константой, поэтому она не может влиять на У. Расчетные значения $У_р$ не отражают функциональную неоднородность. Ошибка модели равна 6,1. По критерию Фишера модель является достоверной.

Таблица 8.4

Исходные данные У и расчетные значения $У_р$, при значениях фактора времени t и изменяющейся фиктивной переменной $Х_ф$

t	$Х_{фt}$	$У_t$	$У_{pt}$
1	1	12	13,411
2	1	13	13,365
3	1	14	13,319
4	1	12	13,273
5	1	15	13,227
6	1	16	13,181
7	1	12	13,135
8	1	13	13,089
9	1	13	13,043
10	1	12	12,996
11	1	15	12,95
12	1	14	12,904

t	$Х_{фt}$	$У_t$	$У_{pt}$
13	2	16	16,151
14	3	18	19,398
15	4	20	22,645
16	5	25	25,892
17	6	30	29,139
18	7	32	32,386
19	8	35	35,632
20	9	41	38,879
21	10	42	42,126
22	11	45	45,373
23	12	50	48,62
24	13	51	51,867

Протокол расчетов по функции “Линейн”.

3,293	-0,046	10,164
0,1655	0,0972	0,7116
0,9901	1,4034	#Н/Д
1045	21	#Н/Д
4116,5	41,36	#Н/Д

Вывод. В табл. 8.4 фиктивная переменная изменяется в соответствии с нарастающей работой по устранению несоответствий в отчетном году. Расчетные значения $У_р$ хорошо отражают функциональную неоднородность. Ошибка модели уменьшилась и стала равной 1.4. Критерий Фишера значительно возрос. По критерию Фишера модель является достоверной.

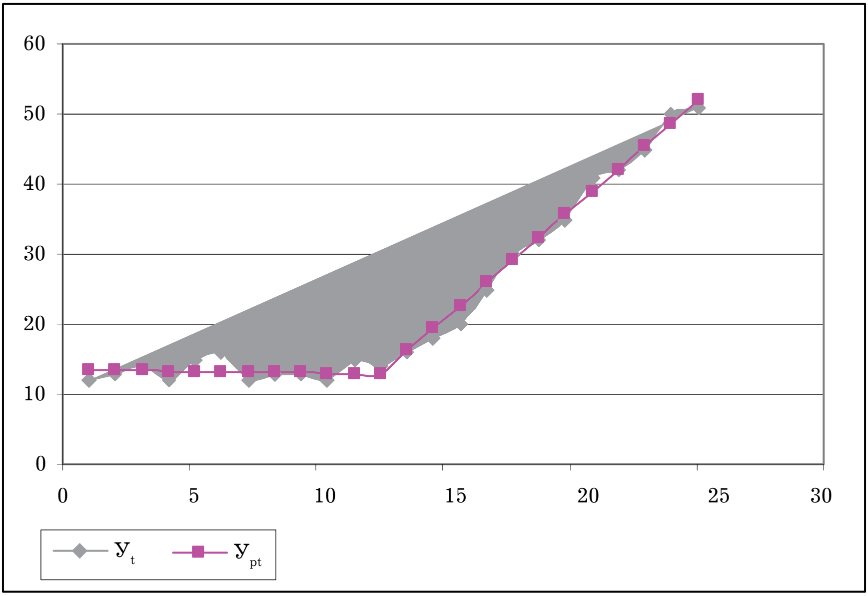


Рис. 8.5. Фактические и расчетные значения У при учете влияния фиктивной переменной

Задание четвертого уровня сложности

Задание. Повторите расчеты задач с помощью пакета прикладных программ Stadia 6.

Выходной тест

Укажите функциональную и структурную неоднородность, изображенных на рис. 8.6 и 8.7.

Нерешенная проблема

Существует сложность экономического обоснования значений фиктивной переменной.

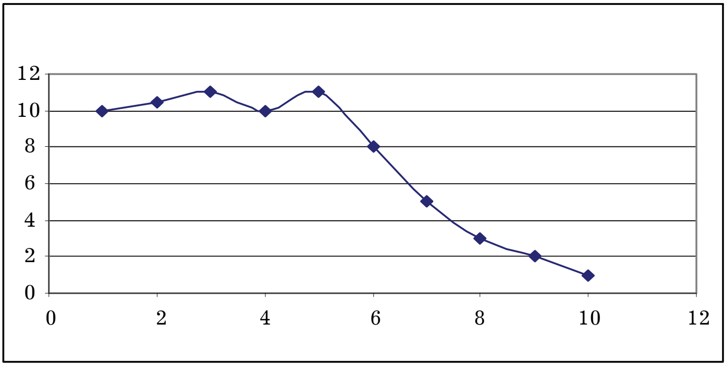


Рис. 8.6. Зависимость У от Х

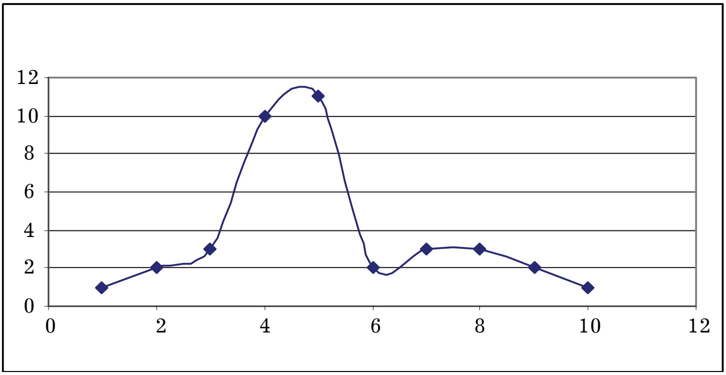


Рис. 8.7. Зависимость У от Х

9. ВРЕМЕННЫЕ РЯДЫ. ХАРАКТЕРИСТИКИ ВРЕМЕННЫХ РЯДОВ

Постановка задачи

Объект — временные данные экономического показателя.

Предмет — регулярности экономического показателя.

Цель — изучить характеристики временных рядов для повышения точности прогноза.

Актуальность — повышение точности прогноза временного ряда всегда высоко ценилось в эконометрическом моделировании.

Рабочая гипотеза — временной ряд имеет определенную структуру.

Метод — для оценки модели временного ряда можно использовать метод наименьших квадратов.

Способ — для расчета характеристик модели можно использовать функцию Excel "Линейн".

Задача — определить характеристики временных рядов.

Ожидаемый результат — характеристики временных рядов должны указать на наличие регулярностей во временном ряду.

Метод для сравнения — сравнение результатов расчетов без наличия и с наличием регулярностей во временном ряду.

Задания первого уровня сложности

Задание. Подготовить устные сообщения по вопросам самостоятельной работы.

1. Дайте определение временного ряда.

2. Почему количество наблюдений временного ряда называют числом уровней, а не объемом выборки временного ряда?

3. Какими свойствами обладают экономические временные ряды?

4. Какие можно выделить составляющие временного ряда?

5. Назовите основные цели анализа временных рядов.

6. Какие имеются характеристики временного ряда?

7. Дайте определение стационарного и нестационарного временного ряда [3, с. 244–264; 7, с. 133–135; 1, т. 2, с. 199–220].

Задание 2. Пройдите входное тестирование

Выберите вариант правильного ответа.

1. 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, называемым уровнем ряда.

а) да;

б) нет.

2. 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 экономи-ческих процессов в течение года;

f_{3t} _ циклическая компонента, которая отражает повторяемость экономи-ческих процессов в течение длительных периодов, например, влияние волн научно-технической революции, демографических спадов, циклов солнечной активности и т. д.;

ε_t — случайное возмущение.

а) да;

б) нет.

3. 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.

а) да;

б) нет.

4. Основные свойства экономического временного ряда

1. Текущее состояние экономической системы испытывает влияние 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.

а) да;

б) нет.

6 Возможные реализации экономического процесса в фиксированный момент времени t имеют определенный закон распределения вероятностей и соответствующие статистические характеристики: среднее арифметическое значение, дисперсию, ковариацию 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.

а) да;

б) нет.

7 Во временных рядах не должно быть пропусков. Если имеется пропуск, то он восстанавливается средним значением, обычно из ближайших к нему чисел.

а) да;

б) нет.

8. Целями анализа временных рядов являются:

— выявление структуры временного ряда для более глубокого понимания динамических экономических процессов;
— получение точечного и интервального прогноза динамического экономического показателя.

а) да;

б) нет.

9. Статистические характеристики временного ряда.

Временные ряды могут иметь следующие статистические характеристики:

— среднее арифметическое значение;
— дисперсия;
— автоковариация;
— автокорреляционная функция (автокоррелограмма);
— периодограмма.

а) да;

б) нет.

10. Среднее арифметическое значение временного ряда $У_t$ вычисляется по следующей формуле:

$$S^2 = \frac{\sum\left(У_t - \overline{У}\right)^2}{n-1};$$

а) да;

б) нет.

$$Cov(У_t, У_{t+k}) = \frac{1}{n-k}\sum_{t=1}^{n-k}(У_t - \overline{У}_1)(У_{t+k} - \overline{У}_2);$$

11. Дисперсия временного ряда $У_t$ вычисляется по формуле

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

а) да;

б) нет.

12. Автоковариация k-го порядка временного ряда $У_t$ вычисляется по формуле

$$a = \frac{2}{n}\sum_{t=1}^{n} У_t Cos(2\pi t/T);$$

а) да;

б) нет.

13. 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 ряда на их последующие значения с временным сдвигом, равным k датам.

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

а) да;
б) нет.

14. 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:

$$У_t = a\mathrm{Cos}(2\pi t/T) + b\mathrm{Sin}(2\pi t/T) + e = c\mathrm{Cos}(2\pi t/T+\varphi) + e,$$

где T — некоторая фиксированная величина, обычно называемая периодом, или длиной волны;

φ — начальная фаза колебания;

a, b, c — амплитуды периодических колебаний;

$$a = \frac{2}{n}\sum_{t=1}^{n} У_t \mathrm{Cos}(2\pi t/T);$$

$$b = \frac{2}{n}\sum_{t=1}^{n} У_t \mathrm{Sin}(2\pi t/T);$$

$$c = \sqrt{a^2 + b^2};$$

$$c^2(T) = a^2(T) + b^2(T).$$

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[1].

а) да;
б) нет.

15. В практических целях можно упростить расчет периодограммы, если построить зависимость ошибки модели

$$У_t = a_0 + a_1 t + a_2\mathrm{Sin}(2\pi t/T) + a_3\mathrm{Cos}(2\pi t/T) + e_t$$

[1] Тюрин Ю. Н., Макаров А. А. Статистический анализ данных на компьютере. — М.: ИНФА — М,1998, с. 374.

от периода Т, где Т принимает значения от 3 до удвоенной длины временного ряда. Коэффициенты модели рассчитываются методом наименьших квадратов.

а) да;

б) нет.

Задание 2. Решите задачу.

Имеются данные затрат на устранение брака в сборочном цехе, вызванные ошибками в чертежах, составленных конструкторским отделом завода (табл. 9.1).

Таблица 9.1

База данных затрат на устранение брака в интервале 10 рабочих дней.

t	$У_t$	t	$У_t$
1	12	6	15
2	15	7	12
3	16	8	16
4	12	9	14
5	13	10	15

где t — время (дни);

У — расходы на устранение брака (тыс. руб.).

Необходимо определить основные характеристики временного ряда:

— среднее значение,

— критерий восходящих и нисходящих серий,

— дисперсию значений $У_t$,

— парную линейную регрессию,

— автокоррелограмму остатков,

— критерий Дарбина-Уотсона,

— периодограмму.

Решение

Свойства временного ряда нужны для построения моделей, которые их воспроизводят.

Выбор переменных

Расходы на устранение брака зависят от шести основных факторов: *сырье* и *материалы* (количество ошибок в заготовках, вызванных ошибками операторов и в чертежах, допущенных конструкторами); *машины* (технические средства, используемые при разработке и деталировке изделий); *методы* (технология разработки чертежей, программные средства разработки чертежей); *люди* (уровень квалификации конструкторов, образование, возраст, деловые характеристики); *физическая* и *абстрактная* среда (температура, влажность, освещенность в конструкторском отделе; традиции, психологический климат существующие в конструкторском отделе; правовая среда существования конструкторского отдела, информационное обеспечение конструкторского отдела); *время* (сезонность, день недели).

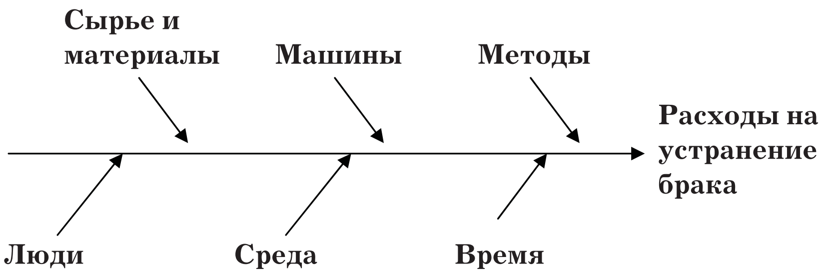


Рис. 9.1. Причинно-следственная диаграмма Исикавы

Выделение входных и выходных переменных

Анализ рис. 9.1 показывает, что зависимой переменной являются расходы на устранение брака, объясняемой переменной служит время.

Сбор статистических данных

Была проведена группировка затрат на устранение брака, вызванного ошибками конструкторского отдела в интервале 10 дней. Данные представлены в табл. 9.1, и на рис. 9.2

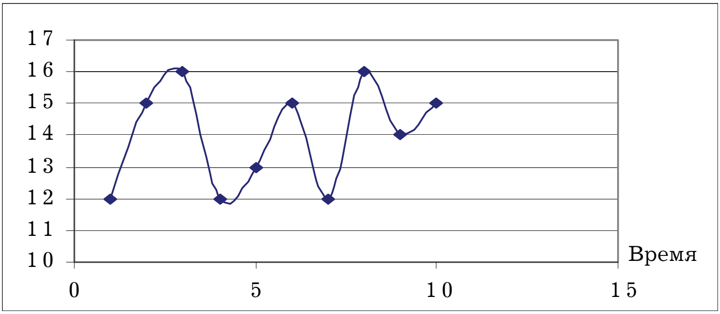


Рис. 9.2. Динамика затрат на устранение брака

Предмодельный анализ

Выдвижение гипотез

Анализ рис. 9.2 показывает, что временной ряд: не содержит четко выраженной монотонной тенденции; имеется слабо выраженная циклическая составляющая, равная трем дням.

Предполагаем, что временной ряд является стационарным.

Оценками характеристик процесса, который генерирует значения временного ряда, могут быть характеристики наблюдаемого временного ряда.

Наблюдаемые значения временного ряда могут иметь следующие характеристики:

— среднее значение;
— критерий восходящих и нисходящих серий;
— дисперсия значений $У_t$;
— парная линейная регрессия;
— автокорреляграмма остатков;
— критерий Дарбина-Уотсона;
— периодограмма.

Среднее значение временного ряда вычисляется по формуле

$$\overline{У} = \frac{\sum_{t=1}^{n} У_t}{n}.$$

Критерий восходящих и нисходящих серий служит для проверки наличия во временном ряду $У_t$ неслучайной составляющей или проверки гипотезы о том, что H_o: "$МУ_t = const$".

Если нарушится хотя бы одно из двух неравенств:

$$\begin{cases} К_{об}(n) > \left[\frac{1}{3}(2n-1) - 1{,}96\sqrt{\frac{16n-29}{90}}\right] \\ К_{дл}(n) < К_o(n), \end{cases}$$

то нулевая гипотеза отвергается на уровне значимости α от 0,05 до 0,097 и утверждается, что временной ряд $У_t$ содержит неслучайную составляющую, зависящую от времени t,

где $Кo(n<=26) = 5$;

$К_o(26 < n <=153) = 6$;

$К_o(153 < n <= 1170) = 7$;

$К_{об}(n)$ — общее число серий подряд идущих положительных и отрицательных первых разностей временного ряда $У_t$;

$К_{дл}(n)$ — количество подряд идущих положительных или отрицательных значений первых разностей временного ряда $У_t$ в самой длинной серии;

n — число уровней временного ряда $У_t$.

Таблица 9.2

Вычисление критерия восходящих и нисходящих серий

t	$У_t$	$У_{t+1}$	$У_{t+1} — У_t$	Наличие серии	Длина серии
1	12	15	3	1	2
2	15	16	1		
3	16	12	-4	1	1
4	12	13	1	1	2
5	13	15	2		
6	15	12	-3	1	1
7	12	16	4	1	1
8	16	14	-2	1	1
9	14	15	1	1	1
10	15				
$К_{об}(10)$				7	

Так как условия:

$К_{об}(10) = 7 > 3,9687$,

$К_{дл}(10) = 2 < К_о(n= 10<=26) = 5$,

$n = 10$

выполняются, то на уровне значимости α от 0,05 до 0,097 можно утверждать, что временной ряд содержит неслучайную составляющую.

Дисперсия временного ряда вычисляется по следующей формуле:

$$S^2 = \frac{\sum_{t=1}^{n}(У_t - \overline{У})^2}{n-1} = 2,6667.$$

Парная линейная регрессия. Коэффициенты и характеристики парной линейной регрессии $У = a_0 + a_1 t + e$, вычислим с помощью функции "Линейн".

0,1333	13,267
0,1848	1,1465
0,0611	1,6783
0,5207	8
1,4667	22,533

a_0=	13,267
a_1 =	0,1333

Остатки вычисляются (табл. 9.3) по формуле

$e = У — (a_0+a_1t)$.

Таблица 9.3

Расчет остатков линейной модели

t	$У_t$	e_t	e_{t-1}	e_{t-2}	e_{t-3}	e_{t-4}
1	12	-1,4				
2	15	1,4667	-1,4			
3	16	2,3333	1,4667	-1,4		
4	12	-1,8	2,3333	1,4667	-1,4	
5	13	-0,9333	-1,8	2,3333	1,4667	-1,4
6	15	0,9333	-0,933	-1,8	2,3333	1,46667
7	12	-2,2	0,9333	-0,933	-1,8	2,33333
8	16	1,6667	-2,2	0,9333	-0,933	-1,8
9	14	-0,4667	1,6667	-2,2	0,9333	-0,9333
10	15	0,4	-0,467	1,6667	-2,2	0,93333

Задание второго уровня сложности

Задание. Рассчитать автокоррелограмму остатков и критерий Дарбина-Уотсона.

Автокорреляграмма остатков — график зависимости автокорреляции (порядка d) остатков модели от величины d, где d = 0, 1, 2, ..., n/2.

Таблица 9.4

Значения коэффициентов автокорреляции разного порядка

Значения автокорреляций		Порядок автокорреляции
$r(e_t,e_t) = r_0 =$	1	0 1 2 3 4
$r(e_t,e_{t-1}) = r_1 =$	-0,405	
$r(e_t,e_{t-2}) = r_2 =$	-0,221	
$r(e_t,e_{t-3}) = r_3 =$	0,2545	
$r(e_t,e_{t-4}) = r_4 =$	-0,388	

Рис. 9.3. Автокоррелограмма. Зависимость коэффициентов автокорреляции остатков от их порядков

Как правило, автокоррелограмма повторяет тенденцию временного ряда.

Анализ рис. 9.3 автокоррелограммы показывает, что во временном ряду имеется периодическая составляющая.

Критерий Дарбина-Уотсона (DW) служит для проверки достоверности коэффициента автокорреляции остатков первого порядка.

Свойства критерия DW:

— если $e_t = e_{t-1}$, то $r_1 = 1$, DW = 0;

— если $e_t = — e_{t-1}$, то $r_1 = — 1$, DW = 4;

— если $e_t \neq e_{t-1}$, или $e_t \neq — e_{t-1}$, $r_1 = 0$, то DW = 2;

— во всех остальных случаях 0 < DW < 4.

Ограничения использования критерия DW.

- Критерий DW применяется только для тех моделей, которые содержат свободный коэффициент.
- Предполагается, что случайные отклонения e определяются по итерационной схеме: $e_t = pe_{t-1} + u_t$.
- Статистические данные не должны содержать пропусков в наблюдениях.
- Критерий DW не применим, если в качестве объясняемой переменной стоит лаговая зависимая переменная первого порядка, так как при этом лаговая объясняемая переменная и остатки будут связаны между собой [10, с. 258].

Определим критерий DW по формуле

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \approx 2(1 - r_1)$$

[7, с. 170−171].

Так как $r_1 = — 0{,}4052$, то DW = 2 (I — (−0,4052)) = 2,8105.

Из табл. 9.5 определим $d_н$ и $d_в$ при n = 10, k = 1

Таблица 9.5

Численные значения нижних $d_н$ и верхних $d_в$ границ критерия Дарбина-Уотсона на уровне значимости α = 0,05

n	k = 1		k = 2	
	$d_н$	$d_в$	$d_н$	$d_в$
6	0,61	1,4		
7	0,7	1,36	0,47	1,9

n	k = 1		k = 2	
	$d_н$	$d_в$	$d_н$	$d_в$
8	0,76	1,33	0,56	1,78
9	0,82	1,32	0,63	1,7
10	0,88	1,32	0,7	1,64
11	0,93	1,32	0,76	1,6
12	0,97	1,33	0,81	1,58
13	1,01	1,34	0,86	1,56
14	1,05	1,35	0,91	1,55
15	1,08	1,36	0,95	1,54
16	1,11	1,37	0,98	1,54
17	1,13	1,38	1,02	1,54
18	1,16	1,39	1,05	1,54
19	1,18	1,4	1,07	1,54
20	1,2	1,41	1,1	1,54
21	1,22	1,42	1,13	1,54
22	1,24	1,43	1,15	1,54
23	1,26	1,44	1,17	1,54
24	1,27	1,45	1,19	1,55
25	1,29	1,45	1,21	1,55
30	1,35	1,49	1,28	1,57
50	1,5	1,59	1,46	1,63
100	1,65	1,69	1,63	1,72
200	1,76	1,78	1,75	1,79

[10, с. 388).

где n — длина ряда наблюдений;

k — количество объясняющих переменных в уравнении регрессии. Например, для модели

$У_t = a_0 + a_1X_t + e_t$, k = 1.

$d_н$ (α = 0,05; n =10, k = 1) = 0,88,

$d_в$ (α = 0,05; n = 10, k = 1) = 1,32,

Таблица 9.6

Числовая ось значений критерия DW

min =	0	Минимальное значение
dн =	0,88	Первое нижнее граничное значение
dв =	1,32	Первое верхнее граничное значение
Середина =	2	Среднее значение
4 dв =	2,68	Второе нижнее граничное значение
4 dн =	3,12	Второе верхнее граничное значение
max =	4	Максимальное значение
DW =	2,81	Значение критерия DW

Зоны положительного, отрицательного и нулевого значения коэффициента автокорреляции первого порядка, а также зона неопределенности (?) показаны на рис. 9.4.

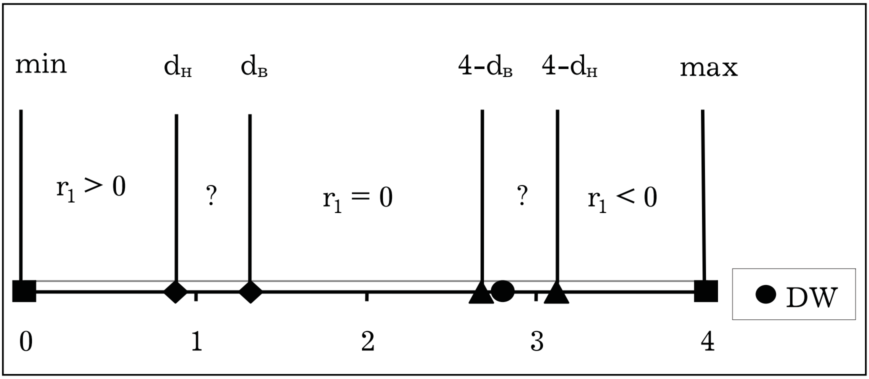


Рис. 9.4. Числовая ось возможных значений критерия DW

Анализ рис. 9.4 показывает, что критерий DW находится в области неопределенности. Это означает, что статистически, с определенной вероятностью мы не можем утверждать о достоверном значении коэффициента автокорреляции первого порядка

Задание третьего уровня сложности

Задание. Рассчитать периодограмму.

Периодограмма вычисляется в следующей последовательности:

— методом наименьших квадратов вычисляются коэффициенты модели:

$$У_t = a_0 + a_1 t + a_2 Sin(2\pi t/T) + a_3 Cos(2\pi t/T) + e_t,\ T = 3, ..., 2n.$$

— для каждого значения периода T вычисляется ошибка модели E;

— характеристики модели вычисляются по функции “Линейн”;

— строится график зависимости E от T, который называется *периодограммой*. Анализируется график периодограммы и определяются циклические составляющие с периодами T, при которых наблюдаются локальные минимумы значений E.

В дальнейшем эти циклические составляющие с соответствующими периодами используются для построения множественной модели.

Таблица 9.7

Расчет периодограммы с протоколом расчетов по функции “Линейн” (для T = 2)

$У_t$	t	Sin(2πt/T)	Cos(2πt/T)	T	E	-1,91E-05	-62,79	0,10
12	1	0,0016	-1	2	1,77	1,22	124,25	0,19
15	2	-0,0032	1	3	1,41	0,21	1,77	#Н/Д
16	3	0,0048	-1	4	1,78	0,55	6	#Н/Д
12	4	-0,0064	1	5	1,77	5,19	18,80	#Н/Д
13	5	0,008	-1	6	1,73			
15	6	-0,0096	1	7	1,78			
12	7	0,0111	-1	8	1,82			
16	8	-0,0127	0,9999	9	1,84			
14	9	0,0143	-1	10	1,85			
15	10	-0,0159	0,9999	11	1,859			
T =	2							

Изменяйте T и определяйте ошибку модели E.

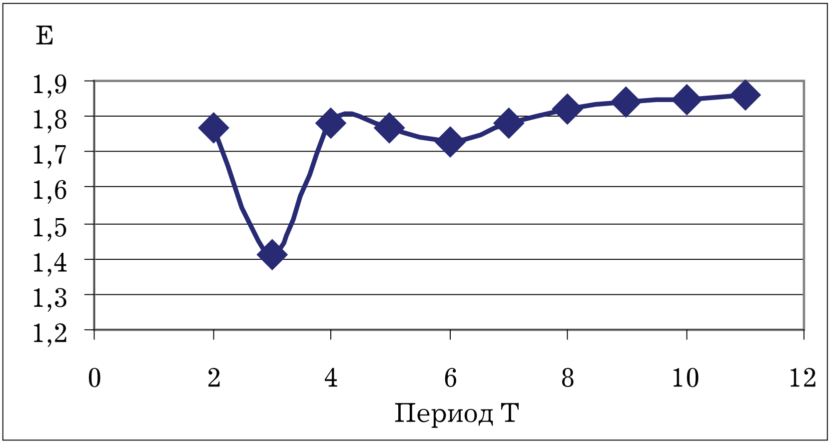


Рис. 9.5. Периодограмма. Зависимость ошибки модели Е от периода Т

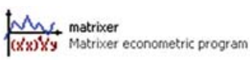
Анализ периодограммы на рис. 9.5 показывает наличие двух периодов 3 и 6 дней, при которых ошибка модели является минимальной. Следовательно, во временном ряду имеются две периодические составляющие с периодами 3 и 6 дней. Однако период в 6 дней является эхом периода в 3 дня (так как период в 6 дней кратен периоду в 3 дня) и, как правило, эхо в модель не включается.

Задание четвертого уровня сложности

Задание. Вычислите основные характеристики временного ряда с помощью любого пакета прикладных программ. Например, с помощью программы Matrixer.

Пояснения.

1. Для запуска программы Matrixer дважды щелкните по изображению файла



На экране появится главное меню программы (рис. 9.6).

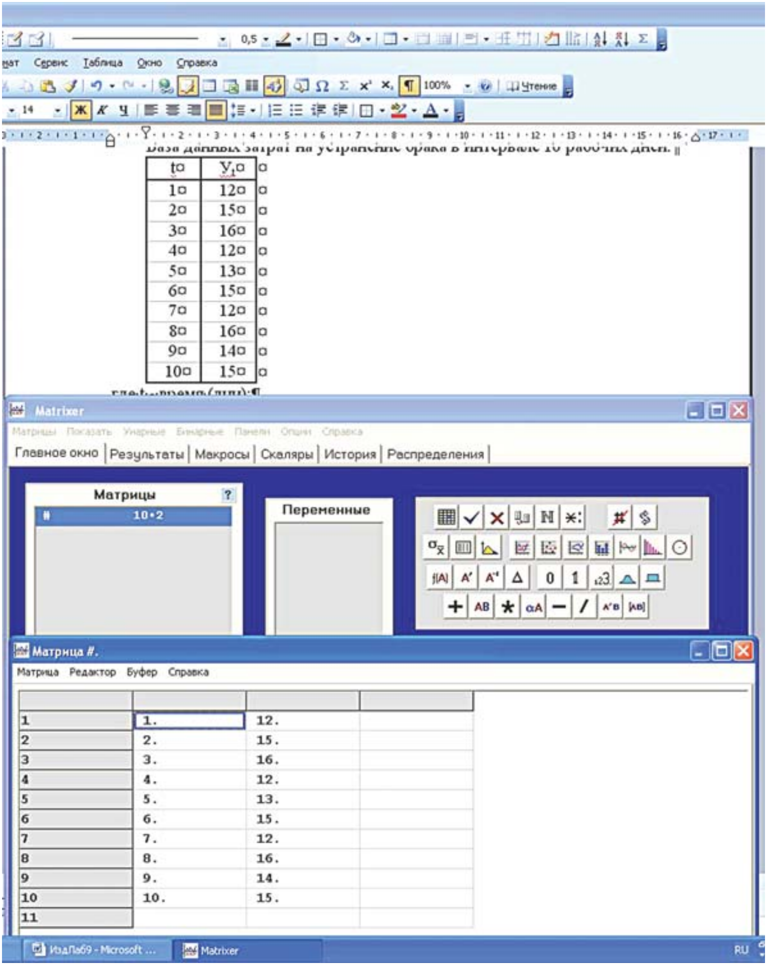


Рис. 9.6. Главное меню и содержимое исходной матрицы программы Matrixer

2. Активизируйте матрицу и введите значения переменных (см. рис. 9.6).

3. Выделите переменную 2 и вычислите для нее описательные статистики ()

Результаты расчетов представлены на рис. 9.7.

```
Переменная #[2], 10 набл.
Минимум 12.
Максимум 16.
Среднее 14. [0.0000]
Медиана 14.5
Дисперсия  2.4 (смещенная оценка)
Дисперсия  2.6666666667 (несмещенная оценка)
Среднеквадратическое  отклонение  1.5491933385 (смещенная оценка)
Среднеквадратическое  отклонение  1.6329931619 (несмещенная оценка)
Асимметрия -0.1613743061 [0.8350]
Эксцесс  -1.5416666667 [0.3197]
Коэффициент вариации 0.1166423687
Сумма 140.
Сумма квадратов  отклонений  от среднего  24.
Сумма квадратов  1984.
Автокорреляция  1-го  порядка  -0,375 [0,2357]
```

Рис. 9.7. Протокол расчетов

Выходное тестирование

Имеется периодограмма временного ряда (рис. 9.8). Укажите периоды периодических составляющих временного ряда.

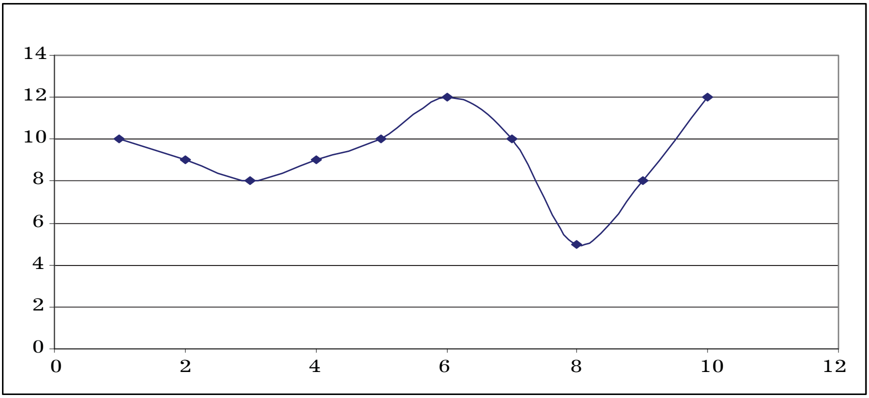


Рис. 9.8. Периодограмма временного ряда

Нерешенная проблема

Нет критерия проверки достоверности автокорреляции остатков больше первого порядка.

10. ЛИНЕЙНЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ С АВТОКОРРЕЛИРОВАННЫМИ ОСТАТКАМИ

Постановка задачи

Объект — временные данные экономического показателя.

Предмет — регулярности экономического показателя.

Цель — изучить свойства характеристик временного ряда, предназначенных для повышения точности прогноза.

Актуальность — повышение точности прогноза временного ряда всегда высоко ценилось в эконометрическом моделировании.

Рабочая гипотеза — временной ряд имеет определенную структуру.

Метод — для устранения автокорреляции остатков используются такие методы:

— наименьших квадратов;
— Эйткена,
— Дарбина;
— Кочрена-Оркатто;
— Хилдрета-Лу.

Способ — для расчета характеристик модели можно использовать функцию Excel "Линейн".

Задача — использовать автокорреляцию остатков для повышения точности прогнозов.

Ожидаемый результат — использование автокорреляции остатков должно повысить точность прогнозирования.

Метод для сравнения — прогнозы, полученные без помощи и с помощью автокорреляции остатков.

Задания первого уровня сложности

Задание 1. Подготовьте устные сообщения по вопросам самостоятельной работы.

1. Дайте определение автокорреляции остатков.

2. Перечислите возможные причины возникновения автокорреляции остатков.

3. Перечислите последствия наличия в модели автокорреляции остатков,

4. Укажите основные направления устранения автокорреляции остатков.

5. Опишите критерии обнаружения автокорреляции остатков.

6. Опишите суть методов устранения автокорреляции остатков:

— метод наименьших квадратов;

— метод Эйткена,

— метод Дарбина;

— метод Кочрена-Оркатто;

— метод Хилдрета-Лу.

7. Опишите сущность прогнозирования с учетом автокорреляции остатков [3, с. 277–286; 7, с. 150–155, 167–190; 1, т. 2, с. 247–261; 10, с. 250–264; 4, 2001 с. 272–282; 8, с.176–177].

Задание 2. Пройдите входное тестирование.

Выберите вариант правильного ответа.

1. ***Временной ряд*** (динамический ряд или ряд динамики) — последовательность наблюдений некоторого признака (случайной величины) У в последовательные моменты времени.

Отдельные наблюдения называются уровнями ряда, которые будем обозначать $У_t$ ($t = 1, 2, \dots, n$), где n — число уровней (размер ряда, или длина ряда, но не объем выборки). Различают три вида данных: пространственные, временные и пространственно-временные.

а) да;

б) нет.

2. ***Пространственные данные*** характеризуются тем, что за определенный интервал времени (месяц, год) собираются данные показателей по нескольким объектам.

а) да;

б) нет.

3. *Временные данные* характеризуются тем, что собираются данные показателей по одному объекту в последовательные интервалы времени.

а) да;

б) нет.

4. ***Пространственно*** *— временные данные* характеризуются тем, что собираются данные показателей нескольких объектов в последовательные интервалы времени.

а) да;

б) нет.

5. ***Аксиома первая***-численное значение показателя экономического объекта в момент времени t является одним из возможных его значений, который получился под воздействием влияния своих же значений, а также всех других факторов, которые были в прошлом, настоящем и будут в будущем.

а) да;

б) нет.

6. ***Аксиома вторая*** — "свежие" значения фактора оказывают более сильное влияние на текущее значение показателя временного ряда, чем "поздние".

а) да;

б) нет.

7. ***Аксиома третья*** — текущие показатели временного ряда не могут оказывать влияния на их прошлые значения (время не может двигаться в обратном направлении).

На основании этих аксиом строятся все эконометрические модели временных рядов. Одними из основных характеристик временного ряда являются: автокорреляция временного ряда и автокорреляция остатков.

а) да;

б) нет.

8. ***Автокорреляция*** = авто (греч. autos — сам)+ корреляция: корреляция с самим собой.

а) да;

б) нет.

9. ***Автокорреляция*** временного ряда порядка d равна коэффициенту корреляции, рассчитанному между исходным временным рядом и его лаговой переменной порядка d.

а) да;

б) нет.

10. ***Лаговая переменная*** порядка d получается от исходного временного ряда методом перемещения его по времени вперед на d дат.

а) да;

б) нет.

11. ***Автокорреляция остатков***

Предположим, имеются выборочные временные данные, которые можно воспроизвести с помощью выборочной регрессионной моделью

$У_t = У_{pt} + e_t = a_0 + a_1 t + e_t$,

где t — порядковый номер времени,

$e_t = У_t - (a_0 + a_1 t)$ — остатки модели.

Автокорреляция остатков порядка m равна коэффициенту корреляции, рассчитанному между исходными остатками и их лаговой переменной порядка m.

а) да;

б) нет.

12. Коэффициент автокорреляции остатков первого порядка обладает тем свойством, что может принимать значения от -1 до +1.

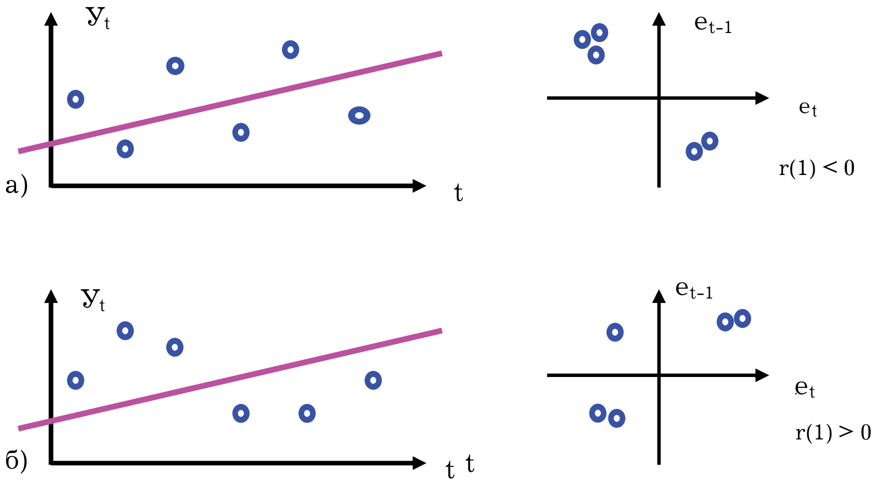


Рис. 10.1. Примеры отрицательного и положительного коэффициентов автокорреляции остатков

Анализ рис.10.1 показывает, что если остатки часто меняют свой знак, то коэффициент автокорреляции первого порядка будет отрицательным. Если имеются серии положительных и отрицательных значений остатков, то он будет положительным.

а) да;

б) нет.

13. *Методы оценки параметров модели* с автокоррелированными остатками.

Эконометристами были предложены несколько методов определения коэффициентов модели

$У_t = a_0 + a_1X_t + pe_{t-1} + v_t$.

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

$$e_t = У_t — (a_0 + a_1X_t)$$

Определяется коэффициент р модели

$$e_t = pe_{t-1} + v_t.$$

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

$$У_t = a_0(1\text{-}p) + pУ_{t-1} + a_1X_t — a_1pX_{t-1} + v_t =$$
$$= b_0 + pУ_{t-1} + a_1X_t — a_1pX_{t-1} + v_t.$$

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

(на основании того, что $B_0 = a_0$ (1-р). $a_0 = b_0/(1\text{-}p)$,

Метод Дарбина позволяет одновременно определить коэффициенты a_0, a_1, р.

Метод Эйткена (в форме преобразований данных) вычисления коэффициентов модели

$$У_t = a_0 + a_1X_t + pe_{t-1} + v_t,$$

где $pe_{t-1} + v_t = e_t$ 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, р при условии

$\mathbf{v'v} \to \min$.

Метод Эйткена выполняется в такой последовательности.

Шаг 1. Вводится предполагаемый коэффициент р.

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]

$$У_{нt} = \sqrt{1-p^2}\,У_t\,;$$

$$X_{н_1t} = \sqrt{1-p^2}\,X_t\,;$$

$$X_{н_2t} = \sqrt{1-p^2}\,X_t\,;$$

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

$$У_t = a_0 + a_1X_t + pe_{t-1} + v_t.$$

Необходимо рассчитать коэффициенты a_0, a_1, р при условии

$\mathbf{v'v} \to \min.$

Метод Дарбина выполняется в такой последовательности.

Шаг 1. Вычисляется коэффициент р, 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

$$У_{нt} = d_0X_{н1t} + d_1X_{н2t} + z_t,$$

---

[1] Доугерти К. Введение в эконометрику. — М.: НИФРА-М, 1997, с. 223.

[2] Замков О. О., Толстопятенко А. В., Черемных Ю. Н. Матемаитческие метды в экономике. — М.: МГУ им. М. В. Ломоносова, Дис, 1997, с. 361.

где новые переменные преобразуются по следующим формулам:

для t = 1:

$$У_{нt} = \sqrt{1 - p^2}\, У_t;$$

$$X_{н_1 t} = \sqrt{1 - p^2}\, X_t;$$

$$X_{н_2 t} = \sqrt{1 - p^2}\, X_t;$$

для t ≥ 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

$$У = d_0 + d_1 X + pe_{t-1} + v_t.$$

Метод Кочрена-Оркатто 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 Кочрена-Оркатта использует метод Эйткена и выполняется в такой последовательности.

Шаг 1. Задается произвольное значение p_1, выполняется метод Эйткена.

Шаг 2. Вычисляется p_2 при использовании коэффициентов Эйткена.

Шаг 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-м 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 2. Определяется при каком значении p ошибка модели S_v будет минимальной [1].

а) да;

б) нет.

14. ***Особенность прогнозирования с учетом автокорреляции остатков*** заключается в том, что удается улучшить прогноз, но только на одну дату.

Прогнозирование по модели с коэффициентом автокорреляции производится по формуле

$$У_{пр(n+1)} = a_0 + a_1 X_{n+1} + pe_n,$$

где $У_{пр(n+1)}$ — 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.

а) да;

б) нет.

Задание 3. Решите задачу 1.

Имеется временной ряд $У_t$ розничного товарооборота (тыс. руб.) киоска розничной торговли в течение недели (табл. 10.1).

[1] Доугерти К. Введение в эконометрику. — М.: НИФРА-М, 1997, с. 224.

Таблица 10.1

База данных временного ряда

t	$У_t$
1	5
2	8
3	6
4	9
5	5
6	10
7	8

где t — время (1 — понедельник, 2 — вторник и т. д.);

$У_t$ — розничный товарооборот (тыс. руб.).

Необходимо определить автокорреляцию первого порядка для временного ряда $У_t$.

Решение

Составим базу данных для расчета (табл. 10.2).

Таблица 10.2

База данных для расчета коэффициента автокорреляции

t	$У_t$	$У_{t-1}$	$У_t$	$У_{t-1}$
1	$У_1 = 5$	$У_{1-1} = У_0$ = нет данных	5	—
2	$У_2 = 8$	$У_{2-1} = У_1 = 5$	8	5
3	$У_3 = 6$	$У_{3-1} = У_2 = 8$	6	8
4	$У_4 = 9$	$У_{4-1} = У_3 = 6$	9	6
5	У5 = 5	$У_{5-1} = У_4 = 9$	5	9
6	$У_6 = 10$	$У_{6-1} = У_5 = 5$	10	5
7	$У_7 = 8$	$У_{7-1} = У_6 = 10$	8	10
8		$У_{8-1} = У_7 = 8$		8

Для изучения свойств временного ряда построим график зависимости $У_t$ от t (рис. 10.2).

Анализ графика на рис. 10.2 показывает, что временной ряд имеет положительную линейную тенденцию с частыми изменениями около ее тенденции.

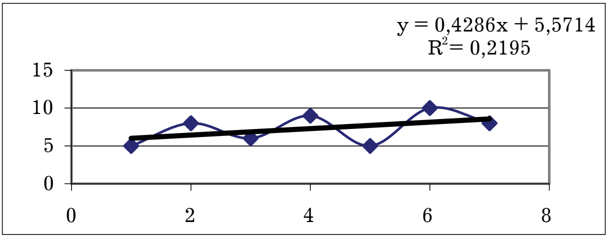


Рис. 10.2 График зависимости $У_t$ от t

Для изучения свойств автокорреляции временного ряда на основании табл. 10.2 построим график зависимости $У_t$ от $У_{t-1}$ (рис. 10.3).

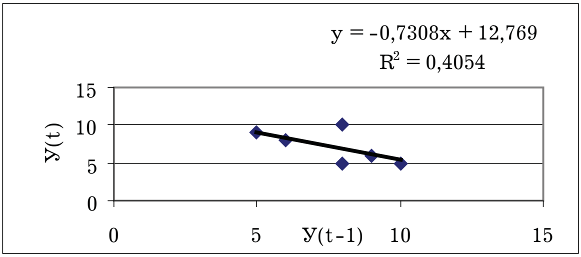


Рис. 10.3. График зависимости $У_t$ от $У_{t-1}$

Анализ рис. 10.3 показывает, что между $У_t$ и $У_{t-1}$ существует отрицательная линейная зависимость.

Выделим в табл. 10.2 те измерения, для которых имеются пары $У_t$ и $У_{t-1}$ и вычислим между ними коэффициент автокорреляции первого порядка

$$r_{d=1} = -\,0{,}6367.$$

Примечание. Расчет коэффициента корреляции можно произвести с помощью программы Ecel “Корреляция”, расположенной по адресу: сервис, анализ данных, корреляция. При расчете коэффициента автокорреляции первого порядка длина временного ряда уменьшится на одно измерение.

Анализ коэффициента автокорреляции первого порядка показывает, что он имеет отрицательное значение (при часто изменяющихся значениях временного ряда). Это означает, что увеличение значения временного ряда приводит на следующий день к его уменьшению (эффект маятника).

Коэффициент автокорреляции может изменяться от −1 до +1.

Комментарии. Существуют две трактовки коэффициента автокорреляции прямая и обратная.

Прямая трактовка — коэффициент автокорреляции показывает степень влияния прошлых значений временного ряда на его будущее значение.

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.

Коэффициент автокорреляции первого порядка будет иметь положительное значение, если временной ряд будет иметь серии возрастающих и серии убывающих значений.

Задание 4. Решите задачу.

Необходимо определить коэффициент и достоверность автокорреляции остатков для задачи 1. База данных представлена в табл. 10.3.

Таблица 10.3

База данных временного ряда

t	$У_t$	$У_{pt}$	e_t	e_{t-1}	e_t^2	$(e_t-e_{t-1})^2$
1	5	6	-1		1	
2	8	6,4286	1,5714	-1	2,4694	6,6122
3	6	6,8571	-0,8571	1,5714	0,7347	5,898
4	9	7,2857	1,7143	-0,8571	2,9388	6,6122
5	5	7,7143	-2,7143	1,7143	7,3673	19,612
6	10	8,1429	1,8571	-2,7143	3,449	20,898
7	8	8,5714	-0,5714	1,8571	0,3265	5,898
8				-0,5714		
Сумма					18,286	65,531

где t — время (1 — понедельник, 2 — вторник и т. д.);

$У_t$ — розничный товарооборот (тыс. руб.).

Решение

Вычислим коэффициенты выборочного уравнения регрессии

$$У_t = У_{pt} + e_t = a_0 + a_1 t + e_t$$

0,4286	5,5714
0,3614	1,6162
0,2195	1,9124
1,4063	5

$$У_t = У_{pt} + e_t = a_0 + a_1 t + e_t = 5{,}5714 + 0{,}4286\ t + e_t$$

Определим остатки и внесем в табл. 10.3.

$e_t = У_t — (a_0 + a_1 t)$ — остатки модели.

Получим лаговые остатки e_{t-1} со сдвигом на одну дату и внесем в табл. 5.3

Для анализа остатков с использованием данных табл. 10.3 построим графики зависимости остатков от времени и зависимость e_t от e_{t-1} (рис. 10.4).

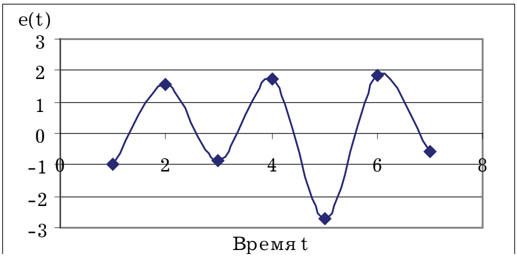


Рис. 10.4. Зависимость остатков e_t от t

Анализ остатков показывает, что остатки часто меняют свои знаки.

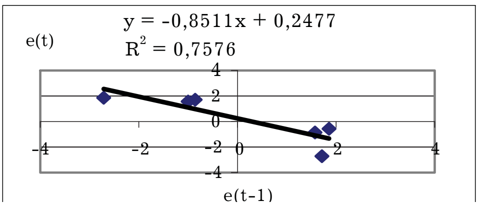


Рис. 10.5. Зависимость e_t от e_{t-1}

Анализ рис. 10.5 показывает, что e_t имеет отрицательную зависимость от e_{t-1}.

Определим коэффициент автокорреляции первого порядка для остатков модели.

Автокорреляция остатков порядка m равна коэффициенту корреляции, рассчитанному между исходными остатками модели и лаговыми остатками порядка m. Коэффициент автокорреляции рассчитывается по обычной формуле коэффициента корреляции, только в качестве первой переменной является исходные остатки, а в качестве второй переменной используется лаговые остатки порядка d.

$$r_{d=1} = r_1 = -0,87$$

Анализ автокорреляции остатков первого порядка показывает, что он имеет отрицательное значение, при этом остатки часто меняют свой знак.

Если остатки имеют серии положительных и отрицательных знаков, то коэффициент автокорреляции остатков первого порядка будет положительным.

Причиной появления автокорреляции остатков 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;
- дисперсии оценок являются смещенными. Часто дисперсии, вычисленные по стандартным формулам, являются заниженными, что влечет за собой увеличение t статистики. Это может привести к признанию статистически значимыми объясняющие переменные, которые в действительности такими не являются [10, с. 253].

Критерии проверки достоверности автокорреляции остатков модели

В статистической литературе известен критерий Дарбина-Уотсона (DW) проверки достоверности автокорреляции остатков только первого порядка. Для проверки достоверности коэффициентов автокорреляции более высокого порядка пока не существует соответствующего статистического критерия. (Дерзайте, наука ждет Вас).

Проверку достоверности коэффициента автокорреляции остатков первого порядка можно проводить по упрощенной или по более точной методике.

Критерий Стьюдента

Упрощенная методика проверки коэффициента автокорреляции совпадает с методикой проверки достоверности обычного парного коэффициента корреляции, которая выполняется в следующей 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 S_{rd} коэффициента автокореляции r_d порядка d

$$S_{rd} = \sqrt{\frac{1 - r_d^2}{n - d}}$$

n =	6
d =	1
r_1 =	-0,8704
S_{r1} =	0,2202

Шаг 4. Вычисляется критерий Стьюдента

$$t = \frac{r_d}{S_{rd}}$$

r_1 =	-0,8704
S_r =	0,2202
t =	3,9526

Шаг 5. Сравнивается t с $t_{\alpha/2}(\alpha = 0,05;\ m = n — d — 2)$

t =	3,9526
$t_{\alpha/2}(\alpha = 0,05;\ m = n — d — 2 = 6-1-2=3)$ =	3,1824

Так как $|t| > t_{\alpha/2}(\alpha = 0{,}05;\ m = n — d — 2)$, то нулевая гипотеза отвергается с вероятностью 1-α,

где t — критерий Стьюдента;

α — уровень значимости критерия.

Следовательно, с вероятностью 0,95 можно утверждать, что коэффициент автокорреляции остатков первого порядка является достоверным.

Критерий Дарбина-Уотсона

Для более точной проверки достоверности коэффициента автокорреляции остатков первого порядка используется критерий Дарбина- Уотсона (DW) в следующей последовательности:

Шаг 1. Вычисляется критерий Дарбина-Уотсона (DW) (табл. 10.3).

Сумма$((e_t\text{-}e_{t-1})^2)$ =	65,531
Сумма(e_t^2) =	18,286
DW =	3,5837

Шаг 2. Определяются по таблице 10.4 нижний $d_н$ и верхний $d_в$ пороговые значения критерия DW, зависящие от уровня значимости, числа измерений и числа объясняющих факторов в модели.

Таблица 10.4

Численные значения нижних $d_н$ и верхних $d_в$ границ критерия DW

n	k = 1		k =2	
	$d_н$	$d_в$	$d_н$	$d_в$
6	0,61	1,4		
7	0,7	1,36	0,47	1,9
8	0,76	1,33	0,56	1,78
9	0,82	1,32	0,63	1,7
10	0,88	1,32	0,7	1,64
11	0,93	1,32	0,76	1,6
12	0,97	1,33	0,81	1,58
13	1,01	1,34	0,86	1,56
14	1,05	1,35	0,91	1,55
15	1,08	1,36	0,95	1,54

n	k = 1		k =2	
	$d_н$	$d_в$	$d_н$	$d_в$
16	1,11	1,37	0,98	1,54
17	1,13	1,38	1,02	1,54
18	1,16	1,39	1,05	1,54
19	1,18	1,4	1,07	1,54
20	1,2	1,41	1,1	1,54
21	1,22	1,42	1,13	1,54
22	1,24	1,43	1,15	1,54
23	1,26	1,44	1,17	1,54
24	1,27	1,45	1,19	1,55
25	1,29	1,45	1,21	1,55
30	1,35	1,49	1,28	1,57
50	1,5	1,59	1,46	1,63
100	1,65	1,69	1,63	1,72
200	1,76	1,78	1,75	1,79

где n- длина ряда наблюдений;

k — количество объясняющих переменных в модели.

$d_н$ (α = 0,05, n = 6, k = 1) = 0,61;

$d_в$ (α = 0,05, n = 6, k = 1) = 1,4.

Шаг 3. Определяется место расположения критерия DW на числовой оси возможных его значений (рис. 10.6).

Таблица 10.5

Числовая ось значений критерия DW

min =	0
$d_н$ =	0,61
$d_в$ =	1,4
Середина =	2
4 — $d_в$ =	2,6
4 — $d_н$ =	3,39
max =	4
DW =	3,584

Рис. 10.6 Числовая ось возможных значений критерия DW

Анализ рис. 10.6 показывает, что критерий DW находится в области достоверной (с вероятностью 0,95) отрицательной автокорреляции первого порядка.

Задание 5. Решите задачу.

Имеется временной ряд $У_t$ розничного товарооборота (тыс. руб.) киоска розничной торговли, полученный за семь дней (табл. 10.6).

Таблица 10.6

База данных временного ряда

t	$У_t$
1	5
2	8
3	6
4	9
5	5
6	10
7	8
8	?

где t — время (1 — понедельник, 2 — вторник и т. д.);

$У_t$ — розничный товарооборот (тыс. руб.).

Условие задачи 3 соответствует задаче 1, которое надо дополнить получением точечного прогноза с учетом автокорреляции остатков.

Решение задачи

Целью моделирования является повышение точности прогноза показателя временного ряда за счет учета автокорреляции остатков для улучшения планирования и снижения издержек, вызванных как дефицитом так и избытком товаров.

Выбор переменных

В качестве временного ряда выбран розничный товарооборот.

Величина товарооборота зависит от шести основных факторов: *сырье и материалы* (ассортимент товаров и их качество); *машины* (оборудование и дизайн магазина); *методы* (форма обслуживания покупателей: форма самообслуживания, культура обслуживания); *люди* (количество продавцов, уровень квалификации, образование, возраст, деловые характеристики); *физическая и абстрактная среда* (температура, влажность; традиции, психологический климат в магазине; правовая среда существования магазина, информационное обеспечение магазина); *время* (линейные и нелинейные тенденции, сезонность, длинноволновая периодичность, влияние прошлых, текущих и будущих значений факторов и зависимой переменной, а также остатков модели, день недели) (рис. 10.7).

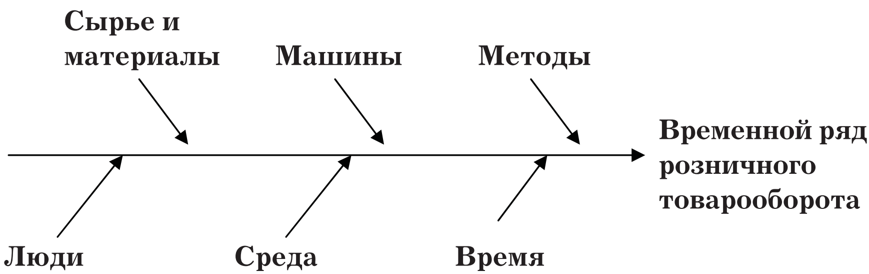


Рис. 10.7 Причинно-следственная диаграмма Исикавы

Выделение входных и выходных переменных

Анализ рис. 10.7 методом мозговой атаки специалистами, в состав которых входил хозяин киоска, показывает, что временной ряд розничного товарооборота зависит в большей степени от времени (дня недели).

Сбор статистических данных

Был проведен сбор данных временного ряда розничного товарооборота киоска в динамике за семь дней, представленный в табл. 6. Динамика временного ряда показана на рис. 10.8.

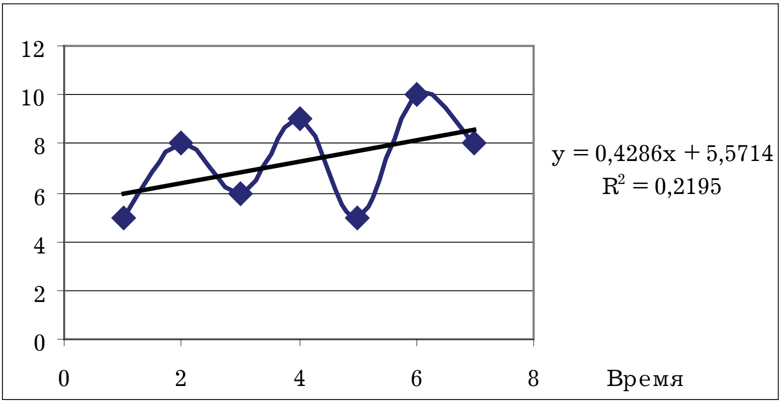


Рис. 10.8. Динамика временного ряда

Предмодельный анализ

Выдвижение гипотез

Анализ рис. 10.8 показывает, что временной ряд имеет линейную тенденцию, остатки модели гомоскедастичны и часто меняют свой знак, следовательно, в остатках имеется отрицательная автокорреляция.

Временной ряд имеет линейную тенденцию.

Формулировка допущений

Предполагаем, что в будущем и временной ряд сохранит свою тенденцию.

Спецификация

Обобщенная линейная модель парной регрессии

Предполагаем, что зависимость У от X можно представить в виде обобщенной линейной модели парной регрессии. Приводим спецификацию данной модели:

$$\mathbf{У = XA} + \varepsilon$$

при условии, что $M(\varepsilon) = 0$,

Предположим, что в обобщенной линейной модели парной регрессии остатки гомоскедастичны и имеется достоверная их автокорреляция, тогда их можно представить в виде следующей модели

$$\varepsilon_t = \rho\varepsilon_t + \upsilon_t$$

где ε_t — возмущения, которые воздействуют на У,

υ_t — случайная составляющая.

ρ — параметр авторегрессии остатков.

Идентифицируемость модели

Обобщенная линейная модель парной регрессии соответствует объекту исследования и может быть применена к изучению временного ряда.

Идентификация модели

Для оценки параметров модели с автокоррелированными остатками воспользуемся выборочным уравнением регрессии следующего вида:

$$У_t = a_0 + a_1 t + e_t = в_0 + в_1 t + pe_{t-1} + v_t, (*)$$

где $pe_{t-1} + v_t = e_t$ — остатки модели,

p — выборочный коэффициент авторегрессии остатков.

$M(\mathbf{v}) = 0$

$M(\mathbf{vv'}) = s^2\,\mathbf{I}$

Модель (*) обладает следующими свойствами:

— коэффициенты a_0 и a_1, определенные методом наименьших квадратов будут несмещенными, состоятельными, но неэффективными.

— нельзя одновременно определить коэффициенты $в_0$, $в_1$ и p.

Для определения коэффициентов $в_1$, $в_2$, р можно использовать метод наименьших квадратов.

Шаг 1. Вычисление коэффициентов a_0, a_1 модели $У_t = a_0 + a_1 t + e_t$ методом наименьших квадратов.

Коэффициенты a_0 и a_1 мы уже находили (см. табл. 10.3)

$a_0 = 5,5714$

$a_1 = 0,4286$

Шаг 2. Вычисление остатков модели $e_t = У_t — (a_0+a_1 t)$.

Остатки модели мы уже находили (см. табл. 10.3).

Шаг 3. Определение коэффициента автокорреляции р.

Возьмем данные из табл. 10.3, ????????????

Таблица 10.7

База данных временного ряда (см. табл. 10.3)

t	$У_t$	$У_{pt}$	e_t	e_{t-1}	$У_{p1t}$
1	5	6	-1		
2	8	6,4286	1,5714	-1	7,2718
3	6	6,8571	-0,8571	1,5714	5,5321
4	9	7,2857	1,7143	-0,8571	8,0084
5	5	7,7143	-2,7143	1,7143	6,2688
6	10	8,1429	1,8571	-2,7143	10,431
7	8	8,5714	-0,5714	1,8571	7,0055
8		9		-0,5714	9,4818

где t — время (1 — понедельник, 2 — вторник, и т. д.);

$У_t$ — розничный товарооборот (тыс. руб.);

$У_{pt} = a_0+a_1 t$;

$У_{p1t} = a_0+a_1 t + р e_{t-1}$.

Вычислим коэффициент р из уравнения $e_t = р e_{t-1} + v_t$, не содержащего свободного коэффициента, где зависимая переменная $У_t = e_t$, объясняемая переменная $X = e_{t-1}$.

Для расчета коэффициента р воспользуемся функцией “Линейн” без свободного коэффициента при учете того, что количество измерений будет на одно значение меньше.

-0,8432	0
0,2243	#Н/Д
0,7361	0,95053
13,947	5

$$e_t = -0{,}8432\ e_{t-1} + v_t$$

Шаг 4. Вычислим точечный прогноз без учета автокорреляции остатков

$$У_{пр} = a_0 + a_1 t = 5.5714 + 0.4286 \times 8 = 9$$

$$a_0 = 5{,}5714$$

$$a_1 = 0{,}4286$$

$$t = 8.$$

Шаг 5. Вычислим расчетные значения $У_{р1}$ с учетом автокорреляции остатков

$$У_р = a_0 + a_1\ t + p\ e_{t-1}.$$

Результаты расчетов занесем в табл. 10.7 и построим график зависимости $У_t$, $У_{p1t}$ от t.

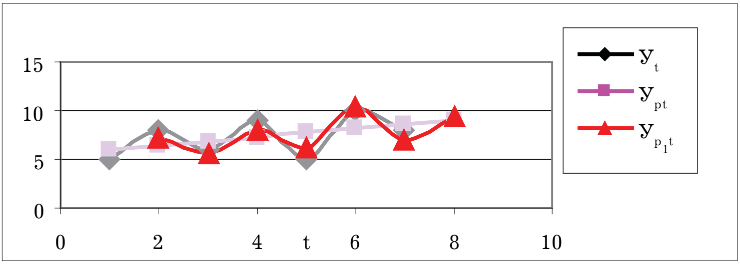


Рис. 10.9. График зависимости $У_t$ — фактические значения, $У_{pt} = a_0+a_1t$, $У_{p1t} = a_0+a_1t+pe_{t-1}$ от t.

Анализ рис. 10.9 показывает хорошую воспроизводимость колебаний исходного ряда $У_t$ с помощью $У_{p1t}$, учитывающую автокорреляцию остатков.

Модели, учитывающие автокорреляцию остатков, имеют два достоинства:

— с помощью линейной функции удается воспроизвести нелинейную тенденцию временного ряда;

— имеется повысить точность краткосрочного прогноза.

Шаг 5. Вычислим точечный прогноз с учетом автокорреляции остатков.

Предположим, что $a_0 = в_0$, $a_1 = в_1$, тогда

$$У_{пр} = a_0 + a_1 t + p\, e_{t-1} = 9{,}4818,$$

где $a_0 = 5{,}5714$;

$a_1 = 0{,}4286$;

$t = 8$;

$e_{t-1} = e_7 = -0{,}5714$;

$p = -0{,}8432$.

Анализ точечного прогноза с учетом автокорреляции остатков показывает, что полученное значение прогноза более точное, чем обычный прогноз без учета автокорреляции остатков. Однако повышение точности прогноза возможно только на одну дату вперед, для которой имеется предшествующее значение остатка.

Задание 6. Сравните характеристики модели автокорреляции остатков, вычисленные разными методами.

Для получения более точных значений коэффициентов b_0, b_1 и p модели

$$У_t = a_0 + a_1 t + e_t = b_0 + b_1 t + pe_{t-1} + v_t.$$

используются методы:

— наименьших квадратов;

— Эйткена;

— Дарбина;

— Кочрена-Оркатто;

— Хилдрета-Лу.

Перечисленные методы позволяют уточнить коэффициенты b_0, b_1 и p, которые, как правило, мало отличаются от коэффициентов, определенных методом наименьших квадратов, однако, являются хорошей иллюстрацией теоретических эконометрических исследований. Уточненные коэффициенты используются для получения точечного прогноза по выше изложенной методике. Для получения уточненных прогнозов, в первом приближении, с использованием модели автокорреляции остатков вполне достаточно метода наименьших квадратов.

Приводим сводную характеристику методов расчета коэффициентов модели с учетом автокорреляции остатков для данных, приведенных в табл. 10.8 и рис. 10.10.

Таблица 10.8

Исходные данные для расчетов разными методами

t	$У_t$
1	5
2	13
3	22
4	18
5	8
6	10
7	15
8	16
9	19
10	8
11	22
12	10

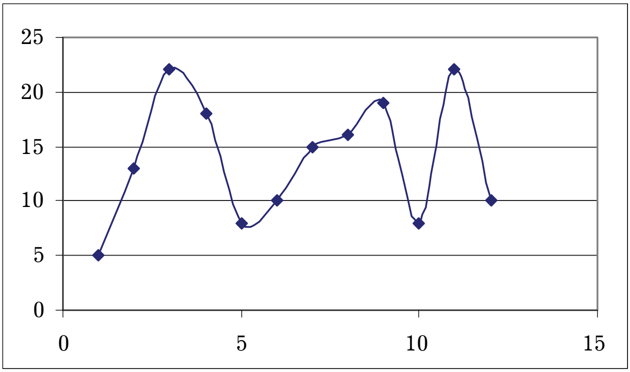


Рис. 10.10 Динамика временного ряда

где t — номер месяца;

$У_t$ — объем производства одной фирмы (тыс. руб.).

Таблица 10.9

Основные характеристики модели $У_t = a_0 + a_1 t + e_t = b_0 + b_1 t + pe_{t-1} + v_t$, произведенные разными методами

Параметр	Оценки параметров модели, полученные методами				
	$м_1$	$м_2$	$м_3$	$м_4$	$м_5$
b_0	12,197	12,547	12,489	12,494	12,494
b_1	0,2517	0,2306	0,2341	0,2338	0,2338
p	-0,2421	-0,2421	-0,196	-0,2	-0,2
S_v	6,1055	6,1	6,1331	6,1324	6,1324
$p_м$	-0,2421	-0,23	-0,23	-0,23	-0,23
$У_{пр}$	16,733	16,831	16,572	16,594	16,594

где $м_1$ — метод наименьших квадратов;

$м_2$ — метод Эйткена;

$м_3$ — метод Дарбина;

$м_4$ — метод Кочрена-Оркатто;

$м_5$ — метод Хилдрета-Лу;

р — коэффициент автокорреляции остатков, используемый моделью

$$У_t = b_0 + b_1t + pe_{t-1} + v_t;$$

S_v — ошибка модели, учитывающая автокорреляцию р остатков модели;

$р_м$ — коэффициент модели $e_t = p_м\, e_{t-1} + v_t$;

$У_{пр}$ — прогнозное значение У с учетом автокорреляции остатков.

Примечание. При реализации метода Дарбина переменная t-1 была исключена из модели

$$У_t = a_0(1\text{-}p) + pУ_{t-1} + a_1t — pa_1(t\text{-}1) + v_t,$$

так как она линейно связана с t.

Анализ табл. 10.9 показывает, что коэффициенты $в_0$, $в_1$ и р для всех моделей примерно равны между собой. По ошибке модели лучшими являются метод Эйткена иметод наименьших квадратов. Однако предпочтение нужно отдать методу наименьших квадратов, так как он требует меньше расчетов.

Вывод. Метод наименьших квадратов позволяет получить лучшую модель с учетом автокорреляции остатков.

Задание второго уровня сложности

Задание. Выполните расчеты характеристик модели автокорреляции остатков разными методами: наименьших квадратов, Эйткена, Дарбина, Кочрена-Оркатто, Хилдрета-Лу, используя данные табл. 10.8.

Имеется модель автокорреляции остатков первого порядка

$$У_t = \alpha_0 + \alpha_1X_t + \rho\varepsilon_{t-1} + \upsilon_t,$$

где $\varepsilon_t = \rho\varepsilon_{t-1} + \upsilon_t$, — возмущение, подчиняющееся авторегрессионной модели первого порядка и для всех t выполняются условия:

$M(\upsilon) = 0$,

$M(\upsilon_t, \upsilon_{t+m}) = \sigma^2$ для $m = 0$,

$M(\upsilon_t, \upsilon_{t+m}) = 0$ для $m \neq 0$,

$У_t$ _ зависимая переменная временного ряда,

X_t _ объясняемая переменная,

t- порядковый номер времени,

M — знак математического ожидания,

$$\varepsilon_t = У_t — (\alpha_0 + \alpha_1 X_t).$$

Необходимо оценить параметры α_0, α_1, ρ при условии $\upsilon'\upsilon$ -> min.

Примечание. Оценку параметров этой модели можно производить с помощью данных выборочной совокупности по модели

$$У_t = a_0 + a_1 X_t + pe_{t-1} + v_t.$$

Метод наименьших квадратов

Метод наименьших квадратов выполняется в такой последовательности.

1. Вычисляются коэффициенты a_0 и a_1 модели

$$У_t = a_0 + a_1 X_t + e_t$$

и проводится эконометрический анализ полученной модели.

Динамика зависимой переменной $У_t$ представлена на рис. 10.10.

Произведем эконометрический анализ линейной модели

$$У_t = a_0 + a_1 X_t + e_t.$$

Таблица 10.10

База данных для оценки параметров модели

t	X_t	$У_t$	$У_{pt}$	$(X_t — \bar{X})^2$	$У_{min.м}$	$У_{max.м}$	$e_t = У_t — У_{pt}$
1	1	5	12,449	30,25	5,2845	19,6129	-7,449
2	2	13	12,7	20,25	6,4432	18,9577	0,2995
3	3	22	12,952	12,25	7,5286	18,3759	9,0478
4	4	18	13,204	6,25	8,3634	18,0445	4,796
5	5	8	13,456	2,25	8,9674	17,944	-5,456
6	6	10	13,707	0,25	9,2753	18,1397	-3,707
7	7	15	13,959	0,25	9,2622	18,6562	1,0408
8	8	16	14,211	2,25	9,2616	19,1603	1,789

t	X_t	$У_t$	$У_{pt}$	$(X_t — \overline{X})^2$	$У_{min.м}$	$У_{max.м}$	$e_t = У_t — У_{pt}$
9	9	19	14,463	6,25	9,0437	19,8817	4,5373
10	10	8	14,714	12,25	8,659	20,7699	-6,714
11	11	22	14,966	20,25	8,154	21,7784	7,0338
12	12	10	15,218	30,25	7,5643	22,8716	-5,218
Сумма	78			143			
Среднее	6,5						

где $У_{min\ м}$, $У_{max\ м}$ — 95%-ный нижний и верхний доверительные интервалы математического ожидания У, которые вычисляются по формуле

$$У_{мi} = У_{pi} \pm t_{таб} E \sqrt{\frac{1}{n} + \frac{\left(X_i - \overline{X}\right)^2}{\sum (X_i - \overline{X})^2}}.$$

Характеристики модели вычислялись с помощью функции “Линейн”.

0,4952	3,6442
0,0252	5,9212
0,2585	10
9,0629	350,6

Представим результаты расчетов в общепринятом обозначении.

У=	12,197	0,2517	X	
S_a=	3,6442	0,4952		
t_a=	3,3469	0,5084		
E=	5,9212			
R^2=	0,0252		$t_{таб}$.=	2,2281
F=	0,2585		$F_{таб}$ =	4,9646

Представляем результаты эконометрического анализа:

— 2,52% исходных данных имеют линейную тенденцию,

— по критерию Фишера достоверность модели статистически не доказана, так как $F < F_{таб}$ ($\alpha = 0{,}05$, $m_1 = k — 1$, $m_2 = n — k$).

— по критерию Стьюдента проверим достоверность коэффициентов модели. Так как $t_{a0} > t_{таб}$, то коэффициент a_0 с вероятностью 0,95 отличается от нуля.

Так как $t_{a1} < t_{таб}$, то коэффициент a_1 не отличается от нуля;

— при ожидаемом значении объясняемой переменной $X_{ож} = 13$

среднее прогнозное значение зависимой переменной будет равно

$$У_{пр} = 15{,}47.$$

Прогнозное фактическое значение зависимой переменной с вероятность 0,95 будет находиться в интервале от 2,2765 до 28,663;

— графическое представление результатов расчета имеет вид, представленный на рис. 10.11

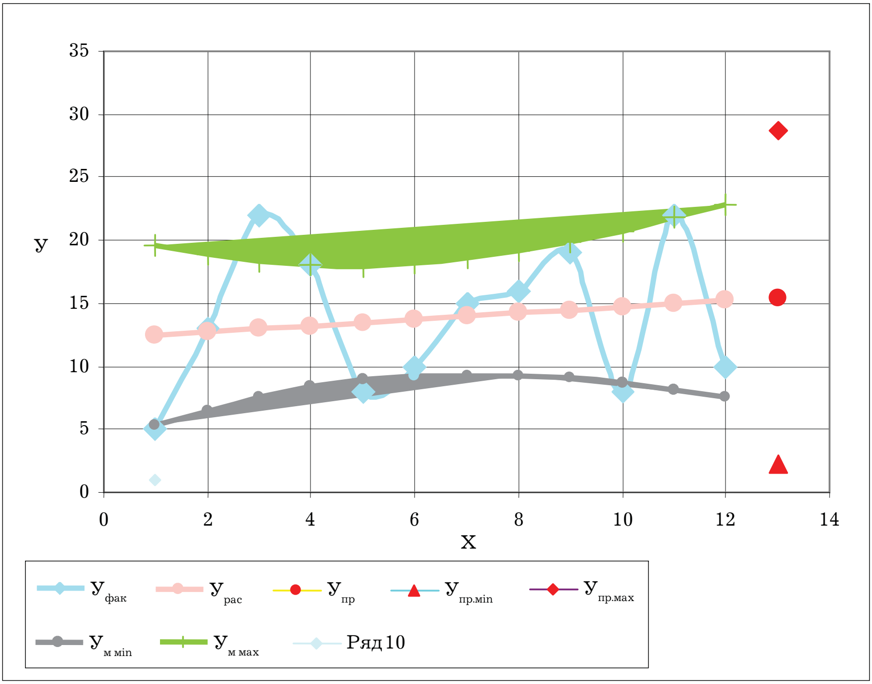


Рис. 10.11 Характеристики модели без учета автокорреляции остатков

Построим график остатков e_t модели (рис. 10.12).

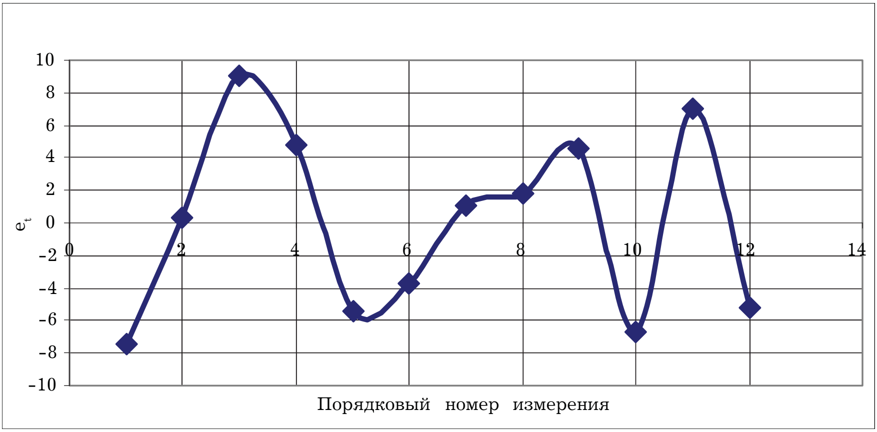


Рис. 10.12. График остатков

Вывод. Остатки содержат серии положительных и отрицательных значений.

2. Вычислим коэффициент p модели

$e_t = pe_{t-1} + v_t$,

где $e_t = У_t — (a_0 + a_1X_t)$.

Коэффициент p равен коэффициенту автокорреляции первого порядка остатков e_t, вычисленному с помощью функции Excel "коррел"

$$p = -0,242.$$

Получим прогнозные значения $У_{пр}$ при ожидаемом значении $X_{ож} = 13$ с учетом автокорреляции остатков.

В результате расчетов была получена модель, учитывающая автокорреляцию остатков.

$$У_t = a_0 + a_1X_t + pe_{t-1} + v_t = 12,197 + 0,2517X_t — 0,2421e_{t-1} + v_t.$$

Выполним прогнозирование зависимого признака У с учетом автокорреляции остатков на период $t = n + 1 = 13$ по формуле

$$У_{пр(n+1)}) = 12,197 + 0,2517X_{n+1} — 0,2421e_n = 16,7331,$$

где ожидаемое значение независимой переменной $X_{t=n+1} = 13$, предыдущее значение остатков в момент времени t = n равно

$$e_n = -5{,}218.$$

Вывод. Особенность прогнозирования с автокоррелированными остатками заключается в том, что к прогнозному значению У, равному

$$a_0 + a_1X_{n+1} = 15{,}47,$$

определенному обычным способом, добавляется поправка на автокорреляцию остатков, равная

$$pe_n = -\, 0{,}24 \times (-\, 2{,}218) = 1{,}2634.$$

Коррекция прогноза на автокорреляцию остатков позволяет получить более обоснованный прогноз, к сожалению, только на одну дату вперед.

3. Вычислим характеристики модели

$$У_t = a_0 + a_1X_t + pe_{t-1} + v_{t.}$$

Таблица 10.11

Расчет характеристик модели с учетом автокорреляции остатков

X_t	$У_{1t}$	e_t	$У_{2t}$	v_t
1	12,449	-7,449	12,449	-7,449
2	12,7	0,2995	14,504	-1,504
3	12,952	9,0478	12,88	9,1203
4	13,204	4,796	11,013	6,9868
5	13,456	-5,456	12,294	-4,294
6	13,707	-3,707	15,028	-5,028
7	13,959	1,0408	14,857	0,1431
8	14,211	1,789	13,959	2,0411
9	14,463	4,5373	14,03	4,9705
10	14,714	-6,714	13,616	-5,616
11	14,966	7,0338	16,592	5,408
12	15,218	-5,218	13,515	-3,515
Сумма				1,2634

где $У_{1t} = c_0 + c_1X_t$;

$e_t = У_t - У_{1t}$;

$У_{2t} = a_0 + a_1X_t + pe_{t-1}$;

$v_t = У_t - У_{2t}$.

$$S_v = \sqrt{\frac{\sum v_t^{\,2}}{n-3}} = 6{,}1055,$$

где S_v — ошибка модели, учитывающая автокорреляцию остатка,

n — 3 — число степеней свободы для остаточной дисперсии (число 3 означает количество коэффициентов модели $У_{2t} = a_0 + a_1X_t + pe_{t-1}$ с учетом коэффициента p).

Вычисляем коэффициент p_1 модели

$$e_t = p_1e(t\text{-}1) + v_t,$$

с помощью функции “Линейн”,

где $e_t = У_t - (a_0 + a_1X_t)$.

$p_1 = -0{,}242$

Вычисляем коэффициент p_2 модели

$$v_t = p_2v_{t-1} + z_t,$$

где $v_t = У_t - (a_0 + a_1X_t + pe_{t-1})$.

$$p_2 = -0{,}063.$$

Вывод. Автокорреляция остатков уменьшилась в 3,8629 раза. Ошибка модели уменьшилась в 0,9698 раза.

4. Выходными характеристиками модели $У_t = a_0 + a_1X_t + pe_{t-1} + v_t$, вычисленными методом наименьших квадратов, являются следующие величины a_0, a_1, p, S_v:

$$a_0 = b_0 = 12{,}197,\ a_1 = b_1 = 0{,}2517,\ p = -0{,}242,\ S_v = 6{,}1055.$$

Метод Эйткена

Оценим параметры модели с помощью метода Эйткена в форме преобразований данных.

Метод Эйткена используется при условии, что известно значение коэффициента p модели

$У_t = a_0 + a_1X_t + pe_{t-1} + v_t$.

Метод Эйткена выполняется в следующей последовательности.

1. Вводится коэффициент p.

Введите любое разумное значение автокорреляции остатков, используя априорную информацию.

$$p = -0{,}242$$

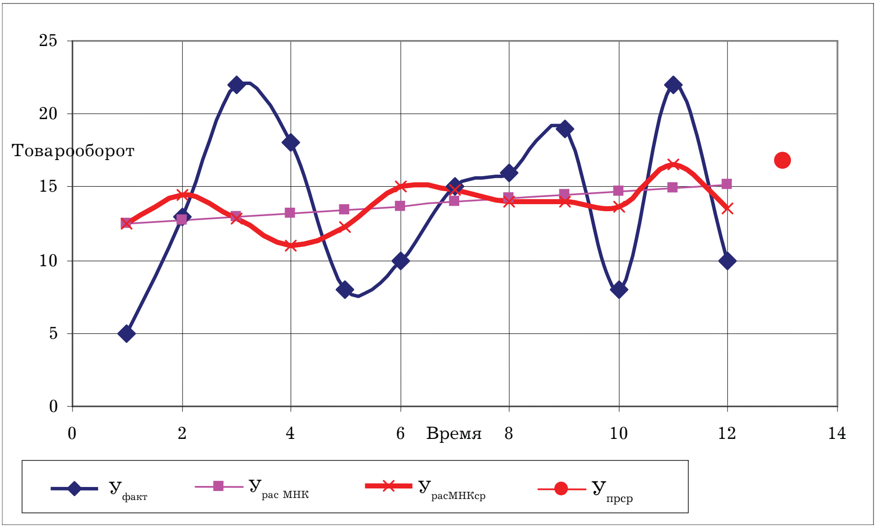


Рис. 10.13. Расчетные значение У без учета и с учетом автокорреляции остатков, определенные МНК

2. Вычисляются методом наименьших квадратов характеристики модели

$$У_{нt} = b_0 X_{н1t} + b_1 X_{н2t} + z_t,$$

где b_0, b_1 - коэффициенты Эйткена.

Новые переменные вычисляются по следующим формулам:

$t = 1$	$t \geq 2$
$У_{нt} = \sqrt{1 - p^2}\, У_t$	$У_{нt} = У_t - pУ_{t-1}$
$X_{н1t} = \sqrt{1 - p^2}$	$X_{н1t} = 1 - p$
$X_{н2t} = \sqrt{1 - p^2}\, X_t$	$X_{н2t} = X_t - pX_{t-1}$

Таблица 10.12

База данных преобразованных переменных

t	$X_{н1t}$	$X_{н2t}$	$У_{нt}$
1	0,97	0,97	4,85
2	1,24	2,24	14,21
3	1,24	3,48	25,14
4	1,24	4,72	23,32
5	1,24	5,96	12,35
6	1,24	7,21	11,93
7	1,24	8,45	17,42
8	1,24	9,69	19,63
9	1,24	10,93	22,87
10	1,24	12,17	12,60
11	1,24	13,42	23,93
12	1,24	14,66	15,32
Сумма	14,63	93,95	203,62
Среднее		7,82	

Вычислим коэффициенты модели

$$У_{нt}= b_0X_{н1t} + b_1X_{н2t} + z_t$$

с использованием функции “Линейн” (без свободного коэффициента)

0,2306	12,547
0,4073	2,9761
0,9145	5,7563
53,49	10
3544,7	331,35

b_0=12,547
b_1=0,2306

3. Вычисляются характеристики модели по исходным данным

$$У_t = b_0 + b_1X_t + pe_{t-1} + v_t.$$

Примечание. Коэффициенты Эйткена b_0 и b_1 вычислялись они используются по преобразованным данным, а для расчета характеристик модели по исходным данным.

Таблица 10.13

Расчет характеристик модели с учетом автокорреляции остатков

X_t	$У_{1t}$	e_t	$У_{2t}$	v_t
1	12,778	-7,778	12,778	-7,778
2	13,008	-0,008	14,892	-1,892
3	13,239	8,761	13,241	8,759
4	13,47	4,5304	11,348	6,6518
5	13,7	-5,7	12,603	-4,603
6	13,931	-3,931	15,311	-5,311
7	14,161	0,8387	15,113	-0,113
8	14,392	1,6081	14,189	1,8112
9	14,622	4,3776	14,233	4,7669
10	14,853	-6,853	13,793	-5,793
11	15,084	6,9164	16,743	5,2571
12	15,314	-5,314	13,639	-3,639
Сумма				-1,883

где $У_{1t} = b_0 + b_1X_t$;

$$e_t = У_t — У_{1t};$$

$$У_{2t} = b_0 + b_1X_t + pe_{t-1};$$

$$v_t = У_t — У_{2t}.$$

$$S_v = \sqrt{\frac{\sum v_t^2}{n-3}} = 6{,}1;$$

где S_v — ошибка модели, учитывающая автокорреляцию остатков,

n — 3 — число степеней свободы для остаточной дисперсии.

Вычислим прогнозное значение У по модели Эйткена с учетом автокорреляции остатков.

$$У_{пр(n+1)} = b_0 + b_1X_{n+1} + pe_n = 16{,}831,$$

где $b_0 = 12{,}547$, $b_1 = 0{,}2306$, $X_{n+1} = 13$, $p = -0{,}242$, $e_n = -5{,}314$.

Вычисляем коэффициент p_1 модели

$$e_t = p_1e_{t-1} + z_t,$$

где $e_t = У_t — (b_0 + b_1X_t)$.

$$p_1 = -0{,}23.$$

Вывод. Коэффициенты автокорреляции остатков, вычисленные для модели по исходным данным и методом Эйткена, совпадают между собой.

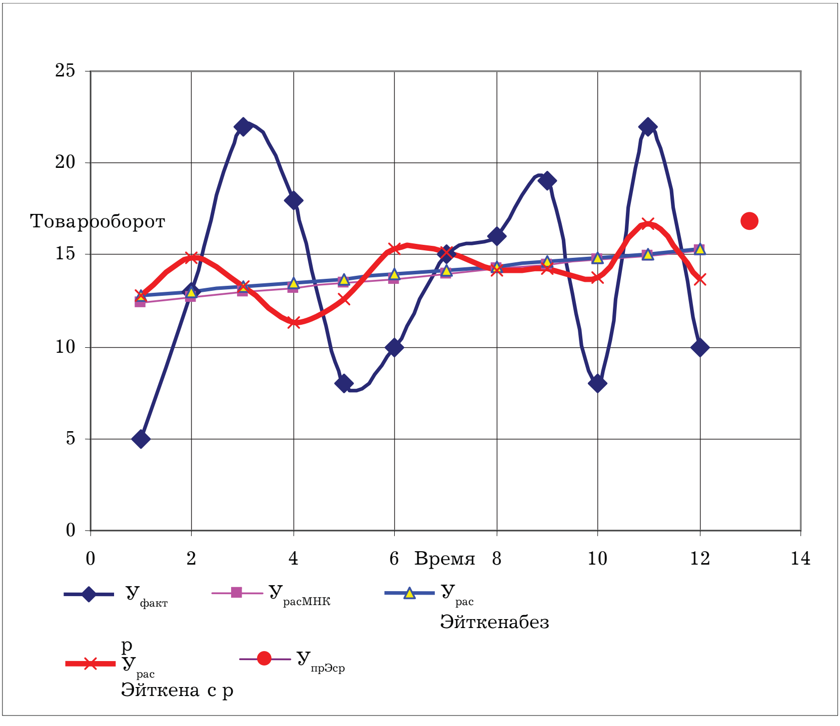


Рис. 10.14 Расчетные значения У без учета и с учетом автокорреляции остатков, определенные методом Эйткена

4. Выходными характеристиками блока метода Эйткена будут являться следующие величины: b_0, b_1, p, S_v, p_1.

$b_0 = 12{,}547$, $b_1 = 0{,}2306$ — коэффициенты Эйткена модели

$$У_t = b_0 + b_1X_t + pe_{t-1} + v_t,$$

$p = -0{,}242$ — используемый коэффициент модели $У_t = b_0 + b_1X_t + pe_{t-1} + v_t$,

$S_v = 6{,}1$ — ошибка модели $У_t = b_0 + b_1X_t + pe_{t-1} + v_t$,

$p_1 = -0{,}23$ — коэффициент модели $e_t = p_1e_{t-1} + z_t$,

Метод Дарбина 1

Оценим параметры модели с помощью метода Дарбина 1, который выполняется в следующей последовательности.

Вычисляется коэффициент p, входящий в состав модели

$$У_t = a_0(1 - p) + pУ_{t-1} + a_1X_t - pa_1X_{t-1} + z_t$$

Таблица 10.14

База данных для расчета коэффициента p

t	$У_{t-1}$	X_t	$-X_{t-1}$	$У_t$
1	5	2	-1	13
2	13	3	-2	22
3	22	4	-3	18
4	18	5	-4	8
5	8	6	-5	10
6	10	7	-6	15
7	15	8	-7	16
8	16	9	-8	19
9	19	10	-9	8
10	8	11	-10	22
11	22	12	-11	10

Расчет характеристик модели

$$У_t = a_0(1 - p) + pУ_{t-1} + a_1X_t - pa_1X_{t-1} + z_t$$

с помощью функции “Линейн”

0	-0,049	-0,196	17,762
0	0,5704	0,3227	5,3619
0,0532	5,7001	#Н/Д	#Н/Д
0,2249	8	#Н/Д	#Н/Д
14,614	259,93	#Н/Д	#Н/Д

Коэффициент перед переменной ($-X_{t-1}$) равен нулю, так как переменные X_t и — X_{t-1} линейно связаны между собой.

Примечание. Объясняемые переменные X_t и ($-X_{t-1}$) имеют одинаковый шаг изменения и являются взаимосвязанными, поэтому оценить коэффициенты методом Дарбина не удается.

Предложим следующий метод решения задачи — исключим из модели переменную ($-X_{t-1}$) и вычислим характеристики модели

$$У_t = a_0(1 - p) + pУ_{t-1} + a_1X_t + v_t$$

с помощью функции "Линейн".

-0,049	-0,196	17,762
0,5704	0,3227	5,3619
0,0532	5,7001	#Н/Д
0,2249	8	#Н/Д
14,614	259,93	#Н/Д

$p = -0{,}196$;

$a_0(1 - p) = 17{,}762$, откуда $a_0 = 17{,}762/(1 + 0{,}196) = 14{,}85117$

$a_1 = -0{,}049$;

$S_v = 5{,}7001$ — ошибка модели.

На этом метод Дарбина 1 заканчивается, но можно продолжить решение задачи с использованием метода Эйткена, который называется методом Дарбина 2.

Метод Дарбина 2

С помощью метода Дарбина 1 был вычислен коэффициент автокорреляции $p = -0{,}196$, который используется для построения модели методом Эйткена.

Метод Эйткена уже был реализован при значении коэффициента автокорреляции $p = -0{,}242$. Заменим коэффициент $-0{,}242$ на $-0{,}196$, произведем аналогичные расчеты и получим следующие результаты без подробных пояснений (табл. 10.15).

Таблица 10.15

База данных преобразованных переменных

t	$X_{н1t}$	$X_{н2t}$	$У_{нt}$
1	0,9806	0,9806	4,903
2	1,196	2,196	13,98
3	1,196	3,392	24,548
4	1,196	4,588	22,312
5	1,196	5,7841	11,528
6	1,196	6,9801	11,568
7	1,196	8,1761	16,96
8	1,196	9,3721	18,94
9	1,196	10,568	22,136
10	1,196	11,764	11,724
11	1,196	12,96	23,568
12	1,196	14,156	14,312
Сумма	14,137	90,918	196,48
Среднее		7,5765	

Вычислим коэффициенты модели

$$У_{нt} = b_0 X_{н1t} + b_1 X_{н2t} + z_t$$

с использованием функции "Линейн" (без свободного коэффициента)

0,2341	12,489
0,4205	3,076
0,9081	5,7683
49,41	10
3288,1	332,73

$b_0 = 12{,}489$

$b_1 = 0{,}2341$

Представим результаты расчетов в общепринятом обозначении.

$У_{нрt} = 12{,}489\ X_{н1t} + 0{,}2341\ X_{н2t}$

$S_b = 3{,}076$ $\quad$ 0,4205

$t_b = 4{,}0602$ $\quad$ 0,5569

$E = 5{,}7683$

$R^2 = 0{,}9081$ $\quad$ $t_{таб} = 2{,}2281$

$F = 49{,}41$ $\quad$ $F_{таб} = 4{,}9646$

Вычислим характеристики модели

$$У_t = d_0 + d_1X_t + pe_{t-1} + v_t$$

Таблица 10.16

Расчет характеристик модели с учетом автокорреляции остатков

t	$У_{1t}$	e_t	$У_{2t}$	v_t
1	12,723	-7,723	12,723	-7,723
2	12,958	0,0425	14,237	-1,237
3	13,192	8,8083	12,949	9,0508
4	13,426	4,5742	11,465	6,5349
5	13,66	-5,66	12,529	-4,529
6	13,894	-3,894	14,769	-4,769
7	14,128	0,8717	14,657	0,3426
8	14,362	1,6376	13,957	2,0426
9	14,597	4,4035	14,041	4,9586
10	14,831	-6,831	13,733	-5,733
11	15,065	6,9352	16,17	5,8304
12	15,299	-5,299	13,705	-3,705
Сумма				1,0617

где $У_{1t} = d_0 + d_1X_t$,

$$e_t = У_t — У_{1t}$$

$$У_{2t} = d_0 + d_1X_t + pe_{t-1}$$

$$v_t = У_t — У_{2t}$$

$$S_v = \sqrt{\frac{\sum v_t^{\,2}}{n-3}} = 6{,}1331$$

где S_v — ошибка модели, учитывающая автокорреляцию остатка,

n — 3 — число степеней свободы для остаточной

Вычислим прогнозное значение $У_{пр}$ по модели Дарбина 2 с учетом автокорреляции остатков:

$$У_{пр(n+1)} = b_0 + b_1X_{n+1} + pe_n = 16{,}572$$

где $b_0 = 12{,}489$, $b_1 = 0{,}2341$, $X_{n+1} = 13$, $p = -0{,}196$, $e_n = -5{,}299$

Вычисляем коэффициент p_1 модели

$$e_t = p_1e_{t-1} + z_t,$$

где $e_t = У_t — (b_0 + b_1X_t)$.

$$p_1 = -0{,}23$$

Выходными характеристиками блока метода Дарбина 2 являются следующие величины: b_0, b_1, p, S_v, p_1.

$b_0 = 12{,}489$, $b_1 = 0{,}2341$, $p = -0{,}196$, $S_v = 6{,}1331$ — коэффициенты Дарбина и ошибка модели $У_t = b_0 + b_1X_t + pe_{t-1} + v_t$,

$p_1 = -0{,}23$ — коэффициент модели $e_t = p_1e_{t-1} + z_t$.

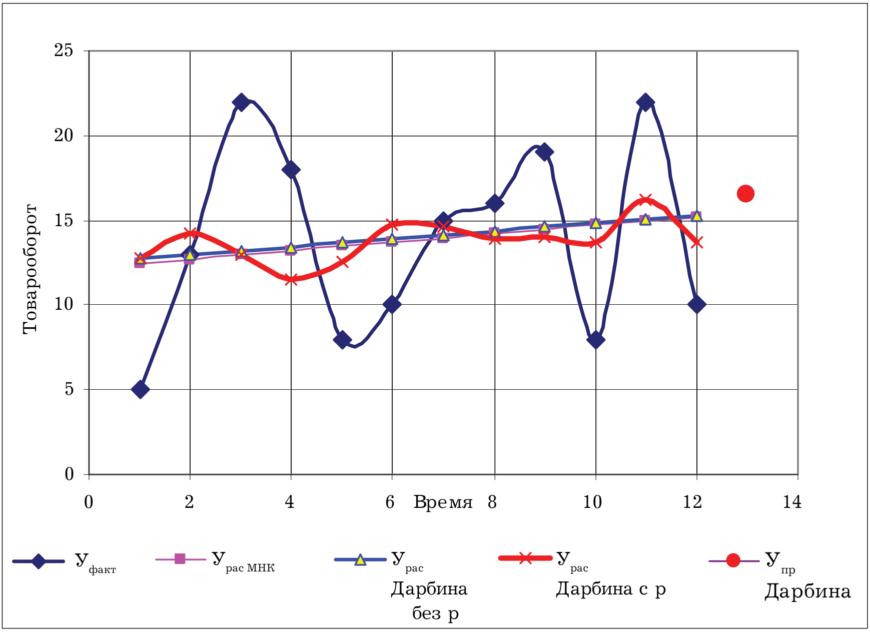


Рис. 10.15 Расчетные значения У без учета и с учетом автокорреляции остатков, определенные методом Дарбина 2

Метод Кочрена-Оркатто

Оценим параметры модели с помощью метода Кочрена Оркатто.

Метод Кочрена-Оркатта основан на методе Эйткена и выполняется в такой последовательности:

1. Вводится любое значение р, вычисляется значение p_1.

2. Если p_1 отличается от p в пределах заданной ошибки, то выполняется пункт первый и вместо p вводится значение p_1 и вычисляется новое значение p_1.

Данная процедура продолжается до тех пор, пока вводимое значение p и расчетное значение p_1 будут равны в пределах заданной ошибки

После первой итерации были получены характеристики модели, вычисленные методом Кочрена-Оркатто:

$b_0 = 12{,}494$, $b_1 = 0{,}2338$, $p_1 = -0{,}23$, $S_v = 6{,}1324$ — коэффициенты метода Кочрена-Оркатто и ошибка модели $У_t = b_0 + b_1X_t + p_1e_{t-1} + v_t$,

p = -0,2 — начальное значение коэффициента автокорреляции остатков

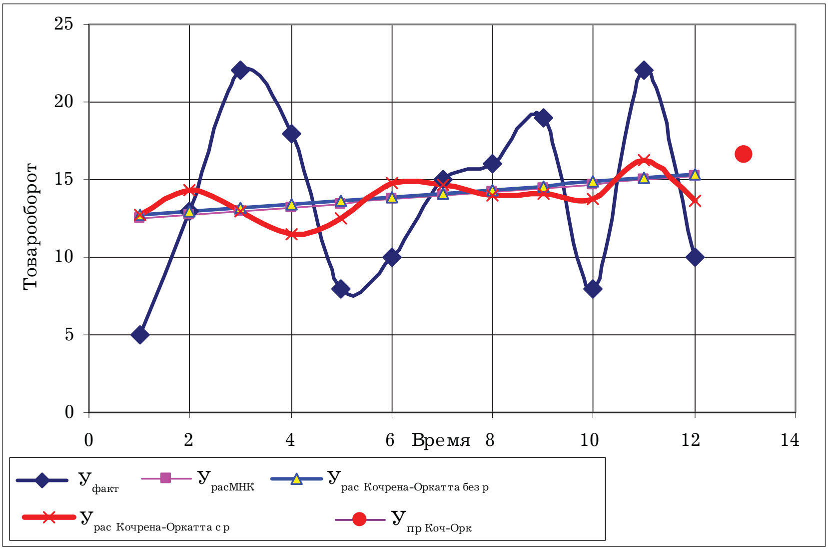


Рис. 10.16. Расчетные значения У без учета и с учетом автокорреляции остатков, определенные методом Кочрена-Оркатта

Метод Хилдрета-Лу

Оценим параметры модели с помощью метода Хилдрета-Лу, который основан на методе Эйткена и выполняется в такой последовательности:

1. Вводятся последовательные значения p из диапазона -1 до +1 с заданным шагом.

2. Для каждого значения p методом Эйткена вычисляются коэффициенты и ошибка модели S_v.

3. Определяется такое оптимальное значение p, при котором ошибка модели будет минимальной.

Ошибка модели получилась минимальной при значении p = -0,23.

Приводим характеристики модели, вычисленные методом Хилдрета-Лу:

$b_0 = 12{,}494$, $b_1 = 0{,}2338$, $p_1 = -0{,}23$, $S_v = 6{,}1324$ — коэффициенты метода Хилдрета-Лу и ошибка модели $У_t = b_0 + b_1X_t + p_1e_{t-1} + v_t$.

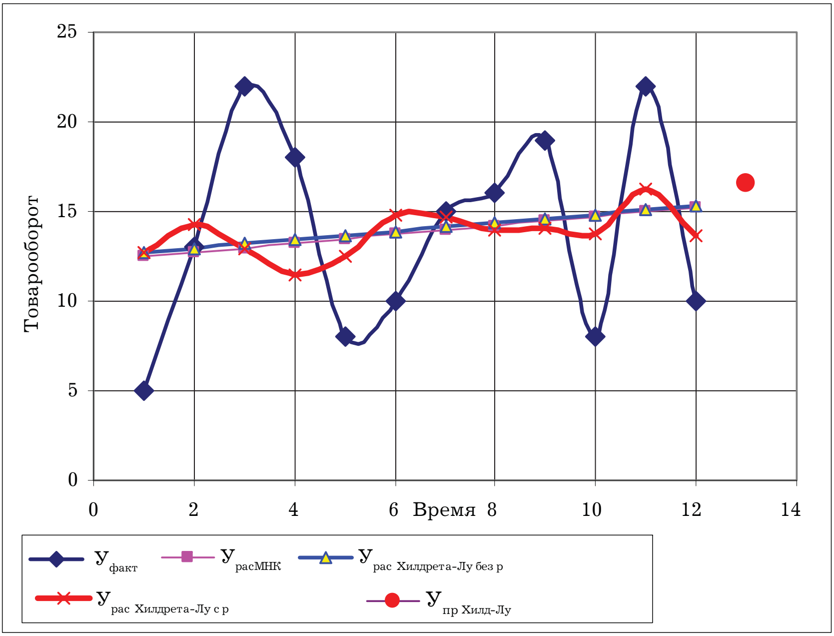


Рис. 10.17. Расчетные значения У без учета и с учетом автокорреляции остатков, определенные методом Хилдрета-Лу

Задания третьего уровня сложности

Задание 1. Создайте блок проверки автокорреляции остатков с помощью критерия Фон Неймана. (13 с. 248–249).

Пояснение. Критерий Фон Неймана.

Проверка достоверности наличия автокорреляции первого порядка осуществляется по критерию Неймана в следующей последовательности:

1. Вычисляется критерий

H = DWn/(n-1),

где DW — критерий Дарбина-Уотсона;

n — объем выборки;

H — критерий Фон Неймана.

Для больших значений n критерий H можно считать распределенным приблизительно по нормальному закону, для которого

$$M(H) = \frac{2n}{n-1}, \quad S^2(H) = \frac{4n^2(n-2)}{(n+1)(n-1)^3},$$

где M — знак математического ожидания;

S^2 — дисперсия критерия Фон Неймана.

2. Проводится проверка существования автокорреляции для большой выборки может быть проведена путем сопоставления эмпирического значения критерия Неймана с предварительно выбранным критическим интервалом для нормального закона распределения, обладающего соответствующими средним и дисперсией [13, с. 248–249).

Задание 2. Создать блок проверки достоверности автокорреляции остатков по критерию знаков.

Пояснение. С помощью критерия знаков проверяют нулевую гипотезу H_0:” Автокорреляция остатков первого порядка отсутствует”

в следующей последовательности:

— остатки e_i модели располагают во временной последовательности;
— выписывают их знаки;
— подсчитывают число положительных знаков n_1;

— подсчитывают число отрицательных знаков n_2;
— подсчитывают число групп остатков, имеющих одинаковый знак n_0;
— по таблицам определяют нижние $m_1(\alpha = 0{,}05; n_1, n_2)$ и верхние $m_2(\alpha = 0{,}05; n_1, n_2)$ критические значения числа групп остатков, имеющих одинаковый знак;
— если $n_0 < m_1$ или $n_0 > m_2$, то Но: $r(1) = 0$ отвергается с вероятностью $P = 1 - \alpha = 0{,}95$ и утверждается, что автокорреляция первого порядка является достоверной и остатки модели взаимозависимы;
— если $n_0 > m_1$ и $n_0 < m_2$, то Но: $r(1) = 0$ принимается и утверждается, что остатки не взаимозависимы[1].

Задание четвертого уровня сложности

Задание. Используйте метод Эйткена для устранения гетероскедастичности и автокорреляции остатков одновременно.

Выходной тест

Определите знак коэффициента автокорреляции остатков первого порядка по данным, представленным на рис. 10.18.

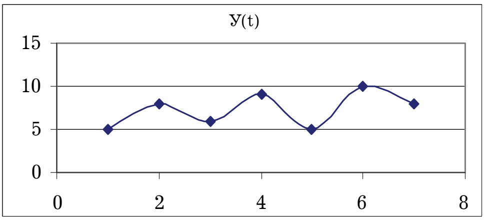


Рис. 10.18. Данные временного ряда

Нерешенная проблема

Если объясняемая переменная имеет постоянный шаг изменения, то метод Дарбина 1 не имеет решения. Предложите методы решения данной проблемы.

[1] Кейн Э. Экономическая статистика и эконометрия. — М.: Статистика, 1977. стр. 187.

11. МЕТОДЫ ПРЕОБРАЗОВАНИЯ НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ В СТАЦИОНАРНЫЕ. АНАЛИТИЧЕСКИЕ И АЛГОРИТМИЧЕСКИЕ МЕТОДЫ ВЫДЕЛЕНИЯ НЕСЛУЧАЙНОЙ СОСТАВЛЯЮЩЕЙ ВРЕМЕННОГО РЯДА

Постановка задачи

Объект — временные данные экономического показателя.

Предмет — регулярности экономического показателя.

Цель — изучить модели нестационарных временных рядов, предназначенных для повышения точности прогноза.

Актуальность — повышение точности прогноза временного ряда всегда высоко ценилось в эконометрическом моделировании.

Рабочая гипотеза — временной ряд имеет определенную структуру.

Моделями нестационарных временных рядов могут быть:

— модели периодических колебаний;

— модель авторегрессии временного ряда.

Методы:

— аналитический метод моделирования периодических колебаний;

— алгоритмические методы выделения неслучайной составляющей временного ряда:

— метод скользящего среднего;

— метод экспоненциально взвешенного скользящего среднего (метод Брауна)

— метод последовательных разностей;

— авторегрессии временного ряда.

Способ — для расчета характеристик модели можно использовать функцию Excel “Линейн”.

Задача — выделение всех составляющих временного ряда.

Ожидаемый результат — использование моделей нестационарных временных рядов должно повысить точность прогнозирования.

Метод для сравнения — парная линейная модель.

Задания первого уровня сложности

Задание 1. Подготовить устные сообщения по вопросам самостоятельной работы.

1. Приведите структуру временного ряда.

2. Приведите шаги реализации аналитического метода выделения неслучайной составляющей временного ряда.

2. Перечислите алгоритмические методы выделения неслучайной составляющей временного ряда.

3. Приведите алгоритм скользящего среднего.

4. Приведите алгоритм получения прогноза по экспоненциально взвешенной средней (метод Брауна).

5. Приведите алгоритм выделения неслучайной составляющей с помощью последовательных разниц. [3, с. 265–276, 1, Т. 2 с. 227–244, 287–291, 296–312].

Задание 2. Пройдите входное тестирование.

Выберите вариант правильного ответа

1. 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 экономи-ческих процессов в течение года;

f_{3t} — циклическая компонента, которая отражает повторяемость экономи-ческих процессов в течение длительных периодов, например влияние волн научно-технической революции, демографических спадов, циклов солнечной активности и т. д.;

ε_t — случайное возмущение.

а) да;

б) нет.

2. 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.

а) да;

б) нет.

3. Реализация аналитического метода выделения неслучайной составляющей временного ряда предполагает выполнение таких шагов.

Шаг 1. Воспроизведение тренда f_{1t} 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);

$$f_{1t} = a_0 + a_1 / t.$$

Обычно выбирается линейная функция, которая меньше всего искажает периодическую составляющую временного ряда.

Коэффициенты модели $У_t = f_{1t} + e_t$ определяются методом наименьших квадратов.

Шаг 2. Строится периодограмма в виде зависимости ошибки E модели

$$У_t = a_0 + a_1t + a_2Sin(2\pi t/T) + a_3Cos(2\pi t/T) + e_t,$$

от периода T,

где T — период, изменяющийся от 3 до удвоенной длины временного ряда.

Шаг 3. Определение на периодограмме локальных минимумов и соответствующих им периодов циклических составляющих.

Шаг 4. Воспроизведение составляющих временного ряда:

f_{1t} — тренда,

f_{2t} — сезонной компоненты,

f_{3t} — циклической компоненты

с помощью многофакторной модели

$$У_t = a_0 + a_1t + a_2Sin(2\pi t/T_1) + a_3Cos(2\pi t/T_1) + a_4Sin(2\pi t/T_2) + \\ + a_5Cos(2\pi t/T_2) + e_t,$$

где T_1, T_2 — периоды для сезонной и циклической составляющей временного ряда.

Шаг 5. Расчет коэффициентов и характеристик модели с помощью функции Excel “Линейн”.

Шаг 6. Получение точечных и интервальных прогнозов.

а) да;

б) нет.

4. Для построения доверительных интервалов регрессии и прогноза необходимо считать более точным не среднее значение временного интервала, а его последнее, более свежее значение.

а) да;

б) нет.

5. Авторегрессия временного ряда представляет собой модель зависимости текущих значений временного ряда от его же предыдущих значений. Модель авторегрессии первого порядка имеет вид

$$У_t = a_0 + a_1t + a_2У_{t-1} + e_t$$

Расчет коэффициентов модели производится с помощью функции Линейн. С помощью этой модели удается повысить точность прогнозирования на одну дату вперед, так как для прогноза известны ожидаемое значение времени и предыдущее значение временного ряда.

а) да;

б) нет.

6. Алгоритмические методы выделения неслучайной составляющей временного ряда могут быть следующих видов:

— метод скользящего среднего;
— метод экспоненциально взвешенного скользящего среднего (метод Брауна);
— метод последовательных разностей.

а) да;

б) нет.

7. 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 следующей 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
где a_0, a_1 — коэффициенты уравнения $У_t = a_0 + a_1 t$, определенные МНК по данным временного ряда $У_t$, входящим 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 $У_{n-2m}, У_{n-2m+1}, ..., У_n$.

а) да;

б) нет.

8. Прогноз временного ряда на дату t+1 равен экспоненциальной взвешенной средней (метод Брауна) на момент времени t, или

$$П_{t+1} = Z_t = \lambda У_t + (1 - \lambda)Z_{t-1} = \lambda У_t + (1 - \lambda)П_t,$$
$$П_{t+1} = П_t + \lambda(У_t - П_t),$$

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;

$П_1$ $П_t$, при t = 1, — первое прогнозное значение определяют экспертным способом, рассчитывают с помощью регрессионного анализа оно или равно первому значению временного ряда;

λ — коэффициент сглаживания.

а) да;

б) нет.

9. Выделение неслучайной составляющей с помощью последовательных разниц. 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 $\Delta(\Delta У_t) = \Delta У_t — \Delta У_{t-1}$.

После вычисления последовательных разностей их анализируют на стационарность. Если тест на стационарность выполняется, то расчеты заканчиваются, если нет, то вычисляются последовательные разности более высокого порядка.

а) да;

б) нет.

Задание 3. Решите задачу 1.

Имеются показатели деятельности хлебокомбината, представленные в табл. 11.1.

Таблица 11.1

База данных показателей деятельности хлебокомбината в действующих ценах (копия табл. 1.3)

t	X_{1t}	X_{2t}	X_{3t}	X_{4t}	$У_t$
1	65,2	2,9	34	68,26	24549
2	63,02	3,47	34	70,17	25045
3	40,67	1,62	35	81,51	36090
4	45,75	1,58	35	90,96	36478
5	69,69	4,22	35	64,1	25386
6	59,85	1,83	36	68,65	35992
7	59,9	3,15	36	70,93	35867
8	52,72	5,17	36	77,88	36143
9	53,52	3,24	36	78,66	35227
10	66,23	7,89	36	90,22	23248
11	48,98	6,23	36	75,89	27346
12	64,99	3,69	34	65,16	19444
13	53,63	2,41	34	67,03	18640
14	45,93	3,14	34	68,01	17661
15	41,82	3,43	34	69,51	19032
16	34,72	1,97	34	88,35	25164
17	42,28	2,63	35	81,68	18998
18	38,69	2,32	36	73,24	17035
19	50,74	3,02	36	65,12	18166
20	38,12	3,01	36	81,52	23204
21	35,75	2,55	36	78,66	21117
22	19,16	1,74	36	80,81	20995
23	35,08	3,51	35	80,06	17863
24	30,99	3,37	35	77,95	17525

где t — время (месяцы) за два условных года;

X_1 — удельный вес сырья в общей себестоимости продукции (%);

X_2 — удельный вес суммы заработной платы производственных рабочих в общей себестоимости продукции (%);

X_3 — количество рабочих (чел.);

X_4 — уровень затрат (%);

У — себестоимость продукции (руб.).

Необходимо выделить неслучайную составляющую аналитическими методами. Получить точечный и интервальный прогноз для переменной X_4 — уровень затрат (%).

Решение

Решение задачи будем проводить в такой последовательности:

1. Построим график динамики переменной X_4.
2. Визуальным методом определим имеющиеся тенденции.
3. Построим многофакторные модели для реализации аналитического метода выделения неслучайной составляющей:
   - модель периодических составляющих;
   - модель авторегрессии временного ряда.
4. Получим прогноз.
5. Реализуем алгоритмические методы выделения неслучайной составляющей временного ряда, которые могут быть следующих видов:
   - метод скользящего среднего;
   - метод экспоненциально взвешенного скользящего среднего (метод Брауна);
   - метод последовательных разностей.

1. Построим график динамики переменной X_4 (рис. 11.1).
2. Визуальным методом определим имеющиеся тенденции.

Анализ рис. 11.1 показывает, что динамика показателя X_4 имеет четко выраженную линейную снижающуюся тенденцию, на которую наложена периодическая сезонная волна с периодом в 6 месяцев.

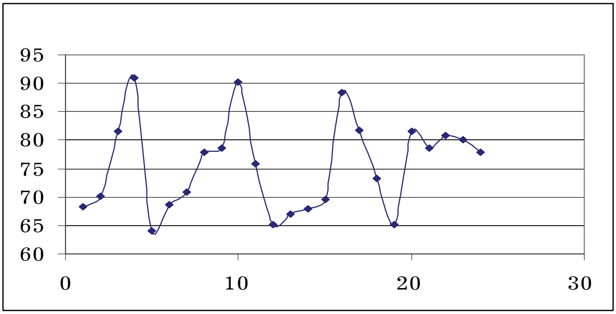


Рис. 11.1. Динамика показателя X_4

3. Построим многофакторную модель для реализации аналитического метода выделения неслучайной составляющей.

Модель периодических состовляющих.

Вычислим периодограмму в виде зависимости ошибки Е модели

$У_t = a_0 + a_1t + a_2Sin(2\pi t/T) + a_3Cos(2\pi t/T) + e_t$, от периода Т, где Т — период, изменяющийся от 3 до удвоенной длины временного ряда;

t — время (1, ..., 24 номера месяцев).

Все расчеты представим в табл. 11.2.

Таблица 11.2

Исходные и расчетные переменные

t	Sin(6,28t/T)	Cos(6.28t/T)	$У_t = X_{4t}$	$У_{pi}$	T	E
1	0,866	0,500	68,260	66,447	3,000	8,263
2	0,867	-0,499	70,170	71,816	4,000	8,260
3	0,002	-1,000	81,510	79,543	5,000	8,435
4	-0,865	-0,502	90,960	82,044	6,000	5,968
5	-0,867	0,498	64,100	76,954	7,000	6,845
6	-0,003	1,000	68,650	69,494	8,000	8,060
7	0,864	0,503	70,930	67,251	9,000	8,388
8	0,868	-0,496	77,880	72,601	10,000	8,435
9	0,005	-1,000	78,660	80,334	11,000	8,426
10	-0,863	-0,505	90,220	82,858	12,000	8,394
11	-0,869	0,495	75,890	77,787	13,000	8,353

t	Sin(6,28t/T)	Cos(6.28t/T)	$У_t = X_{4t}$	$У_{pi}$	T	E
12	-0,006	1,000	65,160	70,322	14,000	8,314
13	0,863	0,506	67,030	68,056	15,000	8,289
14	0,870	-0,494	68,010	73,387	16,000	8,280
15	0,008	-1,000	69,510	81,125	17,000	8,280
16	-0,862	-0,507	88,350	83,673	18,000	8,285
17	-0,871	0,492	81,680	78,620	19,000	8,291
18	-0,010	1,000	73,240	71,150	20,000	8,297
19	0,861	0,509	65,120	68,860	21,000	8,303
20	0,871	-0,491	81,520	74,173	22,000	8,308
21	0,011	-1,000	78,660	81,916	30,000	8,329
22	-0,860	-0,510	80,810	84,487	40,000	8,338
23	-0,872	0,489	80,060	79,453	50,000	8,341
24	-0,013	1,000	77,950	71,978	60,000	8,343

Характеристики модели

$$У_t = a_0 + a_1t + a_2\mathrm{Sin}(2\pi t/T) + a_3\mathrm{Cos}(2\pi t/T) + e_t$$

представлены в протоколе расчетов, выполненных с помощью функции Линейн, при T = 6 (рис. 11.2).

-5,24047	-5,742830204	0,134902116	73,90687974
1,732486	1,750950562	0,179809933	2,556618533
0,516587	5,968885542	#Н/Д	#Н/Д
7,124158	20	#Н/Д	#Н/Д
761,4498	712,5518922	#Н/Д	#Н/Д

Периодограмма зависимости E (ошибки модели) от T (периода) представлена на рис. 11.2.

Анализ периодограммы показывает наличие двух локальных минимумов ошибок модели при периоде, равном 6 и 16 месяцев. Локальный минимум E при периоде, равном 6 месяцев, очень четко выражен и подтверждает результаты визуального анализа. Локальный минимум E при периоде 16 месяцев выражен очень слабо и не имеет очевидного объяснения.

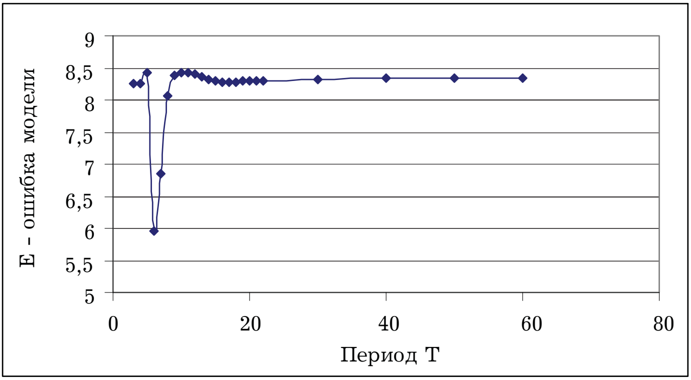


Рис. 11.2. Периодограмма. Зависимость Е от Т

Приводим протокол расчетов при Т=16.

-1,78532	1,561588723	0,168994017	73,08373023
2,408985	2,503113017	0,245622489	3,545464255
0,069714	8,280226913	#Н/Д	#Н/Д
0,499588	20	#Н/Д	#Н/Д
102,7585	1371,243155	#Н/Д	#Н/Д

Анализ протокола расчетов при Т = 16 показывает, что по критерию Фишера модель является недостоверной, по критерию Стьюдента все факторы не оказывают достоверного влияния на зависимую переменную. Следовательно, периодическую составляющую с периодом Т = 16 исключаем из модели.

Модель временного ряда должна включать линейную тенденцию и периодическую составляющую с периодом Т = 6 месяцев. По критерию Фишера модель является достоверной, по критерию Стьюдента все коэффициенты достоверно отличаются от нуля. График фактических и расчетных значений представлен на рис. 11.3.

Вывод. С помощью аналитического метода была выявлена сезонная периодическая составляющая с периодом Т = 6 месяцев и построена достоверная многофакторная модель, которая достоверно описывает регулярности временно ряда.

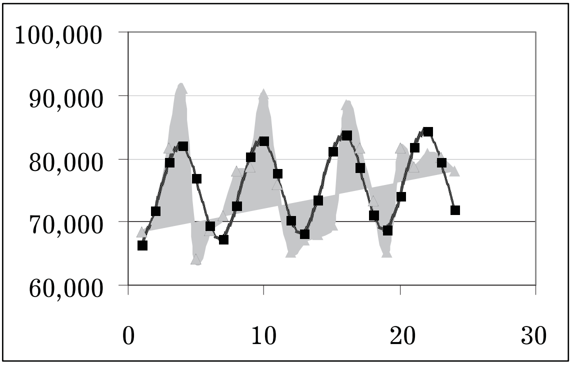


Рис. 11.3. Фактические (Δ) и расчетные ( ) значения X_4 , вычисленные по модели $У_{pt} = X_{4t} = a_0 + a_1 t + a_2 Sin(2\pi t/T) + a_3 Cos(2\pi t/T) =$ $= 73{,}906+0{,}1349\ t — 5{,}7428\ Sin(2\pi t/6) — 5{,}2404\ Cos(2\pi t/6)$

Получим прогноз

Точечный прогноз при t = 25, будет равен

$$У_{пр} = 73{,}906+0{,}1349 \times 25 — 5{,}7428 Sin(2\pi 25/6) — 5{,}2404 Cos(2\pi 25/6) = 69{,}664$$

Фактическое значение $У_ф$ с вероятностью 1 — α = 0,05 будет находиться в интервале

$$У_ф \in У_{пр} \pm t_{\alpha/2}\ (m = n — k = 24 — 4 = 20)\ E = 69{,}664 \pm \pm 2{,}085 \times 5{,}968 = 69{,}664 \pm 12{,}443,$$

где E — ошибка модели взята из протокола расчетов при T= 6

$$t_{\alpha/2}\ (m = n — k = 24 — 4 = 20) = = СТЬЮДРАСПОБР(0{,}05;20) = 2{,}085.$$

Модель авторегрессии временного ряда

Модели авторегрессии временного ряда первого порядка хорошо воспроизводят плавно изменяющиеся данные, но плохо воспроизводит выбросы, так как выбросы сильно влияют на последующие расчетные значения временного ряда. Эти закономерности можно проверить самостоятельно.

Однако, если выбросы происходят с определенной периодичностью, то можно получить достоверную модель авторегрессии и значительно улучшить точность прогноза. Реализуем эту модель на исходных данных временного ряда X_{4t}, содержащего периодическую составляющую с периодом T = 6 месяцев в такой последовательности:

а) создадим базу объясняемых факторов, состоящих из t лаговых факторов зависимой переменной от первого до шестого порядков (табл. 11.3);

б) вычислим характеристики многофакторной модели с помощью функции "Линейн", проверим достоверность коэффициентов модели и определим, какие факторы достоверно влияют на зависимую переменную;

в) построим модель, содержащую только достоверно влияющие факторы;

г) получим точечный и интервальные прогнозы.

Таблица 11.3

База данных фактических и лаговых факторов

t	$У_{t-1}$	$У_{t-2}$	$У_{t-3}$	$У_{t-4}$	$У_{t-5}$	$У_{t-6}$	$У_t = X_{4t}$
1							68,26
2	68,26						70,17
3	70,17	68,26					81,51
4	81,51	70,17	68,26				90,96
5	90,96	81,51	70,17	68,26			64,1
6	64,1	90,96	81,51	70,17	68,26		68,65
7	68,65	64,1	90,96	81,51	70,17	68,26	70,93
8	70,93	68,65	64,1	90,96	81,51	70,17	77,88
9	77,88	70,93	68,65	64,1	90,96	81,51	78,66
10	78,66	77,88	70,93	68,65	64,1	90,96	90,22
11	90,22	78,66	77,88	70,93	68,65	64,1	75,89
12	75,89	90,22	78,66	77,88	70,93	68,65	65,16
13	65,16	75,89	90,22	78,66	77,88	70,93	67,03
14	67,03	65,16	75,89	90,22	78,66	77,88	68,01
15	68,01	67,03	65,16	75,89	90,22	78,66	69,51
16	69,51	68,01	67,03	65,16	75,89	90,22	88,35
17	88,35	69,51	68,01	67,03	65,16	75,89	81,68
18	81,68	88,35	69,51	68,01	67,03	65,16	73,24
19	73,24	81,68	88,35	69,51	68,01	67,03	65,12

t	$У_{t-1}$	$У_{t-2}$	$У_{t-3}$	$У_{t-4}$	$У_{t-5}$	$У_{t-6}$	$У_t = X_{4t}$
20	65,12	73,24	81,68	88,35	69,51	68,01	81,52
21	81,52	65,12	73,24	81,68	88,35	69,51	78,66
22	78,66	81,52	65,12	73,24	81,68	88,35	80,81
23	80,81	78,66	81,52	65,12	73,24	81,68	80,06
24	80,06	80,81	78,66	81,52	65,12	73,24	77,95
	77,95	80,06	80,81	78,66	81,52	65,12	
		77,95	80,06	80,81	78,66	81,52	
			77,95	80,06	80,81	78,66	
				77,95	80,06	80,81	
					77,95	80,06	
						77,95	

С помощью функции “Линейн” рассчитаем характеристики модели

$У_t = X_{4t} = a_0 + a_1 t + a_2 У_{t-1} + a_3 У_{t-2} + a_4 У_{t-3} + a_5 У_{t-4} + a_6 У_{t-5} + a_7 У_{t-6} + e_t$

по данным табл. 11.3, где первые и последние шесть строчек исключим из анализа, так они содержат пропуски значений.

Приводим протокол расчетов.

0,4452	-0,364825	0,0251	-0,258	-0,332	0,2357	0,153	92,239293
0,234	0,2035057	0,2301	0,2243	0,2199	0,2715	0,2762	69,800365
0,6599	5,6352927	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
2,7722	10	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
616,24	317,56524	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д

Анализ расчетов показывает, что по критерию Стьюдента только фактор $У_{t-6}$ оказывает достоверное влияние на $У_t = X_{4t}$, что подтверждает визуальный анализ и модель периодических колебаний.

Оставляем в модели факторы t и $У_{t-6}$.

С помощью функции “Линейн” рассчитаем характеристики модели

$$У_t = X_{4t} = a_0 + a_1 t + a_2 У_{t-6} + e_t$$

Приводим протокол расчетов

0,5257	0,1401769	34,54
0,1737	0,2801672	13,514
0,3922	6,151332	#Н/Д
4,8393	15	#Н/Д
366,22	567,58329	#Н/Д

Анализ расчетов показывает, что по критерию Фишера модель является достоверной, по критерию Стьюдента время t не оказывает достоверного влияния, лаговый фактор $У_{t-6}$ оказывает достоверное влияние на зависимую переменную.

Результирующая авторегрессионная модель имеет такой вид

$$У_t = X_{4t} = a_0 + a_1 t + a_2 У_{t-6} + e_t = 34{,}54 + 0{,}1401t + 0{,}5257У_{t-6} + e_t$$

Для получения прогноза на шесть дат вперед воспользуемся данными, представленными в табл. 11.4.

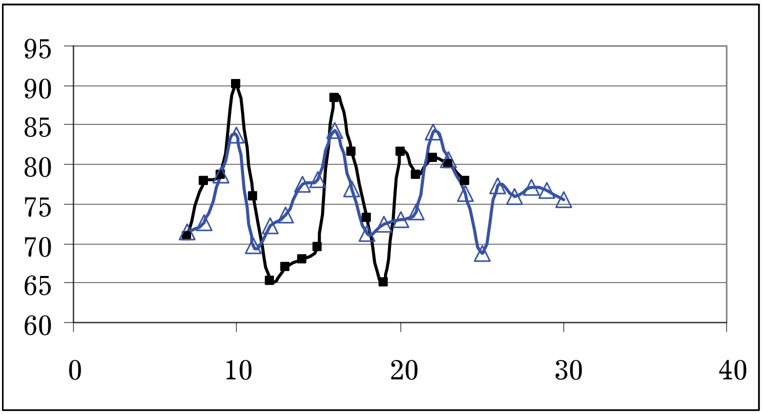


Рис. 11.4. Фактические ( ) и расчетные (Δ) и прогнозные (Δ) значения временного ряда $У_t = X_{4t}$

Анализ рис. 11.4 показывает, что шесть прогнозных значений повторяют предыдущую тенденцию временного ряда, что и следовало ожидать от авторегрессионной модели.

Таблица 11.4

Ожидаемые значение факторов и прогноз зависимой переменной

t	$У_{t-6}$	$У_{прt}$
25	65,12	68,776
26	81,52	77,398
27	78,66	75,894
28	80,81	77,025
29	80,06	76,63
30	77,95	75,521

Задание второго уровня сложности

Задание. Для временного ряда X_4 — уровень затрат (%) (см. задачу 1) реализуйте алгоритмические методы выделения неслучайной составляющей временного ряда, которые могут быть следующих видов:

— метод скользящего среднего;

— метод экспоненциально взвешенного скользящего среднего (метод Брауна);

— метод последовательных разностей.

Метод скользящего среднего

Установим длину усреднения m = 3. Расчеты представлены в табл. 11.5.

Расчетные значения X_4, вычисленные по методу скользящего среднего, при m = 3

Таблица 11.5

t	X_{4t}	Среднее при m = 3	t	X_{4t}	Среднее при m = 3
1	68,26		13	67,03	66,73333333
2	70,17	73,31333333	14	68,01	68,18333333
3	81,51	80,88	15	69,51	75,29
4	90,96	78,85666667	16	88,35	79,84666667
5	64,1	74,57	17	81,68	81,09
6	68,65	67,89333333	18	73,24	73,34666667
7	70,93	72,48666667	19	65,12	73,29333333
8	77,88	75,82333333	20	81,52	75,1
9	78,66	82,25333333	21	78,66	80,33
10	90,22	81,59	22	80,81	79,84333333
11	75,89	77,09	23	80,06	79,60666667
12	65,16	69,36	24	77,95	

На рис. 11.5 построен график фактических и расчетных значений

Рис. 11.5. Фактические и расчетные значения X_{4t}, определенные методом скользящей средней при m = 3

Анализ рис. 11.5 показывает, что метод скользящей средней хорошо воспроизводит регулярности временного ряда, но нет экономической интерпретации полученных расчетов.

Метод экспоненциально взвешенного скользящего среднего (метод Брауна)

онреализуется по формуле

$$П_{t+1} = П_t + \lambda(У_t — П_t),$$

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;

$П_t = П_1$, при t = 1, — первое прогнозное значение определяют экспертным способом, рассчитывают с помощью регрессионного анализа или оно равно первому значению временного ряда;

λ — коэффициент сглаживания.

Результаты расчетов представлены в табл. 11.6.

Таблица 11.6

Результаты расчетов по методу экспоненциального взвешенного скользящего среднего (метод Брауна)

t	X_{4t}	$П_{(t+1)}$	e_t^2
1	68,26	68,26	0
2	70,17	68,26	3,6481
3	81,51	68,642	165,59
4	90,96	71,216	389,84
5	64,1	75,164	122,42
6	68,65	72,952	18,504
7	70,93	72,091	1,3485
8	77,88	71,859	36,252
9	78,66	73,063	31,324
10	90,22	74,183	257,2
11	75,89	77,39	2,2502
12	65,16	77,09	142,33
13	67,03	74,704	58,891
14	68,01	73,169	26,618

t	X_{4t}	$П_{(t+1)}$	e_t^2
15	69,51	72,137	6,9031
16	88,35	71,612	280,16
17	81,68	74,96	45,165
18	73,24	76,304	9,3858
19	65,12	75,691	111,74
20	81,52	73,577	63,096
21	78,66	75,165	12,212
22	80,81	75,864	24,46
23	80,06	76,853	10,282
24	77,95	77,495	0,2073
Сумма			1819,8

E =	9,095
λ =	0,2

На рис. 11.6 построен график фактических и расчетных значений

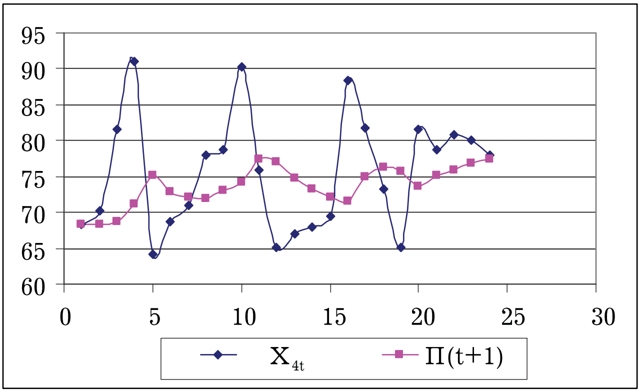


Рис. 11.6. Фактические и расчетные значения X_4, вычисленные по методу экспоненциального взвешенного скользящего среднего, при $\lambda = 0,2$

Примечание. Коэффициент сглаживания λ выбирается оптимальным, обеспечивающим минимальную ошибку модели

$$E = \sqrt{\frac{\sum e_t^2}{n-2}}.$$

Для этой цели изменяйте значение λ от 0 до 1 с шагом 0,1 и наблюдайте за ошибкой модели. При оптимальном значении λ ошибка модели будет минимальной.

Для данного примера оптимальное значение λ получилось равным 0,2.

Вывод. Метод экспоненциального взвешенного скользящего среднего плохо воспроизводит выбросы и резко изменяющиеся значения.

Метод последовательных разностей

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 $\Delta(\Delta У_t) = \Delta У_t — \Delta У_{t-1}$.

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.

Результаты расчетов представлены в табл. 11.7.

Таблица 11.7

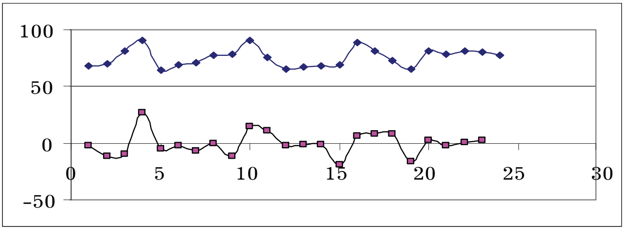
Расчеты первых (ΔУ), вторых (ΔΔУ) и третьих (ΔΔΔУ) разностей

t	X_{4t}	ΔУ	ΔΔУ	ΔΔΔУ
1	68,26	-1,91	9,43	11,32
2	70,17	-11,34	-1,89	34,42
3	81,51	-9,45	-36,31	-67,72
4	90,96	26,86	31,41	33,68
5	64,1	-4,55	-2,27	-6,94
6	68,65	-2,28	4,67	10,84
7	70,93	-6,95	-6,17	-16,95
8	77,88	-0,78	10,78	36,67

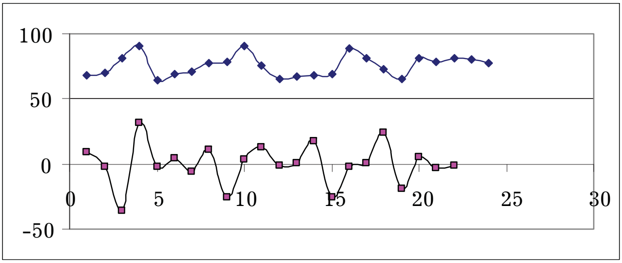
t	X_{4t}	ΔУ	ΔΔУ	ΔΔΔУ
9	78,66	-11,56	-25,89	-29,49
10	90,22	14,33	3,6	-9
11	75,89	10,73	12,6	13,49
12	65,16	-1,87	-0,89	-1,41
13	67,03	-0,98	0,52	-16,82
14	68,01	-1,5	17,34	42,85
15	69,51	-18,84	-25,51	-23,74
16	88,35	6,67	-1,77	-2,09

t	X_{4t}	ΔУ	ΔΔУ	ΔΔΔУ
17	81,68	8,44	0,32	-24,2
18	73,24	8,12	24,52	43,78
19	65,12	-16,4	-19,26	-24,27
20	81,52	2,86	5,01	7,91
21	78,66	-2,15	-2,9	-1,54
22	80,81	0,75	-1,36	
23	80,06	2,11		
24	77,95			

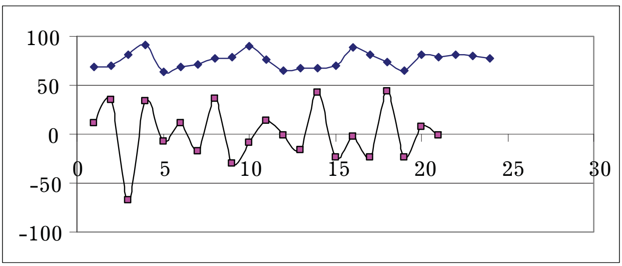
На рис. 11.7 построен график фактических и расчетных значений и первых, вторых и третьих разностей



а) Фактические значения и разности первого порядка



б) Фактические значения и разности второго порядка



в) Фактические значения и разности третьего порядка

Рис. 11.7 Фактические значения и разности первого, второго и третьего порядков

Анализ рис. 11.7. показывает, что с увеличением порядка разниц уменьшается количество восходящих и нисходящих серий. Можно считать, что вторые разности являются стационарными.

Примечание. Проверьте это утверждение с помощью критерия восходящих и нисходящих серий.

Критерий восходящий и нисходящих серий служит для проверки наличия во временном ряду $У_t$ неслучайной составляющей или проверки гипотезы о том, что

$$H_o: \text{“}МУ_t = const\text{”}.$$

Если нарушится хотя бы одно из двух неравенств:

$$\begin{cases} K_{об}(n) > \left[\frac{1}{3}(2n-1) - 1{,}96\sqrt{\frac{16n-29}{90}}\right] \\ K_{дл}(n) < K_o(n), \end{cases}$$

то нулевая гипотеза отвергается на уровне значимости α от 0,05 до 0,097 и утверждается, что временной ряд $У_t$ содержит неслучайную составляющую, зависящую от времени t,

где $K_o(n \le 26) = 5$;

$K_o(26 < n \le 153) = 6$;

$K_o(153 < n \le 1170) = 7$;

$K_{об}(n)$ — общее число серий идущих подряд положительных и отрицательных первых разностей временного ряда $У_t$;

$K_{дл}(n)$ — количество идущих подряд положительных или отрицательных значений первых разниц временного ряда $У_t$ в самой длинной серии;

n — число уровней временного ряда $У_t$.

Задание третьего уровня сложности

Задние. Реализуйте метод скользящего среднего в следующей последовательности:

Шаг 1. Выбирается нечетная длина 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 линейной функции

$$Z_t = a_0 + a_1 t,\ t = m + 1, m + 2, \ldots, n - m,$$

где a_0, a_1 — коэффициенты уравнения $У_t = a_0 + a_1 t$, определенные МНК (функция Линейн) по данным временного ряда $У_t$, входящих в длину усреднения.

Шаг 3. Вычисляются сглаженные значения в краевых точках по следующей формуле:

$$Z_t = a_0 + a_1 t,\ t = 1, 2, \ldots, m,$$

где a_0, a_1 — коэффициенты уравнения $У_t = a_0 + a_1 t$, определенные МНК по данным временного ряда $У_1, У_2, \ldots, У_{2m+1}$.

$$Z_t = a_0 + a_1 t,\ t = n — m + 1, \ldots, n,$$

где a_0, a_1 — коэффициенты уравнения $У_t = a_0 + a_1 t$, определенные МНК по данным временного ряда $У_{n-2m}, У_{n-2m+1}, \ldots, У_n$.

Задание четвертого уровня сложности

Задние. Реализуйте аналитический метод выделения неслучайной составляющей временного ряда, взятого из статистических сборников, журналов и др. источников информации, посвященных вашей профессиональной деятельности.

Выходное тестирование

Вычислите точечный прогноз по модели

$У_{pt} = a_0 + a_1 t + a_2 \mathrm{Sin}(2\pi t/T_1) + a_3 \mathrm{Cos}(2\pi t/T_1) + a_4 \mathrm{Sin}(2\pi t/T_2) + a_5 \mathrm{Cos}(2\pi t/T_2)$,

где $a_0 = 0{,}2$; $a_1 = 0{,}21$; $a_2 = 0{,}35$; $a_3 = 0{,}4$; $a_4 = 0{,}1$; $a_5 = -0{,}5$; $T_1 = 6$; $T_2 = 30$; $t = 25$.

Нерешенная проблема

Нет методики определения разрешающей способности выделения по периодограмме (зависимость E от T) двух периодических составляющих с близкими периодами.

12. МОДЕЛИ СТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ И ИХ ИДЕНТИФИКАЦИЯ

Постановка задачи

Объект — временные данные экономического показателя.

Предмет — регулярности экономического показателя.

Цель — изучить свойства временного ряда для повышения точности прогноза.

Актуальность — повышение точности прогноза временного ряда всегда высоко ценилось в эконометрическом моделировании.

Рабочая гипотеза — временной ряд имеет определенную структуру.

Модель — скользящей средней остатков.

Способ — для расчета характеристик модели можно использовать функцию Excel "Линейн".

Задача — оценить коэффициенты модели скользящей средней.

Ожидаемый результат — использование моделей остатков должно повысить точность прогнозирования.

Метод для сравнения — прогнозы, полученные с использование и без использования модели скользящей средней остатков.

Задания первого уровня сложности

Задание 1. Подготовить устные сообщения по вопросам самостоятельной работы.

1. Приведите общий вид модели скользящей средней остатков порядка q [3, с. 72–79, 1, т. 2 с. 246–261].

Задание 2. Пройдите входное тестирование.

Выберите вариант правильного ответа.

1. Допустим, имеется временной ряд экономического показателя $У_t$, который обладает следующими свойствами:

а) Зависимость У от t можно представить в виде следующей модели:

$$У_t = \alpha_0 + \alpha_1 t + \varepsilon_t,$$

где Уt — зависимая переменная, имеющая ограничение на вид тенденции;

α_0, α_1 — параметры модели;

$\varepsilon_t = \rho\varepsilon_{t-1} + \delta_t$ — модель авторегрессии остатков первого порядка;

ρ — коэффициент авторегрессии остатков первого порядка;

ε_t — слабо стационарный временной ряд со средним значением, равным нулю, при наличии тенденции, напоминающей гиперболу;

δ_t — строго стационарный временной ряд (белый шум) со средним значением, равным нулю, без наличия автоковариации при постоянной дисперсии, не зависящей от t.

$M\delta_t = 0$,

0 при $d \neq 0$,

$M(\delta_t \delta_{t+d}) =$

S^2 при $d = 0$.

б) ε_t и δ_t связаны следующим соотношением:

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

где $\varepsilon_t = \rho\varepsilon_{t-1} + \delta_t$ — модель авторегрессии первого порядка АР(1);

$\varepsilon_t = \delta_t + \rho\delta_{t-1} + \rho^2\delta_{t-2} + \rho^3\delta_{t-3} + \ldots + \rho^\infty\delta_{t-\infty}$ — модель скользящей средней остатков бесконечного порядка, которая в природе не существует и заменяется на модель скользящей средней остатков конечного порядка, например, второго порядка

$\varepsilon_t = \delta_t + \rho\delta_{t-1} + \rho^2\delta_{t-2} + \upsilon_t$ — модель скользящей средней остатков второго порядка СС(2).

а) да;

б) нет.

2. Белый шум δ_t можно получить с помощью генератора случайных чисел, функции Excel — СЛЧИС.

а) да;

б) нет.

3. Параметры модели

$$У_t = \alpha_0 + \alpha_1 t + \varepsilon_t = \alpha_0 + \alpha_1 t + \delta_t + \rho\delta_{t-1} + \rho^2\delta_{t-2} + \upsilon_t$$

можно оценить с помощью коэффициентов модели для выборочной совокупности

$$У_t = a_0 + a_1 t + e_t = a_0 + a_1 t + 1\delta_t + p\delta_{t-1} + p^2\delta_{t-2} + v_t,$$

которые можно вычислить с помощью программы Excel “Поиск решения”, при заданных ограничениях на коэффициенты и минимизации суммы квадратов остатков v_t.

а) да;

б) нет.

Задание 2. Решите задачу 1.

Для реализации модели скользящей средней в ее классическом виде необходимо, чтобы временной ряд обладал определенными свойствами, которые изложены во входном тестировании.

Для экспериментов в качестве зависимой переменной был взят показатель X_{4t} (табл. 1.3) с целью продолжения расчетов, начатых в теме 11.3 задание 3. К сожалению, существенно улучшить модель не удалось, так как временной ряд X_{4t} не обладает необходимыми свойствами. Расчеты проводились по методике, изложенной во входном тестировании.

Поэтому упростим задачу. Необходимо получить стационарные остатки (без заметных регулярностей), которые требуется воспроизвести с помощью белого шума, а также улучшенное прогнозное значение зависимой переменной с учетом белого шума.

Решение

Воспользуемся базой данных показателей деятельности хлебокомбината (табл. 1.3) и составили табл. 12.1.

Таблица 12.1

База данных показателей деятельности хлебокомбината (в действующих ценах) за два условных года.

t	Уt	t	Уt
1	65,2	13	53,63
2	63,02	14	45,93
3	40,67	15	41,82
4	45,75	16	34,72
5	69,69	17	42,28
6	59,85	18	38,69
7	59,9	19	50,74
8	52,72	20	38,12
9	53,52	21	35,75
10	66,23	22	19,16
11	48,98	23	35,08
12	64,99	24	30,99

где t — время (месяцы) за два условных года;

X_{1t} — удельный вес сырья в общей себестоимости продукции (%);

Решение задачи выполним в такой последовательности:

а) построим график временного ряда и визуально определим имеющиеся регулярности;

б) получим стационарные остатки, которые воспроизведем лаговыми факторами белого шума второго порядка;

в) уточним прогноз с учетом лаговых факторов белого шума.

В табл. 12.2 представлен протокол расчетов.

Таблица 12.2

Протокол расчетов

t	δ_t	δ_{t-1}	δ_{t-2}	$У_t$	$У_{pt}$	e_t	$У_{1pt}$
1	0,1571			65,2	63,955	1,245	
2	0,3195	0,1571		63,02	62,587	0,433	
3	0,9761	0,3195	0,1571	40,67	61,219	-20,549	56,617
4	0,0997	0,9761	0,3195	45,75	59,852	-14,102	56,372
5	0,1962	0,0997	0,9761	69,69	58,484	11,206	67,973
6	0,1176	0,1962	0,0997	59,85	57,116	2,734	57,041
7	0,1319	0,1176	0,1962	59,9	55,749	4,151	57,206
8	0,9035	0,1319	0,1176	52,72	54,381	-1,661	51,115
9	0,6647	0,9035	0,1319	53,52	53,013	0,507	45,324
10	0,9371	0,6647	0,9035	66,23	51,645	14,585	52,524
11	0,7275	0,9371	0,6647	48,98	50,278	-1,298	47,642
12	0,1008	0,7275	0,9371	64,99	48,910	16,080	53,781
13	0,0085	0,1008	0,7275	53,63	47,542	6,088	55,326
14	0,7384	0,0085	0,1008	45,93	46,175	-0,245	44,461
15	0,8123	0,7384	0,0085	41,82	44,807	-2,987	36,324
16	0,3313	0,8123	0,7384	34,72	43,439	-8,719	44,454
17	0,6421	0,3313	0,8123	42,28	42,072	0,208	45,922
18	0,3367	0,6421	0,3313	38,69	40,704	-2,014	38,671
19	0,0297	0,3367	0,6421	50,74	39,336	11,404	44,359
20	0,717	0,0297	0,3367	38,12	37,969	0,151	38,688
21	0,6751	0,717	0,0297	35,75	36,601	-0,851	29,173
22	0,6048	0,6751	0,7170	19,16	35,233	-16,073	35,702
23	0,4936	0,6048	0,6751	35,08	33,866	1,214	34,963
24	0,1087	0,4936	0,6048	30,99	32,498	-1,508	35,572

где δ_t — белый шум, полученный с помощью функции Excel “СЛЧИС”;

δ_{t-1} — лаговый белый шум первого порядка;

δ_{t-2} — лаговый белый шум второго порядка;

$У_{pt} = 65,322 - 1,368t$ — коэффициенты вычислены с помощью функции “Линейн” (рис. 12.3).

-1,368	65,322
0,2644	3,7774
0,5489	8,965
26,765	22
2151,1	1768,2

10,533	-7,448	-4,909001	0,9135
5,9883	5,8404	5,95929	5,27059
0,2683	8,4737	#Н/Д	#Н/Д
2,1998	18	#Н/Д	#Н/Д
473,85	1292,5	#Н/Д	#Н/Д

$e_t = У_t — У_{pt}$ — остатки модели.

$e_t = 0{,}9135 - 4{,}909001\ \delta_t\ -7{,}448\ \delta_{t-1} + 10{,}533\ \delta_{t-2}$ — коэффициенты вычислены с помощью функции “Линейн” см. рис. 12.4 (размер ряда уменьшен на две даты);

$$У_{1pt} = 65{,}322\ -1{,}368t + 0{,}9135\ -4{,}909001\ \delta_t — 7{,}448\ \delta_{t-1} + 10{,}533\ \delta_{t-2}.$$

Результаты расчетов представлены на рис. 12.3.

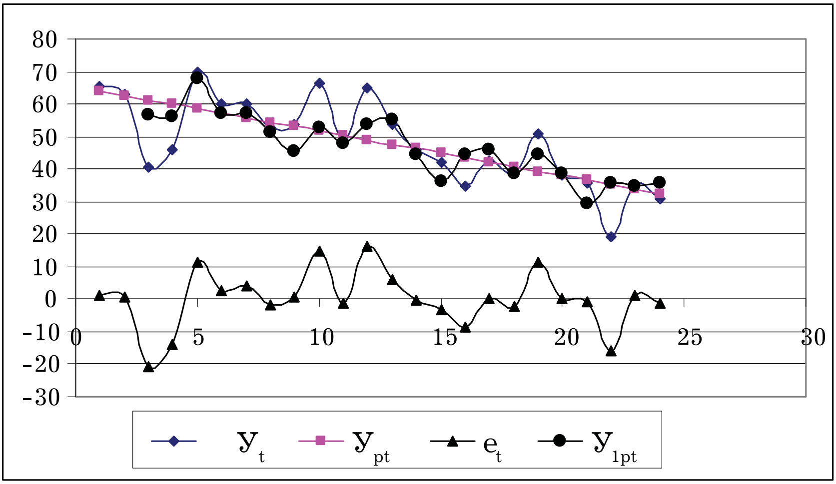


Рис. 12.3. Результаты расчетов по модели линейной регрессии и лаговых факторов белого шума второго порядка

Точечный прогноз равен:

$$У_{пр} = 65{,}322 - 1{,}368t + 0{,}9135 - 4{,}909001\,\delta_t — 7{,}448\,\delta_{t-1} + 10{,}533\,\delta_{t-2} =$$
$$= 65{,}322 - 1{,}368 \times 25 + 0{,}9135 - 4{,}909001 \times 0{,}19 - 7{,}448 \times 0{,}6115 +$$
$$+ 10{,}533 \times 0{,}0011 = 26{,}6584$$

Примечание. Расчетные значения зависимой переменной хорошо воспроизводят регулярности временного ряда, однако эти значения получены с помощью перебора подходящего белого шума. При любом изменении содержимого любой ячейки Excel происходит обновление случайных чисел. Этот набор значений белого шума получен одиннадцатым. Остальные наборы белого шума не существенно улучшали модель. Если белый шум отдаленно напоминает регулярности остатков, то с его помощью можно получить достоверную модель.

Вывод. Использование белого шума в предложенной модели является нежелательным, так как результаты расчетов зависят от субъективного выбора набора значений белого шума.

Задание второго уровня сложности

Задание. Определите вид временного ряда, который генерируется по уравнению

$$У_t = рУ_{t-1}.$$

Решение

Временные ряды, полученные при различных значениях р, представлены в табл. 12.3.

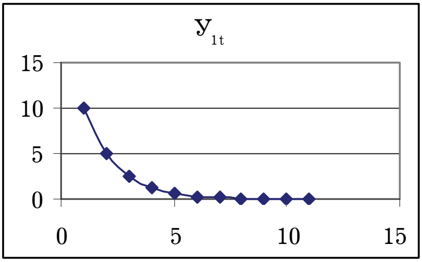
Таблица 12.3

Генерация временных рядов по формуле $У_t = рУ_{t-1}$

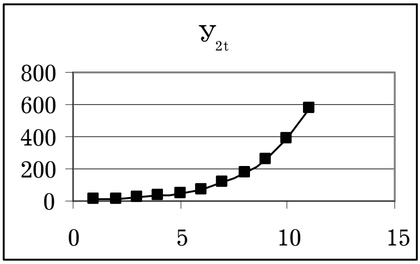
t	$У_{1t}$	$У_{2t}$	$У_{3t}$	$У_{4t}$
1	10	10	10	10
2	5	15	-5	-15
3	2,5	22,5	2,5	22,5
4	1,25	33,75	-1,25	-33,75
5	0,625	50,625	0,625	50,625
6	0,3125	75,938	-0,313	-75,94
7	0,1563	113,91	0,1563	113,91
8	0,0781	170,86	-0,078	-170,9
9	0,0391	256,29	0,0391	256,29

t	$У_{1t}$	$У_{2t}$	$У_{3t}$	$У_{4t}$
10	0,0195	384,43	-0,02	-384,4
11	0,0098	576,65	0,0098	576,65
Сумма	19,99	1710	6,6699	349,99
p =	0,5	1,5	-0,5	-1,5

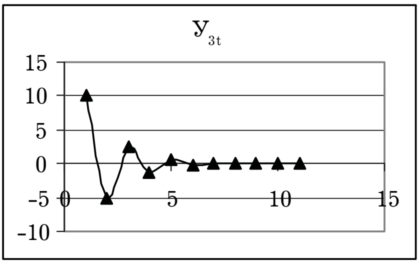
На рис. 12.4 представлены графики временных рядов, полученных при различных значениях Р.



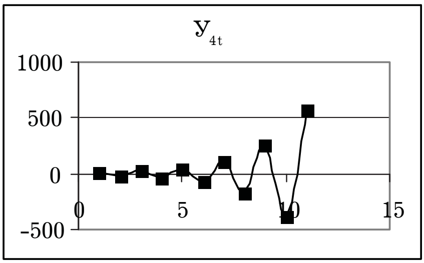
а) p = 0,5



б) p = 1,5



в) p = -0,5



г) p = -1,5

Рис. 12.4. Графики временных рядов, полученных при различных значениях p

Задание третьего уровня сложности

Задание. Определите вид временного ряда, для которого выполняются такие условия:

$У_t = \alpha_0 + \alpha_1 t + \varepsilon_t,$

$\varepsilon_t = \rho\varepsilon_{t-1} + \delta_t .$

На временной ряд остатков ε_t накладывается ограничение, состоящее в том, что сумма остатков должна быть равна нулю.

В качестве исходного временного ряда можно взять данные, представленные в табл. 12.4.

Таблица 12.4

Исходные данные для анализа

t	$У_t$
1	10
2	-5
3	2,5
4	-1,25
5	0,625
6	-0,313
7	0,1563
8	-0,078
9	0,0391
10	-0,02
11	0,0098

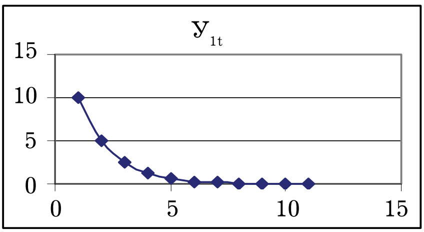
Задания четвертого уровня сложности

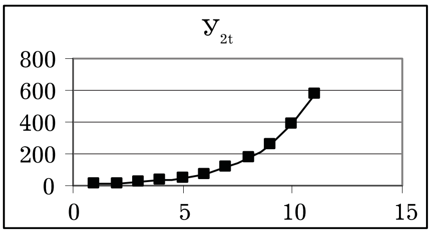
Используя данные табл. 12.4, получите прогноз с использованием модели скользящей средней.

Примечание. Алгоритм расчета представлен во входном тестировании.

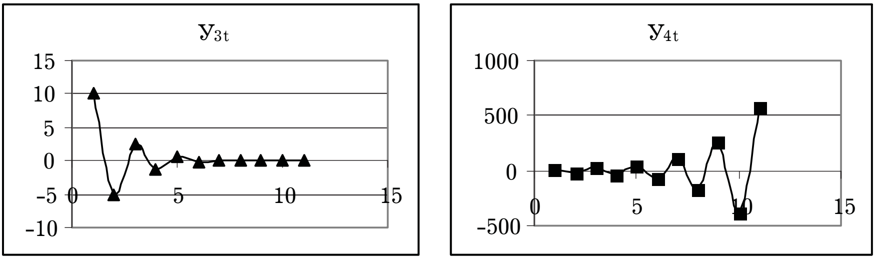
Выходное тестирование

Имеются графики временных рядов, которые сгенерированы по формуле $У_t = рУ_{t-1}$. Требуется определить знак и примерную величину коэффициента р.





а) б)



в) г)

Рис. 12.4. Графики временных рядов, сгенерированные по формуле

$$У_t = рУ_{t-1}$$

Нерешенная проблема

Нет критерия выбора белого шума, который можно использовать для построения модели скользящей средней.

13. СИСТЕМА ОДНОВРЕМЕННЫХ УРАВНЕНИЙ. ПОНЯТИЕ О КОСВЕННОМ, ДВУХШАГОВОМ МЕТОДЕ НАИМЕНЬШИХ КВАДРАТОВ ДЛЯ РЕШЕНИЯ СИСТЕМ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ

Постановка задачи

Объект — показатели экономической системы.

Предмет — взаимосвязи экономических показателй.

Цель — изучить свойства связей показателей экономического объекта.

Актуальность — необходимость получения прогноза в динамической системе.

Рабочая гипотеза — в систему экономического объекта входят внутренние и внешние факторы.

Модель — система одновременных уравнений;

Способ — для расчета характеристик модели можно использовать функцию Excel “Линейн”.

Задача — получить прогноз эндогенных переменных.

Ожидаемый результат — получение прогнозных значений эндогенных переменных.

Задания первого уровня сложности

Задание 1. Подготовить устные сообщения по вопросам самостоятельной работы.

1. Дайте определение эндогенной и экзогенной переменной

2. Укажите на основные свойства эндогенных и экзогенных переменных

3. С помощью какой модели отражаются свойства эндогенных и экзогенных переменных?

4. Дайте определение структурной и приведенной системы одновременных уравнений.

5. Приведите пример структурной системы одновременных уравнений

6. Укажите способ получения приведенной системы одновременных уравнений.

7. Укажите основное назначение приведенной системы одновременных уравнений.

8. Укажите признаки идентифицируемости структурной системы одновременных уравнений.

9. Укажите на недостатки МНК при вычислении коэффициентов структурной системы одновременных уравнений.

10. Опишите двухшаговый метод определения коэффициентов структурной системы одновременных уравнений.

10. Опишите косвенный метод определения коэффициентов структурной системы одновременных уравнений [3, с. 329–339; 1, т. 2 с. 331–364]

Задание 2. Пройдите входное тестирование.

Выберите вариант правильного ответа.

1. Эндогенные переменные (У) зависят от переменных системы и могут влиять на остальные переменные.

а) да;

б) нет.

2. Экзогенные переменные (X) не зависят от деятельности системы, но могут влиять на эндогенные переменные. Экзогенными переменными можно условно считать инвестиции в экономику, гуманитарную помощь. Истинными экзогенными переменными можно считать лаговые эндогенные переменные, природные, космические факторы.

а) да;

б) нет.

3. Свойства эндогенных переменных.

Свойство 1. Эндогенные переменные имеют обратные связи, которые порождают проблему устойчивости экономической системы. Возможны следующие сценарии изменения эндогенной переменной:

— она плавно монотонно или циклически вернется к своему исходному значению; при наличии отрицательной обратной связи;
— она плавно монотонно или циклически примет минимальное или максимальное возможное значение; при наличии положительной обратной связи.

Свойство 2. Если экономическая система находится в устойчивом состоянии, то изменение эндогенной переменной не способно изменить средние динамические равновесные значения эндогенных переменных.

а) да;
б) нет.

4. 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.

а) да;
б) нет.

5. Взаимосвязь экзогенных и эндогенных переменных описывается системой одновременных уравнений, которые принято называть эконометрической моделью.

а) да;
б) нет.

6. 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.

а) да;

б) нет.

7. Приведем структурную систему одновременных уравнений

$$\begin{cases} У_1 = a_0 + a_1 У_2 + a_2 X_1 + e_1; \\ У_2 = в_0 + в_1 У_1 + в_2 X_2 + e_2. \end{cases}$$

а) да;

б) нет.

8. Если в первом уравнении структурной системы одновременных уравнений 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.

$$\begin{cases} У_1 = c_0 + c_1 X_1 + c_2 X_2 + e_3; \\ У_2 = d_0 + d_1 X_1 + d_2 X_2 + e_4. \end{cases}$$

Однозначный переход от структурной системы к приведенной системе одновременных уравнений можно было произвести при условии идентифицируемости структурной системы одновременных уравнений.

а) да;

б) нет.

9. 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.

а) да;

б) нет.

10. Структурная 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.

а) да;

б) нет.

11. Структурная система одновременных уравнений неидентифицируема, 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.

а) да;

б) нет.

12. Структурная система одновременных уравнений сверхидентифицируемая, если хотя бы для одного уравнения системы выполняется неравенство

$$n > (n_i + d_i).$$

а) да;

б) нет.

13. Эндогенные переменные могут быть случайными величинами, поэтому, находясь в системе одновременных уравнений в качестве объясняемой переменной, они могут быть связаны с остатками. Взаимосвязь объясняемых переменных с остатками модели является нарушением предпосылок Гаусса-Маркова для метода наименьших квадратов, что делает оценки параметров модели смещенными и неэффективными.

а) да;

б) нет.

14. Двухшаговый метод наименьших квадратов

Задача. 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, 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.

Шаг 1. Устраняется зависимость $У_{2t}$ от e_t с помощью уравнения приведенной системы одновременных уравнений

$$У_{2pt} = b_0 + b_1 X_{1t} + b_2 X_{2t}.$$

Переменная $У_{2pt}$ не содержит случайной составляющей e_t.

Шаг 2. Рассчитываются коэффициенты модели

$$У_{1t} = a_0 + a_1 У_{2pt} + a_2 X_{1t} + e_t$$

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.

а) да;

б) нет.

15. Косвенный метод расчета коэффициентов структурных уравнений.

Если известны коэффициенты приведенных уравнений, то методом подстановки необходимо от приведенных уравнений

$$У_{1t} = c_0 + c1X_{1t} + c_2 X_{2t} + e_{3t}$$
$$У_{2t} = d_0 + d_1 X_{1t} + d_2 X_{2t} + e_{4t}$$

перейти к структурным уравнениям

$$У_{1t} = a_0 + a_1 У_{2t} + a_2 X_{1t} + e_{1t},$$
$$У_{2t} = в_0 + в_1 У_{1t} + в_2 X_{2t} + e_{2t}.$$

В этом и заключается косвенный метод расчета коэффициентов структурных уравнений.

а) да;

б) нет.

Задание 3. Решите задачу.

Имеется база данных бюджетного обследования семьи за каждый месяц текущего года (табл. 13.1).

Таблица 13.1

Динамика численных значений переменных бюджета семьи.

№ п/п	t	$У_{1t}$	X_{1t}	X_{2t}	$У_{2t}$
1	1	5		15	10
2	2	4	5	12	8
3	3	6	4	10	7
4	4	8	6	25	12
5	5	13	8	31	18

№ п/п	t	$У_{1t}$	X_{1t}	X_{2t}	$У_{2t}$
6	6	10	13	42	20
7	7	5	10	12	7
8	8	4	5	15	7
9	9	8	4	23	12
10	10	10	8	45	18
11	11	5	10	13	8
12	12	8	5	29	13
Ожидаем	13	7	8	25	10

где t — порядковый номер месяца текущего года,

$У_{1t}$ — объем покупок потребительских товаров, производимых из заработной платы главы семьи (тыс. руб.),

$X_{1t} = У_{1(t-1)}$ — объем покупок потребительских товаров в предшествующий период (лаговая эндогенная переменная) (тыс. руб.); семьей

X_{2t} — доход семьи (тыс. руб.);

$У_{2t}$ — заработная плата главы семьи (тыс. руб.).

Предложенные переменные входят в состав динамической микроэкономической эконометрической модели бюджета семьи, представленной в виде структурной формы системы одновременных уравнений:

$$У_{1t} = a_0 + a_1 У_{2t} + a_2 X_{1t} + e_{1t},$$
$$У_{2t} = в_0 + в_1 У_{1t} + в_2 X_{2t} + e_{2t}.$$

Необходимо:

— получить прогнозные значения $У_1$ и $У_2$, если в следующем прогнозном месяце ожидаются значения объясняемых переменных:

$$X_1(13) = 8 \text{ (тыс. руб.)},$$
$$X_2(13) = 25 \text{ (тыс. руб.)};$$

— вычислить коэффициенты структурной формы системы одновременных уравнений двухшаговым методом наименьших квадратов.

— вычислить коэффициенты структурной формы системы одновременных уравнений косвенным методом.

Решение задачи

В структурной системе одновременных уравнений все уравнения связаны между собой. При этом прогноз эндогенной переменной по одному уравнению зависит от прогнозных значений эндогенных переменных, полученных по другому уравнению. Экономическая система в конкретный момент времени может находиться в двух состояниях: в равновесном или переходном. При равновестном состоянии все эндогенные переменные находятся в среднем динамическом равновесном состоянии. Вспомните о равновесном состоянии трех эндогенных переменных: спрос, цена, предложения.

Если экономическая система находится в переходном состоянии, то численные значения эндогенных переменных, вычисленные по структурным уравнениям, характеризуют экономическую систему в одном из возможных неустойчивых состояний.

Целью моделирования является определение средних динамических равновесных значений эндогенных переменных экономической системы.

Можно продолжить изучение экономической системы и поставить следующие интересные цели например, если нам известны значения эндогенных переменных в устойчивом состоянии, то при изменении эндогенной или экзогенной переменной по экономической системе пройдут циклические волны. При этом необходимо определить, при каких условиях и через сколько циклов экономическая система перейдет в устойчивое стабильное состояние или в неустойчивое состояние разорения и самоуничтожения (пример явления гиперинфляции). Иными словами, необходимо определить условия переходных процессов, происходящих циклически и ведущих к затуханию или к увеличению значений эндогенных переменных при изменении значений эндогенной или экзогенной переменной. Реализация этих целей может превратиться в хорошую научную работу.

Выбор переменных

Объектом изучения является бюджет семьи. Для построения систем одновременных уравнений выделим следующие переменные:

$У_{1t}$ — объем покупок потребительских товаров, производимых из заработной платы главы семьи (тыс. руб.),

$X_{1t} = У_{1(t-1)}$ — объем покупок потребительских товаров в предшествующий период (лаговая эндогенная переменная) (тыс. руб.),

X_{2t} — доход семьи (тыс. руб.),

$У_{2t}$ — заработная плата главы семьи (тыс. руб.)

Выделение входных и выходных переменных

Для первого структурного уравнения выходной переменной является $У_{1t}$.

Входными переменными являются: $У_{2t}$ X_{1t}

Для второго структурного уравнения выходной переменной является $У_{2t}$

Входными переменными являются: $У_{1t}$ и X_{2t}

Сбор статистических данных

Исходные данные представлены в табл. 13.1, динамика всех переменных показана на рис. 13.1.

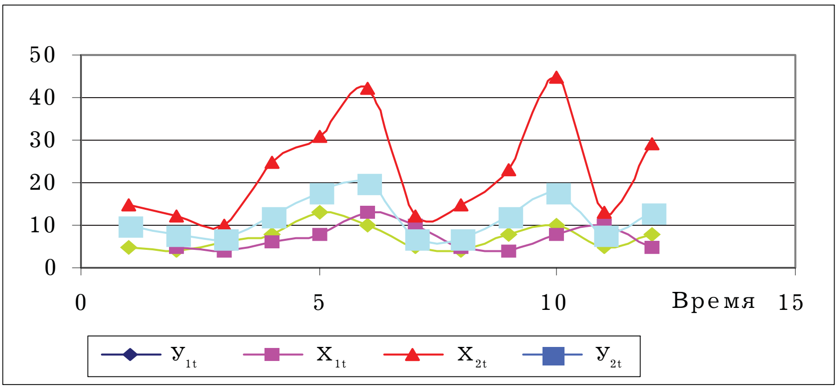


Рис. 13.1. Динамика всех переменных.

Предмодельный анализ

Выдвижение гипотез

Предположим, что все зависимости между переменными имеют линейный вид.

Формулировка допущений

Предполагаем, что при изменении эндогенных переменных экономическая система через несколько циклов вернется в исходное среднее динамическое устойчивое состояние.

Спецификация

Первое структурное уравнение будет иметь следующий вид:

$$У_{1t} = \alpha_0 + \alpha_1 У_{2t} + \alpha_2 X_{1t} + \varepsilon_{1t}$$
$$M(\varepsilon) = 0,\ M(\varepsilon\varepsilon') = \sigma_\varepsilon^2 I$$

Случайная составляющая должна быть распределена независимо от объясняющих переменных:

$$M(У_2\varepsilon_1) = 0,\ M(X_1\varepsilon_1) = 0.$$

При нарушении последнего условия оценки параметров модели методом наименьших квадратов будут обладать свойством смещенности и несостоятельности. [7, с. 194].

Второе структурное уравнение будет иметь следующий вид:

$$У_{2t} = \beta_0 + \beta_1 У_{1t} + \beta_2 X_{2t} + \varepsilon_{2t}$$

$$M(\varepsilon) = 0,\ M(\varepsilon\varepsilon') = \sigma_\varepsilon^2 I$$

Случайная составляющая должна быть распределена независимо от объясняющих переменных:

$M(У_1\varepsilon_2) = 0,\ M(X_2\varepsilon_2) = 0.$

При нарушении последнего условия оценки параметров модели методом наименьших квадратов будут обладать свойством смещенности и несостоятельности [7, с. 194].

Представим взаимосвязь между структурными уравнениями в виде блок-схемы (рис. 13.2).

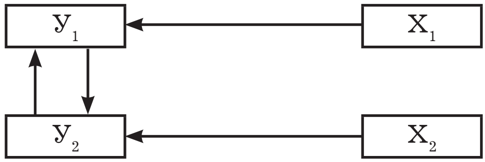


Рис. 13.2. Блок-схема связей между эндогенными и экзогенными переменными

Идентифицируемость модели

(Можно ли в принципе однозначно восстановить значения неизвестных параметров по выбранной модели).

Имеются следующие условия строгой идентифицируемости структурной системы одновременных уравнений:

— количество всех эндогенных и экзогенных переменных должно равняться между собой;

— для каждого структурного уравнения должны выполняться следующие условия: количество объясняемых переменных должно равняться количеству экзогенных переменных;

— комбинации объясняемых переменных для каждого уравнения не должны повторяться.

Имеется еще один признак строгой идентифицируемости структурной системы одновременных уравнений:

— если удается перейти от структурной системы одновременных уравнений к их приведенной форме, то структурная система одновременных уравнений является строго идентифицируемой.

Проверим структурную систему одновременных уравнений на идентифицируемость.

— количество эндогенных и экзогенных переменных равно двум; Следовательно, условие равенства количества всех эндогенных и экзогенных переменных соблюдается;

— в каждом структурном уравнении количество объясняемых переменных равно количеству экзогенных (или эндогенных);

— в каждом структурном уравнении не повторяются комбинации объясняемых переменных.

Следовательно, все условия строгой идентифицируемости структурной системы одновременных уравнений соблюдены и мы можем получить приведенную систему одновременных уравнений.

Идентификация модели

(Предложить и реализовать процедуру оценивания параметров модели).

Для прогноза $У_1$ и $У_2$ используют приведенную форму системы одновременных уравнений, которая получается заменой в структурном уравнении объясняющей эндогенной переменной ее выражением, взятым из соответствующего структурного уравнения.

Подставим в выражение

$$У_{1t} = a_0 + a_1 У_{2t} + a_2 X_{1t} + e_{1t}$$

эндогенную переменную $У_{2t}$

$$У_{2t} = в_0 + в_1 У_{1t} + в_2 X_{2t} + e_{2t}.$$

После преобразований получим приведенное уравнение

$$У_{1t} = c_0 + c_1 X_{1t} + c_2 X_{2t} + e_{3t}.$$

Подставим в выражение

$$У_{2t} = в_0 + в_1 У_{1t} + в_2 X_{2t} + e_{2t}$$

эндогенную переменную $У_{1t}$

$$У_{1t} = a_0 + a_1 У_{2t} + a_2 X_{1t} + e_{1t}$$

После преобразований получим приведенное уравнение

$$У_{2t} = d_0 + d_1 X_{1t} + d_2 X_{2t} + e_{4t}.$$

Составим приведенную систему одновременных уравнений

$$У_{1t} = c_0 + c_1 X_{1t} + c_2 X_{2t} + e_{3t}.$$
$$У_{2t} = d_0 + d_1 X_{1t} + d_2 X_{2t} + e_{4t}.$$

Оценим параметры приведенной системы одновременных уравнений методом наименьших квадратов с помощью функции “Линейн” для каждого уравнения в отдельности.

Вычислим коэффициенты приведенного уравнения

$У_{1t} = c_0 + c_1X_{1t} + c_2X_{2t} + e_{3t}$

0,1949	-0,0502	3,16562
0,05153	0,21567	1,62613
0,6693	1,84721	#Н/Д
8,09568	8	#Н/Д
55,248	27,2975	#Н/Д

$У_{2t} = d_0 + d_1X_{1t} + d_2X_{2t} + e_{4t}$

0,37089	0,12551	2,2628
0,0435	0,18206	1,37271
0,92016	1,55933	#Н/Д
46,0996	8	#Н/Д
224,184	19,4522	#Н/Д

$$У_1 = 3,16562 — 0,0502\ X_1 + 0,1949\ X_2 + e_{3t}$$
$$У_2 = 2,2628 + 0,12551\ X_1 + 0,37089\ X_2 + e_{4t}$$

Получим прогнозные величины средних динамических эндогенных переменных при условии, что ожидаемые экзогенные переменные имеют следующие значения:

$X_{1ож} = 8$ (тыс. руб.)

$X_{2ож} = 25$ (тыс. руб.)

Получим прогнозные значения $У_{1п}$ и $У_{2п}$

$У_{1п} = 3,16562 - 0,0502\ X_{1ож} + 0,1949\ X_{2ож} = 7,63697$ (тыс. руб.);

$У_{2п} = 2,2628 + 0,12551\ X_{1ож} + 0,37089\ X_{2ож} = 12,5392$ (тыс. руб.).

Вычислим коэффициенты уравнения

$$У_{1t} = a_0 + a_1У_{2t} + a_2X_{1t} + e_{1t}$$

двухшаговым методом наименьших квадратов.

Пояснения.

1. Основная проблема вычисления коэффициентов данного уравнения заключается в зависимости объясняемой переменной $У_{2t}$ от величины остатков e_{1t}, что является нарушением предпосылок метода наименьших квадратов и определенные коэффициенты не являются оптимальными. Основным признаком оптимальности коэффициентов модели является ортогональность объясняемых переменных с остатками модели. Объясняемая переменная и остатки будут ортогональны (или независимы), если выполняется условие:

$$\Sigma\,(У_{2t}e_{1t}) = 0$$

2. Устранить зависимость $У_{2t}$ от e_{1t} можно, если в базу данных вместо $У_{2t}$ фактического поставить $У_{2t}$ расчетное, определенное по приведенному уравнению системы одновременных

уравнений. Поэтому данный способ получил название — двухшаговый метод наименьших квадратов.

На первом шаге вычисляются расчетные значения $У_{2pt}$ по приведенному уравнению системы одновременных уравнений.

На втором шаге вычисляются коэффициенты уравнения структурной системы одновременных уравнений по базе данных, где вместо $У_{2t}$ фактического должно находиться $У_{2pt}$ расчетное.

Таблица 13.2

Данные для расчета коэффициентов структурной системы двухшаговым методом наименьших квадратов

№ п/п	t	$У_{1t}$	$У_{2t}$	$У_{2pt}$	X_{1t}	$У_{1pt}$	X_{2t}
1	1	5	10				15
2	2	4	8	7,34107	5	5,25369	12
3	3	6	7	6,47377	4	4,91404	10
4	4	8	12	12,2882	6	7,73728	25
5	5	13	18	14,7645	8	8,8064	31
6	6	10	20	19,4719	13	10,6996	42
7	7	5	7	7,96864	10	5,00292	12
8	8	4	7	8,45374	5	5,8384	15
9	9	8	12	11,2954	4	7,44778	23
10	10	10	18	19,957	8	11,535	45
11	11	5	8	8,33953	10	5,19782	13
12	12	8	13	13,6462	5	8,56705	29
Ожидаем	13	7	10		8		25

Реализуем двухшаговый метод наименьших квадратов для расчета коэффициентов структурного уравнения

$$У_{1t} = a_0 + a_1 У_{2t} + a_2 X_{1t} + e_{1t}$$

Шаг 1. Вычисляем $У_{2pt}$ по приведенному уравнению

$$У_{2pt} = 2{,}2628\ 0{,}12551\ X_{1t} + 0{,}37089\ X_{2t}.$$

Результаты расчетов занесем в табл. 13.2.

Шаг 2. Вычисляем коэффициенты структурного уравнения

$$У_{1t} = a_0 + a_1 У_{2pt} + a_2 X_{1t} + e_{1t}$$

методом наименьших квадратов с помощью функции “Линейн”.

-0,1161	0,5255	1,97652
0,22313	0,13893	1,76646
0,6693	1,84721	#Н/Д
8,09568	8	#Н/Д
55,248	27,2975	#Н/Д

Структурное уравнение регрессии, коэффициенты которого определены двухшаговым методом наименьших квадратов, имеет следующий вид:

$$У_{1t} = 1{,}97652 + 0{,}5255\ У_{2t} - 0{,}1161\ X_{1t} + e_{1t}.$$

Аналогично двухшаговым методом наименьших квадратов вычисляются коэффициенты структурного уравнения

$$У_{2t} = в_0 + в_1 У_{1t} + в_2 X_{2t} + e_{2t}.$$

Шаг 1. Вычисляем $У_{1tp}$ по приведенному уравнению

$$У_{1pt} = 3{,}16562 - 0{,}0502\ X_{1t} + 0{,}1949\ X_{2t}.$$

Результаты расчетов занесем в табл. 13.2.

Шаг 2. Вычисляем коэффициенты структурного уравнения

$$У_{2t} = в_0 + в_1 У_{1tp} + в_2 X_{2t} + e_{2t}.$$

методом наименьших квадратов с помощью функции “Линейн”.

0,85865	-2,5026	10,185
0,69148	3,62999	10,6529
0,92016	1,55933	#Н/Д
46,0996	8	#Н/Д
224,184	19,4522	#Н/Д

Структурное уравнение регрессии, коэффициенты которого определены двухшаговым методом наименьших квадратов, имеет следующий вид:

$$У_{2t} = 10{,}185 - 2{,}5026\ У_{1t} + 0{,}85865\ X_{2t} + e_{2t}$$

Косвенный метод расчета коэффициентов структурных уравнений

Если известны коэффициенты приведенных уравнений, то методом подстановки необходимо от приведенных уравнений

$$У_{1t} = c_0 + c1X_{1t} + c_2X_{2t} + e_{3t}$$
$$У_{2t} = d_0 + d_1X_{1t} + d_2X_{2t} + e_{4t}$$

перейти к структурным уравнениям

$$У_{1t} = a_0 + a_1У_{2t} + a_2X_{1t} + e_{1t},$$
$$У_{2t} = в_0 + в_1У_{1t} + в_2X_{2t} + e_{2t}.$$

В этом и заключается косвенный метод расчета коэффициентов структурных уравнений.

Например, для расчета коэффициентов структурного уравнения

$$У_{1t} = a_0 + a_1У_{2t} + a_2X_{1t} + e_{1t}$$

необходимо из уравнения

$$У_{2t} = d_0 + d_1X_{1t} + d_2X_{2t} + e_{4t}$$

выразить переменную X_{1t} и вставить в уравнение

$$У_{1t} = c_0 + c_1X_{1t} + c_2X_{2t} + e_{3t}$$

После несложных преобразований получим структурное уравнение

$$У_{1t} = a_0 + a_1У_{2t} + a_2X_{1t} + e_{1t}.$$

Примечание. Косвенный метод определения коэффициентов структурных уравнений возможен только при наличии строгой идентифицируемости структурной системы одновременных уравнений.

Задание второго уровня сложности

Задание. Постройте действующую динамическую модель, реализующую структурную систему одновременных уравнения.

Задание третьего уровня сложности

Задание. Приведите матричную формулу косвенного метода определения коэффициентов структурной системы одновременных уравнений

Задание четвертого уровня сложности

Задание. Укажите признаки устойчивости экономической системы

Выходное тестирование

Укажите факторы, которые могут изменить средне динамические равновесные значения эндогенных переменных.

Нерешенные проблемы

1. При вычислении коэффициентов структурной системы одновременных уравнений МНК не учитывается мультиколлинеарность между объясняющими переменными.

2. Плохо изучены условия, признаки устойчивости экономической системы.

14. АДАПТИВНЫЕ МОДЕЛИ ПРОГНОЗИРОВАНИЯ

Постановка задачи

Объект — временные данные экономического показателя.

Предмет — регулярности экономического показателя.

Цель — изучить адаптивные модели прогнозирования временных рядов, предназначенных для повышения точности прогноза.

Актуальность — повышение точности прогноза временного ряда всегда высоко ценилось в эконометрическом моделировании.

Рабочая гипотеза — последние значения временного ряда оказывают большее влияние на прогноз, чем первые.

Модели:

— адаптивная модель Брауна;
— адаптивная модель прогнозирования Хольта;
— адаптивная модель прогнозирования Бокса-Дженкинса;
— адаптивная модель Уинтерса;
— адаптивная ????????????? модель сезонности Тейла-Вейджа.

Способ — для расчета характеристик модели можно использовать функцию Excel “Линейн”.

Задача — выделение всех составляющих временного ряда.

Ожидаемый результат — использование моделей нестационарных временных рядов должно повысить точность прогнозирования.

Метод для сравнения — парная линейная модель.

Задания первого уровня сложности

Задание 1. Подготовить устные сообщения по вопросам самостоятельной работы.

1. Укажите характерные черты адаптивных моделей прогнозирования
2. Укажите достоинства и недостатки адаптивных моделей прогнозирования.
3. Приведите расчетные формулы адаптивной модели прогнозирования Брауна.
4. Приведите основные свойства адаптивной модели прогнозирования Брауна.
5. Приведите расчетные формулы адаптивной модели прогнозирования Хольта.
6. Приведите расчетные формулы адаптивной модели прогнозирования Бокса-Дженкинса
7. Приведите расчетные формулы адаптивной модели прогнозирования Уинтерса
8. Приведите расчетные формулы адаптивной аддитивной модели сезонности Тейла-Вейджа [3, с. 323–327)

Задание 2. Пройдите входное тестирование.

Выберите вариант правильного ответа

1. Характерной чертой адаптивных моделей 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.

а)да;

б) нет.

2. Все адаптивные модели хорошо воспроизводят плавно изменяющиеся значения временного ряда и плохо аппроксимируют резко изменяющиеся данные и выбросы. Аппроксимированный временной ряд с помощью адаптивных моделей повторяет резко изменяющиеся данные, но сдвинутыми во

времени на одну дату, что значительно ухудшает качество модели. Это негативное свойство адаптивных моделей прогнозирования следует учитывать при подборе коэффициентов сглаживания.

а)да;

б) нет.

3. Адаптивная модель прогнозирования Брауна

Экспоненциально взвешенное среднее значение вычисляется по формуле

$$Z_t = \lambda У_t + (1 — \lambda)Z_{t-1}.$$

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:

$$П_{t+1} = Z_t = \lambda У_t + (1 — \lambda)Z_{t-1} = \lambda У_t + (1 — \lambda)П_t;$$
$$П_{t+1} = П_t + \lambda\,(У_t\text{-}П_t),$$

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;

$П_1 = П_t$, при t = 1, — первое прогнозное значение определяется экспертным способом, рассчитывают его с помощью регрессионного анализа или считают равным первому значению временного ряда;

λ — коэффициент сглаживания.

а)да;

б) нет.

4. Адаптивная модель прогнозирования Брауна обладает следующими свойствами:

— для построения прогноза по экспоненциально взвешенному среднему необходимо знать только начальную оценку прогноза, дальнейшее прогнозирование возможно при 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 $0 < \lambda < 0{,}3$, то временной ряд является стационарным, если $1 > \lambda > 0.3$, 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].

а)да;
б) нет.

5. Адаптивная модель прогнозирования Хольта.

Экспоненциальная средняя приводит к смещенным прогнозам, т. е. дает систематическую ошибку, когда временной ряд имеет тенденцию линейного роста. Для устранения этой систематической ошибки разработано несколько вариантов

---

[1] Льюис К. Д. Методы прогнозирования экономических показателей. — М.: Финансы и статистика, 1986.

Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования. — М.: Статистика, 1979.

адаптивных моделей, также использующих процедуру экспоненциального сглаживания. В основе моделей лежит гипотеза о том, что прогноз может быть получен по уравнению

$$П_{t+k} = a_{0,\,t} + da_{1,\,t},$$

где $a_{0,\,t}$, $a_{1,\,t}$ — текущие оценки коэффициентов адаптивного полинома первого порядка;

d = 1 — глубина прогноза.

Одной из первых моделей этого типа была двухпараметрическая модель Хольта, в которой оценка коэффициентов производится следующим образом:

$$П_{t+1} = a_{0,t} + da_{1,t} = a_{0,t} + a_{1,t};$$
$$a_{0,\,t} = \lambda_1 У_t + (1\text{-}\lambda_1)\,(a_{0,t-1} + a_{1,\,t-1}) = \lambda_1 У_t + (1\text{-}\lambda_1)\,П_t;$$
$$a_{1,t} = \lambda_2(a_{0,t} — a_{0,\,t-1}) + (1 — \lambda_2)a_{1\,t-1},$$

где λ_1, λ_2 — параметры экспоненциального сглаживания ($0 < \lambda_1$, $\lambda_2 < 1$).

При t = 0:

$a_{0,0}$ и $a_{1,0}$ вычисляются МНК по регрессионному уравнению

$$У_t = a_{0,0} + a_{1,0}t + e_t.$$

Оптимальное значение параметров λ_1, λ_2 можно определить по минимальной ошибке модели.

а)да;

б) нет.

6. Адаптивные модели прогнозирования Бокса-Дженкинса

Если модель Хольта усовершенствовать путем включения разности ошибок, то получится полная трехпараметрическая модель Дж. Бокса и Г. Дженкинса:

$$П_{t+1} = a_{0,\,t} + a_{1,\,t};$$
$$a_{0,\,t} = \lambda_1 У_t + (1\text{-}\lambda_1)(a_{0,\,t-1} + a_{1,\,t-1}) + \lambda_3(e_t — e_{t-1});$$
$$a_{1,\,t} = \lambda_2(a_{0,t}\text{-}a_{0,\,t-1}) + (1 — \lambda_2)a_{1,\,t-1},$$

где λ_1, λ_2, λ_3 — параметры экспоненциального сглаживания ($0<\lambda_1$, λ_2, $\lambda_3<1$).

а)да;

б) нет.

7. 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 d тактов времени вперед, равен:

$$П_{t+d} = (a_{0,\,t} + da_{1,\,t}) + w_{t+d-T}\,,$$

при d = 1

$$П_{t+1} = (a_{0,\,t} + 1a_{1,\,t}) + w_{t+1-T},$$

где w_{t+d-T} — коэффициент сезонности;

T — число временных тактов, содержащихся в полном сезонном цикле (если сезонная волна равна 3 месяцам, то T = 3).

Модель Уинтерса содержит три параметра сглаживания $\lambda_1, \lambda_2, \lambda_3$ $(0<\lambda_1, \lambda_2, \lambda_3 <1)$.

Коэффициенты модели вычисляются по следующим формулам:

$$a_{0,\,t} = \lambda_1 \frac{У_t}{w_{t-T}} + (1-\lambda_1)(a_{0,\,t-1} + a_{1,\,t-1}),$$

$$w_t = \lambda_2 \frac{У_t}{a_{0,\,t}} + (1-\lambda_2) w_{t-T}, \text{ при } t \geq T,$$

$$a_{1,\,t} = \lambda_3 (a_{0,\,t} - a_{0,t\text{-}1}) + (1 \quad \lambda_3) a_{1,\,t\,\text{-}}\ [1,.\ \text{с. } 325—327].$$

а)да;

б) нет.

8. Адаптивная аддитивная модель сезонности Тейла-Вейджа

Предположим, что исходный временной ряд был преобразован так, чтобы его можно аппроксимировать аддитивной моделью:

$$У_t = a_{0,t} + w_t + \delta_t;$$
$$a_{0,t} = a_{0,t\text{-}1} + a_{1,t},$$

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 d временных тактов вперед, вычисляется по следующей формуле:

$$П_{t+d} = a_{0,t} + da_{1,t} + w_{t+d-T},$$

где $a_{0,t} = a_{0,t-1} + a_{1,t-1} + \lambda_1(У_t — П_t)$;

$a_{1,t} = a_{1,t-1} + \lambda_1\lambda_2(У_t — П_t)$;

$w_t = w_{t-T} + (1-\lambda_1)\lambda_3(У_t — П_t)$;

$\lambda_1, \lambda_2, \lambda_3$ $(0<\lambda_1, \lambda_2, \lambda_3 <1)$ — параметры адаптации;

d = 1 — глубина прогноза;

T — число временных тактов, содержащихся в полном сезонном цикле [1, с. 326−327].

а)да;

б) нет.

Задание 3. Решите задачу.

Имеется база данных показателей деятельности хлебокомбината (в действующих ценах). Продолжим изучение временного ряда X_4 — уровень затрат (%), которые содержат сезонную волну (табл. 1.3).

Необходимо получить прогнозные значение уровня затрат на ожидаемую дату t = 25 с помощью адаптивных моделей прогнозирования.

Решение

Модель Брауна

Прогнозные значения адаптивной модели Брауна вычисляются по формуле

$$П_{t+1} = П_t + \lambda\,(У_t — П_t),$$

где $П_{t+1}$ — прогноз временного ряда на момент времени t+1;

$У_t — П_t$ — ошибка прогноза;

$П_t = П_1$, при t = 1, — первое прогнозное значение равно первому значению временного ряда;

λ — коэффициент сглаживания.

Адаптивную модель прогнозирования Брауна можно реализовать в такой последовательности:

а) откройте лист Excel, введите значения t и $У_t = X_{4t}$ в таблицу, значения X_{4t} возьмите из табл. 1.3;

б) в свободную ячейку введите значение λ;

в) произведите расчет колонки $П_{t+1}$ таблицы по формуле

$$П_{t+1} = П_t + \lambda\,(У_t - П_t),$$

где $П_{t=1} = У_1$,

$t = 1, 2, ..., n$;

г) произведите расчет колонки с квадратом ошибки модели по формуле

$$e_t^2 = (У_t - П_t)^2;$$

д) вычислите ошибку модели по формуле

$$E = \sqrt{\frac{e_t^2}{n-2}};$$

е) постройте графики зависимости $У_t$ и $П_t$ от t;

ж) определите оптимальное значение λ — коэффициента сглаживания способом перебора: изменяйте значение λ от 0 до 1 с шагом 0,1 и определяйте, при каком его значении ошибка модели E будет минимальной, а также наблюдайте, как на графике расчетные значения $П_t$ воспроизводят временной ряд.

Признаком минимума ошибки модели E является увеличение ошибки при значениях λ больше или меньше его оптимального значения;

з) зафиксируйте оптимальное значение λ;

и) получите прогноз на дату n + 1;

к) выходной характеристикой адаптивной модели прогнозирования Брауна будет прогноз $П_{n+1}$ на дату n + 1 и ошибка модели E.

Протокол расчетов представлен в табл. 14.1.

Таблица 14.1

Реализация модели Брауна

t	$У_t = X_{4t}$	$П_{t+1}$	e_t^2	t	$У_t = X_{4t}$	$П_{t+1}$	e_t^2
1	68,26	68,26	0	5	64,1	75,164	122,42
2	70,17	68,26	3,6481	6	68,65	72,952	18,504
3	81,51	68,642	165,59	7	70,93	72,091	1,3485
4	90,96	71,216	389,84	8	77,88	71,859	36,252

t	$У_t = X_{4t}$	$П_{t+1}$	e_t^2
9	78,66	73,063	31,324
10	90,22	74,183	257,2
11	75,89	77,39	2,2502
12	65,16	77,09	142,33
13	67,03	74,704	58,891
14	68,01	73,169	26,618
15	69,51	72,137	6,9031
16	88,35	71,612	280,16
17	81,68	74,96	45,165
18	73,24	76,304	9,3858

t	$У_t = X_{4t}$	$П_{t+1}$	e_t^2
19	65,12	75,691	111,74
20	81,52	73,577	63,096
21	78,66	75,165	12,212
22	80,81	75,864	24,46
23	80,06	76,853	10,282
24	77,95	77,495	0,2073
25		77,586	
Сумма			1819,8

$\lambda =$ 0,2 $\qquad$ E = 9,095

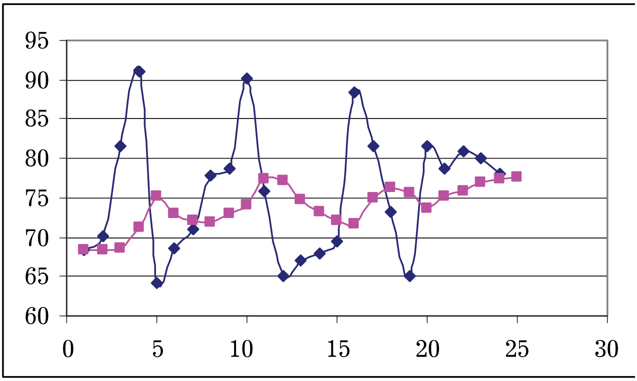


Рис. 14.1. Результаты расчетов модели Брауна,
где ◆ — фактические значения X_4, — расчетные значения X_4.

Анализ рис. 14.1 показывает, что расчетные значения повторяют выбросы со сдвижкой на одну дату.

Вывод. При оптимальном значении коэффициента сглаживания $\lambda = 0{,}2$ прогноз на ожидаемую дату равен $У_{прт = 25} = 77{,}586$, ошибка модели равна $E = 9{,}095$.

Модель Хольта

Прогноз вычисляется по формуле

$$П_{t+1} = a_{0,\,t} + a_{1,t};$$

$a_{0,t} = \lambda_1 У_t + (1-\lambda_1)(a_{0,t-1} + a_{1,t-1})$;
$a_{1,t} = \lambda_2(a_{0,t} — a_{0,t-1}) + (1 — \lambda_2)a_{1,t-1}$,
При $t = 0$:
$a_{0,0}$ и $a_{1,0}$ вычисляются МНК по регрессионному уравнению

$$У_t = a_{0,0} + a_{1,0}t + e_t.$$

Адаптивную модель прогнозирования Хольта можно реализовать в такой последовательности:

а) откройте лист Excel, введите значения t от 0 до $n = 24$ и $У_t = X_{4t}$ в таблицу (значения X_{4t} возьмите из табл. 1.3);

б) в свободные ячейки введите значения λ_1 и λ_2, взятые из диапазона от 0 до 1;

в) выделите две колонки для значений $a_{0,t}$ и $a_{1,t}$

г) с помощью функции Линейн определите первые значения коэффициентов $a_{0,0}$ и $a_{1,0}$ уравнения

$$У_t = a_{0,0} + a_{1,0}t + e_t$$

и введите их в первые значения $a_{0,0}$ и $a_{1,0}$ колонок $a_{0,t}$ и $a_{1,t}$ для $t = 0$;

д) при $t = 1$ в первое значение $П_1$ колонки $П_t$ введите значение

$$П_1 = a_{0,0} + a_{1,0};$$

е) при $t = 2$ вычислите значения коэффициентов $a_{0,1}$ и $a_{1,1}$ по формулам

$$a_{0,1} = \lambda_1 У_1 + (1-\lambda_1)(a_{0,0} + a_{1,0});$$
$$a_{1,1} = \lambda_2(a_{0,1}-a_{0,0}) + (1 — \lambda_2)a_{1,0}$$

введите эти значение в колонки $a_{0,t}$ и $a_{1,t}$ в строчки при $t = 1$;

ж) вычислите прогнозное значение $П_2$ по формуле

$$П_2 = a_{0,1} + a_{1,1};$$

введите полученный прогноз $П_2$ в колонку $П_t$ в строчку при $t = 2$;
и так далее реализуя алгоритм

$$П_{t+1} = a_{0,t} + a_{1,t};$$

$a_{0,t} = \lambda_1 У_t + (1-\lambda_1)(a_{0,t-1} + a_{1,t-1})$;
$a_{1,t} = \lambda_2(a_{0,t}-a_{0,t-1}) + (1 — \lambda_2)a_{1,t-1}$;

и) при $t = n = 24$ можно получить прогноз $П_{25}$ на следующую дату $n+1$.

к) произведите расчет квадрата ошибки модели по формуле

$$e_t^{\,2} = (У_t — П_t)^2$$

и введите их в дополнительную колонку;

л) вычислите ошибку модели по формуле

$$E = \sqrt{\frac{e_t^2}{n-2}};$$

м) постройте графики зависимости $У_t$ и $П_t$ от t;

н) определите оптимальное значение λ_1, λ_2 способом перебора: изменяйте значение λ_1, λ_2 от 0 до 1 с шагом 0,1 и определяйте, при каком их значении ошибка модели Е будет минимальной, а также наблюдайте, как на графике расчетные значения $П_t$ воспроизводят временной ряд.

Признаком минимума ошибки модели Е является увеличение ошибки при значениях λ_1, λ_2 больше или меньше его оптимального значения.

Примечание. Оптимальные значений λ_1, λ_2 можно определить двумя способами.

Первый способ. Оптимальные значения λ_1, λ_2 должны выбираться из всех их возможных комбинаций. Поэтому для каждого фиксированного значения λ_1, значения λ_2 должны последовательно принимать все допустимые значения. В результате получится очень большая матрица всех возможных комбинаций значений λ_1, λ_2 .

Второй способ. Оптимальные значения λ_1, λ_2 можно определить с помощью программы “Поиск решения”, где изменяя ячейки со значениями λ_1, $\lambda_{2,}$ необходимо получить минимум ошибки модели как целевой функции при ограничениях на значения λ_1, λ_2 .

о) зафиксируйте оптимальные значения λ_1, λ_2;

п) получите прогноз на дату $n + 1$;

р) выходной характеристикой адаптивной модели прогнозирования Хольта будет прогноз $П_{n+1}$ на дату $n + 1$ и ошибка модели Е.

Расчеты по модели Хольта представлены в таб. 14.2.

Таблица 14.2

Расчеты по модели Хольта для данных X_4

t	$У_t = X_{4t}$	$П_t$	a_{0t}	a_{1t}	e_t^2
0			73,284	0,1851	
1	68,26	73,469	72,427	0,0809	27,13
2	70,17	72,508	72,04	0,0342	5,4654
3	81,51	72,074	73,962	0,2229	89,03
4	90,96	74,184	77,54	0,5584	281,42
5	64,1	78,098	75,298	0,2784	195,94
6	68,65	75,577	74,191	0,1399	47,98
7	70,93	74,331	73,651	0,0719	11,569
8	77,88	73,723	74,554	0,155	17,282
9	78,66	74,709	75,499	0,234	15,608
10	90,22	75,733	78,631	0,5237	209,86
11	75,89	79,155	78,502	0,4585	10,657
12	65,16	78,96	76,2	0,1825	190,44
13	67,03	76,383	74,512	-0,005	87,469
14	68,01	74,507	73,208	-0,135	42,216
15	69,51	73,073	72,361	-0,206	12,698
16	88,35	72,155	75,394	0,1181	262,28
17	81,68	75,512	76,746	0,2415	38,044
18	73,24	76,987	76,238	0,1665	14,04
19	65,12	76,404	74,147	-0,059	127,33
20	81,52	74,088	75,575	0,0895	55,232
21	78,66	75,664	76,263	0,1494	8,9761
22	80,81	76,413	77,292	0,2373	19,337
23	80,06	77,529	78,036	0,2879	6,404
24	77,95	78,323	78,249	0,2805	0,1395
25		78,529			
Сумма					1776,6

$\lambda_1 = 0,2$

$\lambda_2 = 0,1$

E = 8,9862

Протокол расчетов по функции Линейн.

0,1851	73,2836
0,2381	3,4025
0,0267	8,07524
0,6041	22
39,394	1434,61

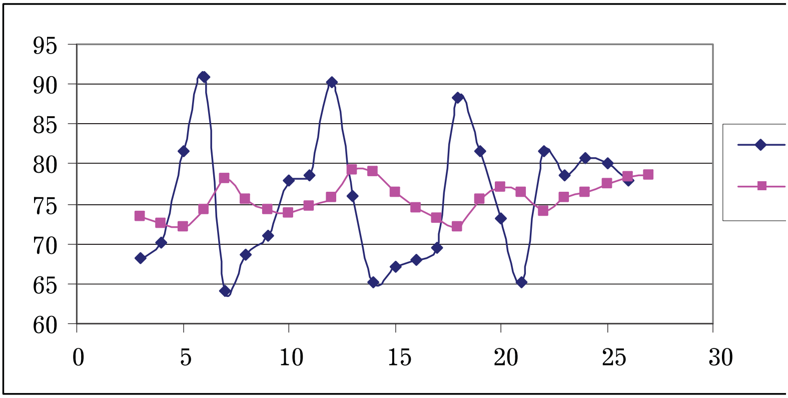


Рис. 14.2. Результаты расчетов по модели Хольта

Анализ проведенных расчетов показывает, что оптимальными λ_1, λ_2, определенными по программе "Поиск решения", являются значения, равные нулю, и модель Хольта превращается в линейную модель, которая оказывается лучше, чем модель с параметрами адаптации λ_1 и λ_1. Этот факт можно объяснить тем, что исходные данные содержат выбросы, которые плохо воспроизводятся моделью Хольта. В протоколе расчетов приведены значения λ_1 и λ_2, отличающиеся от нуля, для сравнения с другими адаптивными моделями. При измени исходных данных на более плавные значения модель Хольта позволила получить хорошие результаты и программа "Поиск решения" определила оптимальные значения λ_1, λ_2 отличающиеся от нуля.

Вывод. Модель Хольта для данных, содержащих сезонную волну, не позволила улучшить модель по сравнению с линейной моделью.

Модель Бокса-Дженкинса

Модель Бокса-Дженкинса является измененной моделью Хольта путем включения разности ошибок $\lambda_3(e_t — e_{t-1})$:

$$П_{t+1} = a_{0,t} + a_{1,t};$$

$$a_{0,t} = \lambda_1 У_t + (1\text{-}\lambda_1)\,(a_{0,t-1} + a_{1,t-1}) + \lambda_3\,(e_t — e_{t-1});$$

$$a_{1,t} = \lambda_2\,(a_{0,t}\text{-}a_{0,t-1}) + (1 — \lambda_2)a_{1,t-1}.$$

Повторим расчеты модели Хольта с измененным значением $a_{0,t}$, начиная с t = 2.

Расчеты по модели Бокса-Дженкинса представлены в табл. 14.3.

Таблица 14.3

Расчеты по модели Бокса-Дженкинса для данных X_4

t	$У_t = X_{4t}$	$П_t$	a_{0t}	a_{1t}	e_t	e_t^2
0			73,284	0,1851		
1	68,26	73,469	73,469	0,1851	-5,209	27,13
2	70,17	73,654	74,092	0,1851	-3,484	12,136
3	81,51	74,277	76,997	0,1851	7,2334	52,322
4	90,96	77,182	78,843	0,1851	13,778	189,84
5	64,1	79,028	71,742	0,1851	-14,93	222,85
6	68,65	71,927	74,884	0,1851	-3,277	10,74
7	70,93	75,069	74,851	0,1851	-4,139	17,134
8	77,88	75,036	76,808	0,1851	2,8444	8,0906
9	78,66	76,993	76,694	0,1851	1,6668	2,7781
10	90,22	76,879	79,842	0,1851	13,341	177,97
11	75,89	80,027	75,591	0,1851	-4,137	17,119
12	65,16	75,776	74,132	0,1851	-10,62	112,71
13	67,03	74,317	75,162	0,1851	-7,287	53,101
14	68,01	75,347	75,334	0,1851	-7,337	53,834
15	69,51	75,52	75,857	0,1851	-6,01	36,115
16	88,35	76,042	80,691	0,1851	12,308	151,5
17	81,68	80,876	77,956	0,1851	0,804	0,6464
18	73,24	78,141	76,693	0,1851	-4,901	24,021

t	$У_t = X_{4t}$	$П_t$	a_{0t}	a_{1t}	e_t	e_t^2
19	65,12	76,878	75,138	0,1851	-11,76	138,25
20	81,52	75,323	79,88	0,1851	6,1971	38,405
21	78,66	80,065	78,136	0,1851	-1,405	1,9746
22	80,81	78,321	79,309	0,1851	2,4893	6,1965
23	80,06	79,494	79,006	0,1851	0,5657	0,3201
24	77,95	79,191	78,733	0,1851	-1,241	1,5404
25		78,918				
Сумма						1356,7

$\lambda_1 =$ 0

$\lambda_2 =$ 0

$\lambda_3 =$ 0,25381

E = 7,853

Протокол расчетов по функции “Линейн”

0,1851	73,2836
0,2381	3,4025
0,0267	8,07524
0,6041	22
39,394	1434,61

На рис. 14.3 представлен результат расчетов по моделям Бокса-Жденкинса

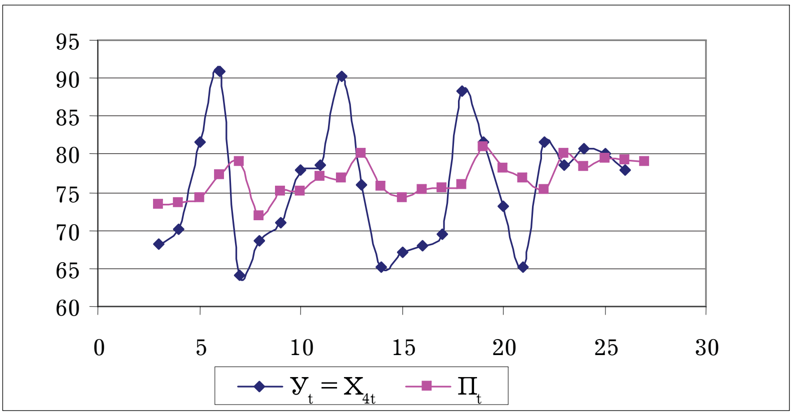


Рис. 14.3. Результаты расчетов по модели Бокса-Дженкинса

Анализ проведенных расчетов показывает, что коэффициенты сглаживания λ_1, λ_2, λ_3, определенные по программе “Поиск решения”, являются оптимальными, при этом значение коэффициента λ_3 отличается от нуля. Модель-Бокса — Дженкинса плохо воспроизводит выбросы, ошибка модели уменьшилась по србнению с рассмотренными моделями.

Вывод. Модель Бокса-Дженкинса для данных, содержащих сезонную волну, позволила улучшить модель по сравнению с линейной моделью.

Модель Уинтерса

Модель Уинтерса реализуем по схеме расчетов модели Хольда с такими добавлениями:

— необходимо визуально по графику временного ряда определить его период Т сезонной (Т = 3 месяца) или иной периодической составляющей;

— добавить столбец коэффициентов сезонности $w_{t\text{-}T}$, при изменении t от 1 до Т

$$w_t = \lambda_2 \frac{У_t}{a_{0,t}},$$

$$a_{0,t} = \lambda_1 У_t + (1-\lambda_1)(a_{0,t-1} + a_{1,t-1}),$$

$$a_{1,t} = \lambda_3(a_{0,t} - a_{0,t-1}) + (1 \quad \lambda_3)a_{1,t-1};$$

при t > Т

$$w_t = \lambda_2 \frac{У_t}{a_{0,t}} + (1-\lambda_2)w_{t\text{-}T},$$

$$a_{0,t} = \lambda_1 \frac{У_t}{w_{t\text{-}T}} + (1-\lambda_1)(a_{0,t-1} + a_{1,t-1}),$$

$$a_{1,t} = \lambda_3(a_{0,t} - a_{0,t-1}) + (1 \quad \lambda_3)a_{1,t-1};$$

— прогноз, сделанный в момент t на один такт времени вперед, равен:

$$П_{t+1} = (a_{0,t} + a_{1,t}) + w_{t+1\text{-}T}.$$

Модель Уинтерса содержит три параметра сглаживания λ_1, λ_2, λ_3 ($0<\lambda_1$, λ_2, $\lambda_3 <1$).

Примечание. Практическая реализации модели Уинтерса не позволила учесть сезонную волну и улучшить модель по сравнению с линейной моделью, поэтому не приводим результатов расчетов.

Модель сезонности Тейла-Вейджа

Модель Тейла-Вейджа реализуем по схеме расчетов модели Хольда с такими добавлениями:

— необходимо визуально по графику временного ряда определить его период T сезонной (T = 3 месяца) или иной периодической составляющей;

— добавить столбец коэффициентов сезонности w_{t-T}, при изменении t от 1 до T

$$w_t = (1-\lambda_1)\lambda_3(У_t — П_t);$$

при t > T

$$w_t = w_{t-T} + (1-\lambda_1)\lambda_3(У_t — П_t)$$

— прогноз, сделанный в момент t на один такт времени вперед, равен:

$$П_{t+1} = (a_{0,t} + a_{1,t}) + w_{t+1-T},$$

где $a_{0,t} = a_{0,t-1} + a_{1,\,t-1} + \lambda_1 (У_t — П_t);$

$$a_{1,t} = a_{1,t-1} + \lambda_1\lambda_2 (У_t — П_t);$$

$\lambda_1, \lambda_2, \lambda_3$ $(0<\lambda_1, \lambda_2, \lambda_3 <1)$ — параметры адаптации.

Расчеты по модели Тейла-Вейджа представлены в табл. 14.4.

Таблица 14.4

Расчеты по модели Тейла-Вейджа для данных X_4

t	$У_t = X_{4t}$	$П_t$	a_{0t}	a_{1t}	W_t	e_t^2
0			73,284	0,1851		
1	68,26	73,469	73,469	0,1851	-1,964	27,13
2	70,17	73,654	73,654	0,1851	-1,314	12,136
3	81,51	73,839	73,839	0,1851	2,8931	58,847
4	90,96	74,024	74,024	0,1851	6,3872	286,83
5	64,1	74,209	74,209	0,1851	-3,812	102,19
6	68,65	74,394	74,394	0,1851	-2,166	32,994
7	70,93	72,615	74,579	0,1851	-2,6	2,8385

t	$У_t = X_{4t}$	$П_t$	a_{0t}	a_{1t}	W_t	e_t^2
8	77,88	73,45	74,764	0,1851	0,3567	19,621
9	78,66	77,842	74,949	0,1851	3,2014	0,6685
10	90,22	81,522	75,134	0,1851	9,6676	75,663
11	75,89	71,507	75,319	0,1851	-2,159	19,21
12	65,16	73,338	75,505	0,1851	-5,251	66,884
13	67,03	73,09	75,69	0,1851	-4,885	36,722
14	68,01	76,231	75,875	0,1851	-2,744	67,592
15	69,51	79,261	76,06	0,1851	-0,476	95,086
16	88,35	85,913	76,245	0,1851	10,587	5,9413
17	81,68	74,27	76,43	0,1851	0,6349	54,901
18	73,24	71,364	76,615	0,1851	-4,543	3,5176
19	65,12	71,915	76,8	0,1851	-7,448	46,172
20	81,52	74,241	76,985	0,1851	0,0012	52,979
21	78,66	76,694	77,17	0,1851	0,2653	3,8644
22	80,81	87,942	77,355	0,1851	7,8971	50,869
23	80,06	78,175	77,54	0,1851	1,3457	3,5519
24	77,95	73,182	77,726	0,1851	-2,745	22,731
25		70,463				
Сумма						1148,9

$\lambda_1 =$	0
$\lambda_2 =$	0,0247
$\lambda_3 =$	0,37713

Протокол расчетов по функции “Линейн”.

0,1851	73,2836
0,2381	3,4025
0,0267	8,07524
0,6041	22
39,394	1434,61

E =	7,2267,
T =	6

На рис. 14.4 представлены результаты расчетов по моделям **Тейла-Вейджа**

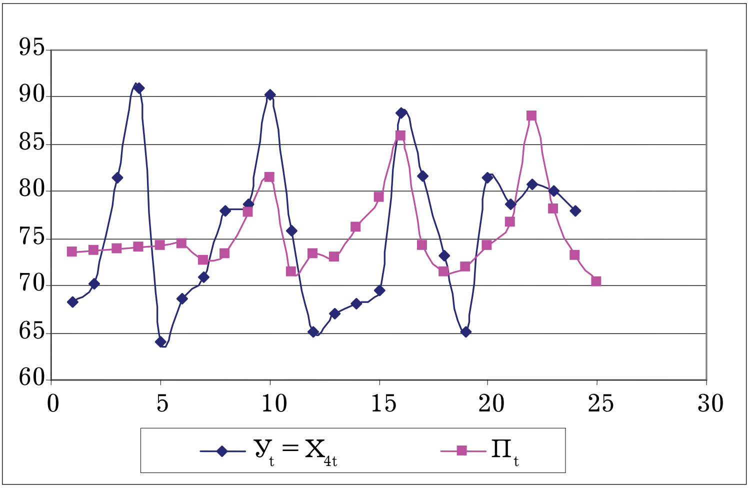


Рис. 14.4. Результаты расчетов по модели Тейла-Вейджа

Анализ проведенных расчетов показывает, что коэффициенты сглаживания λ_1, λ_2, λ_3, определенные по программе “Поиск решения”, являются оптимальными, при этом значения коэффициентов λ_2 и λ_3 отличается от нуля. Модель Тейла-Вейджа хорошо воспроизводит периодические выбросы, ошибка модели уменьшилась по сравнению с рассмотренными моделями.

Примечание. Модель Тейла-Вейджа хорошо воспроизводит выбросы, только с одним ограничением: выбросы должны регулярно повторяться с одинаковым периодом. Это ограничение снижает ценность данной модели, однако не уменьшает ее достоинства, так как эта одна из немногих адаптивных моделей прогнозирования, которая лишена основного недостатка всех адаптивных методов прогнозирования и авторегрессионных моделей, состоящее в плохом воспроизведении выбросов.

Вывод. Модель Тейла-Вейджа для данных, содержащие сезонную волну, позволила улучшить модель по сравнению с линейной моделью.

Сравнение характеристик полученных адаптивных моделей представленно в табл. 14.5.

Таблица 14.5

Сводная характеристика адаптивных методов прогнозирования

№ п/п	Модель	Ошибка модели (E)	Прогноз (t=25)
1	Линейная	8,07524	77,911
2	Хольта	8,9862	78,529
3	Бокса-Дженкинса	7,853	78,918
4	Уинтерса	-	-
5	Тейла-Вейджа	7.2267	70,463

Анализ табл. 14.5 показывает, что для исходных данных, содержащих сезонную волну с периодом T = 6 месяцам, наилучшей адаптивной моделью является модель Тейла-Вейджа.

Вывод. Следует обратить особое внимание на модель Тейла-Вейждa и попытаться применить ее для других временных рядов.

Задание второго уровня сложности

Задание. Сравните результаты расчетов по лучшей адаптивной модели Тейла-Вейджа с реализацией аналитической модели периодических колебаний.

Задание третьего уровня сложности

Задание. Попытайтесь реализовать модель Уинтерса для воспроизведения сезонной составляющей временного ряда.

Задание четвертого уровня сложности

Задание. Модифицируйте модель Уинтерса для воспроизведения сезонной составляющей временного ряда.

Выходной тест

Почему адаптивные методы прогнозирования, кроме тех, которые учитывают сезонность, плохо воспроизводят выбросы?

15. ПРОВЕДЕНИЕ РАСЧЕТОВ С ПОМОЩЬЮ ПРОГРАММЫ МДИПЛОМ

Постановка задачи

Объект — временные ряды экономических показателей.

Предмет — регулярности экономического показателя.

Цель — выбрать наилучшую модель, позволяющую получить наиболее обоснованный прогноз.

Актуальность — повышение точности прогноза временного ряда всегда высоко ценилось в эконометрическом моделировании.

Рабочая гипотеза — временной ряд имеет определенную структуру.

Модели и методы:

Однофакторный прогноз

1. Визуальный метод анализа определения регулярностей временного ряда.
2. Определение тенденции динамики экономического показателя.
3. Эконометрический анализ линейной модели.
4. Адаптивный метод прогнозирования (метод Брауна).
5. Прогнозирование по модели автокорреляции остатков
6. Прогнозирование по модели авторегрессии зависимой переменной.
7. Прогнозирование по модели гетероскедастичности остатков.
8. Прогнозирование по модели цикличности экономических процессов.
9. Прогнозирование по модели взвешенной регрессии.

10. Прогнозирование по модели с участием качественной переменной.

11. Сравнительный метод анализа методов прогнозирования.

Многофакторный прогноз

1. Модель производственной функции.
2. Построение многофакторной модели.
3. Модель сосредоточенного лага.
4. Модель распределенного лага.
5. Системы одновременных уравнений.
6. Анализ качественных переменных

Способ — использование программы МДиплом

Задачи:

— построить график динамики зависимой переменной;
— произвести анализ регулярностей временного ряда зависимой переменной;
— построить различные модели, учитывающие обнаруженные регулярности;
— определить оптимальные значения основного коэффициента модели, в качестве критерия оптимальности можно выбрать коэффициент детерминации, или ошибку модели, или визуальный анализ реальности согласованности фактических и расчетных значений зависимой переменной;
— получить точечный и интервальный прогноз;
— представить результаты расчетов в графическом виде;
— сравнить результаты прогноза по разным моделям и сделать вывод по реально ожидаемому значению прогноза зависимой переменной;
— построить многофакторные модели.

Ожидаемый результат — использование наилучшей моделей временного ряда позволит получить наиболее обоснованный прогноз.

Модель для сравнения — линейная модель временного ряда.

Задания первого уровня сложности

Задание 1. Подготовить устные сообщения по вопросам самостоятельной работы.

1. Укажите структуру динамических моделей.
2. Изложите последовательность выполнения действий для реализации динамической модели.
3. Укажите область эффективного использования динамических моделей. [3, с. 397–399]

Задание 2. Пройдите входное тестирование.

Выберите вариант правильного ответа

1. На рис. 15.1 представлена структура динамической модели, созданной в Excel.

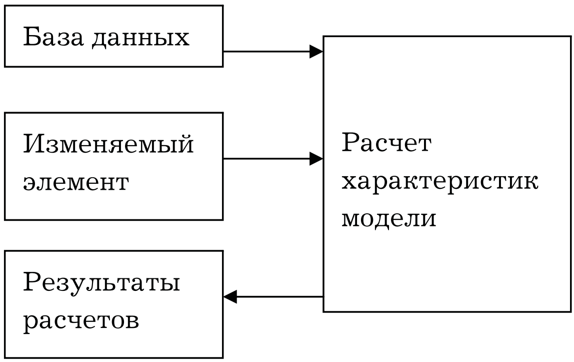


Рис. 15.1. Структура динамической модели

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.

а) да;

б) нет.

2. Пример динамической модели адаптивного прогнозирования по методу Брауна представлен на рис. 15.2.

Динамическая модель приводится в действие кнопкой изменения коэффициента дисконтирования. Изменение коэффициента дисконтирования приводит к изменению всех характеристик: ошибки модели, расчетных значений, точечного и интервального прогноза, 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 и интервальный прогноз будет наименьший.

При этом пользователю ясно, куда нажимать, куда смотреть, что надо получить — необходимые составляющие при выполнении любых работ.

Рис. 15.2. Динамическая модель адаптивного прогнозирования Брауна, реализованная в Excel

а) да;

б) нет.

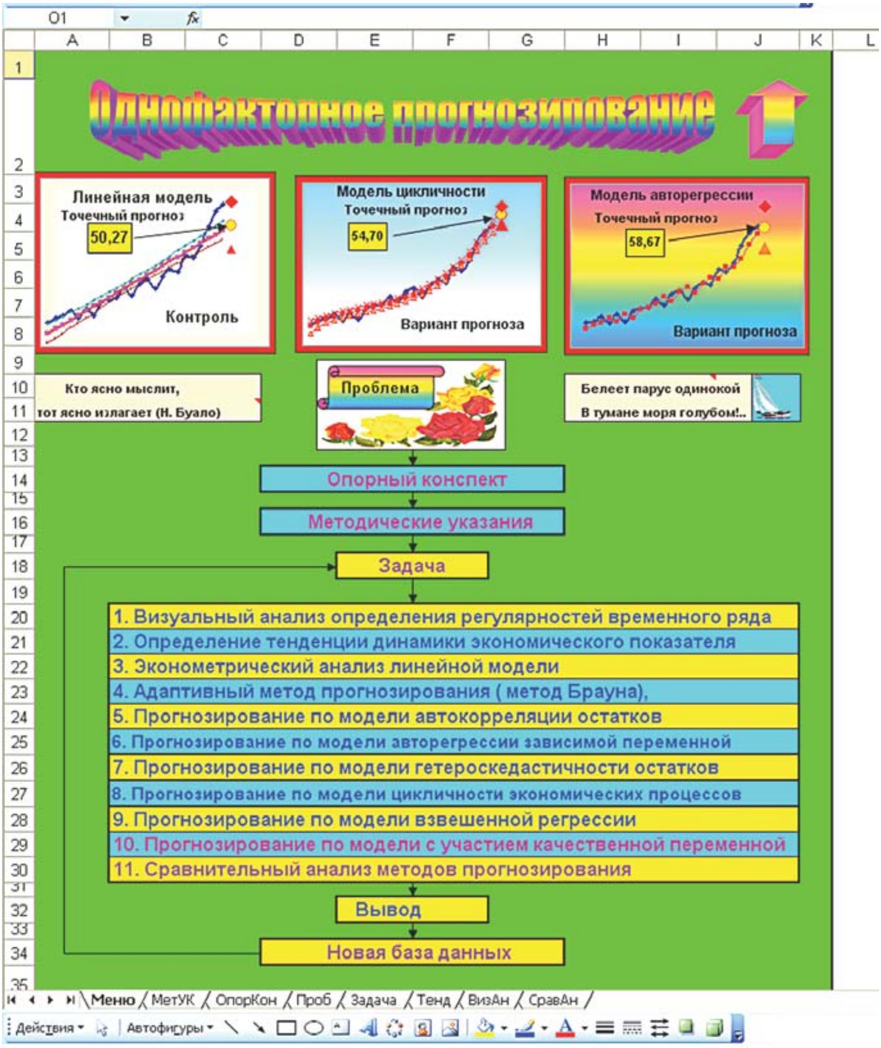
3. Динамические модели, реализованные в Excel, можно эффективно использовать по следующим темам: метод

наименьших квадратов (определение зависимости ошибки модели от ее коэффициентов), спецификация модели (зависимость графика математической функции от ее коэффициентов); автокорреляция, авторегрессия, адаптивные методы прогнозирования (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 в зависимости 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).

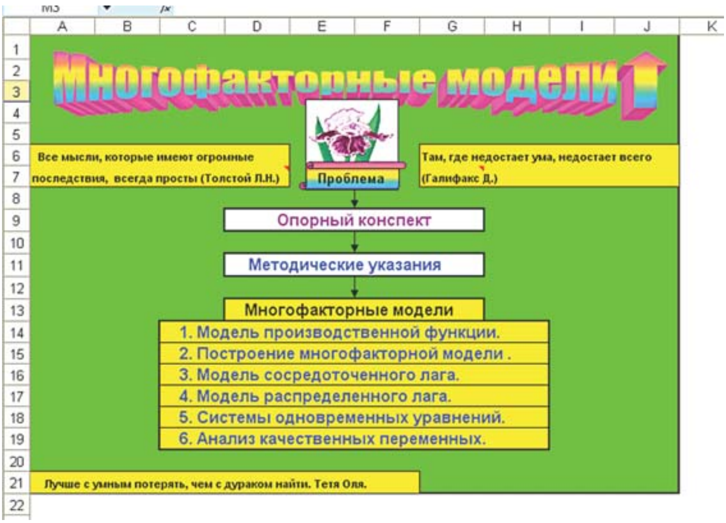
а) да;

б) нет.

4. Меню программы МДиплом имеет следующий вид для однофакторного и многофакторного прогнозирования (рис. 15.3).


Однофакторное прогнозирование
Линейная модель
Точечный прогноз
50,27
Контроль
Модель цикличности
Точечный прогноз
54,70
Вариант прогноза
Модель авторегрессии
Точечный прогноз
58,67
Вариант прогноза
Кто ясно мыслит,
тот ясно излагает (Н. Буало)
Проблема
Белеет парус одинокой
В тумане моря голубом!..
Опорный конспект
Методические указания
Задача
1. Визуальный анализ определения регулярностей временного ряда
2. Определение тенденции динамики экономического показателя
3. Эконометрический анализ линейной модели
4. Адаптивный метод прогнозирования (метод Брауна),
5. Прогнозирование по модели автокорреляции остатков
6. Прогнозирование по модели авторегрессии зависимой переменной
7. Прогнозирование по модели гетероскедастичности остатков
8. Прогнозирование по модели цикличности экономических процессов
9. Прогнозирование по модели взвешенной регрессии
10. Прогнозирование по модели с участием качественной переменной
11. Сравнительный анализ методов прогнозирования
Вывод
Новая база данных
Меню
МетУК
ОпорКон
Проб
Задача
Тенд
ВизАн
СравАн
Действия
Автофигуры

а)



б)

Рис. 15.3. Мен ю программы МДиплом для однофакторного (а) и многофакторного (б) прогнозирования

а) да;

б) нет.

Задание 3. Решите задачу.

Задача. Имеются значения временного ряда, представленные в табл. 15.1.

Таблица 15.1.

Данные временного ряда

N% п/п	t	$У_t$	N% п/п	t	$У_t$
1	1	10	9	9	19
2	2	12	10	10	20
3	3	13	11	11	24
4	4	15	12	12	21
5	5	12	13	13	26
6	6	16	14	14	28
7	7	13	15	15	24
8	8	18	16	16	29

N% п/п	t	$У_t$
17	17	32
18	18	31
19	19	38
20	20	37
21	21	42

N% п/п	t	$У_t$
22	22	50
23	23	56
24	24	59
Ожидаем	25	?

где t — время (месяцы за два года);

$У_t$ — значения показателя экономического объекта (усл. ед.).

Примечание. В зависимости от специальности (прикладная информатика; финансы и кредит; бухгалтерский учет, анализ и аудит; мировая экономика изучаемыми показателями соответственно могут быть: количество документов, обработанных информационной системой предприятия; размер инфляции или размер поступления денежных средств от вклада в сберегательный банк; размер дебиторской задолженности или размер перерасхода денежных средств, выделенных на зароботную плату; размер экспорта товаров по ассортименту зарубеж.

Необходимо:

1. Определить регулярности временного ряда.

2. Выполнить эконометрический анализ линейной модели экономического показателя.

3. Улучшить прогноз, полученный по линейной модели, с использованием различных моделей.

4. Произведите сравнительный анализ полученных прогнозов с использованием всех изученных моделей.

5. Самостоятельно повторите расчеты для новой базы данных.

Решение задачи выполнить с помощью программы МДиплом.

Алгоритмы реализации моделей уже были изучены, поэтому необходимо привести протоколы расчетов.

Решение

Решение задачи будем проводить в такой последовательности.

1. Запустите программу МДиплом. Если возникнут затруднения, то обратитесь к преподавателю.

2. Войдите в меню программы (рис. 15.3) и выполните пункты: “Проблема”, “Опорный конспект”, “Методические указания”, “Задача”.

В пункте “Задача” введите исходные данные, взятые из табл. 15.1.

3. Вернитесь в меню и выполните пункт

“1. Визуальный анализ определения регулярностей временного ряда”.

Результат выполнения задания изображен на рис. 15.4.

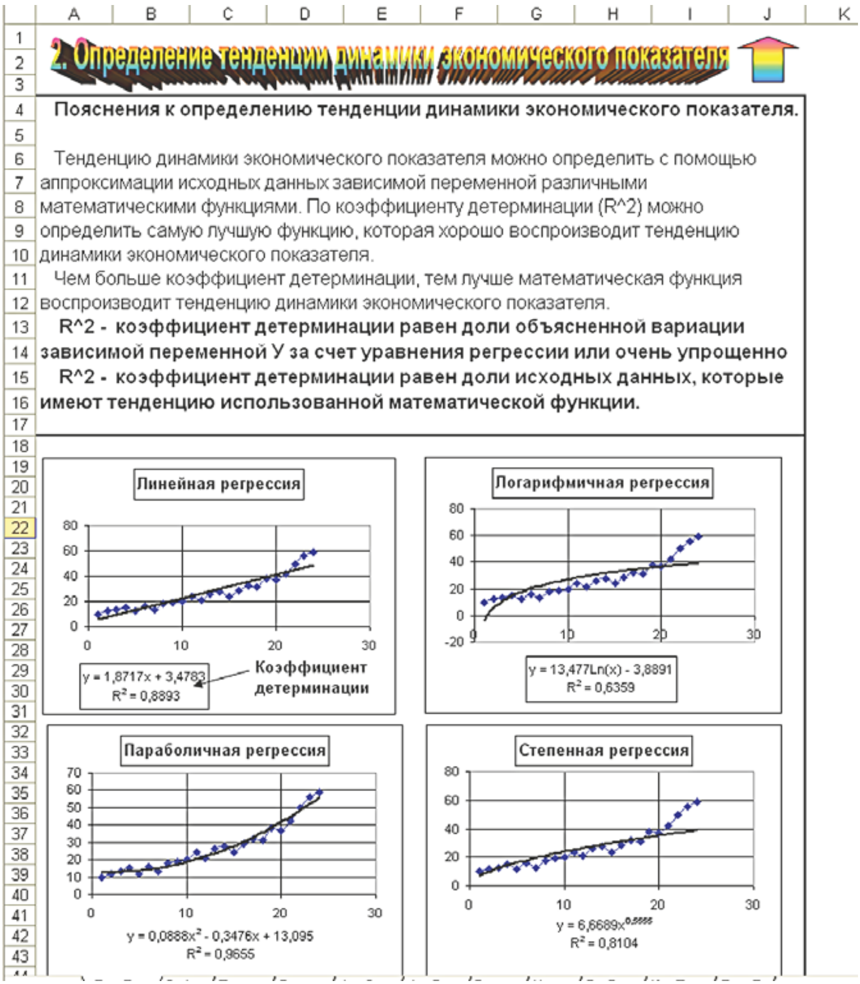
Рис. 15.4. Результат выполнения первого пункта меню программы МДиплом

На рис. 15.4 горизонтальные стрелки указывают на дополнительную информацию, вертикальные стрелки на возврат в меню.

4. Вернитесь в меню и выполните пункт

“2. Определение тенденции динамики экономического показателя”.

Результаты выполнения задания представлены на рис. 15.5.



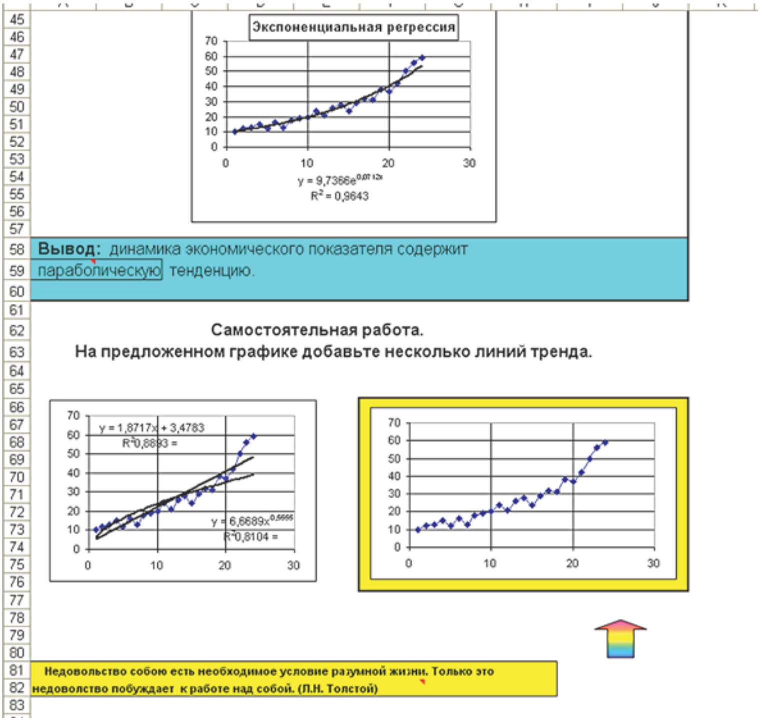


Рис. 15.5. Результат выполнения второго пункта меню программы МДиплом

5. Вернитесь в меню и выполните пункт
“3. Эконометрический анализ линейной модели”.
Результаты выполнения задания представлены на рис. 15.6.

3. Эконометрический анализ линейной модели

Пояснения к эконометрическому анализу линейной модели.

Необходимо специфицировать модель линейной функцией вида

Ур= а0 + а1*Х

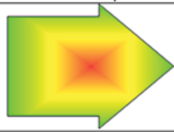
Расчеты характеристик линейной модели с помощью функции **Линейн.**

База данных условия задачи.

Х	У
1	10
2	12
3	13
4	15
5	12
6	16
7	13
8	18
9	19
10	20
11	24
12	21
13	26
14	28
15	24
16	29
17	32
18	31
19	38
20	37
21	42
22	50
23	56
24	59

1,8717	3,4783
0,1408	2,0121
0,8893	4,7754
176,67	22
4028,9	501,71

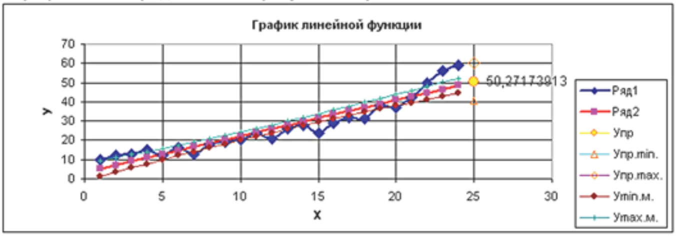
Запишем результаты расчетов в общепринятом обозначении.

	A–G	H–J
37	Протокол расчетов.	Пояснения к расчетам.
38	Yi= 3,47826 + 1,87174 * Xi + ei	
39	Sa0= 2,01213 Sa1= 0,14082	
40	ta0= 1,72864 ta1= 13,2917 tтаб. = 2,07387	
41	E= 4,77544 E= 17,7691 (%)	
42	R^2= 0,88926	
43	F= 176,669 Fтаб = 4,30095	

44	Где Sa - ошибка коэффициента a,
45	E - ошибка модели
46	R^2 - коэффициент детерминации,
47	ta - критерий Стьюдента для коэффициента a,
48	tтаб - табличное значение критерия Стьюдента,
49	F - критерий Фишера.
50	F таб - табличное значение критерия Фишера.
51	Представляем результаты эконометрического анализа.
52	1. Вычисление доли объясненной вариации.
53	88,92633 % исходных данных имеют тенденцию выбранной функции.
54	2. Проверка достоверности модели.
55	По критерию Фишера проверяется достоверность модели.
56	Так как F= 176,669 > Fтаб(альфа=0,05, M1=k - 1, M2= n - k) = 4,30095 , то модель
57	является достоверной с вероятностью 0,95
58	3. Проверка достоверности коэффициентов модели.
59	По критерию Стьюдента проверяется достоверность коэффициентов модели.
60	Так как ta0 = 1,72864 < tтаб. = 2,07387 , то коэффициент a0
61	статистически не отличается от нуля
62	Так як ta1 = 13,2917 > tтаб. = 2,07387 , то коэффициент a1
63	с вероятностью 0,95 отличается от нуля
64	4. Вычисление точечного и интервального прогноза.
65	Наиболее вероятное значение зависимой переменной будет равно Yпр= 50,2717
66	Фактическое прогнозное значение зависимой переменной с вероятностью 0,95 будет
67	находиться в интервале от 40,3681 до 60,1754

находиться в интервале от 40,3681 до 60,1754

5. Графическое представление результатов расчетов.



Где Ymin. г. - 95% нижний доверительный интервал математического ожидания У,
Ymax. г. - 95% верхний доверительный интервал математического ожидания У.

Вывод: линейная регрессия имеет следующие характеристики
ошибка модели равна E= 17,769 %, точечный прогноз равен Упр = 50,272

Примечания.

Коэффициенты линейной модели, определенные методом наименьших квадратов являются оптимальными и позволяют получить наименьшую ошибку.

Проверим данное утверждение. Для этой цели изменим найденные оптимальные коэффициенты в небольших пределах и будем наблюдать за ошибкой модели и графиком расчетных значений зависимой переменной, вычисленные по изменяемым коэффициентам.

Ожидаемый результат- если изменять оптимальные значения коэффициентов, то ошибка модели будет возрастать.

При решении контрольного примера были получены следующие оптимальные значения коэффициентов и ошибка линейной модели

Ур = a0 + a1*X

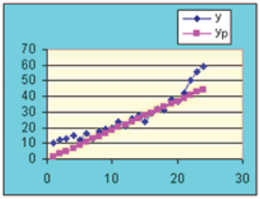
a0 = 3,4783 **a1=** 1,8717 **ошибка модели E% =** 17,769 **%**

Изменяйте значения коэффициентов **a0, a1** и наблюдайте за ошибкой модели и графиком Ур.

a0 = -0,522 **a1=** 1,8717 **E% =** 23,609 **%**

Таблица расчета ошибки модели.

X	У	Ур	e^2	Ур	e^2
1	10	1,35	74,823		
2	12	3,2217	77,058		
3	13	5,0935	62,513		
4	15	6,9652	64,558		
5	12	8,837	10,005		
6	16	10,709	27,998		
7	13	12,58	0,176		
8	18	14,452	12,587		
9	19	16,324	7,1614		
10	20	18,196	3,2557		
11	24	20,067	15,465		
12	21	21,939	0,882		
13	26	23,811	4,7923		
14	28	25,683	5,3703		
15	24	27,554	12,633		
16	29	29,426	0,1816		
17	32	31,298	0,493		
18	31	33,17	4,707		
19	38	35,041	8,7539		
20	37	36,913	0,0076		
21	42	38,785	10,338		
22	50	40,657	87,301		
23	56	42,528	181,49		
24	59	44,4	213,16		
Сумма	645		885,71		
Среднее	26,875				



Поставь над собой сто учителей - они окажутся бессильными, если ты не можешь сам заставлять себя и сам требовать от себя. (В.А. Сухомлинский)

Рис. 15.6. Результат выполнения третьего пункта меню программы МДиплом

5. Вернитесь в меню и выполните пункт
“4. Адаптивный метод прогнозирования (метод Брауна)”.
Результаты выполнения задания представлены на рис. 15.7.

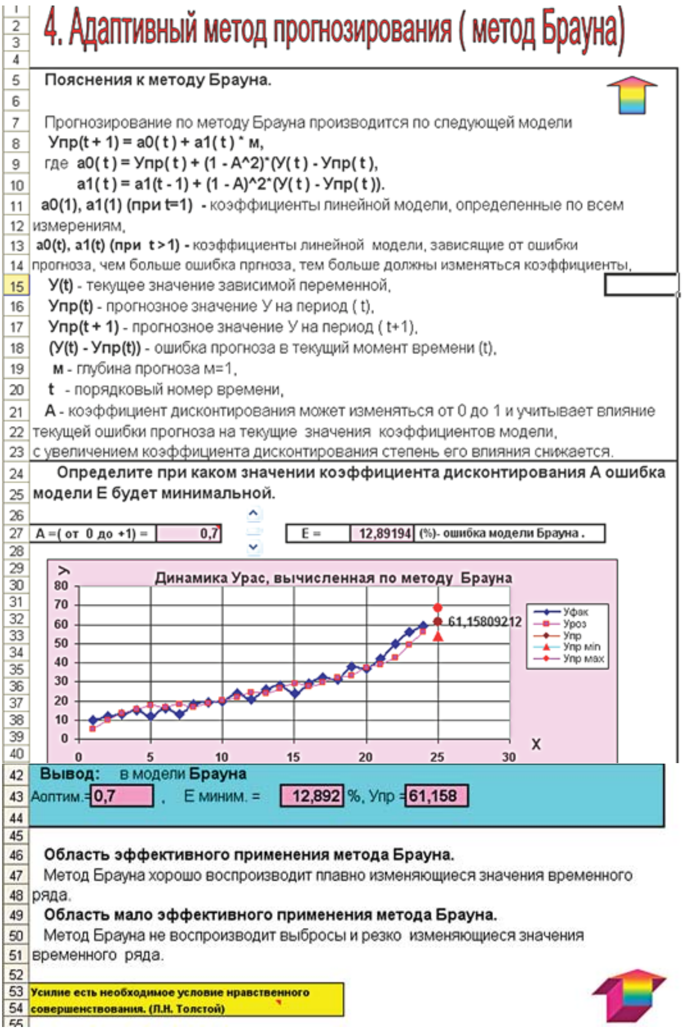


Рис. 15.7. Результат выполнения четвертого пункта меню программы МДиплом

6. Вернитесь в меню и выполните пункт
“5. Прогнозирование по модели автокорреляции остатков”.
Результаты выполнения задания представлены на рис. 15.8.

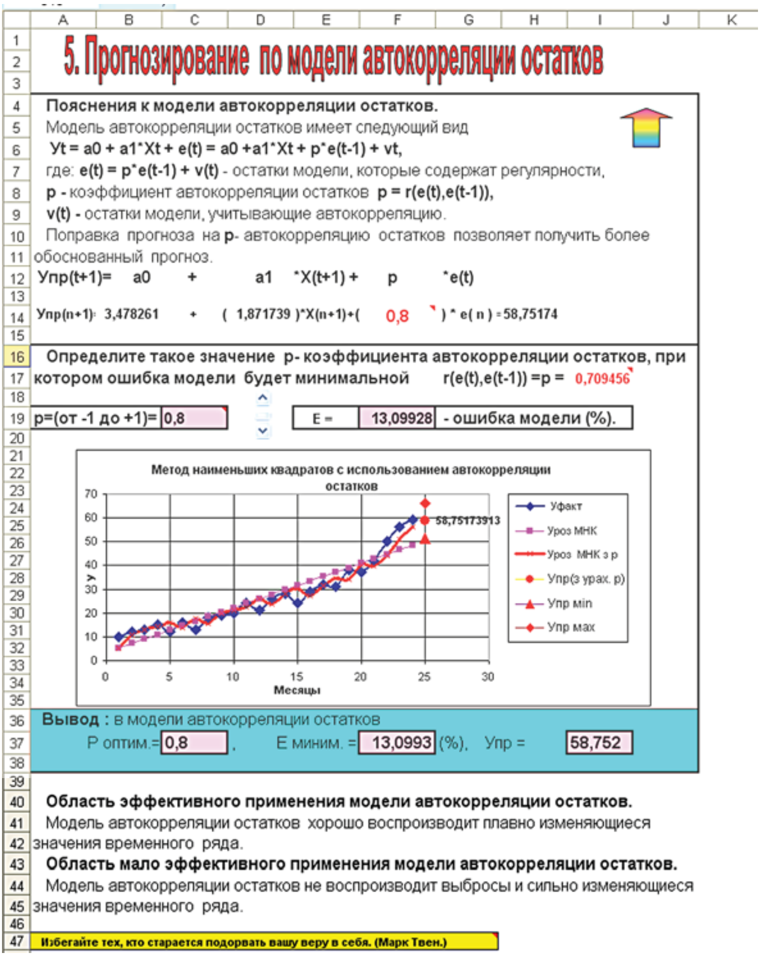


Рис. 15.8. Результат выполнения четвертого пункта меню программы МДиплом

7. Вернитесь в меню и выполните пункт

"6. Прогнозирование по модели авторегрессии зависимой переменной".

Результаты выполнения задания представлены на рис. 15.9.

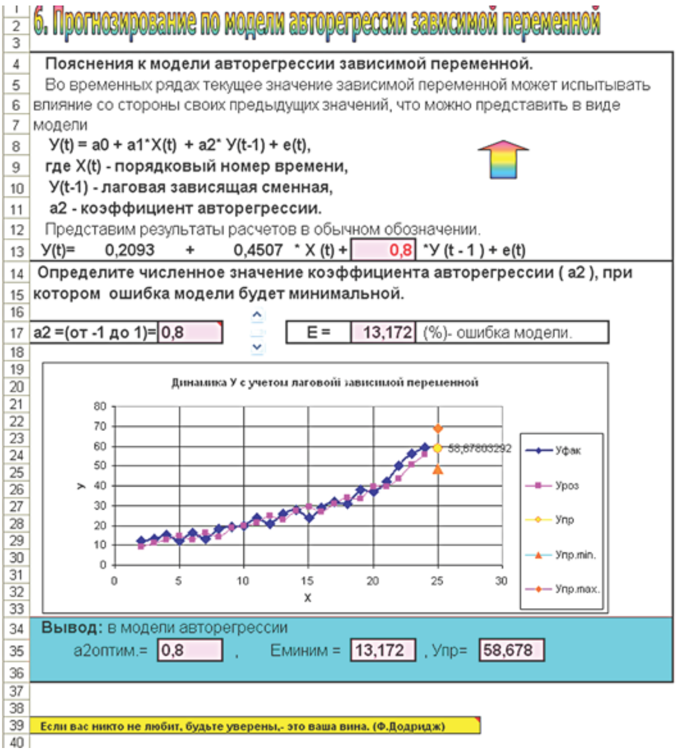


Рис. 15.9. Результат выполнения шестого пункта меню программы МДиплом

8. Вернитесь в меню и выполните пункт

“7. Прогнозирование по модели гетероскедастичнгости остатков остатков”.

Результаты выполнения задания представлены на рис. 15.10.

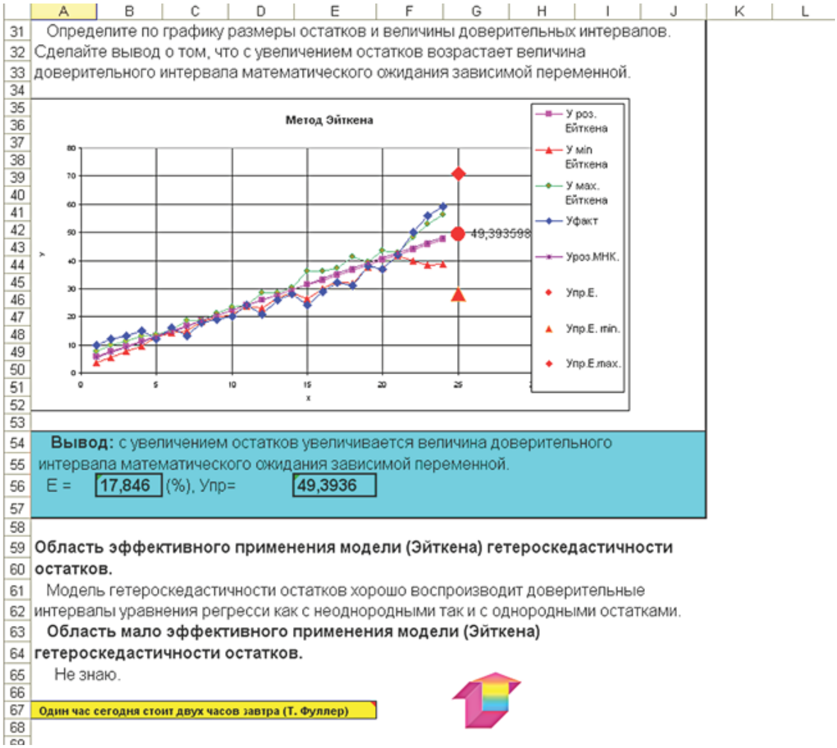


Рис. 15.10. Результат выполнения седьмого пункта меню программы МДиплом

9. Вернитесь в меню и выполните пункт

“8. Прогнозирование по модели цикличности экономических процессов”.

8. Прогнозирование по модели цикличности экономических процессов

Пояснения к модели циклических процессов.

Анализ циклических процессов временного ряда проводится в такой последовательности:

- приводится временный ряд к стационарному виду с помощью линейной тенденции,
- определяются периодические составляющие стационарного временного ряда методом фильтрации,
- определяются характеристики многофакторной модели с учетом периодических составляющих временного ряда.

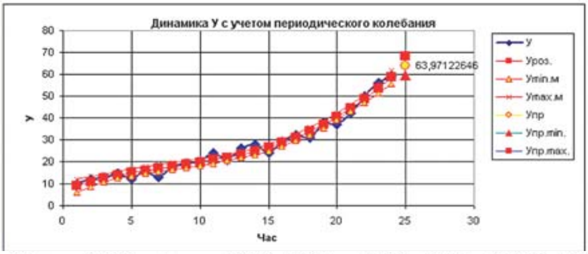
Метод фильтрации реализуем с помощью явления резонанса регрессионной модели, настроенной на определенную частоту.

Построим график периодограммы зависимости ошибки модели от периода циклической составляющей регрессионной модели.

Анализ ошибок модели по графику периодограммы позволит выявить локальные минимумы, которые отвечают периодам периодических составляющих временного ряда.

Введите период Т периодической составляющей в интервале допустимых значений от 3 до 2*n (n - объем выборки временного ряда), определите при каком значении Т ошибка модели Е будет минимальной.

Т=(от 3 до 40) = 48 E = **8,48029** (%) - ошибка модели.



У = -12,83 + 3,799 * X1 + -10,96 * X2 + 19,756 * X3

Где X1 - Время, $X2 = SIN\left(\frac{2*\pi*X1i}{T}\right)$, $X3 = COS\left(\frac{2*\pi*X1i}{T}\right)$

Вывод: в модели циклических процессов

Топтим. 48 , Еминим. = 8,4803 (%), Упр= 63,971

Область эффективного применения модели цикличности временного ряда.

Модель цикличности временного ряда хорошо воспроизводит периодические составляющие экономического процесса, если периоды циклических составляющих не сильно отличаются от длины временного ряда.

Область мало эффективного применения модели цикличности временного ряда.

Модель цикличности временного ряда плохо воспроизводит периодические составляющие экономического процесса, если периоды циклических составляющих значительно больше длины временного ряда.

Нужно любить то, что делаешь, и тогда труд- даже свмый грубый - возвышается до творчества. (М. Горький)

Рис. 15.11. Результат выполнения восьмого пункта меню программы МДиплом

Результаты выполнения задания представлены на рис. 15.11.

10. Вернитесь в меню и выполните пункт
“9. Прогнозирование по модели взвешенной регрессии”.
Результаты выполнения задания представлены на рис. 15.12.

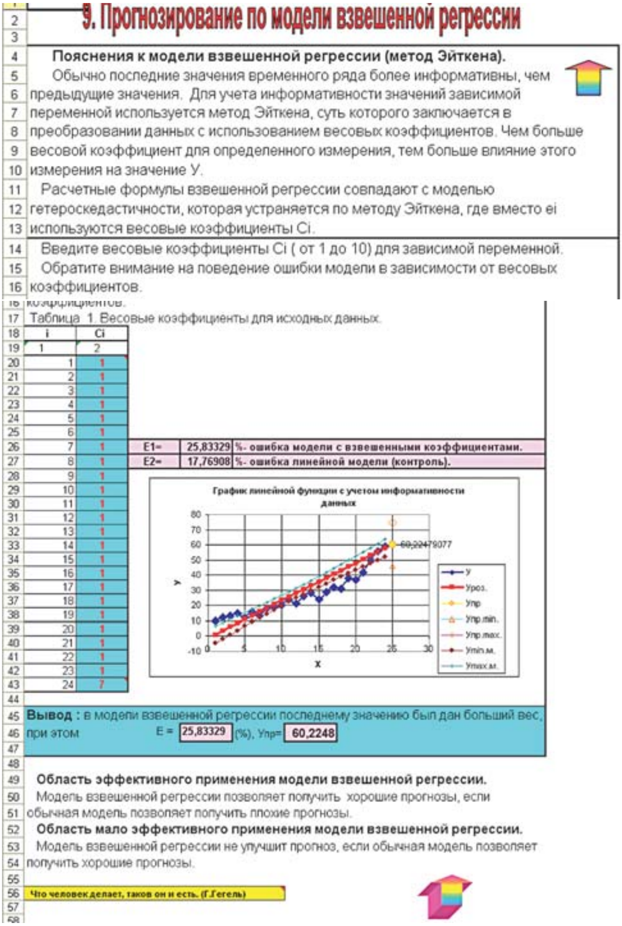
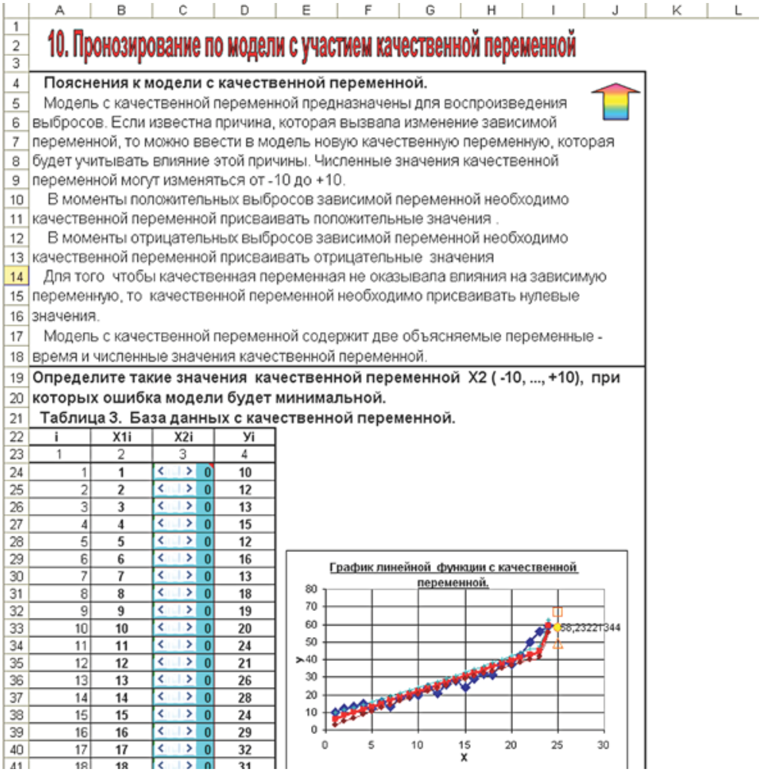


Рис. 15.12. Результат выполнения девятого пункта меню программы МДиплом

11. Вернитесь в меню и выполните пункт

“10. Прогнозирование по модели с участием качественной переменной”.

Результаты выполнения задания представлены на рис. 15.13.



10. Прогнозирование по модели с участием качественной переменной

Пояснения к модели с качественной переменной.

Модель с качественной переменной предназначены для воспроизведения выбросов. Если известна причина, которая вызвала изменение зависимой переменной, то можно ввести в модель новую качественную переменную, которая будет учитывать влияние этой причины. Численные значения качественной переменной могут изменяться от -10 до +10.

В моменты положительных выбросов зависимой переменной необходимо качественной переменной присваивать положительные значения .

В моменты отрицательных выбросов зависимой переменной необходимо качественной переменной присваивать отрицательные значения

Для того чтобы качественная переменная не оказывала влияния на зависимую переменную, то качественной переменной необходимо присваивать нулевые значения.

Модель с качественной переменной содержит две объясняемые переменные - время и численные значения качественной переменной.

Определите такие значения качественной переменной X2 (-10, ..., +10), при которых ошибка модели будет минимальной.

Таблица 3. База данных с качественной переменной.

i	X1i	X2i	Yi
1	2	3	4
1	1	0	10
2	2	0	12
3	3	0	13
4	4	0	15
5	5	0	12
6	6	0	16
7	7	0	13
8	8	0	18
9	9	0	19
10	10	0	20
11	11	0	24
12	12	0	21
13	13	0	26
14	14	0	28
15	15	0	24
16	16	0	29
17	17	0	32
18	18	0	31

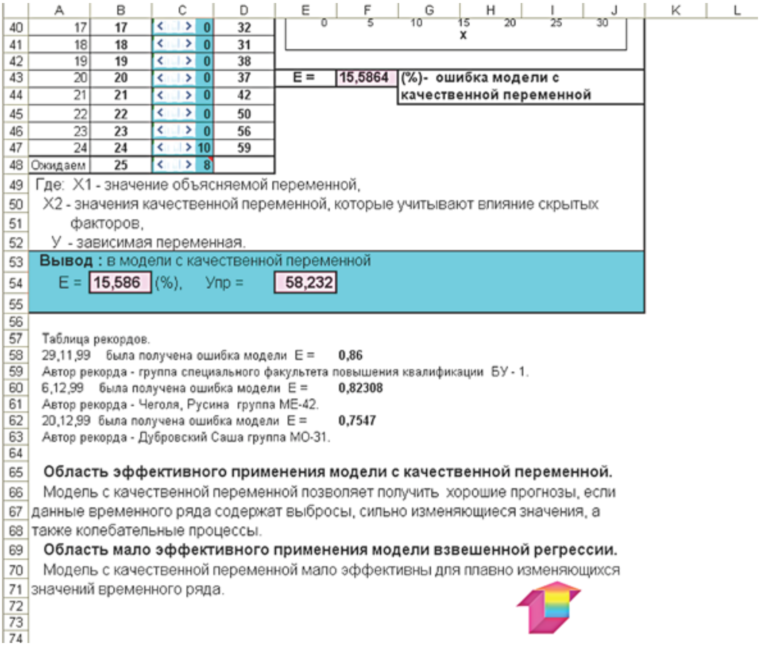


Рис. 15.13. Результат выполнения десятого пункта меню программы МДиплом

12. Вернитесь в меню и выполните пункт "11. Сравнительный анализ методов прогнозирования".

Результаты выполнения задания представлены на рис. 15.14.

11. Сравнительный анализ методов прогнозирования

Задача.

Имеется база данных временного ряда показателя деятельности экономического объекта. Необходимо получить точечный и интервальный прогноз на одну дату вперед с использованием различных эконометрических моделей, провести сравнительный анализ полученных моделей.

Сравнительный анализ точности прогнозирования по всем моделям.

Выводы по анализу прогнозных моделей.

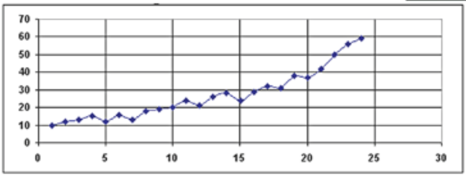
1. Визуальный анализ.

Вывод : динамика экономического показателя содержит:

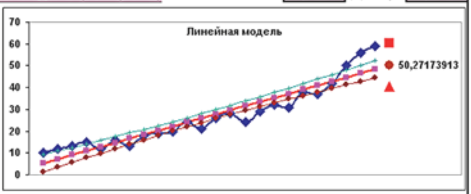
- параболическую тенденцию с

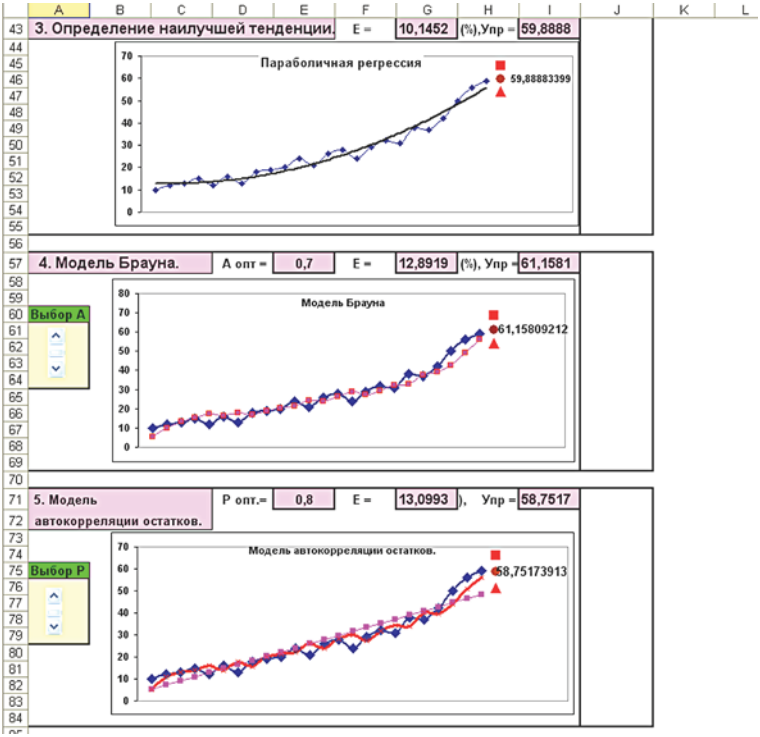
- однородными остатками

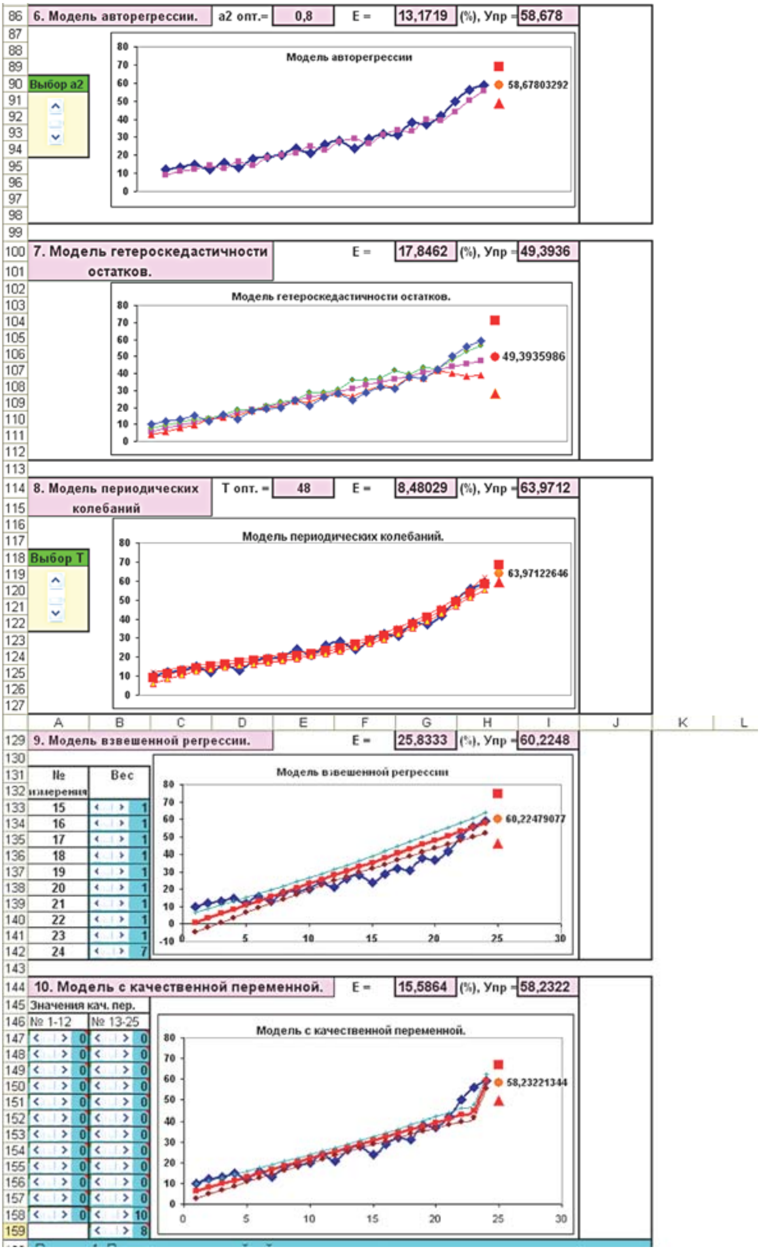
- без существенных выбросов Упр= 56,5



2. Линейная модель (контроль). E = 17,7691 (%), Упр = 50,2717




3. Определение наилучшей тенденции.
E =
10,1452
(%),Упр =
59,8888
Параболичная регрессия
59,88883399
4. Модель Брауна.
A опт =
0,7
E =
12,8919
(%), Упр =
61,1581
Выбор A
Модель Брауна
61,15809212
5. Модель автокорреляции остатков.
P опт.=
0,8
E =
13,0993
Упр =
58,7517
Выбор P
Модель автокорреляции остатков.
58,75173913



6. Модель авторегрессии.
a2 опт.=
0,8
E =
13,1719
(%), Упр =
58,678
Выбор a2
Модель авторегрессии
58,67803292
7. Модель гетероскедастичности остатков.
E =
17,8462
(%), Упр =
49,3936
Модель гетероскедастичности остатков.
49,3935986
8. Модель периодических колебаний
T опт. =
48
E =
8,48029
(%), Упр =
63,9712
Выбор T
Модель периодических колебаний.
63,97122646
9. Модель взвешенной регрессии.
E =
25,8333
(%), Упр =
60,2248
№ измерения
Вес
Модель взвешенной регрессии
60,22479077
10. Модель с качественной переменной.
E =
15,5864
(%), Упр =
58,2322
Значения кач. пер.
№ 1-12
№ 13-25
Модель с качественной переменной.
58,23221344

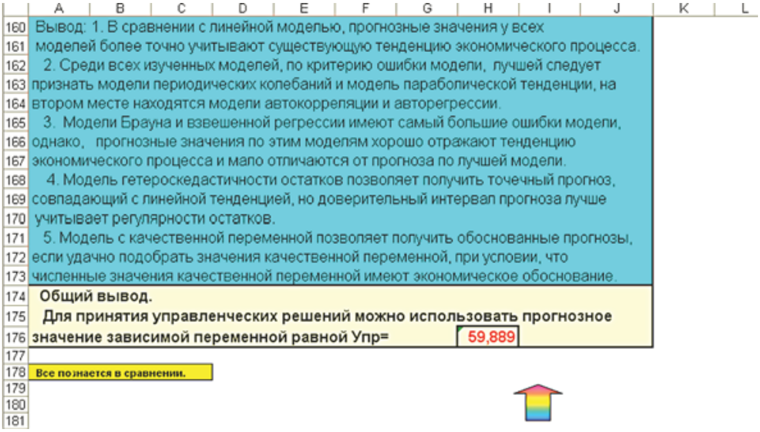


Рис. 15.14. Результат выполнения одиннадцатого пункта меню программы МДиплом

Задание второго уровня сложности

Задание. Выполните многофакторный прогноз с помощью программы МДиплом.

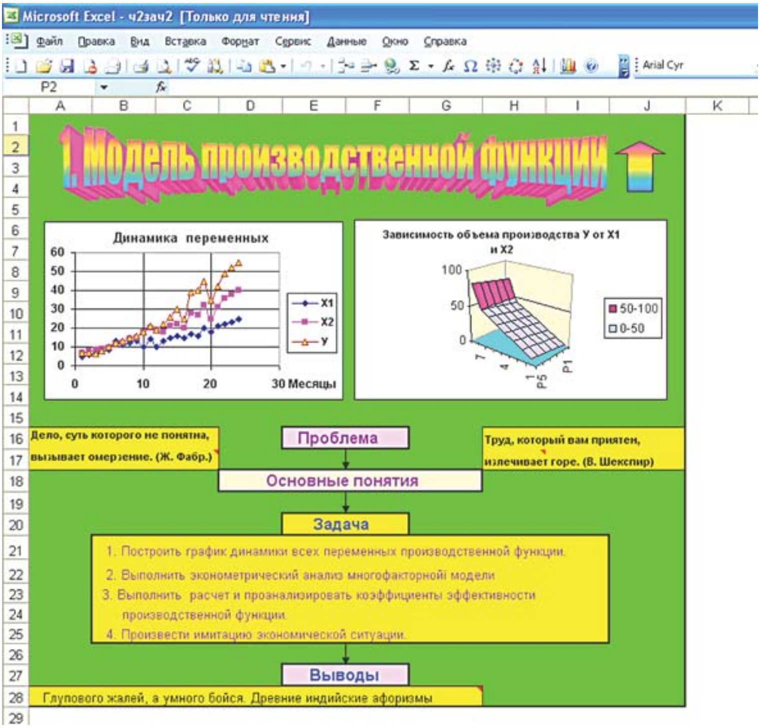
Решение.

Решение задачи будем проводить в такой последовательности.

1. Запустите программу МДиплом. Если возникнуть затруднения, то обратитесь к преподавателю.

2. Войдите в меню программы (см. рис. 15.3 (б) и выполните пункты: “Проблема”, “Опорный конспект”, “Методические указания”, “1. Модель производственной функции”.

Результаты выполнения задания представлены на рис. 15.15.



Microsoft Excel - ч2зач2 [Только для чтения]
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
Arial Cyr
P2
1. Модель производственной функции
Динамика переменных
Месяцы
X1
X2
У
Зависимость объема производства У от X1 и X2
50-100
0-50
Дело, суть которого не понятна, вызывает омерзение. (Ж. Фабр.)
Проблема
Труд, который вам приятен, излечивает горе. (В. Шекспир)
Основные понятия
Задача
1. Построить график динамики всех переменных производственной функции.
2. Выполнить эконометрический анализ многофакторной модели
3. Выполнить расчет и проанализировать коэффициенты эффективности производственной функции.
4. Произвести имитацию экономической ситуации.
Выводы
Глупового жалей, а умного бойся. Древние индийские афоризмы

Microsoft Excel - ч2зач2 [Только для чтения]

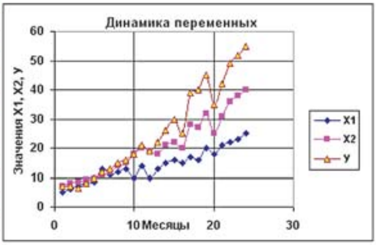
Глупового жалей, а умного бойся. Древние индийские афоризмы

Задача.

Есть среднемесячные данные статистической отчетности по однородным предприятиям, которые выпускают продукцию с использованием некоторых ресурсов.

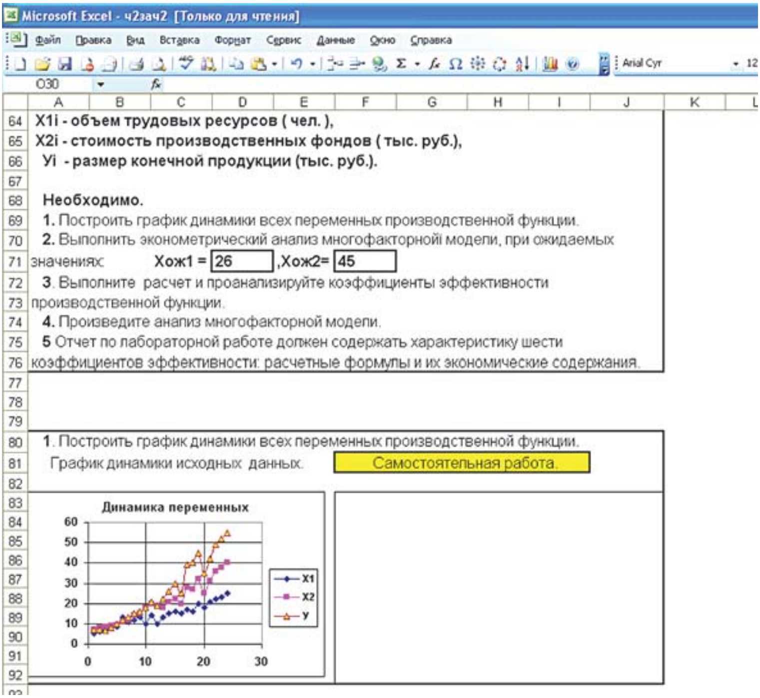
Таблица1 . Условные данные за отчетный год по однородным предприятиям.

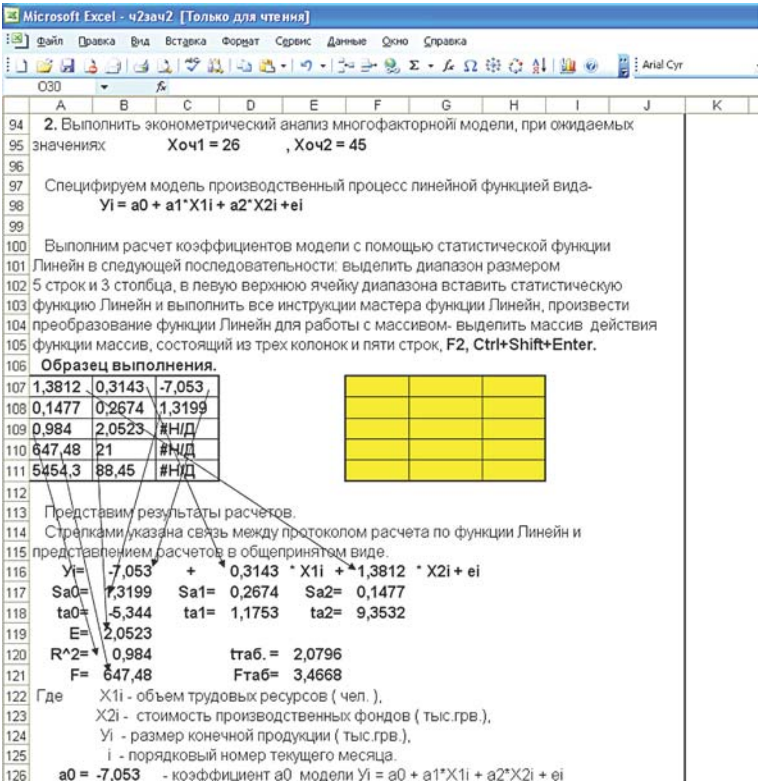
i	X1i	X2i	Yi
1	2	3	4
1	5	7	7
2	6	8	7
3	7	8,5	6,5
4	8	9	8
5	8,5	9,5	10
6	13	11	12
7	11	12	13
8	12	14	15
9	13	15	16
10	10	18	18
11	14	20	21
12	10	19	19
13	13	18	22
14	15	21	26
15	16	22	30
16	15	20	25
17	17	28	39
18	16	27	40
19	20	32	45
20	18	25	35
21	21	31	42
22	22	36	49
23	23	38	52
24	25	40	55
Ожидаем	26	45	
Сумма	338,5	489	612,5
Среднее	14,104	20,375	25,521



Где i - порядковый номер текущего месяца,

X1i - объем трудовых ресурсов (чел.)


Microsoft Excel - ч2зач2 [Только для чтения]
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
Arial Cyr
O30
X1i - объем трудовых ресурсов (чел.),
X2i - стоимость производственных фондов (тыс. руб.),
Yi - размер конечной продукции (тыс. руб.).
Необходимо.
1. Построить график динамики всех переменных производственной функции.
2. Выполнить эконометрический анализ многофакторноий модели, при ожидаемых
значениях: Хож1 = 26 ,Хож2= 45
3. Выполните расчет и проанализируйте коэффициенты эффективности
производственной функции.
4. Произведите анализ многофакторной модели.
5 Отчет по лабораторной работе должен содержать характеристику шести
коэффициентов эффективности: расчетные формулы и их экономические содержания.
1. Построить график динамики всех переменных производственной функции.
График динамики исходных данных.
Самостоятельная работа.
Динамика переменных
X1
X2
У


Microsoft Excel - ч2зач2 [Только для чтения]

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

2. Выполнить эконометрический анализ многофакторной модели, при ожидаемых значениях **Хоч1 = 26 , Хоч2 = 45**

Специфицируем модель производственный процесс линейной функцией вида-

Yi = a0 + a1*X1i + a2*X2i +ei

Выполним расчет коэффициентов модели с помощью статистической функции Линейн в следующей последовательности: выделить диапазон размером 5 строк и 3 столбца, в левую верхнюю ячейку диапазона вставить статистическую функцию Линейн и выполнить все инструкции мастера функции Линейн, произвести преобразование функции Линейн для работы с массивом- выделить массив действия функции массив, состоящий из трех колонок и пяти строк, **F2, Ctrl+Shift+Enter.**

Образец выполнения.

1,3812	0,3143	-7,053
0,1477	0,2674	1,3199
0,984	2,0523	#Н/Д
647,48	21	#Н/Д
5454,3	88,45	#Н/Д

Представим результаты расчетов.

Стрелками указана связь между протоколом расчета по функции Линейн и представлением расчетов в общепринятом виде.

Yi=	**-7,053**	**+**	**0,3143**	*** X1i**	**+**	**1,3812**	*** X2i + ei**
Sa0=	**1,3199**	**Sa1=**	**0,2674**		**Sa2=**	**0,1477**	
ta0=	**-5,344**	**ta1=**	**1,1753**		**ta2=**	**9,3532**	
E=	**2,0523**						
R^2=	**0,984**		**tтаб. =**	**2,0796**			
F=	**647,48**		**Fтаб=**	**3,4668**			

Где X1i - объем трудовых ресурсов (чел.),

X2i - стоимость производственных фондов (тыс.грв.),

Yi - размер конечной продукции (тыс.грв.),

i - порядковый номер текущего месяца.

a0 = -7,053 - коэффициент a0 модели Yi = a0 + a1*X1i + a2*X2i + ei

A	B	C	D	E	F	G	H	I	J	K

a1 = 0,3143 - коэффициент а1 модели Yi = a0 + a1*X1i + a2*X2i + ei

a2 = 1,3812 - коэффициент а2 модели Yi = a0 + a1*X1i + a2*X2i + ei

Sa - ошибка коэффициента а,

E - ошибка модели

R^2 - коэффициент детерминации,

ta - критерий Стьюдента для коэффициента а,

tтаб - табличное значение критерия Стьюдента,

F - критерий Фишера.

Fтаб - табличное значение критерия Фишера.

Предъявляем результаты эконометрического анализа.

1. Доля объясненной вариации.

98,404 % исходных данных имеют выбранную тенденцию;

2. Проверка достоверности модели.

По критерию Фишера проверим достоверность модели.

Так как **F= 647,48 > Fтаб(альфа=0,05, M1=k - 1, M2= n - k) = 3,4668**

то модель является **достоверной с вероятностью 0,95.**

3. Проверка достоверности коэффициентов модели.

По критерию Стьюдента проверим достоверность коэффициентов модели.

Так как **ta0 = 5,3435 > tтаб. = 2,0796** , то

коэффициент **a0 с вероятностью 0,95 достоверно отличается от нуля.**

Так как **ta1 = 1,1753 < tтаб. = 2,0796** , то

коэффициент **a1 статистически не отличается от нуля.**

Так как **ta2 = 9,3532 > tтаб. = 2,0796** , то

коэффициент **a2 с вероятностью 0,95 достоверно отличается от нуля.**

4. Доля влияния каждого фактора.

r^2(X1) = 0,0153 - доля влияния фактора **X1**,

r^2(X2) = 0,9687 - доля влияния фактора **X2**,

Всего = 0,984 - доля влияния всех факторов.

Microsoft Excel - ч2зач2 [Только для чтения]

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 12

P129 fx

A	B	C	D	E	F	G	H	I	J	K	L

5. Точечный и интервальный прогноз.

На ожидаемый период t = **25** при ожидаемых значениях объясняемых переменных **X1ож= 26 X2ож= 45**

наиболее вероятный точечный прогноз будет равен **Yпр = 63,271**

С вероятностью 0,95 фактическое значение прогнозной зависимой переменной будет находиться в интервале от **59,003 до 67,5385**

6. Графическое представление эконометрического анализа.

3. Выполните расчет и проанализируйте коэффициенты эффективности производственной функции.

Существует шесть коэффициентов эффективности производственной функции:

1. **Средняя отдача ресурса.**

$$M1 = \frac{Y_{cp}}{X1_{cp}} = 1{,}8095$$ (тыс.грв./чел.)

M1=	
Уср=	25,521
X1ср=	14,104

M1 - средняя производительность трудовых ресурсов. 1 человек в среднем за месяц производит 1,8095 тыс.грв. валовой продукции.

$$M2 = \frac{Y_{cp}}{X2_{cp}} = 1{,}2526$$ (тыс. грв./тыс. грв.)

M2 - средняя фондоотдача производственных фондов. На одну тыс.грв производственных фондов приходится 1,2526 тыс.грв. объема валовой продукции.

2. **Граничная эффективность ресурса.**

E1 = a1 = 0,3143 (тыс. грв. / чел.)

$$a1 = \frac{dY}{dX1}$$

E1 - предельная эффективность объема трудовых ресурсов.

При увеличении трудовых ресурсов на 1 чел. объем производства возрастает на 0,3143 тыс. грв.

E2 = a2 = 1,3812 (тыс. грв. /тыс. грв.)

$$a2 = \frac{dY}{dX2}$$

E2 - предельная эффективность производственных фондов.

При увеличении производственных фондов на 1 тыс.грв. объем производства возрастает на 1,3812 тыс. грв.

3. Эластичность ресурса.

$$D1 = \frac{E1}{M1} = \boxed{0,1737}\ \%$$

D1=	
E1=	0,3143
M1=	1,8095

$$D1 = \frac{a1}{M1} = \frac{a1 * Xcp}{Ycp} = \frac{dY * Xcp}{dX * Ycp}$$

D1 - эластичность трудовых ресурсов.

При увеличении трудовых ресурсов на 1% объем производства возрастет на 0,1737 %.

$$D2 = \frac{E2}{M2} = \boxed{1,1027}$$

D2 - эластичность производственных фондов.

При увеличении производственных фондов на 1 % объем производства возрастет на 1,1027 %.

4. Мера эффективности ресурса.

$$G1 = \frac{E1}{X1cp} = \boxed{0,0223}$$ (тыс. грв.)

G1=	
E1=	0,3143
X1cp=	14,104

$$G1 = \frac{E1}{X1cp} = \frac{a1}{X1cp} = \frac{dY}{dX * X1cp}$$

G1 - мера эффективности трудовых ресурсов.

0,0223 тыс.грв. дополнительной продукции, произведенной дополнительным одним человеком, приходится на каждого работника.

$$G2 = \frac{E2}{X2cp} = \boxed{0,0678}$$ (тыс. грв.)

G2 - мера эффективности производственных фондов.

0,0678 тыс. грв. дополнительной продукции, произведенной дополнительным производственным фондом в размере 1 тыс. грв., приходится на каждую 1 тыс. грв. производственных фондов

5. Граничная норма замещения.

$$H12 = \frac{-E1}{E2} = \boxed{-0{,}228} \text{ (тыс. грв.)}$$

H12 =	
E1 =	0,3143
E2 =	1,3812

$$H12 = \frac{-E1}{E2} = \frac{-a1}{a2} = \frac{-dY(X1)*dX2}{dX1*dY(X2)}$$

H12 - предельная норма замещения действия первого ресурса с помощью второго ресурса.

Если ресурс X1 изменится на единицу dx1, то объем производства изменится на dy(X1).

Чтобы компенсировать изменение dy(X1), необходимо изменить ресурс X2 на такую величину H12 = dx2, при которой dy(X2) будет равно dy(X1).

На -0,228 тыс. грв. нужно изменить размер производственных фондов, чтобы компенсировать изменение объема выпускаемой продукции, вызванной изменением трудовых ресурсов на 1 чел.

6. Суммарная эластичность.

$$D = D1 + D2 = \boxed{1{,}2764} \ \%$$

D =	
D1=	0,1737
D2=	1,1027

D - суммарная эластичность.

На 1,2764 % увеличится объем выпускаемой продукции, если трудовые и производственные фонды одновременно возрастут на 1%.

4. Произвести имитацию экономической ситуации.

Имеется следующая модель производственного процесса

Yi= -7,053 + 0,3143 * X1i + 1,3812 * X2i + ei

Где X1 - количество рабочих (чел.)

X2 - количество производственных фондов (тыс. грв.),

У - объем производства (тыс. грв.).

Цель: определить при каких значениях X1 и X2 объем производства будет превышать заданную величину.

Способ 1: построить трехмерную поверхность, по которой определить при каких значениях X1 и X2 объем производства будет превышать заданную величину.

Способ 2: в модель производственного процесса подставляем известные значения

Способ 2: в модель производственного процесса подставляем известные значения переменных и определяем неизвестную величину.

Способ 1.

Вычислим матрицу расчетных значений **У** при заданных величинах **Х1** и **Х2**.

Приделы изменения **Х1** от	5	до	30	
Приделы изменения **Х2** от	5	до	40	

Произведем имитацию экономического процесса с помощью модели производственной функции.

Допустим на предприятии имеется **Х1=** 20 работников.

Необходимо определить при каком значении **Х2** - количестве производственных фондов объем производства будет больше **У =** 80 усл. ден. ед.

Изменяйте значения Х2 в таблице2 и по графику определяйте численные значения У, которые удовлетворяют заданному ограничению.

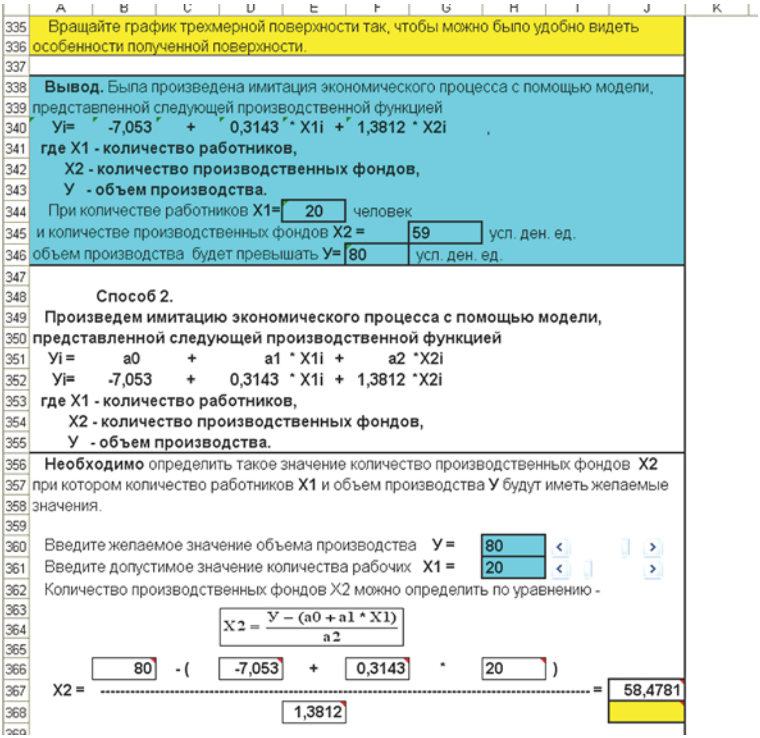
Таблица 2. Зависимость объема производства **У** от **Х1** и **Х2.**

Х1	Х2							
	5	10	15	20	25	30	35	59
5	1,4242	8,3299	15,236	22,141	29,047	35,953	42,859	76,006
10	2,9957	9,9014	16,807	23,713	30,619	37,5245	44,43	77,578
15	4,5672	11,473	18,379	25,284	32,19	39,096	46,002	79,149
20	6,1387	13,044	19,95	26,856	33,762	40,6675	47,573	80,721
25	7,7102	14,616	21,522	28,427	35,333	42,239	49,145	82,292
30	9,2817	16,187	23,093	29,999	36,905	43,8105	50,716	83,864

Определить **Х2,** если

Х1 = 20

У = 80



Вращайте график трехмерной поверхности так, чтобы можно было удобно видеть особенности полученной поверхности.

Вывод. Была произведена имитация экономического процесса с помощью модели, представленной следующей производственной функцией

Yi= -7,053 + 0,3143 * X1i + 1,3812 * X2i

где X1 - количество работников,

X2 - количество производственных фондов,

У - объем производства.

При количестве работников **X1=** 20 человек

и количестве производственных фондов **X2 =** 59 усл. ден. ед.

объем производства будет превышать **У=** 80 усл. ден. ед.

Способ 2.

Произведем имитацию экономического процесса с помощью модели, представленной следующей производственной функцией

Yi = a0 + a1 * X1i + a2 *X2i

Yi= -7,053 + 0,3143 * X1i + 1,3812 *X2i

где X1 - количество работников,

X2 - количество производственных фондов,

У - объем производства.

Необходимо определить такое значение количество производственных фондов **X2** при котором количество работников **X1** и объем производства **У** будут иметь желаемые значения.

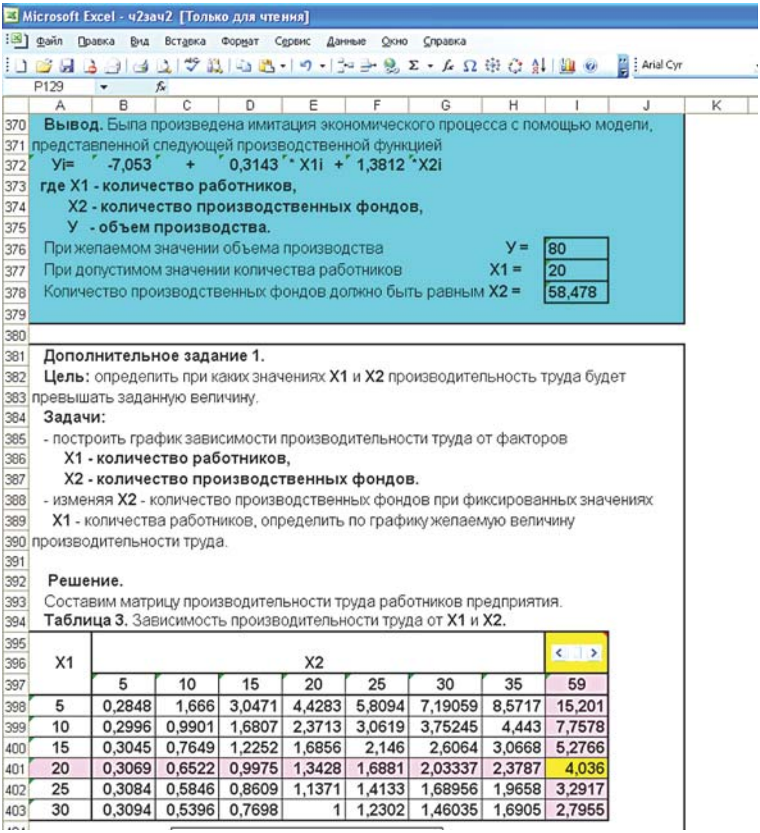
Введите желаемое значение объема производства **У =** 80

Введите допустимое значение количества рабочих **X1 =** 20

Количество производственных фондов X2 можно определить по уравнению -

$$X2 = \frac{У - (a0 + a1 * X1)}{a2}$$

$$X2 = \frac{80 - (-7{,}053 + 0{,}3143 * 20)}{1{,}3812} = 58{,}4781$$



Вывод. Была произведена имитация экономического процесса с помощью модели, представленной следующей производственной функцией

Yi= -7,053 + 0,3143 * X1i + 1,3812 *X2i

где X1 - количество работников,

X2 - количество производственных фондов,

У - объем производства.

При желаемом значении объема производства **У =** 80

При допустимом значении количества работников **X1 =** 20

Количество производственных фондов должно быть равным **X2 =** 58,478

Дополнительное задание 1.

Цель: определить при каких значениях **X1** и **X2** производительность труда будет превышать заданную величину.

Задачи:

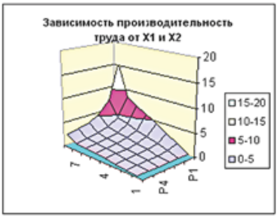
- построить график зависимости производительности труда от факторов
  X1 - количество работников,
  X2 - количество производственных фондов.
- изменяя **X2** - количество производственных фондов при фиксированных значениях **X1** - количества работников, определить по графику желаемую величину производительности труда.

Решение.

Составим матрицу производительности труда работников предприятия.

Таблица 3. Зависимость производительности труда от **X1** и **X2.**

X1	X2							
	5	10	15	20	25	30	35	59
5	0,2848	1,666	3,0471	4,4283	5,8094	7,19059	8,5717	15,201
10	0,2996	0,9901	1,6807	2,3713	3,0619	3,75245	4,443	7,7578
15	0,3045	0,7649	1,2252	1,6856	2,146	2,6064	3,0668	5,2766
20	0,3069	0,6522	0,9975	1,3428	1,6881	2,03337	2,3787	4,036
25	0,3084	0,5846	0,8609	1,1371	1,4133	1,68956	1,9658	3,2917
30	0,3094	0,5396	0,7698	1	1,2302	1,46035	1,6905	2,7955



Произведем имитацию экономического процесса.

Допустим на предприятии имеется **Х1=** 20 работников.

Необходимо определить при каком значении **Х2 -** количестве производственных фондов производительность труда будет больше 4 усл. ед.

Изменяя в таблице3 численное значение **Х2**, определите по графику желаемое значение производительности труда.

Вывод. При количестве работников равных **Х1 =** 20 и количестве сырья **Х2 =** 59 , производительность труда будет больше 4 условных единиц.

Дополнительное задание 2.

Цель: Определить при каких значениях **Х1, Х2, Х3** можно получить желаемую величину **У.**

Задачи:

- построить четырехмерную поверхность зависимости **У от Х1, Х2, Х3,**
- определить по графику при каких значениях **Х1, Х2, Х3** можно получить желаемую величину **У.**

Решение.

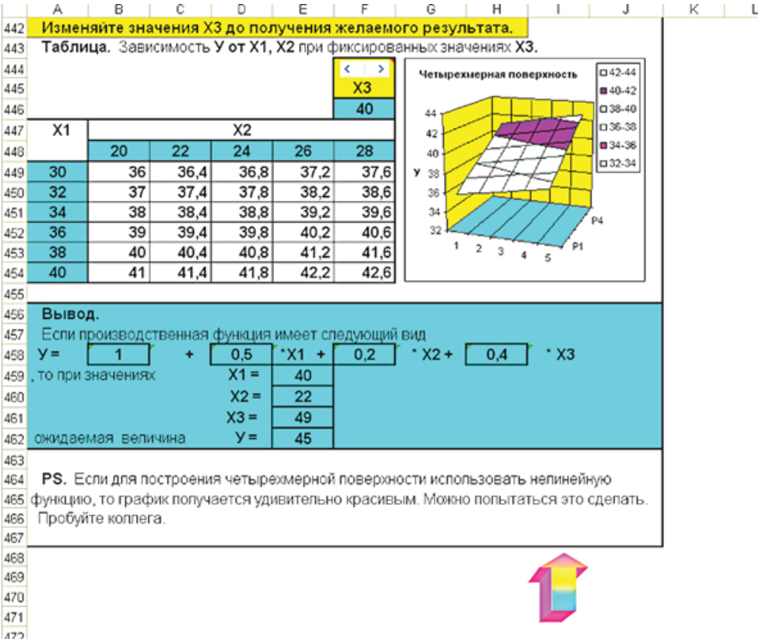
Допустим имеется многофакторная модель следующего вида:

У = 1 **+** 0,5 *Х1 **+** 0,2 * Х2 + 0,4 * Х3

Объясняемые переменные могут изменяться в следующих пределах

Х1	**от**	30	**до**	40
Х2	**от**	20	**до**	32
Х3	**от**	40	**до**	50

Желаемая величина **У =** 45



Изменяйте значения Х3 до получения желаемого результата.

Таблица. Зависимость **У** от **Х1**, **Х2** при фиксированных значениях **Х3**.

Х3
40

Х1	Х2				
	20	22	24	26	28
30	36	36,4	36,8	37,2	37,6
32	37	37,4	37,8	38,2	38,6
34	38	38,4	38,8	39,2	39,6
36	39	39,4	39,8	40,2	40,6
38	40	40,4	40,8	41,2	41,6
40	41	41,4	41,8	42,2	42,6

Вывод.

Если производственная функция имеет следующий вид

У = 1 + 0,5 *Х1 + 0,2 * Х2 + 0,4 * Х3

, то при значениях Х1 = 40, Х2 = 22, Х3 = 49

ожидаемая величина У = 45

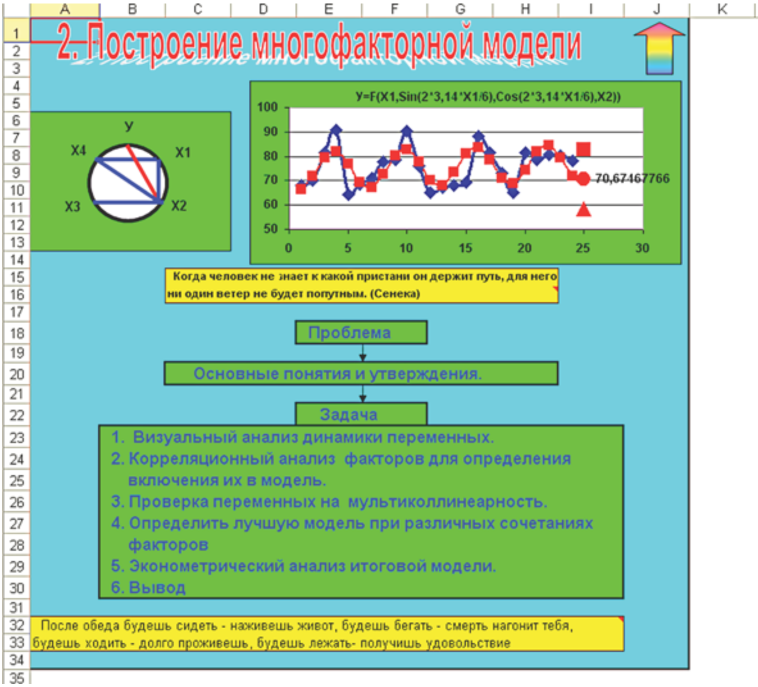
PS. Если для построения четырехмерной поверхности использовать нелинейную функцию, то график получается удивительно красивым. Можно попытаться это сделать. Пробуйте коллега.

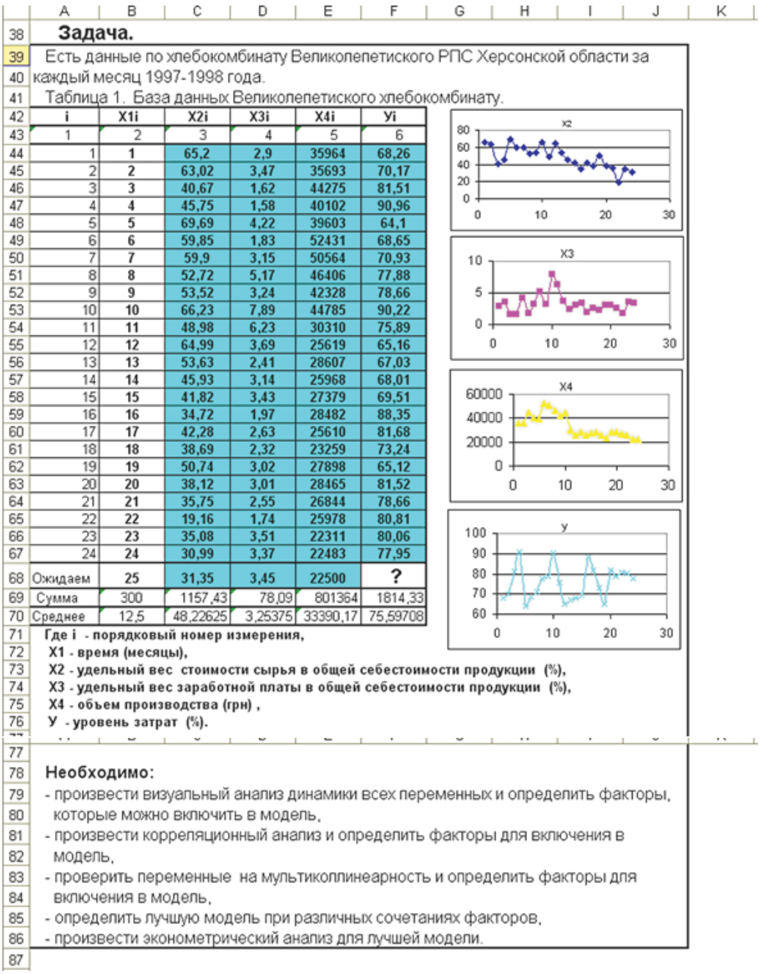
Рис. 15.15. Результат выполнения первого пункта меню программы МДиплом

3. Вернитесь в меню и выполните пункт

“2. Построение многофакторной модели”, “Проблема”, “Основные понятия и утверждения”, “Задача”. При выполнении пункта “Задача” введите свои значения переменных для многофакторной модели. Аналогичное решение задачи уже проводились в теме 11, поэтому для экономии места протокол расчетов приводить не будем.

Результаты выполнения задания представлены на рис. 15.16.


2. Построение многофакторной модели
У
X4
X1
X3
X2
Y=F(X1,Sin(2*3,14*X1/6),Cos(2*3,14*X1/6),X2))
70,67167766
Когда человек не знает к какой пристани он держит путь, для него ни один ветер не будет попутным. (Сенека)
Проблема
Основные понятия и утверждения.
Задача
1. Визуальный анализ динамики переменных.
2. Корреляционный анализ факторов для определения включения их в модель.
3. Проверка переменных на мультиколлинеарность.
4. Определить лучшую модель при различных сочетаниях факторов
5. Эконометрический анализ итоговой модели.
6. Вывод
После обеда будешь сидеть - наживешь живот, будешь бегать - смерть нагонит тебя, будешь ходить - долго проживешь, будешь лежать- получишь удовольствие



Задача.

Есть данные по хлебокомбинату Великолепетиского РПС Херсонской области за каждый месяц 1997-1998 года.

Таблица 1. База данных Великолепетиского хлебокомбинату.

i	X1i	X2i	X3i	X4i	Yi
1	2	3	4	5	6
1	1	65,2	2,9	35964	68,26
2	2	63,02	3,47	35693	70,17
3	3	40,67	1,62	44275	81,51
4	4	45,75	1,58	40102	90,96
5	5	69,69	4,22	39603	64,1
6	6	59,85	1,83	52431	68,65
7	7	59,9	3,15	50564	70,93
8	8	52,72	5,17	46406	77,88
9	9	53,52	3,24	42328	78,66
10	10	66,23	7,89	44785	90,22
11	11	48,98	6,23	30310	75,89
12	12	64,99	3,69	25619	65,16
13	13	53,63	2,41	28607	67,03
14	14	45,93	3,14	25968	68,01
15	15	41,82	3,43	27379	69,51
16	16	34,72	1,97	28482	88,35
17	17	42,28	2,63	25610	81,68
18	18	38,69	2,32	23259	73,24
19	19	50,74	3,02	27898	65,12
20	20	38,12	3,01	28465	81,52
21	21	35,75	2,55	26844	78,66
22	22	19,16	1,74	25978	80,81
23	23	35,08	3,51	22311	80,06
24	24	30,99	3,37	22483	77,95
Ожидаем	25	31,35	3,45	22500	?
Сумма	300	1157,43	78,09	801364	1814,33
Среднее	12,5	48,22625	3,25375	33390,17	75,59708

Где i - порядковый номер измерения,
X1 - время (месяцы),
X2 - удельный вес стоимости сырья в общей себестоимости продукции (%),
X3 - удельный вес заработной платы в общей себестоимости продукции (%),
X4 - объем производства (грн) ,
У - уровень затрат (%).

Необходимо:

- произвести визуальный анализ динамики всех переменных и определить факторы, которые можно включить в модель,
- произвести корреляционный анализ и определить факторы для включения в модель,
- проверить переменные на мультиколлинеарность и определить факторы для включения в модель,
- определить лучшую модель при различных сочетаниях факторов,
- произвести эконометрический анализ для лучшей модели.

Рис. 15.16. Результат выполнения второго пункта меню программы МДиплом

4. Вернитесь в меню и выполните пункт
“3. Модель сосредоточенного лага”.
Результаты выполнения задания представлены на рис. 15.17.


3. Модель сосредоточенного лага.
Динамика переменных X и У
Время (месяцы)
Xt
Yt
Проблема
Основные понятия.
Задача 1
Определение величины временного лага.
Вывод
Задача 2
Прогнозирование временного ряда с учетом сосредоточенных лаговых факторов.
Вывод
С солнцем, царем и с женщиной нельзя быть слишком близким ибо они сожгут тебя и нельзя быть далеко ибо они будут бесполезны для тебя

Задача 1.

Определение величины временного лага.

Рабочая гипотеза. Предполагается, что эмиссия денег приводит через три месяца к росту цен и снижению уровня жизни.

Необходимо проверить рабочую гипотезу.

Метод решения задачи - регрессионный метод.

Способ решения задачи - смещение по времени одного временного ряда относительно другого.

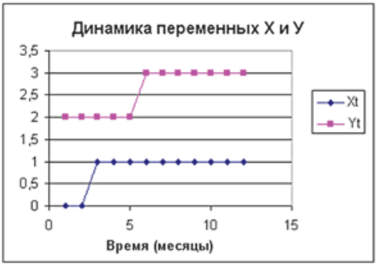
Ожидаемый результат - если смещать один временной ряд относительно другого, то при совпадении по времени причины и следствия ошибка модели резко уменьшится, тогда величина смещения одного временного ряда относительно другого будет равна сосредоточенному лагу между эмиссией денег и ростом цен.

Решение задачи 1.

Имеются условные данные цен на товары народного потребления и эмиссии денег за каждый месяц в течении года.

Таблица. Условные данные цен на товары народного потребления и эмиссии денег.

t	Xt	Yt
1	2	3
1	**0**	**2**
2	**0**	**2**
3	**1**	**2**
4	**1**	**2**
5	**1**	**2**
6	**1**	**3**
7	**1**	**3**
8	**1**	**3**
9	**1**	**3**
10	**1**	**3**
11	**1**	**3**
12	**1**	**3**

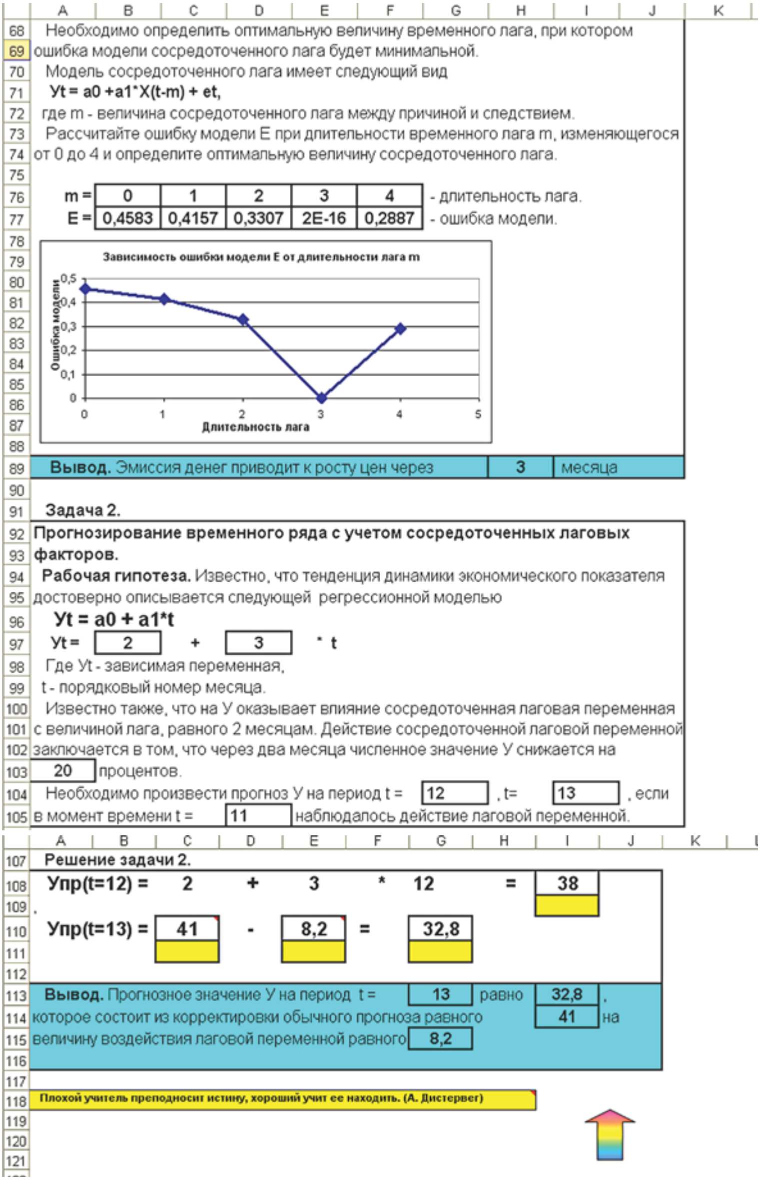


Где: **t** - порядковый номер месяца,

Xt - эмиссия денег (фиктивная переменная, фиксирующая факт эмиссии денег),

Yt - цены на товары народного потребления (фиктивная переменная, фиксирующая факт повышения цен на товары народного потребления).

Необходимо определить оптимальную величину временного лага, при котором



Необходимо определить оптимальную величину временного лага, при котором ошибка модели сосредоточенного лага будет минимальной.

Модель сосредоточенного лага имеет следующий вид

Yt = a0 +a1*X(t-m) + et,

где m - величина сосредоточенного лага между причиной и следствием.

Рассчитайте ошибку модели E при длительности временного лага m, изменяющегося от 0 до 4 и определите оптимальную величину сосредоточенного лага.

m =	0	1	2	3	4	- длительность лага.
E =	0,4583	0,4157	0,3307	2E-16	0,2887	- ошибка модели.

Вывод. Эмиссия денег приводит к росту цен через 3 месяца

Задача 2.

Прогнозирование временного ряда с учетом сосредоточенных лаговых факторов.

Рабочая гипотеза. Известно, что тенденция динамики экономического показателя достоверно описывается следующей регрессионной моделью

Yt = a0 + a1*t

Yt = 2 + 3 * t

Где Yt - зависимая переменная,

t - порядковый номер месяца.

Известно также, что на У оказывает влияние сосредоточенная лаговая переменная с величиной лага, равного 2 месяцам. Действие сосредоточенной лаговой переменной заключается в том, что через два месяца численное значение У снижается на 20 процентов.

Необходимо произвести прогноз У на период t = 12 , t= 13 , если в момент времени t = 11 наблюдалось действие лаговой переменной.

Решение задачи 2.

Упр(t=12) = 2 + 3 * 12 = 38

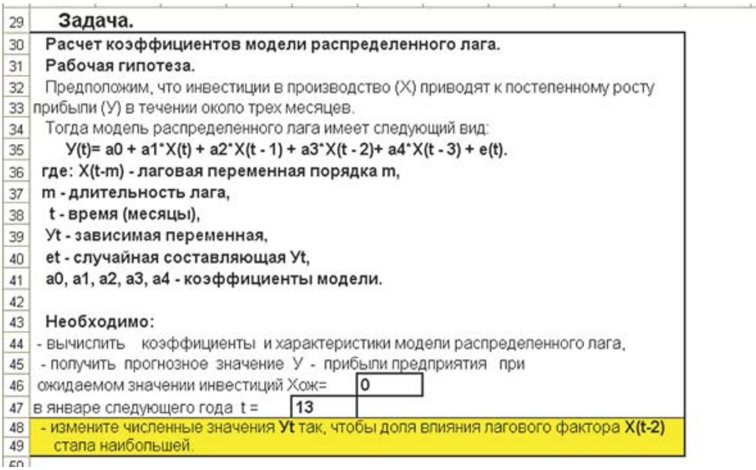
Упр(t=13) = 41 - 8,2 = 32,8

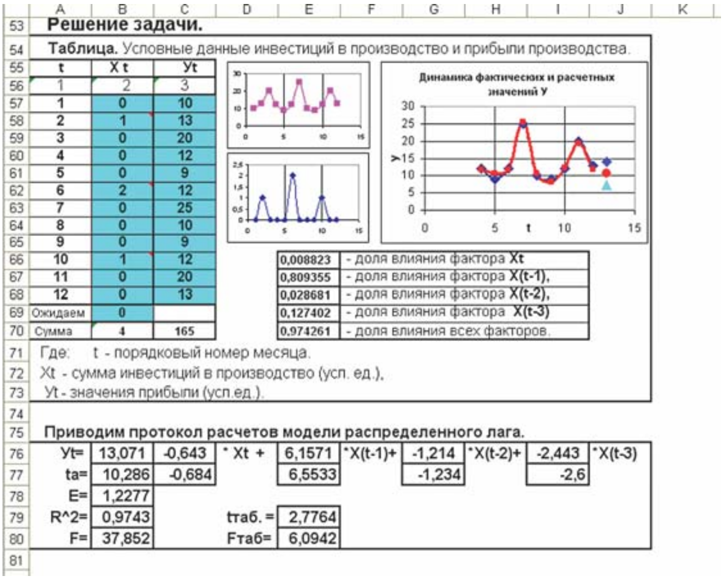
Вывод. Прогнозное значение У на период t = 13 равно 32,8 , которое состоит из корректировки обычного прогноза равного 41 на величину воздействия лаговой переменной равного 8,2

Плохой учитель преподносит истину, хороший учит ее находить. (А. Дистервег)

Рис. 15.17. Результат выполнения третьего пункта меню программы МДиплом

5. Вернитесь в меню и выполните пункт
“4. Модель распределенного лага”.
Результаты выполнения задания представлены на рис. 15.18.





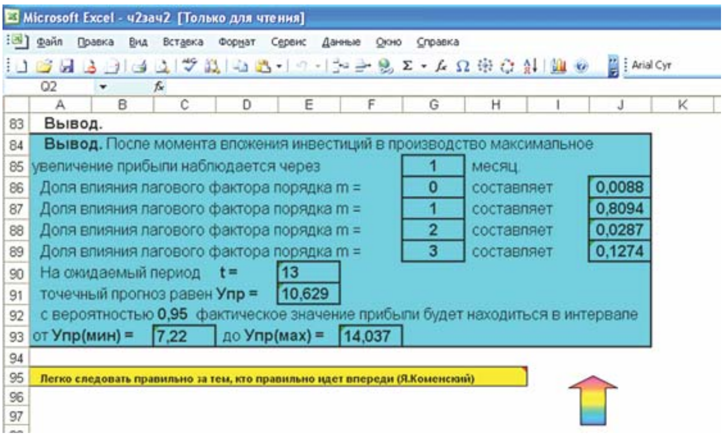
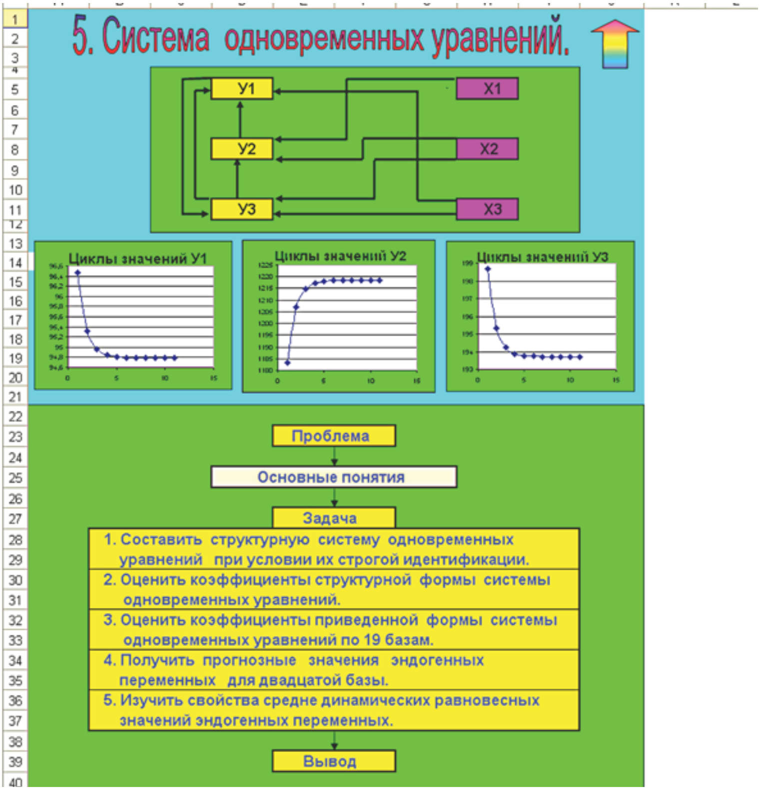


Рис. 15.18. Результат выполнения четвертого пункта меню программы МДиплом

6. Вернитесь в меню и выполните пункт
“5. Система одновременных уравнений”.
Результаты выполнения задания представлены на рис. 15.19.


5. Система одновременных уравнений.
У1
У2
У3
Х1
Х2
Х3
Циклы значений У1
Циклы значений У2
Циклы значений У3
Проблема
Основные понятия
Задача
1. Составить структурную систему одновременных уравнений при условии их строгой идентификации.
2. Оценить коэффициенты структурной формы системы одновременных уравнений.
3. Оценить коэффициенты приведенной формы системы одновременных уравнений по 19 базам.
4. Получить прогнозные значения эндогенных переменных для двадцатой базы.
5. Изучить свойства средне динамических равновесных значений эндогенных переменных.
Вывод

Задача.

Есть база данных по 20 оптовым базам УКООПСОЮЗ за 1998 год, представленные в таблицы.

Таблица. База данных по оптовым базам УКООПСОЮЗА за 1998 год.

i	X1	X2	X3	У1	У2	У3
1	2	3	4	5	6	7
1	83	4	0,1	28	190	28
2	30	6	0,2	27,8	474	28,1
3	84	4	0,3	91,1	642	31,7
4	105	4	0,2	91,7	621	32
5	37	4	0,2	65,3	670	38,6
6	103	12	0,4	60,8	907	38,9
7	34	4	0,7	35,5	747	42,6
8	55	11	0,2	33,3	653	42,9
9	79	14	0,5	21,4	867	54,8
10	77	12	0,2	20	598	54,9
11	150	2	0,1	13,9	858	60,3
12	53	4	0,1	34,7	748	60,4
13	93	31	0,6	47	1666	72,3
14	93	21	0,5	27,7	1319	71,9
15	110	4	0,2	29,3	919	91,5
16	169	31	0,5	26	1486	91,2
17	139	19	0,4	69,8	1116	105
18	174	18	0,4	68,1	1357	105
19	271	33	1,3	96,7	1163	152
20	45	31	0,9	94,6	1683	153
Сумма	1984	269	8	982,7	18684	1355,1

Где X1 - количество рабочих, которое работают на базе (чел.),

X2 - количество механизмов (шт.),

X3 - активная часть основных фондов (млн. руб.),

У1 - доход (тыс. руб.),

У2 - затраты обращения (тыс. руб.),

У3 - оптовый товарооборот (млн. руб.).

Факторы **X(i)** считать экзогенными, **У(i)** - эндогенными.

Microsoft Excel - ч2зач2 [Только для чтения]

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Необходимо:

1. Составить структурную систему одновременных уравнений при условии их строгой идентификации.
2. Оценить коэффициенты структурной формы системы одновременных уравнений.
3. Оценить коэффициенты приведенной формы системы одновременных уравнений по 19 базам.
4. Получить прогнозные значения эндогенных переменных для двадцатой базы.
5. Изучить свойства средне динамических равновесных значений эндогенных переменных.

Решение задачи.

1. Составить структурную систему одновременных уравнений при условии их строгой идентификации.

Взаимосвязь переменных можно вообразить в виде структурной формы системы одновременных уравнений:

У1= a0 + a1*У2 + a2*У3 + a3*X3 + e1,
У2= b0 + b1*У3 + b2*X1 + b3*X2 + e2,
У3= c0 + c1*У1 + c2*X2 + c3*X3 + e3.

Взаимосвязь переменных можно представить в виде блока-схемы

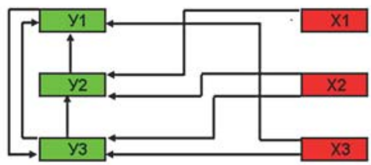


Рис.3. Взаимосвязь переменных структурной формы системы одновременных уравнений.

Вывод. Анализ блок-схемы структурной системы одновременных уравнений показывает, что для каждой эндогенной переменной соблюдено условие строгой идентификации, которое заключается в следующем:

- количество эндогенных переменных должно равняться количеству экзогенных переменных,
- количество объясняемых переменных для каждой эндогенной переменной должно равняться количеству экзогенных переменных и комбинации объясняемых переменных не должны повторятся для различных эндогенных переменных.

2. Оценить коэффициенты структурной формы системы одновременных уравнений.

Приводим коэффициенты структурных систем одновременных уравнений, полученных двух шаговым методом наименьших квадратов.

У1 = 39,2 + -0,023 * У2 + 0,178 * У3 + 48,781 * Х3

У2 = 630,09 + -7,069 * У3 + 2,219 * Х1 + 39,531 * Х2

У3 = -7,883 + 0,9515 * У1 + 2,4464 * Х2 + -3,561 * Х3

3. Оценить коэффициенты приведенной формы системы одновременных уравнений по 19 базам.

Приведенная форма систем одновременных уравнений имеет вид:

У1д= d0 + d1*X1 + d2*X2 + d3*X3 ,

У2д= f0 + f1*X1 + f2*X2 + f3*X3,

У3д= h0 + h1*X1 + h2*X2 + h3*X3.

Вычислим коэффициенты приведенной формы системы одновременных уравнений.

У1д = 26,895 + 0,168 * Х1 + -1,017 * Х2 + 41,317 * Х3

У2д = 520,46 + 0,4491 * Х1 + 31,49 * Х2 + -176,5 * Х3

У3д = 13,128 + 0,3484 * Х1 + 0,7678 * Х2 + 13,29 * Х3

Где Х1 - количество рабочих, которое работают на базе (чел.),

Х2 - количество механизмов (шт.),

Х3 - активная часть основных фондов (млн.крб.),

У1 - доход (тыс.крб.),
У2 - затраты обращения (тыс.крб.),
У3 - оптовый товарооборот (млн.крб.).
Х1, Х2, Х3 - экзогенные переменные,
У1, У2, У3 - эндогенные переменные,
У1д, У2д, У3д - средние динамические значения соответствующих эндогенных переменных.

4. Получить прогнозные значения эндогенных переменных для двадцатой базы.

Оценим коэффициенты приведенной формы системы одновременных уравнений по 19 базам и получим прогноз для двадцатой базы.

Введите ожидаемые значения экзогенных переменных двадцатой базы для получения средних динамических значений эндогенных переменных.

Х1ож =	45	, Х2ож =	31	, Х3ож =	0,9

Приводим величины средних динамических значений эндогенных переменных.

У1д =	40,121	У2д =	1358	У3д =	64,568

Вывод. При изменении экзогенных переменных изменяются средне динамические равновесные значения эндогенных переменных.

5. Изучить свойства средне динамических равновесных значений эндогенных переменных.

Рабочая гипотеза.

Предположим, что

средне динамические значения эндогенных переменных зависят только от значений экзогенных переменных,

тогда будут справедливы следующие утверждения.

Первое утверждение-

если экзогенные переменные не изменяются, то спонтанное отклонение одной или нескольких эндогенных переменных от своих средне динамических равновесных значений приведет к появлению циклических волн, которые восстановят прежние средне динамические равновесные значения эндогенных переменных.

Второе утверждение-

если экзогенные переменные изменятся, то это приведет к появлению других средне динамических равновесных значений эндогенных переменных.

Проверка первого утверждения.

Предположим, что экзогенные переменные не изменяются, величины У1, У2, У3 имеют средние динамические значения, которые соответствуют определенным значениям экзогенных переменных. Допустим, что в определенный момент времени изменилась величина У1, тогда в экономической системе возникнет волна, возбужденная изменением У1, по следующей цепочке логических связей: У1 - У3 - У2 -У1. Ожидается, что циклические изменения всех эндогенных переменных должны привести к восстановлению их средне динамических значений, которые были до изменения переменной У1.

Определите через сколько циклов установятся средние динамические значения эндогенных переменных при условии, что экзогенные переменные остались на прежнем уровне.

Приводим циклы изменений эндогенных переменных, вызванные изменением У1 при заданных значениях экзогенных переменных.

Измените значения экзогенных переменных для получения других значений средне динамических равновесных значений эндогенных переменных **У1д, У2д, У3д.**

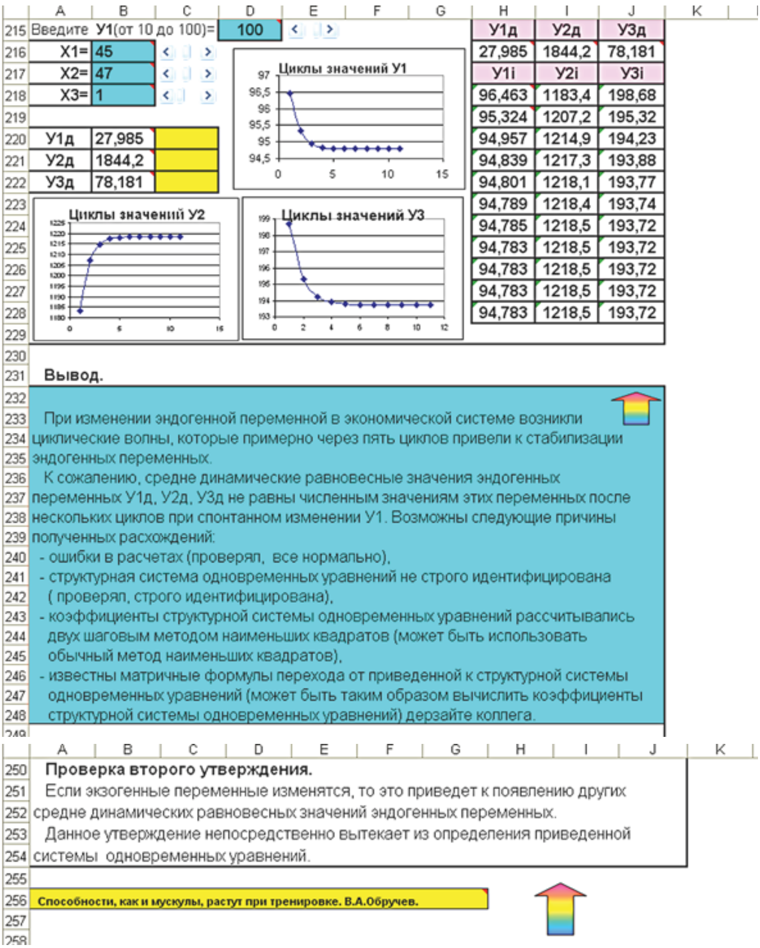


Рис. 15.19. Результат выполнения пятого пункта меню программы МДиплом

7. Вернитесь в меню и выполните пункт "6. Анализ качественных переменных".
Результаты выполнения задания представлены на рис. 15.20.

6. Анализ качественных переменных

i / j	B1	B2	Всего
A1	8	4	12
A2	2	4	6
Всего	10	8	18

Проблема

Основные понятия

Задача

1. Определение коэффициента корреляции между качественными переменными А и В по критерию Юла.
2. Определение достоверности связи между качественными переменными А и В по критерию Хи-квадрат.

Общий вывод

Живи настоящим ибо прошлое не вернешь, а будущее еще не наступило. Тетя Оля.

Задача.

По условной области была взята выборка, которая состоит из нескольких предприятий. Каждое предприятие характеризуется двумя качественными переменными: (А)- состояние приватизации и (В)- состояние рентабельности.

Переменная А имеет два класса: (А1)- приватизированное и (А2)- неприватизированное. Переменная В тоже имеет два класса: (В1)- рентабельное и (В2) - нерентабельное. Количество предприятий aij, которые обладают свойствами Ai, Bj, представлены в таблице сопряженности.

Таблица 1. Перекрестные данные двух качественных переменных.

i / j	B1	B2	Всего
A1	8	4	12
A2	2	4	6
Всего	10	8	18

$$Q = \frac{f11 * f22 - f12 * f21}{f11 * f22 + f12 * f21} = \boxed{0{,}6}$$

Где А - вид приватизации предприятия,
А1 - предприятие приватизировано,
А2 - предприятие не приватизировано,
В - вид рентабельности предприятия,
В1 - предприятие рентабельно,
В2 - предприятие нерентабельно,
i - классы приватизации,
j - классы рентабельности,

f11 =	8	- количество предприятий приватизируемых и рентабельных,
f12 =	4	- количество предприятий приватизируемых и нерентабельных,
f21 =	2	- количество предприятий неприватизированных и рентабельных,
f22 =	4	- количество предприятий неприватизированных и нерентабельных,
f01= f11+f21=	10	- количество рентабельных предприятий,
f02= f12+f22=	8	- количество нерентабельных предприятий,
f10= f11+f12=	12	- количество прибыльных предприятий,
f20= f21+f22=	6	- количество неприбыльных предприятий,
f00=f11+f21+f21+f22=	18	- количество всех предприятий,

Необходимо:

- произвести расчет коэффициента корреляции между А и В по критерию Юла.
- проверить независимость А и В на равные значимости альфа = 0,05 по критерию Хи-квадрат.

Критерий Юла

Произведем расчет коэффициента корреляции между А и В при двух градациях переменных.

Коэффициент корреляции между качественными переменными А и В вычисляют по критерию Юла:

$$Q = \frac{f11 * f22 - f12 * f21}{f11 * f22 + f12 * f21}$$

Достоверность критерия Юла проверяют по критерию Стьюдента.

Проверим нулевую гипотезу. Ho:" Q= 0 " или " между переменными А и В нет связи ".

Если фактический критерий Стьюдента

$$t_ф = \frac{Q}{mQ}$$

больше табличного tт(альфа= 0,05, k= n - 1), то нулевая гипотеза отклоняется.

Где k - число степеней свободы,

n - количество всех объектов,

Q - критерий Юла,

mQ - ошибка Q,

$$Q = \frac{f11*f22 - f12*f21}{f11*f22 + f12*f21} = \boxed{0,6}$$

$$mQ = \frac{1 - Q^2}{2}\sqrt{\frac{1}{f11} + \frac{1}{f12} + \frac{1}{f21} + \frac{1}{f22}} = \boxed{0,3394}$$

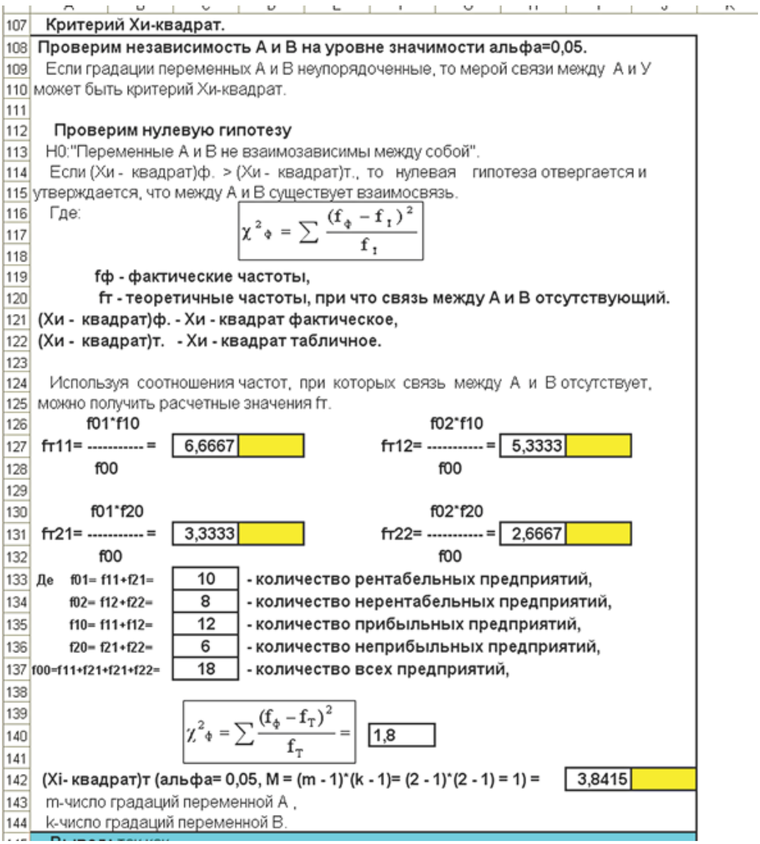
Де f11 =	8	- количество предприятий приватизируемых и рентабельных,
f12 =	4	- количество предприятий приватизируемых и нерентабельных,
f21 =	2	- количество предприятий неприватизированных и рентабельных,
f22 =	4	- количество предприятий неприватизированных и нерентабельных,

Вывод: Так как фактическое значение критерия Стьюдента

$$t_ф = \frac{Q}{m} = \boxed{1,7678} < t\,т(альфа = 0,05, k = n - 1 = 17) = \boxed{2,1098}$$

то нулевая гипотеза H0: "Q= 0"

нулевая гипотеза принимается



Критерий Хи-квадрат.

Проверим независимость А и В на уровне значимости альфа=0,05.

Если градации переменных А и В неупорядоченные, то мерой связи между А и У может быть критерий Хи-квадрат.

Проверим нулевую гипотезу

Н0:"Переменные А и В не взаимозависимы между собой".

Если (Хи - квадрат)ф. > (Хи - квадрат)т., то нулевая гипотеза отвергается и утверждается, что между А и В существует взаимосвязь.

Где:

$$\chi^2_\phi = \sum \frac{(f_\phi - f_T)^2}{f_T}$$

fф - фактические частоты,

fт - теоретичные частоты, при что связь между А и В отсутствующий.

(Хи - квадрат)ф. - Хи - квадрат фактическое,

(Хи - квадрат)т. - Хи - квадрат табличное.

Используя соотношения частот, при которых связь между А и В отсутствует, можно получить расчетные значения fт.

fт11 = f01*f10 / f00 = 6,6667 fт12 = f02*f10 / f00 = 5,3333

fт21 = f01*f20 / f00 = 3,3333 fт22 = f02*f20 / f00 = 2,6667

Де f01= f11+f21=	10	- количество рентабельных предприятий,
f02= f12+f22=	8	- количество нерентабельных предприятий,
f10= f11+f12=	12	- количество прибыльных предприятий,
f20= f21+f22=	6	- количество неприбыльных предприятий,
f00=f11+f21+f21+f22=	18	- количество всех предприятий,

$$\chi^2_\phi = \sum \frac{(f_\phi - f_T)^2}{f_T} = 1{,}8$$

(Xi- квадрат)т (альфа= 0,05, М = (m - 1)*(k - 1)= (2 - 1)*(2 - 1) = 1) = 3,8415

m-число градаций переменной А ,

k-число градаций переменной В.

Задание третьего уровня сложности

Задание. Обратитесь к информационным базам данных профессиональной деятельности по вашей специальности, выберите переменные временных рядов размером в 24 даты и проведите расчеты с использованием программы МДиплом.

Пояснение. В качестве информационных источников могут быть: интернет, статистические справочники, аналитические обзоры, профессиональные журналы, бухгалтерская отчетность.

Выходной тест

Имеется матрица парных коэффициентов корреляции ($r_{\alpha=0.05} = 0{,}8$).

	X_1	X_2	X_3	У
X_1	1			
X_2	0,9	1		
X_3	0,2	0,85	1	
У	0,85	0,9	0,75	1

Необходимо определить, какие факторы можно включить в модель.

ЛИТЕРАТУРА

Основная

1. Айвазян С. А. Основы эконометрики. Том. 2. — М.: ЮНИТИ-ДАНА, 2001.

2. Берндт Э. Р. Практика эконометрики: классика и современность /Пер. с англ. под ред. проф. С. А. Айвазяна. — М.: ЮНИТИ — ДАНА, 2005.

3. Валентинов В. А. Эконометрика. — М.: Дашков и К, 2006.

4. Елисеева И. И. и др. Эконометрика. — М.: Финансы и статистика, 2001.

5. Елисеева И. И. и др. Практикум по эконометрике. — М.: Финансы и статистика, 2002.

6. Катышев П. К. и др. Сборник задач к начальному курсу эконометрики. -М.: Дело, 2002.

7. Кремер Н. Ш., Путко Б. А. Эконометрика. — М.: ЮНИТИ-ДАНА, 2002.

8. Магнус Я. Р. и др. Эконометрика. Начальный курс. — М.: Дело, 2000.

Дополнительная

9. Айвазян С. А.. Мхитарян В. С. Теория вероятностей и прикладная статистика. Том 1. — М.: ЮНИТИ-ДАНА, 2001.

10. Бородич С. А. Эконометрика. — Минск: Новое знание, 2001.

11. Грубер Й. Эконометрия. Том 1. Введение в эконометрию. — К.: Астарта, 1996.

12. Доугерти К. Введение в эконометрику.- М.: НИФРА-М, 1997.

13. *Джонстон Д.* Эконометрические методы. — М.: Статистика, 1980.

14. *Замков О. О.* Эконометрические методы в макроэкономическом анализе. — М.: ГУ ВШЭ, 2001.

15. Информатика / Под ред. Н. В. Макаровой. — М.: Финансы и статистика, 2000.

16. Статистические методы повышения качества / Под ред. Х. Кумэ. — М.: Финансы и статистика, 1990.