

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

Государственное образовательное учреждение
высшего профессионального образования -
«Оренбургский государственный университет»

В.Н. АФАНАСЬЕВ, А.П. ЦЫПИН

Эконометрика в пакете STATISTICA

Рекомендовано Ученым советом
государственного образовательного учреждения
высшего профессионального образования –
«Оренбургский государственный университет»
в качестве учебного пособия для студентов, обучающихся по экономическим
специальностям

Оренбург 2008

УДК 330.43:004

ББК 65в631

А96

Афанасьев В.Н., Цыпин А.П.

Эконометрика в пакете STATISTICA: учебное пособие по выполнению лабораторных работ / В.Н. Афанасьев, А.П. Цыпин, - Оренбург: ГОУ ОГУ, 2008. - 204 с.

Учебное пособие предназначено для преподавания дисциплины специализации студентам очной формы обучения экономических специальностей. Содержит основные разделы курса «Эконометрика», представлены индивидуальные задания для лабораторных работ, разобраны решения типовых задач.

Методическое пособие составлено с учетом Государственного образовательного стандарта высшего профессионального образования по направлению подготовки дипломированных специалистов.

ISBN 978-5-91933-004-2

© Афанасьев В.Н., Цыпин А.П., 2010
© ГОУ ОГУ, 2010

Содержание

Введение	6
1 Организация данных и построение графиков в пакете STATISTICA	7
1.1 Цели и задачи лабораторной работы	7
1.2 Обзор статистических и эконометрических пакетов прикладных программ	7
1.3 Рекомендуемая литература	8
1.4 Интернет источники, посвященные проблемам эконометрики и сбору статистических данных	9
1.5 Запуск пакета STATISTICA	9
1.6 Рабочее окно пакета STATISTICA	10
1.7 Создание файла данных	12
1.8 Форматирование файла данных	18
1.9 Сохранение файла данных	20
1.10 Построение графиков в пакета STATISTICA	21
1.11 Тесты для самоконтроля	25
1.12 Задание для самостоятельного выполнения	26
2 Выявление и измерение корреляционной связи	27
2.1 Цели и задачи лабораторной работы	27
2.2 Понятие корреляции и методы ее выявления	27
2.3 Рекомендуемая литература	28
2.4 Сопоставление двух параллельных рядов	29
2.5 Построение аналитической группировки	30
2.6 Графический метод (построение поля корреляции)	35
2.7 Расчет коэффициента ранговой корреляции Спирмена и Кендала	36
2.8 Расчет парного линейного коэффициента корреляции Пирсона	38
2.9 Тесты для самоконтроля	39
2.10 Задания для самостоятельного выполнения	41
3 Построение классической линейной регрессии	42
3.1 Цели и задачи лабораторной работы	42
3.2 Понятие классической линейной регрессии	42
3.3 Рекомендуемая литература	43
3.4 Расчет описательных (дескриптивных) статик	44
3.5 Построение классической линейной регрессии	47
3.6 Прогнозирование (имитация) неизвестных значений зависимой переменной	55
3.7 Тесты для самоконтроля	57
3.8 Задания для самостоятельного выполнения	58
4 Выявление и устранение мультиколлениарности и гетероскедостичности	60
4.1 Цели и задачи лабораторной работы	60
4.2 Понятие мультиколлениарности и гетероскедостичности, методы	60

выявления и устранения	
4.3 Рекомендуемая литература	62
4.4 Мультиколлениарность, выявление и устранение	62
4.5 Гетероскедстичность выявление и устранение	68
4.6 Анализ нормальности распределения случайного члена уравнения	75
4.7 Тесты для самоконтроля	77
4.8 Задание для самостоятельного выполнения	78
5 Выявление тенденции во временных рядах	80
5.1 Цели и задачи лабораторной работы	80
5.2 Понятие временных рядов и методы выявления тренд составляющей ряда	80
5.3 Рекомендуемая литература	83
5.4 Выявление наличия тенденции в динамическом ряду	83
5.5 Сглаживание уровней ряда	85
5.6 Построение полиномиального тренда	87
5.7 Автокорреляция, выявление и устранение	93
5.8 Исключение влияния автокорреляции	97
5.9 Прогнозирование уровней исследуемого показателя	98
5.10 Тесты для самоконтроля	99
5.11 Задание для самостоятельного выполнения	100
6 Экспоненциальное сглаживание временных рядов	102
6.1 Цели и задачи лабораторной работы	102
6.2. Понятие экспоненциального сглаживания временных рядов	102
6.3 Рекомендуемая литература	103
6.4 Простое экспоненциальное сглаживание (метод Брауна)	103
6.5 Экспоненциальное сглаживание с учетом тренда (метод Хольта)	108
6.6 Экспоненциальное сглаживание с учетом одновременно тренда и сезонности (метод Винтерса)	113
6.7 Тесты для самоконтроля	115
6.8 Задания для самостоятельного выполнения	116
7 Анализ и прогнозирования временных рядов на основе моделей ARMA	118
7.1 Цели и задачи лабораторной работы	118
7.2 Теоретические аспекты адаптивных методов прогнозирования	118
7.3 Рекомендуемая литература	120
7.4 Определение стационарности модели	120
7.5 Приведение динамического ряда к стационарному виду	122
7.6 Идентификация порядка разности модели	124
7.7 Построение ARMA модели	127
7.8 Тесты для самоконтроля	132
7.9 Задания для самостоятельного выполнения	134
8 Выявление и анализ сезонности во временных рядах	135

8.1 Цели и задачи лабораторной работы	135
8.2 Понятие сезонности и методы ее выявления	135
8.3 Рекомендуемая литература	137
8.4 Сезонная корректировка в модуле <i>Seasonal decomposition 1</i>	138
8.5 Сезонная корректировка в модуле <i>X11/Y2k (Census 2) – monthly</i>	143
8.6 Тесты для самоконтроля	147
8.7 Задания для самостоятельного выполнения	148
9 Построение регрессионных моделей с фиктивными переменными	150
9.1 Цели и задачи лабораторной работы	150
9.2 Понятие фиктивных переменных и их применение в пространственных и временных моделях	150
9.3 Рекомендуемая литература	151
9.4 Использование фиктивных переменных при построении классической регрессии	152
9.5 Выявление сезонности с использованием сезонных фиктивных переменных в модуле <i>Multiple regression</i>	154
9.6 Тест (критерий) Г. Чоу	157
9.7 Тесты для самоконтроля	160
9.8 Задания для самостоятельного выполнения	162
10 Оценка нелинейных регрессионных моделей	163
10.1 Цели и задачи лабораторной работы	163
10.2 Определение нелинейной регрессии и основные формы нелинейных моделей	163
10.3 Рекомендуемая литература	164
10.4 Оценка нелинейной модели на основе пространственных данных	164
10.5 Оценка нелинейной модели на основе временного ряда	168
10.6 Тесты для самоконтроля	172
10.7 Задания для самостоятельного выполнения	174
Список используемой литературы	175
Приложения	178

Введение

Цель настоящего учебного пособия - ознакомление с основными методами анализа временных рядов и алгоритмами обработки данных для получения необходимых характеристик. Пособие предназначено для студентов, изучивших теорию вероятностей и математическую статистику или соответствующий курс высшей математики. Оно будет также полезно и специалистам, которые применяют методы анализа временных рядов для обработки данных в своей предметной области.

Пособие состоит из семи глав, посвященных основным направлениям в анализе временных рядов. Теоретический материал служит руководством к практическим работам. Каждая лабораторная работа составлена по следующим принципам: формулируется цель работы, затем определяется порядок ее выполнения, подразумевающий последовательность заданий. Лабораторные работы ориентированы на использование ЭВМ с графическим дисплеем, так как многие результаты полезно (а иногда и необходимо) иллюстрировать графиками. Однако при наличии определенного навыка можно организовать выдачу результатов в виде графика на любой ЭВМ. Наиболее подходящим средством выполнения лабораторных работ являются персональные ЭВМ типа IBM PC с применением соответствующей периферии и программного обеспечения. Лабораторные работы составлены таким образом, что каждая последующая производится с возможным использованием результатов предыдущих (по крайней мере, в пределах одной темы). Поэтому при прохождении практикума необходимо иметь персональную дискету, на которой бы организовывались и сохранялись все файлы данных для использования в случае необходимости.

Основная направленность лабораторных работ - не реализация методов и алгоритмов, а проведение исследования их эффективности. Поэтому следует обязательно производить сравнение применяемых методов и алгоритмов на основе полученных при работе наблюдений. В пособии излагаются, как правило, несколько известных подходов, на основе которых можно провести грамотный сравнительный анализ.

1 Организация данных и построение графиков в пакете STATISTICA

1.1 Цели и задачи лабораторной работы

В данной лабораторной работе последовательно рассмотрим основные ресурсы всемирной сети *Internet* посвященных проблемам эконометрики, а также операции в статистическом пакете программ (СПП) STATISTICA, при этом выделим следующие задачи:

- 1 Ознакомимся с содержанием сайтов посвященных проблемам эконометрики;
- 2 Ознакомимся с возможностями ввода и сохранения информации в пакете STATISTICA используя данные приложения 1;
- 3 Используя данные приведенные в приложении 1 построим двух- и трехмерный графики в STATISTICA.

1.2 Обзор статистических и эконометрических пакетов прикладных программ

Информационные технологии в последние десятилетия XX века прочно вошли в социально-экономическую жизнь общества. Внедрение компьютерных технологий в экономические исследования, позволило в настоящее время использовать сложные математические методы, применительно к экономическим объектам, будь то экономика страны в целом или экономика отдельного предприятия.

Проблема выбора наиболее подходящего пакета для конкретной категории пользователей, круга решаемых задач, типа и возможностей компьютера и т.д. весьма непростая. Из курса общей теории статистики известно, что любое исследование можно условно разделить на три этапа: сбор статистической информации, ее обработка и анализ полученных результатов. В соответствии с этим имеющиеся на сегодняшний момент пакеты программ можно разделить на три группы:

- 1 Программные продукты и организационная техника, способствующие увеличению скорости подготовки исходной информации в электронном виде.
- 2 Пакеты программ, непосредственно отвечающие за обработку числового материала.
- 3 Пакеты программ необходимые для оформления результатов исследования.

Применительно к эконометрике наибольший интерес представляет вторая группа, из которой наиболее известными на сегодняшний момент пакетами программ являются: STATISTICA 6.0, SPSS 12.00, Stata 8, STADIA 6.0, Eviews 4.1, S-Plus, StatGraphics, MathCad 12 и другие.

Перечисленные программы можно разделить на подгруппы по количеству используемых статистических и эконометрических процедур.

- 1 Пакеты STATISTICA 6.0, SPSS 12.00 и STADIA 6.0 отличает большое количество, представленных в распоряжение пользователя методов, начиная от простых дескриптивных (описательных) статистик, заканчивая такими сложными статистико-математическими методами как нейронные сети и многомерное шкалирование. Основным недостатком данных пакетов программ является низкая гибкость алгоритмов построения эконометрических моделей.
- 2 Статистико-эконометрические программы (Eviews 4.1, Stata 8) - рассчитаны в основном на профессионалов в области эконометрического моделирования. Важной особенностью данных пакетов программ является практически полное отсутствие оконного меню и наличие командной строки в качестве основной рабочей области.
- 3 К данной подгруппе можно отнести программы S-Plus и StatGraphics, которые хотя и обладают довольно большим набором математико-статистических методов, но не пригодны для аналитических изысканий, они скорее всего наилучшим образом подойдут для начинающих исследователей, студентов и аспирантов ВУЗ-ов.
- 4 Пакеты Mathcad, Maple и другие - являются программами предназначенные для математических вычислений, рассчитанные на исследователя с обширными математическими знаниями. Данные пакеты имеют возможность реализации статистических методов, но при этом требуют от исследователя знания алгоритмов построения данных процедур.
- 5 С ростом типов информации возникает необходимость в так называемых табличных редакторах, таких как MS Excel, Lotus 1-2-3, Quattro Pro, StarOffice Calc и 602Tab, предназначенных для быстрого ввода числовых данных, их редактирования и преобразования в необходимый формат. К достоинствам данных программ можно отнести возможность импорта информации из баз данных, редактирование информации и проведение разведочного анализа. Недостатком, с точки зрения исследователя, можно назвать незначительное количество статистических алгоритмов и процедур обработки данных, либо неудобство их применения.

1.3 Рекомендуемая литература

Для лучшего понимания материала изложенного в данной главе необходимо дополнительно проанализировать следующие источники литературы (см. список используемой литературы):

- 1 Вуколов, Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: Учебное пособие. – М.: ФОРУМ: ИНФРА-М, 2004. – 464с. - ISBN 5-8199-0141-X, 5-16-002003-9. (см. стр. 8-18)

- 2 Кремер Н, Путко Б. Эконометрика: Учебник для вузов. – М.: ЮНИТИ-ДАНА. 2003. - 311с. (см. стр. 9-24)
- 3 Тюрин Ю.Н., Макарова А.А. Анализ данных на компьютере / Под ред. В.Э. Фигурнова. – 3-е изд., перераб. и доп. – М.: ИНФРА-М, 2003. – 544с. (см. стр. 470-487)
- 4 Халафян А.А. STATISTICA 6. Статистический анализ данных. 3-е изд. Учебник – М.: ООО «Бином-Пресс», 2007 г. – 512 с. ISBN 978-5-9518-0215-6. (см. стр. 13-51, 59-81)

1.4 Интернет источники, посвященные проблемам эконометрики и сбору статистических данных

В настоящее время всемирная сеть Internet предоставляет пользователям огромные возможности для получения справочной информации и «добыче» исходной статистической информации для проведения эконометрического исследования. В связи с этим все многообразие сайтов можно разделить на три группы:

1 Источники данных

Государственный комитет по статистике РФ - www.gks.ru

Межгосударственный статистический комитет СНГ - <http://www.cisstat.com/rus/>

Центробанк РФ - www.cbr.ru

Высшая школа экономики - <http://stat.hse.ru/hse/index.html>

ПРАЙМ-ТАСС - <http://e3.prime-tass.ru/macro/index.aspx>

Всемирный банк - www.worldbank.org

2 Проблемы эконометрического моделирования

Центр статистических технологий - <http://nickart.spb.ru/analysis/links.php> на данной страничке содержатся раздел ПОЛЕЗНЫЕ СТРАНИЦЫ В ИНТЕРНЕТ

Эконометрическая страничка Александра Цыплакова -

<http://www.nsu.ru/ef/tsy/ecmr/index.htm>

Практикум по эконометрике - <http://dist-economics.eu.spb.ru/book/index.html>

Страница И.Н. Молчанова - <http://molchanov.narod.ru/>

3 Производители и распространители программного обеспечения

СПП STATISTICA - www.statistica.ru

СПП SPSS - www.spss.ru

СПП EViews - www.eviews.com

СПП STATA - www.stata.com

1.5 Запуск пакета STATISTICA

Запустить данный программный продукт можно тремя способами:

Во-первых, на *Рабочем столе* необходимо два раза кликнуть по пиктограмме программы (рисунок 1.1):



Рисунок 1.1 – Пиктограмма программы STATISTICA

Во-вторых, можно выбрать кнопку *Пуск* → *Программы* → *STATISTICA* → *STATISTICA*

К третьему способу необходимо прибегать в самых крайних случаях, он заключается в следующем: на диске *C* (или другом диске на который было произведена инсталляция пакета) необходимо найти папку *Program Files* деле выбрать папку *StatSoft* → *statist.exe*

После описанных действий произойдет загрузка требуемой программы.

1.6 Рабочее окно пакета STATISTICA

Рассматриваемая версия программного продукта по сравнению с предыдущими, обладает более удобным пользовательским интерфейсом. Управление программой максимально приближено к семейству *Windows*-приложений.

Обратимся к рисунку 1.2, и рассмотрим возможности пакета. В основу данной программы положена электронная таблица, которая поддерживает основные возможности таких табличных редакторов как MS Excel, Lotus 1-2-3, Quattro Pro и др.

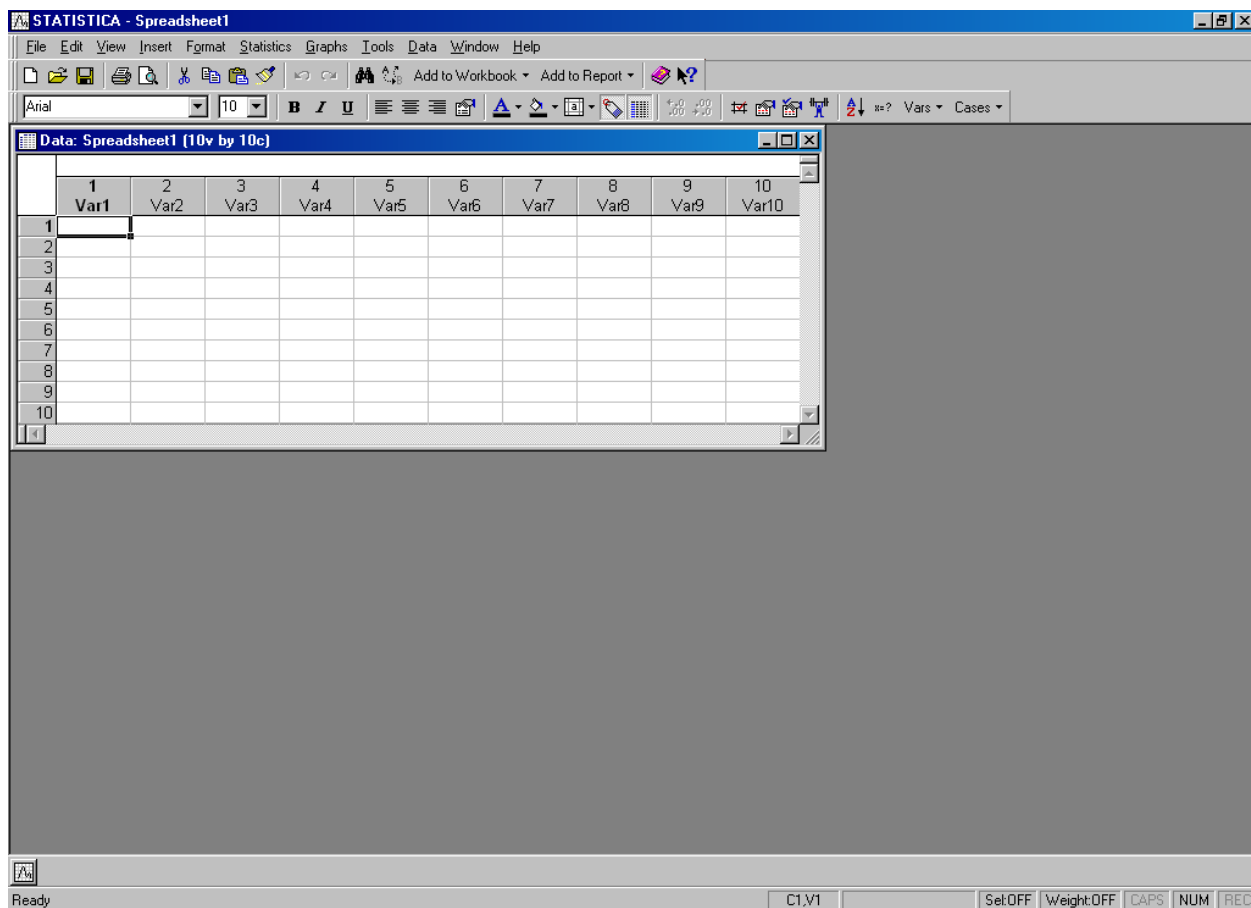


Рисунок 1.2 – Рабочее окно пакета STATISTICA

Имея сходства с табличными редакторами, такие как копирование, вставка, перенос, данных, ввод формул, преобразование данных и т.д. рассматриваемый пакет программ имеет некоторые особенности. Одним из основных отличий является особое расположение данных в рабочем листе, при этом необходимо ввести два понятия. Первое понятие *Variables* (Переменные) – это переменные (совокупности), располагаемые в столбец (на рисунке 1.2 отображены как *Var1*, *var2*, ..., *Var10*). Второе понятие *Cases* (Значения, наблюдения) - это единицы совокупности или уровни ряда (в данном случае отображаются в левой части таблицы как 1, 2, 3, ...10).

Заполняя электронную таблицу пакета фактическими данными, ни когда не стоит забывать об этой конфигурации, в противном случае применяя к неправильно введенным данным эконометрические методы заведомо совершается ошибка.

Также немаловажным является то, что в электронной таблице не допускаются пропущенные (незаполненные) строки, т.к. программа будет расценивать их как ячейки заполненные нулями, соответственно при выполнении какой-либо процедуры будут получены искаженные результаты.

В верхней части рабочего окна пакета (как и в любом *Windows*-приложении) расположено главное меню:

Рисунок 1.3 – Главное меню программы

Как видим из рисунка 1.3, пункты из панели главного меню: *File* (Файл), *Edit* (Правка), *View* (Вид), *Insert* (Вставка), *Format* (Формат), *Tools* (Сервис), *Windows* (Окно) и *Help* (Справка) по своей функциональной принадлежности являются стандартными для *Windows*-приложений.

Специфическими пунктами меню можно считать следующие:

Statistics (Вычисления) – данный пункт меню содержит огромное количество методов статистической обработки данных начиная от расчета описательных статистик (максимум, минимум, средняя и т.д.) до сложнейших многомерных статистических алгоритмов.

Graphs (Графики) – в данном пункте доступны огромное количество разнообразных графиков и диаграмм, как двухмерных так и трехмерных.

Date (Данные) – в данном меню доступны алгоритмы направленные на преобразование имеющихся данных (стандартизация, ранжирование и т.д.).

1.7 Создание файла данных

Создать файл данных в пакете STATISTICA можно двумя способами:

- 1 Импортировать готовые данные из других программ (баз данных, математических и статистических пакетов прикладных программ).
- 2 Ввести необходимую информацию с клавиатуры.

Для создания файла данных первым способом введем в табличном редакторе Excel данные представленные в приложении 1 – таблица 1.1.

Далее сохраним файл с именем *Книга 1.xls* на диске *C* или папке *Мои документы* (или другой папке) после этого необходимо закрыть табличный редактор Excel.

1.7.1 Импорт данных из внешних программ

Для импорта файла с данными в пакет STATISTICA необходимо пройти следующие шаги:

Шаг 1. В главном меню пакета выберем *File* → *Open* (Файл → Открыть).

Шаг 2. В появившемся окне (рисунок 1.4) необходимо выбрать тип файла, в данном случае файл электронной таблицы Excel (т.е. необходимо выбрать расширение *.xls) и имя искомого файла, далее нажать кнопку *Открыть*.

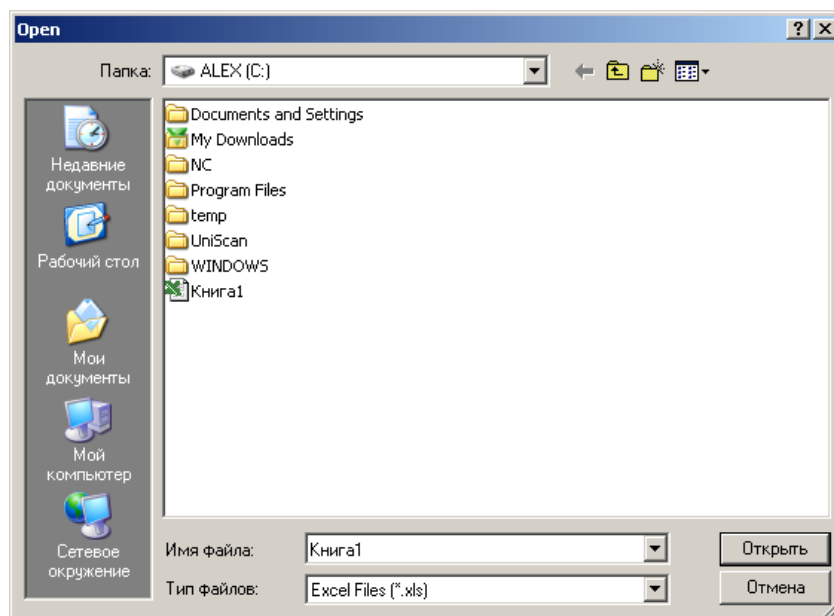


Рисунок 1.4 - Диалоговое окно *Open* (Открытие)

Шаг 3. В открывшемся диалоговом окне (рисунок 1.5) будет предложено импортировать отдельную страницу или все страницы рабочей книги.

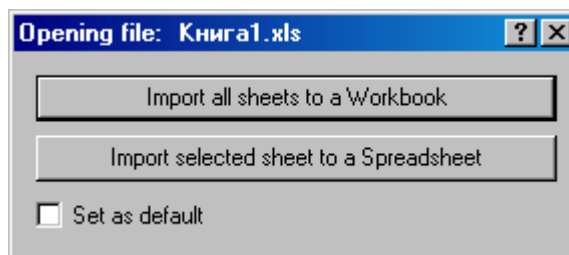


Рисунок 1.5 – Диалоговое окно *Opening file* (Открытие файла)

где: *Import all sheets to a Workbook* - Импорт всех страниц рабочей книги
Import selected sheets to a Spreadsheet - Импорт выбранных страниц рабочей книги

Set as default - Установки отсутствуют

В нашем случае выберем второй вариант и перейдем к следующему шагу.

Шаг 4. В появившемся диалоговом окне *Select Sheet to Import* (Выбор импортируемой страницы) выберем необходимую страницу и нажмем клавишу ОК.

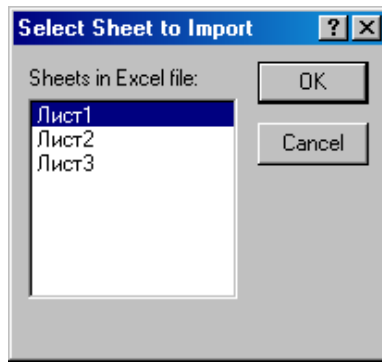


Рисунок 1.6 – Диалоговое окно *Select Sheet to Import*

Шаг 5. В следующем окне будет предложено указать размер таблицы, а также предоставлена возможность оставить имеющиеся имена импортируемых переменных и имена записей. Сделаем установки как показано на рисунке 1.7 и нажмем кнопку ОК.

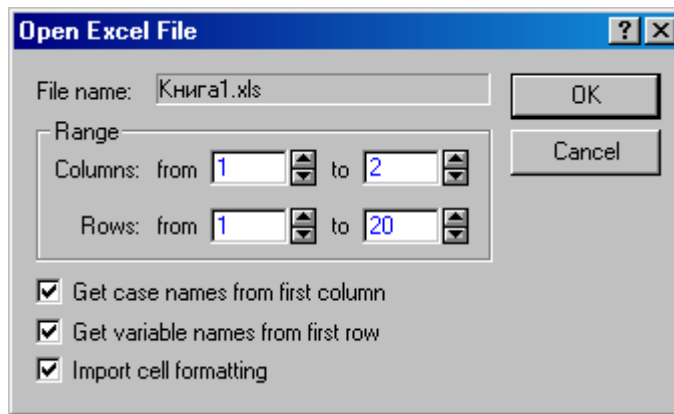


Рисунок 1.7 – Диалоговое окно *Open Excel File*

где: Имя файла: - *File name:*

Диапазон - *Range*

Столбцы: с 1 по 2 - *Columns: from*

Строки: с 1 по 20 - *Rows: from*

Имена наблюдений из первого столбца - *Get cases names from first column*

Имена переменных из первой строки - *Get variable names from first row*

Импорт формата ячеек - *Import cell formatting*

Пройдя вышеперечисленные шаги, получим следующий лист рабочей книги пакета STATISTICA (рисунок 1.8).

	1
	Y
1955г.	13,5
1956г.	11,6
1957г.	14,7
1958г.	16,2
1959г.	15,2
1960г.	15,1
1961г.	16,9
1962г.	16,8
1963г.	12,9
1964г.	13,8
1965г.	16,1
1966г.	20,4
1967г.	17,8
1968г.	18,3
1969г.	18,9
1970г.	22,8
1971г.	23,1
1972г.	19,6
1973г.	27

Рисунок 1.8 – Результат импорта файла табличного редактора Excel в пакет STATISTICA

При желании пользователь может ввести заголовок для создаваемой таблицы, для этого в главном меню необходимо ввести *View* → *Display Header* (Вид→Вывод заголовка таблицы). Далее необходимо установить курсор на белой строке таблицы между ее техническим заголовком (в котором указано имя файла и размерность таблицы) и названием столбцов. Затем дважды щелкнуть левой кнопкой мыши и ввести необходимый заголовок, в нашем случае – Динамика урожайности зерновых культур, ц/га. В результате проделанной процедуры получаем следующий результат (рисунок 1.9).

Динамика урожайности зерновых культур, ц/га	
	1
	Y
1955г.	13,5
1956г.	11,6
1957г.	14,7
1958г.	16,2
1959г.	15,2

Рисунок 1.9 – Результат ввода заголовка рабочей таблицы

1.7.2 Ручной ввод информации

Для иллюстрации создания файла данных вторым способом используем данные приведенные в приложении 1 – таблица 1.2.

Шаг 1. После запуска программы STATISTICA в главном меню выбрать *File*→*New* (Файл→Новый). В появившемся окне *Create New Document* (Создание нового документа), рисунок 1.10, необходимо ввести количество переменных *Number of variables* и объем совокупности *Number of cases*. В данном случае исходная матрица данных 3×16.

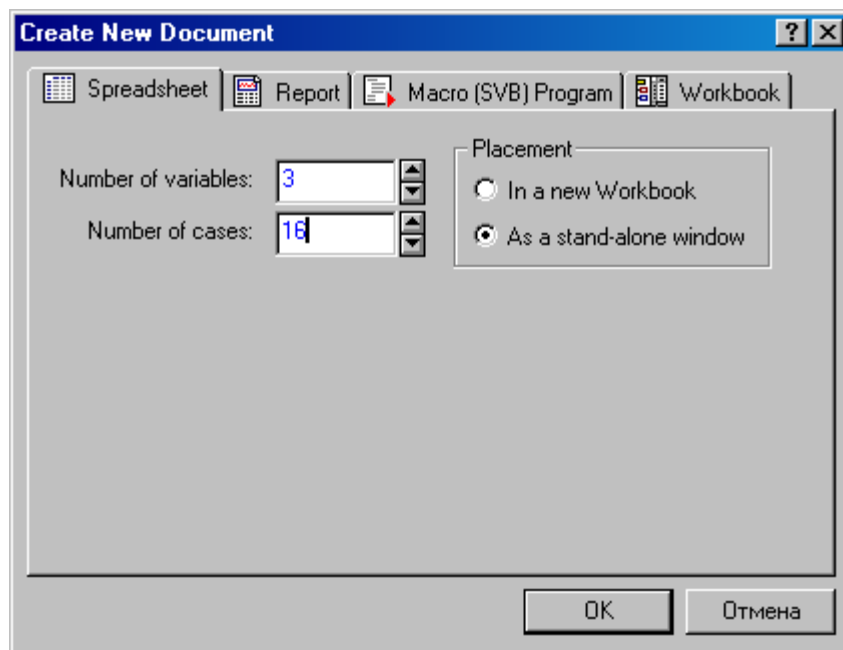


Рисунок 1.10 – Окно задания параметров исходной матрицы данных

где: *In a new Workbook* – Создание новой рабочей книги

As a stand-alone window – Создание листа рабочей книги

Шаг 2. После соответствующих установок переходим к рабочему листу электронной таблицы (рисунок 1.11). В данном пакете переменные (столбцы) имеют стандартные имена *Var1*, *Var2*, *Var3*, ..., как правило, исследователь заменяет их на необходимые обозначения, используемые в анализе. В нашем случае необходимо заменить *Var1* на *Y*, *Var2* на *X1*, *Var3* на *X2*. Для этого, необходимо дважды щелкнуть по заголовку переменной (*Var1*) или выбрать в главном меню *Data* → *Variable Specs...* (Данные → Спецификация переменных)

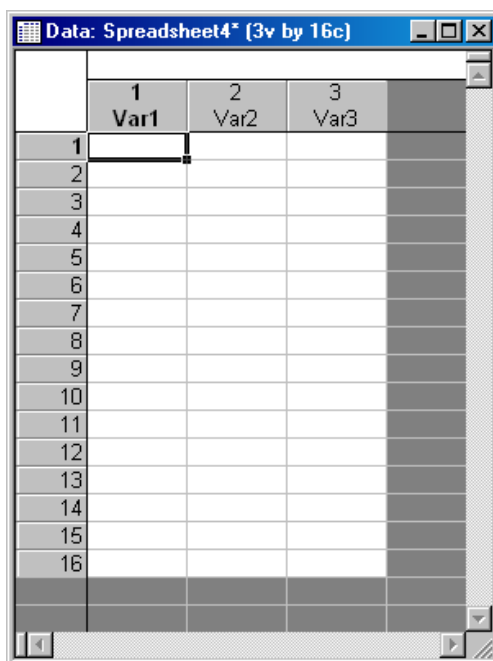


Рисунок 1.11 – Окно рабочей книги электронной таблицы

Шаг 3. В появившемся окне (рисунок 1.12) можно изменить шрифт, его размер и т.д. В группе опций *Display format* (Выводимый формат) можно задать формат данных содержащихся в соответствующем столбце. В поле *Long name (label or formula with Functions)*: (Длинное имя (вставка функции)) можно внести дополнительную информацию по переменной или преобразовать (создать новую) переменную введя формулу.

В данном случае ограничимся внесением в поле *Name:* (Имя:) вместо обозначения *Var1* букву *Y* и нажмем кнопку ОК. Далее вносим данные в поле таблицы получаем результат представленный на рисунке 1.13.

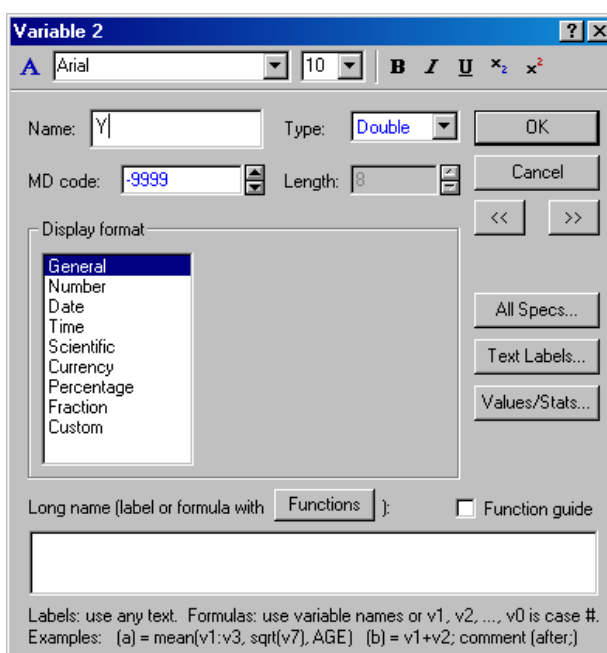


Рисунок 1.12 – Окно индивидуальных установок переменной

	1 Y	2 X1	3 X2
1	20	10	12
2	35	15	10
3	30	20	9
4	45	25	9
5	60	40	8
6	69	37	8
7	75	43	6
8	90	35	4
9	105	38	4
10	110	55	5
11	120	50	3
12	130	35	1
13	130	40	2
14	130	55	3
15	135	45	1
16	140	65	2

Рисунок 1.13 – Результат создания файла данных в пакет STATISTICA

1.8 Форматирование файла данных

Часто при создании рабочей таблицы (рабочей книги) возникает необходимость добавления или удаления строк (столбцов).

Например, в таблицу 1.1 (приложение 1) необходимо добавить новый уровень ряда за 1974г. равный 28,5 ц/га., для этого в главном меню выберем *Insert* → *Add Cases* (Вставка→Добавить значение/строку). В появившемся окне (рисунок 1.14) укажем, сколько строк необходимо ввести (*How many*), в нашем случае одно значение. Также необходимо указать после какой строки произвести вставку нового значения (*Insert after cases*), в нашем случае последним значением является 19.

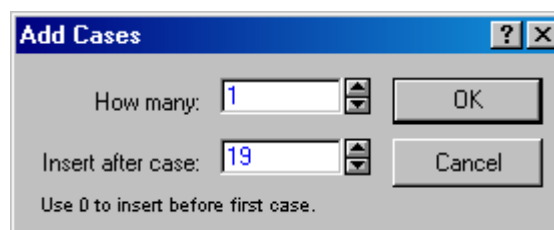


Рисунок 1.14 – Окно установок для вставки строки

В результате проведения данной процедуры получаем следующие результаты (значение 1974г. и 28,5 вводится в ручную):

1969г.	18,9	
1970г.	22,8	
1971г.	23,1	
1972г.	19,6	
1973г.	27	
1974г.	28,5	

Рисунок 1.15 – Результат добавления нового значения (приведена часть исходного окна)

Аналогичным образом производится добавление новой переменной *Insert* → *Add Variables* (Вставка → Добавить переменную).

Чтобы удалить ненужной строки или столбец (индивидуальное значение или переменную) можно поступить двумя способами:

Во-первых, чтобы удалить значение (строку) необходимо навести курсор мыши на заголовок строки, как показано на рисунке 1.16.

1970г.	22,8	
1971г.	23,1	
1972г.	19,6	
1973г.	27	
1974г.	28,5	

Рисунок 1.16 – Выделение удаляемой строки (приведена часть исходного окна)

Далее щелкнуть правой кнопкой мыши, в результате становится доступным меню форматирования строки (рисунок 1.17).

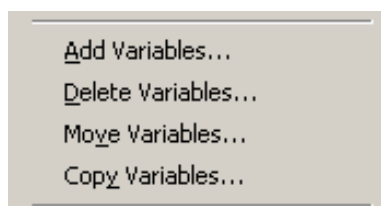


Рисунок 1.17 – Меню форматирования строки (приведена часть исходного окна)

В данном меню выберем пункт *Delete Cases...* (Удалить строку), после этой операции выделенная строка будет удалена из таблицы (аналогичным образом удаляется из таблицы переменная).

Во-вторых, можно воспользоваться кнопками *Vars* и *Cases* на панели *Spreadsheet* (Форматирование). Для удаления уровня ряда за 1974г. необходимо

выбрать *Cases*→*Delete...* (Строка→Удаление). В появившемся окне (рисунок 1.18) указать с какой строки (*From cases*) начинать удаление и по какую строку (*To cases*), в нашем случае удалению подлежит одна строка с номером 20.

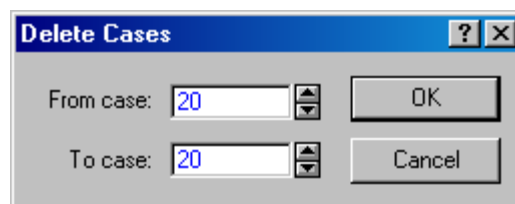


Рисунок 1.18 – Удаление строки (*Delete Cases*)

Процедура удаления переменной осуществляется аналогичным образом.

1.9 Сохранение файла данных

Для того чтобы сохранить созданный файл данных, необходимо выбрать в строке главного меню пакета команду *File* → *Save As...* (Файл → Сохранить как), после этого откроется диалоговое окно *Save As* (Сохранить как).

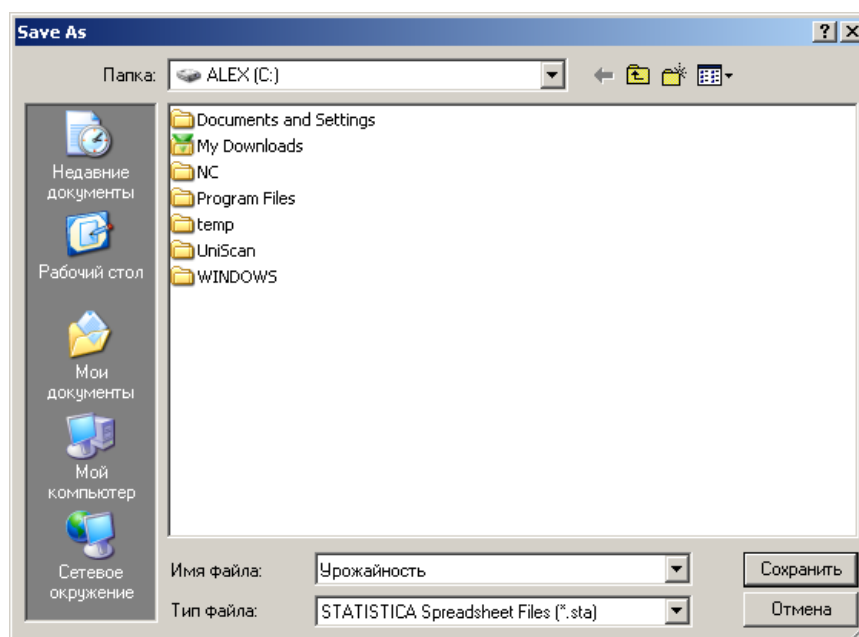


Рисунок 1.19 – Диалоговое окно *Save As* (Сохранить как)

В данном окне укажем тип и название созданного файла, при этом по умолчанию STATISTICA сохраняет файл данных в текущем каталоге с расширением **.sta*. В связи с тем, что в рамках курса «Эконометрика» выполняется 10 работ необходимо четко определить систему каталогов для сохранения результатов анализа. Предлагается следующие обозначения папок:

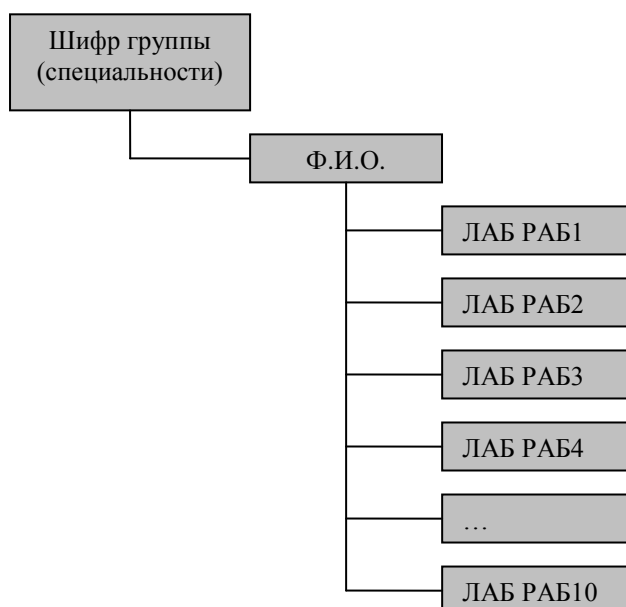


Рисунок 1.20 – Система каталогов для сохранения результатов анализа

Стоит упомянуть о том, что файлы, получаемые при анализе данных в пакете STATISTICA можно разделить на два типа:

- 1 Рабочая таблица (книга). В связи с этим рекомендуется все таблицы сохранять под именем *Таблица X.x.sta*. (X – номер выполняемой работы, x – номер таблицы в данной работе).
- 2 Результаты расчетов и рисунки (графики) выводятся в *Workbook* (Рабочая книга). В связи с этим данные этого типа рекомендуется сохранять под именем *Итоги X.x.stw* (X – номер выполняемой работы, x – номер итоговой таблицы или рисунка в данной работе).

1.10 Построение графиков в пакета STATISTICA

Обширное количество графиков в пакете доступно в пункте меню *Graphs* (Графики). Воспользуемся данными, приведенными в таблице А.1 (приложение 1) и построим график динамики.

Шаг 1. В главном меню выберем *2D Graphs* → *Line Plots (Variables)*... (2DГрафики → Линейный график), в результате будет вызвано следующее окно (рисунок 1.21).

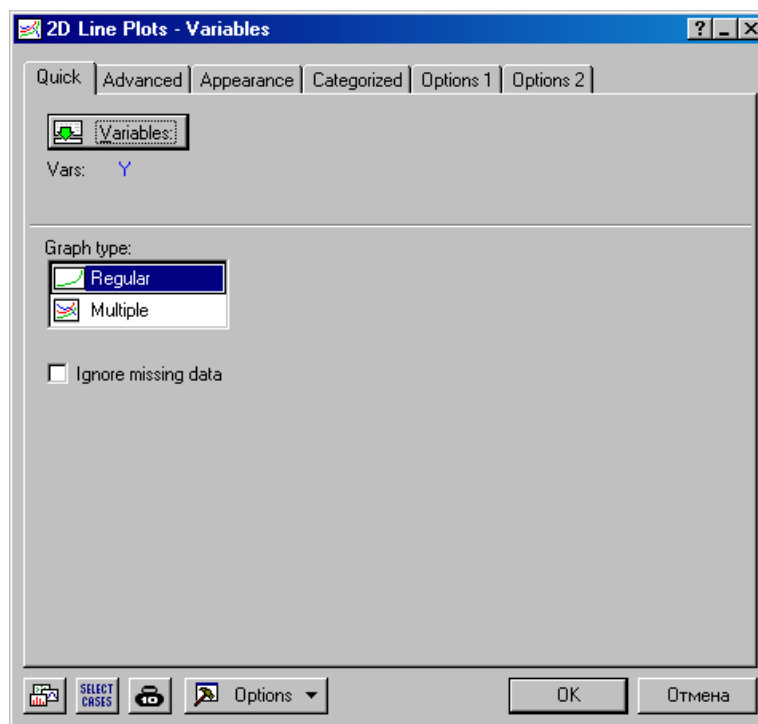


Рисунок 1.21 – Окно установок линейного графика

Шаг 2. Прежде чем приступить непосредственно к построению графиков необходимо выделить одно очень важное замечание: все действия по анализу и визуализации данных необходимо начинать с указания переменных, для этого используется кнопка *Variables:* (Переменные). В появившемся окне *Select Variables for Line Plot* (Выбор переменных для построения линейного графика) выберем переменную, на основе которой будет построен график, в данном случае это *Y*.

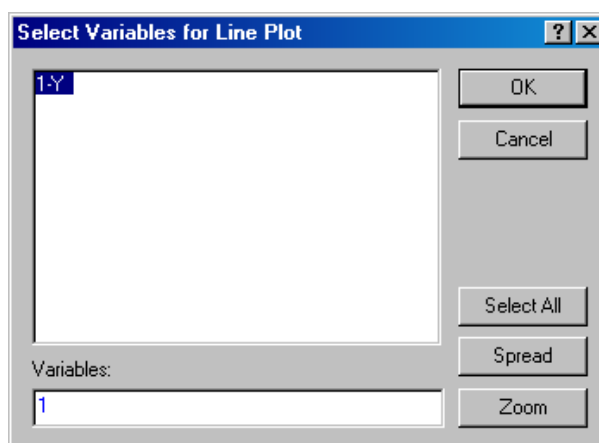


Рисунок 1.22 – Окно выбора переменной для построения линейного графика (*Select Variables for Line Plot*)

После выбора переменной на основе которой необходимо построить график, получаем следующий результат:

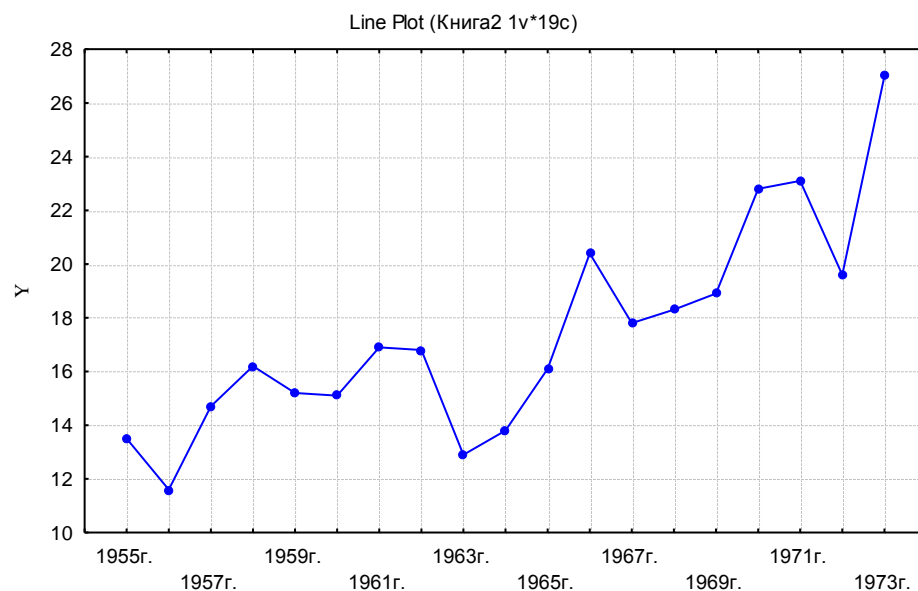


Рисунок 1.23 – Динамика урожайности зерновых культур, ц/га

Воспользуемся данными, приведенными в таблице 1.2 (приложение 1), построим двухмерное и трехмерное корреляционное поле.

Для построения двухмерного корреляционного поля в главном меню выберем *Graphs*→*2D Graphs*→*Scatterplots* (Графики → Двухмерные графики → Точечная диаграмма), в появившемся окне необходимо указать переменные, как показано на рисунке 1.24.

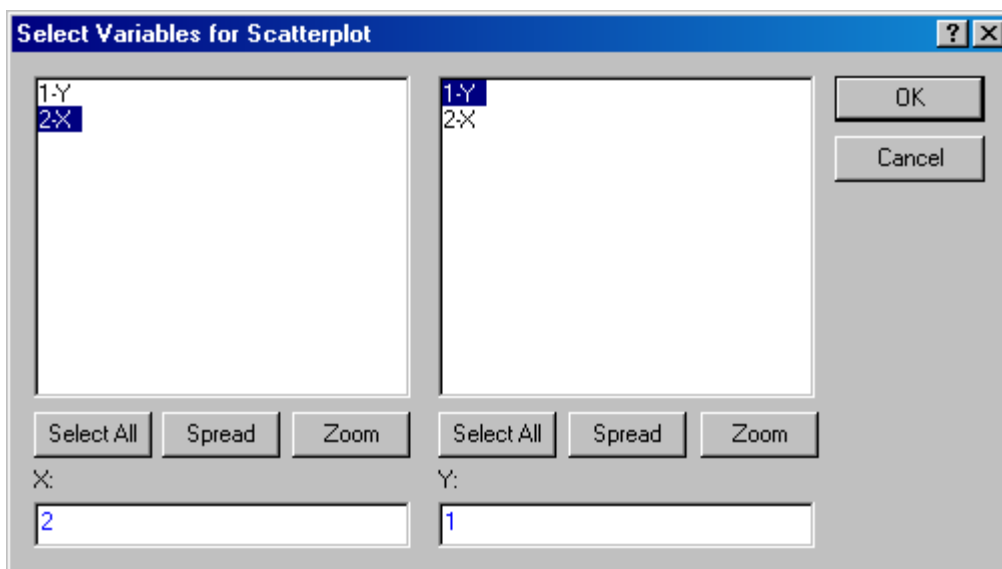


Рисунок 1.24 – Окно выбора переменных для построения поля корреляции (*Select Variables for Scatterplot*)

В результате получаем следующий график:

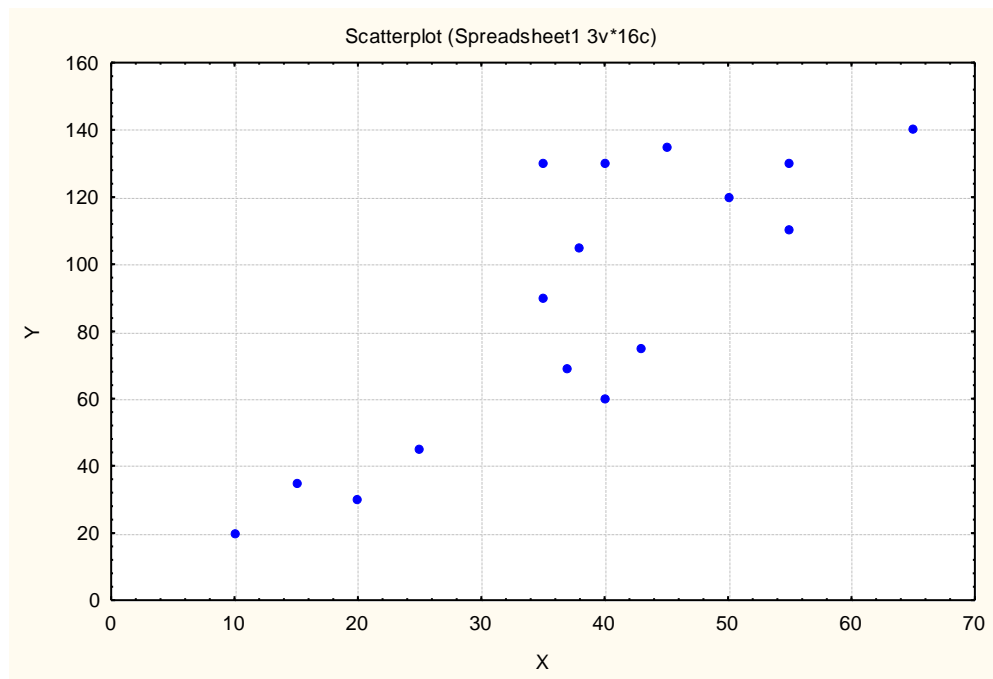


Рисунок 1.25 – Двухмерное корреляционное поле

Для построения трехмерного корреляционного поля в главном меню выберем *Graphs* → *3D XYZ Graphs* → *Scatterplots* (Графики → Трехмерные графики → Точечная диаграмма), в результате получим следующий график.

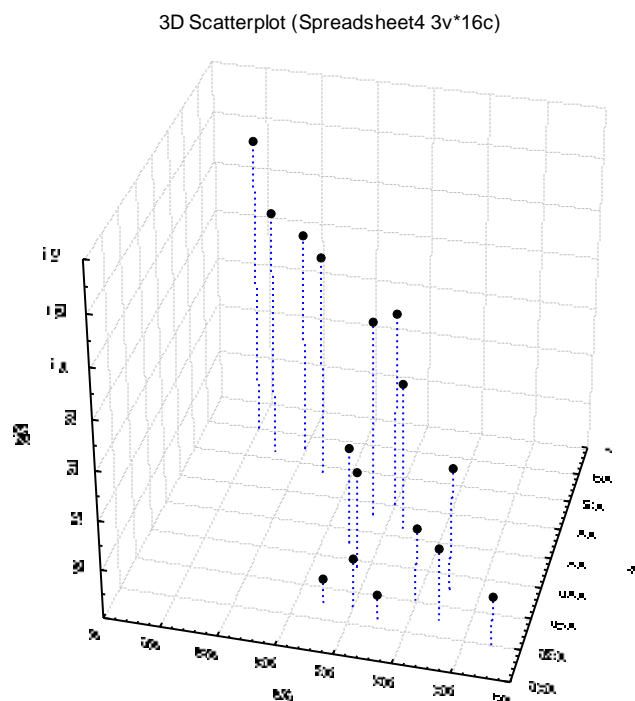



Рисунок 1.26 - Трехмерное корреляционное поле

Для корректировки (вращения) полученного графика необходимо выбрать *View*→*Rotete* (Вид → Вращение) либо использовать кнопку  меню *Graphs Tools* (Настройки графиков).

1.11 Тесты для самоконтроля

1 Выстройте в логической последовательности этапы эконометрического моделирования:

- верификация модели;
- информационный;
- априорный;
- параметризация;
- постановочный;
- идентификация модели.

2 Эконометрика как наука получила свое начало от таких наук как:

- математика, кибернетика, статистика;
- математика, статистика, экономическая теория;
- математика, статистика, экономика предприятия.

3 Выбор вида функциональной зависимости в уравнении регрессии называется:

- агрегированием модели;
- параметризацией модели;
- линеаризацией модели;
- структуризацией модели;
- спецификацией модели.

4 Эконометрическая модель - это модель:

- гипотетического экономического объекта;
- конкретно-существующего экономического объекта, построенная на гипотетических данных;
- конкретно-существующего экономического объекта, построенная на статистических данных.

5 На основе, каких данных можно оценить трендовую модель:

- пространственных данных;
- панельным данным;
- временным рядам.

6 Можно ли совместно включать в множественное уравнение регрессии факторы X_1 и X_2 если коэффициент корреляции между ними равен 0,87:

- можно;
- нельзя;
- корреляция между ними не имеет ни какого значения.

7 Объем изучаемой совокупности равен 15, можно ли включать в множественное уравнение регрессии 5 независимых факторов:

- можно;
- нельзя;

- не влияет на результаты исследования.
- 8 Модель, отражающая положительную зависимость предложения денег от ставки процента, является:
- мезомоделью;
 - макро моделью;
 - микро моделью.
- 9 Чем точнее информация об исследуемом объекте, тем:
- больше доля «черного ящика»;
 - меньше доля «черного ящика»;
 - качество информации не влияет на долю «черного ящика» в моделировании.
- 10 Какие бывают виды зависимости между переменными:
- экономические;
 - функциональные;
 - статистические;
 - сильные.

1.12 Задание для самостоятельного выполнения

Задания для самостоятельной работы составлены в пяти вариантах, номер варианта выбирается в соответствии с последней цифрой зачетной книжки студента:

Таблица 1.1 – Задания для самостоятельной работы по теме «Организация данных и построение графиков в пакете STATISTICA»

Последняя цифра номера зачетной книжки	1 и 6	2 и 7	3 и 8	4 и 9	5 и 0
Вариант №	1	2	3	4	5

Используя данные соответствующего варианта (приложение 2 и 3), необходимо выполнить следующие этапы самостоятельной работы:

- 1 Введите данные в STATISTICA (либо используя табличный редактор, либо вводя данные непосредственно в поле пакета программ);
- 2 Сохраните данные на диске (или ином носителе);
- 3 Постройте линейный график и двухмерное поле корреляции.

2 Выявление и измерение корреляционной связи

2.1 Цели и задачи лабораторной работы

В данной лабораторной работе на фактическом примере ознакомимся с методами выявления и измерения корреляционной связи, при этом будут решаться следующие задачи:

- 1 Провести процедуру сопоставления двух параллельных рядов;
- 2 Провести аналитическую группировку;
- 3 Построить корреляционное поле;
- 4 Рассчитать коэффициенты ранговой корреляции Спирмена и Кендала;
- 5 Рассчитать парный линейный коэффициент корреляции Пирсона.

2.2 Понятие корреляции и методы ее выявления

Теория корреляции начала развиваться во второй половине XIX века и особого расцвета достигла в XX в. Основателями теории корреляции являются английские биометрики Ф. Гальтон и К. Пирсон в России их идеи получили развитие в трудах А.А. Чупрова.

В эконометрике различают следующие варианты зависимостей:

- 1 **парная корреляция** – связь между двумя признаками, один из которых результативный, а другой факторный;
- 2 **частная корреляция** – зависимость между результативным и одним факторным признаком, при фиксированном значении других факторных признаков;
- 3 **множественная корреляция** - зависимость результативного признака от нескольких факторных признаков.
- 4 **каноническая корреляция** – зависимость группы результативного признака от группы факторных признаков.

Использование методов корреляции позволяет решить следующие задачи:

- 1 установить абсолютное изменение результативного признака за счет изменения одного или комплекса факторов;
- 2 определить меру зависимости между результативным признаком и одним из факторов при постоянном значении других;
- 3 установить меру относительного изменения зависимой переменной на единицу относительного изменения фактора или факторов;
- 4 изучить общий объем вариации результативного признака и определить роль каждого фактора в объяснении этого изменения;
- 5 оценить статистическую надежность выборочных показателей корреляционной связи.

Для выявления наличия корреляционной связи можно назвать несколько методов:

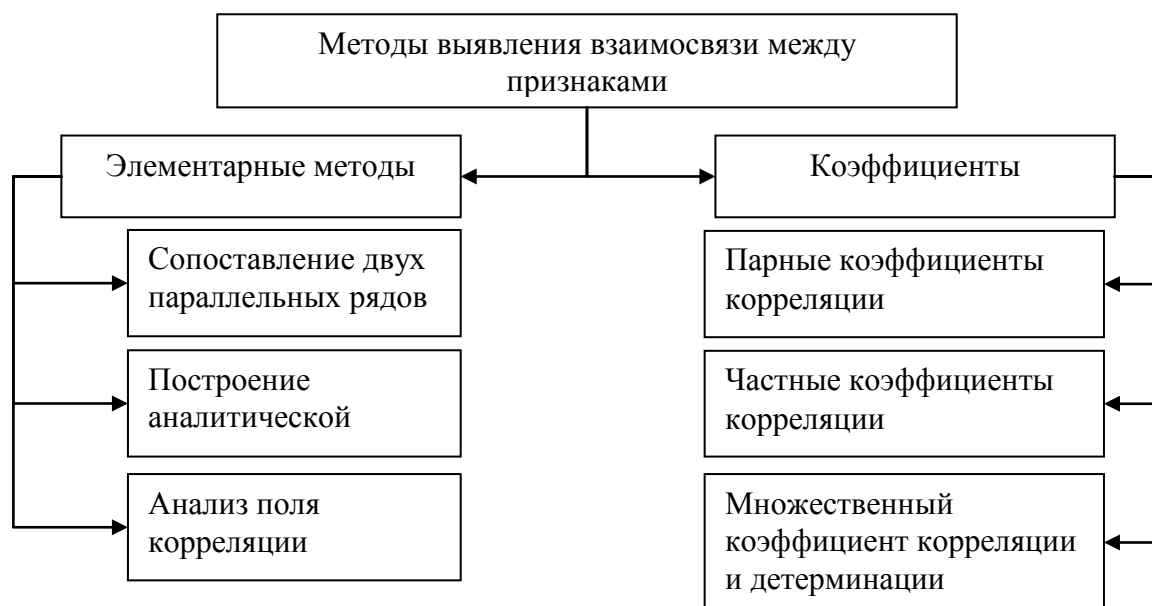


Рисунок 2.1 – Классификация методов выявления и анализа взаимосвязей

В данной работе последовательно рассмотрим реализацию приведенных на рисунке 2.1 методов в пакете STATISTICA.

2.2 Рекомендуемая литература


Для лучшего понимания материала изложенного в данной главе необходимо дополнительно проанализировать следующие источники литературы (см. список используемой литературы):

- 1 Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: Учебное пособие. – М.: ФОРУМ: ИНФРА-М, 2004. – 464с. (см. стр. 115-121, 225-232)
- 2 Дрейнер Н., Смит Г. Прикладной регрессионный анализ. В 2-х кн. Пер. с англ. – М.: Финансы статистика, 1986. – 366 с. (см. Т1 стр. 63-68)
- 3 Практикум по курсу «Статистика» (в системе STATISTICA). Салин В.Н., Чурилова Э.Ю. – М.: «Издательский Дом» Социальные отношения», Издательство «Перспектива». 2002. – 188с. (см. стр. 71-96)
- 4 Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. 2-е изд., испр. - Т. 2: Айвазян С.А. Основы эконометрики. – М.: ЮНИТИ-ДАНА, 2001. -432с. (см. стр. 46-47)
- 5 Эконометрика: Учебник / Под ред. И.И. Елиссеовой. 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2005. – 576 с. (см. стр. 51-63, 99-106, 136-145)

2.3 Сопоставление двух параллельных рядов

Для реализации данного метода в пакете STATISTICA воспользуемся данными по 25 предприятиям в качестве зависимой переменной выберем стоимость произведенной продукции (тыс. руб.), в качестве не зависимой переменной - среднегодовую стоимость основных производственных фондов (тыс. р.), исходные данные содержатся в приложении 2.

Шаг 1. Введем данные в электронную таблицу пакета STATISTICA (см. лабораторная 1).

Шаг 2. Упорядочим имеющиеся данные по возрастанию для этого в главном меню выберем *Date*→*Sort...* (Данные →Сортировка), либо нажмем кнопку  на панели *Spreadsheet* (Форматирование).

Шаг 3. В появившемся окне (рисунок 2.2) необходимо указать по какой переменной производится сортировка, в данном случае это *Y* (как видим на рисунке 2.2, существует также возможность сортировать по нескольким переменным).

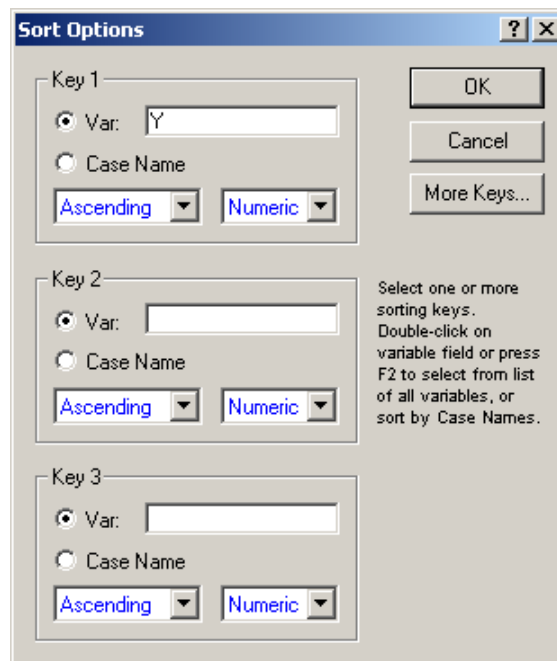


Рисунок 2.2 – Опции сортировки данных

Шаг 4. На заключительном этапе рассматриваемого алгоритма отобразим на одном графике упорядоченные переменные *Y* и *X*. Для этого в главном меню выберем *Graphs* → *2D Graphs*→*Line Plots (Variables)...* (Графики → 2DГрафики → Линейный график). В появившемся окне *2D Line Plots – Variables* обратимся к вкладке *Advanced* (Расширенные), далее в группе опций *Graph type:* (Установки графика:) выделим *Multiple* (Множественный), с целью одновременного выведения обеих переменных на график.

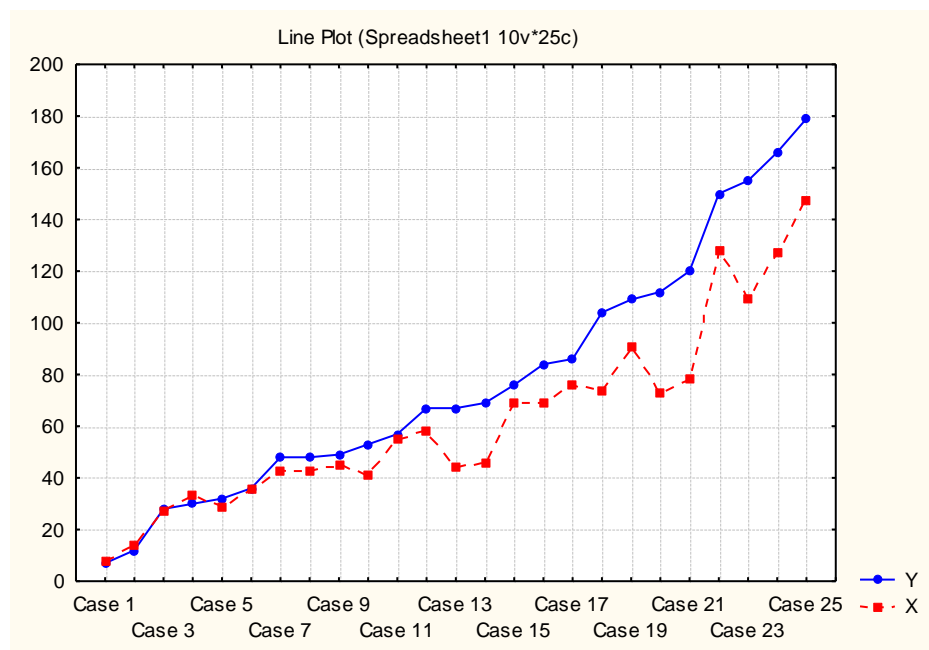


Рисунок 2.3 – Вариация показателей Y и X

Опираясь на полученный график (рисунок 2.3) можно сделать вывод, что значения упорядоченной переменной Y растут, параллельно растут и значения переменной X , соответственно можно сделать предположение о наличии корреляционной связи между анализируемыми признаками.

2.4 Построение аналитической группировки

С целью изучения зависимости между среднегодовой стоимостью основных производственных фондов X и объемом произведенной продукции Y проведем процедуру аналитической группировки. Основанием группировки будет являться стоимость произведенной продукции.

Шаг 1. Определим количество групп, на которые необходимо разбить совокупность, для этого воспользуемся формулой Стерджесса:

$$n = 1 + 3,322 \cdot \lg N \quad (2.1)$$

где: N – число единиц изучаемой совокупности.

в нашем случае получаем: $n = 1 + 3,322 \cdot \lg 25 = 6$

Шаг 2. Найдем величину интервала используя формулу:

$$h = \frac{X_{\max} - X_{\min}}{n} \quad (2.2)$$

где: X_{\max} , X_{\min} – максимальное и минимальное значение единиц в изучаемой совокупности

Для получения максимального и минимального значения в пакете, выберем *Statistics* → *Basic Statistics/Tables* (Статистики → Основные статистики и таблицы). В появившемся окне *Basic Statistics and Tables* (Основные статистики и таблицы) необходимо выбрать пункт *Descriptive statistics* (Описательные статистики).

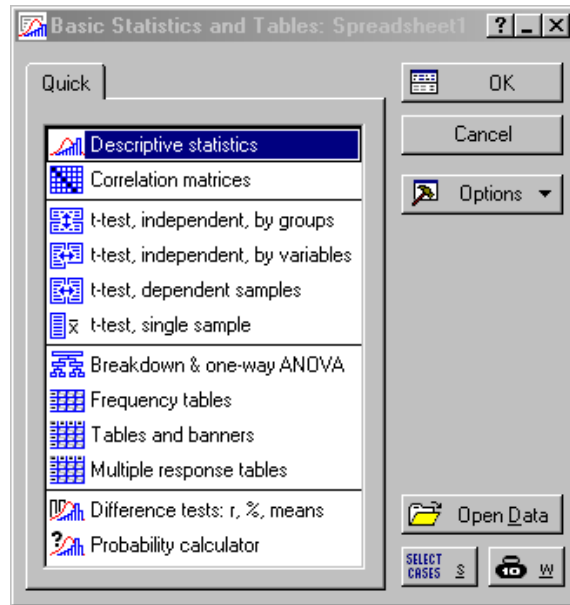


Рисунок 2.4 – Окно выбора основных статистик и таблиц

В окне *Descriptive Statistics* (Описательные статистики) укажем по каким переменным проводится расчет, для этого выберем кнопку *Variables* (Переменные) и выделим *Y* и *X*.

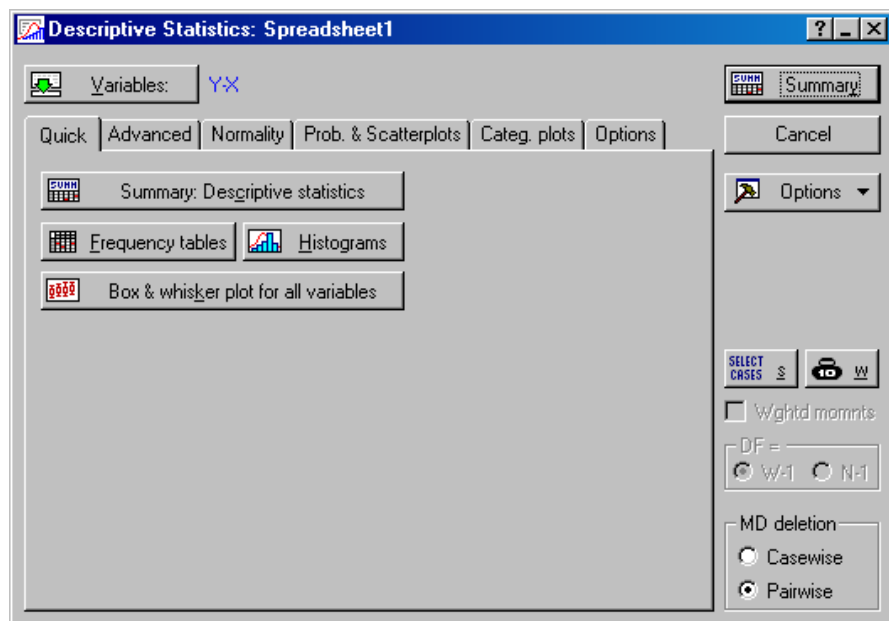


Рисунок 2.5 – Окно установок дескриптивных статистики

Нажав кнопку *Summary* (Вычислить) получим результаты, представленные в таблице 2.1.

Таблица 2.1 – Результаты расчета описательных статистик

	<i>Valid N</i> (N набл.)	<i>Mean</i> (Среднее)	<i>Minimum</i> (Минимум)	<i>Maximum</i> (Максимум)	<i>Std.Dev.</i> (Стд.откл.)
Y	25	77,76	7	179	48,231
X	25	62,52	8	147	35,831

Шаг 3. Согласно полученным данным размах вариации будет равен $h=(179-7)/6=28,6$. Соответственно величины интервалов будут следующие: [7, 35,67); [35,67, 64,3); [64,3, 93,00); [93,00, 121,67); [121,67, 150,33); [150,33, 179,00].

Шаг 4. Прежде чем приступить непосредственно к построению группировки в пакете, образуем новую переменную (следующую после переменной X), для этого выберем *Insert*→*Add Variables* (Вставка →Добавить переменную). Далее в окне *Variable 3* (Переменная 3) заменим *Var 3* на *Y1*, а в поле *Long name (label or formula with Functions)*: (Длинное имя (вставка функции)) введем выражение *=v1* (в результате выполнения данного действия будет выполнен перенос всех данных из первого столбца в текущий) (рисунок 2.6).

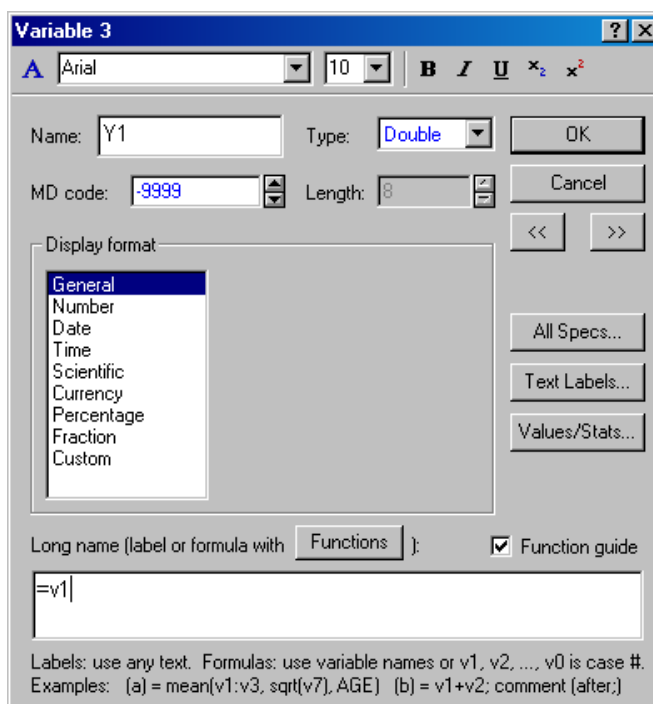


Рисунок 2.6 – Окно установок переменной

Прежде чем перейти к дальнейшему выполнению алгоритма построения аналитической группировки сделаем следующее замечание: так как переменные используемые исследователем могут быть обозначены как угодно, в данной программе все переменные, вводимые в поле *Long name*, обозначаются как *v1*, *v2*, *v3*, и т.д. Соответственно очень важно при

образовании новой переменной или создании формул следить за правильностью ссылки на нужные переменные.

Шаг 5. Чтобы произвести группировку по переменной *Y1* нужно ее выделить, щелкнув по заглавию, далее в главном меню выбрать *Date*→*Recode* (Данные→Перекодировка). В появившемся окне *Recode Values of Variable 3* (Перегруппировка значений переменной 3) укажем интервалы группировки, а в группе опций *New Value* (Новое значение) последовательно введем номера групп в которые попали предприятия - 1, 2, 3, 4, 5, 6 (рисунок 2.7).

То есть должны получить следующее:

- 1 – $v3 \geq 7$ and $v3 < 35,67$;
- 2 – $v3 \geq 35,67$ and $v3 < 64,33$;
- 3 – $v3 \geq 64,33$ and $v3 < 93,00$;
- 4 – $v3 \geq 93,00$ and $v3 < 121,67$;
- 5 – $v3 \geq 121,67$ and $v3 < 150,33$;
- 6 – $v3 \geq 150,33$ and $v3 \leq 179,00$.

При вводе интервалов проследите чтобы все группы были заполнены, для этого используйте прокрутку окна.

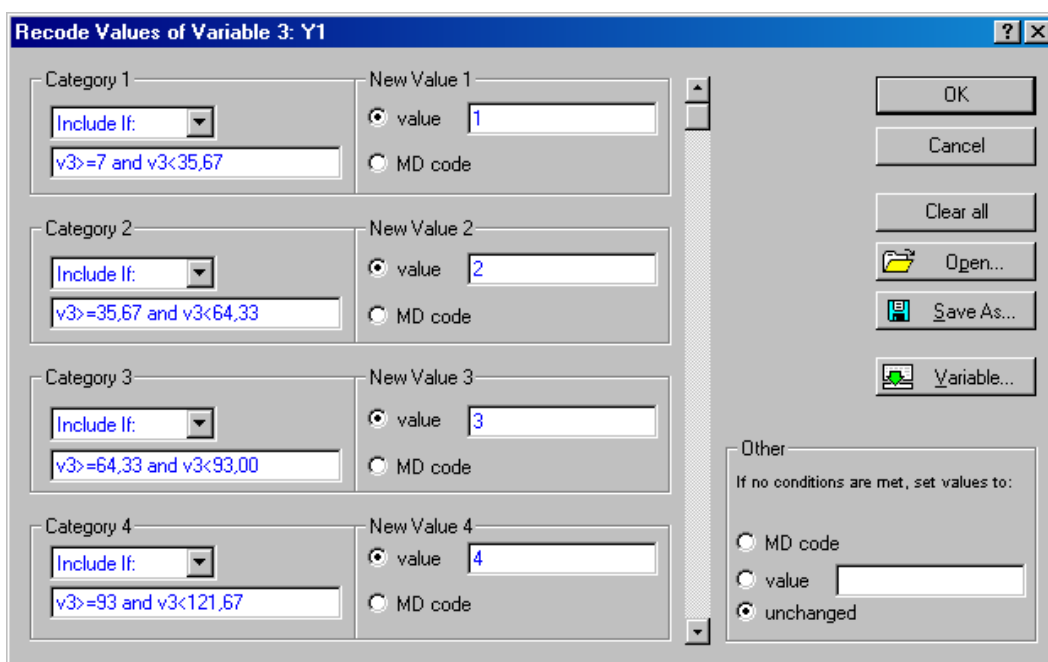


Рисунок 2.7 – Запись границ интервалов

Шаг 6. Полученные данные необходимо упорядочить в порядке возрастания по переменной *Y1*.

Шаг 7. Для определения описательных статистик по группам предприятий вернемся в модуль *Basic Stat/Tables* (рисунок 2.4), в меню стартовой панели модуля выберите опцию *Breakdown and one-way ANOVA* (Разбиение и однофакторный дисперсионный анализ).

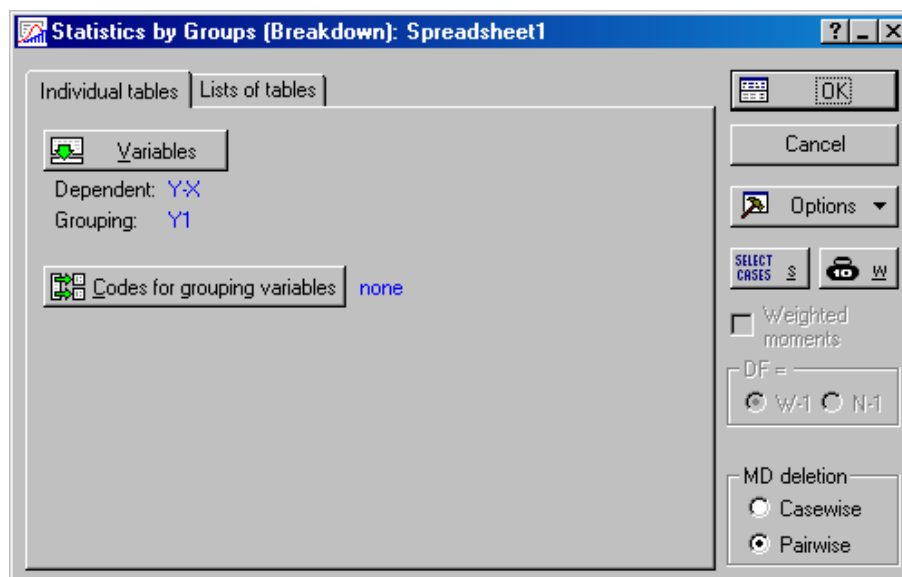


Рисунок 2.8 – Установки для проведения группировки

В качестве группирующей переменной (*Grouping*) выберем *Y1*, а в качестве зависимых переменных (*Dependent*): *Y* и *X*.

Шаг 8. В появившемся окне (рисунок 2.9) необходимо выбрать кнопку *Summary: Table of statistics* (Итоговая таблицы и статистики).

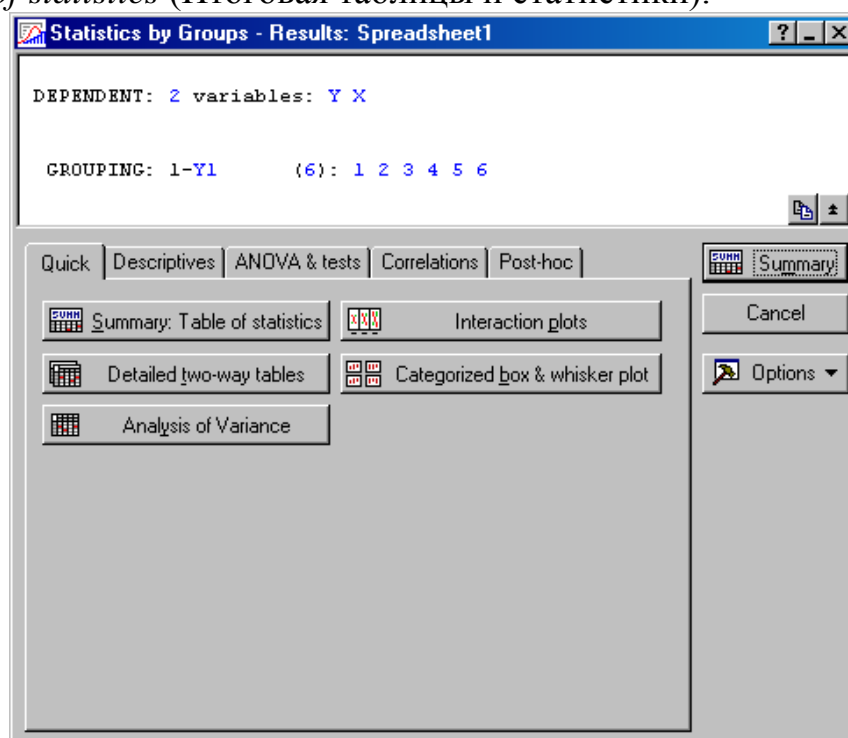


Рисунок 2.9 – Окно результатов классификации и дисперсионного анализа

В полученной таблице 2.2 содержатся все необходимые результаты по группам предприятий для переменных *Y* и *X*.

Таблица 2.2 – Результаты проведения аналитической группировки

<i>YI</i>	<i>Y Means</i> (Средняя)	<i>Y N</i>	<i>Y Std. Dev</i> (Ст. Откл)	<i>X Means</i> (Средняя)	<i>X N</i>	<i>X Std. Dev</i> (Ст. Откл)
1	21,8	5	11,5	22,2	5	10,7
2	48,5	6	7,1	43,8	6	6,3
3	74,8	6	8,6	60,3	6	13,2
4	111,3	4	6,7	79,0	4	8,3
5	150,0	1	0,0	128,0	1	0,0
6	166,7	3	12,0	127,7	3	19,0
<i>All Grup</i>	77,8	25	48,2	62,5	25	35,8

В столбце *Means* (Средние) содержатся значения средних арифметических по каждой группе, *N* – содержится количество индивидуальных значений вошедших в конкретную группу, *Std. Dev.* – стандартное отклонение.

Согласно приведенным в таблице 2.2, данным наблюдается рост групповых средних значений переменной *Y*, также наблюдается рост средних значений по переменной *X*, т.е. можно предположить наличие прямой корреляционной зависимости между рассматриваемыми признаками.

2.5 Графический метод (построение поля корреляции)

Для построения двухмерного корреляционного поля используем данные приложения 4, выберем *Graphs*→*2D Graphs*→*Scatterplots* (см. лабораторная 1). Далее выберем кнопку *Variables* (переменные) в появившемся окне *Select Variables for Scatterplot* (Выделение переменных для корреляционного поля) укажем переменные которые будет располагаться по оси *OY* и *OX*.

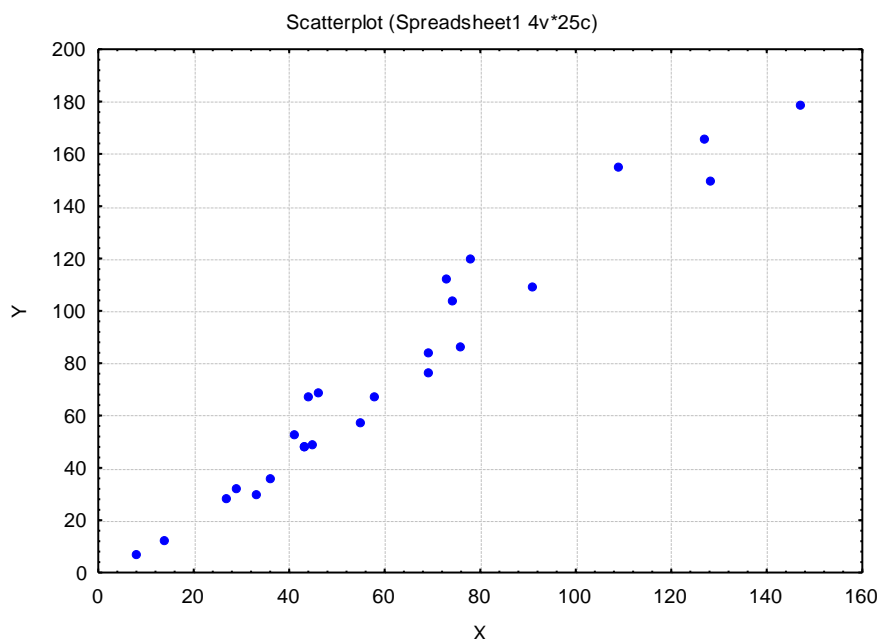


Рисунок 2.10 – Двухмерное корреляционное поле

Согласно приведенному графику, с ростом значений Y наблюдается рост значений X , т.е. имеем прямую взаимосвязь между признаками.

2.6 Расчет коэффициента ранговой корреляции Спирмена и Кендала

Опираясь на данные, приведенные в приложении 4, рассчитаем значения коэффициентов ранговой корреляции.

Шаг 1. Образует, дополнительные переменные $Y2$ и $X2$, в которые переносим значения с первого и второго столбца (Y и X). Сделать это можно двумя способами: во-первых, скопировав данные с помощью команд *Copy* (Копировать) и *Paste* (Вставить). Во-вторых, воспользоваться полем *Long Name*: окна настройки параметров переменных и ввести – для $Y2$ - $=v1$, для $X2$ - $=v2$.

Шаг 2. Так как ранговые коэффициенты строятся на основе ранжированных значений необходимо провести ранжирование имеющейся совокупности. Для этого в главном меню выберем *Date*→*Rank*→*Variables* (Данные→Ранг→Переменные) в окне *Rank Order Values* (Ранжирование совокупности).

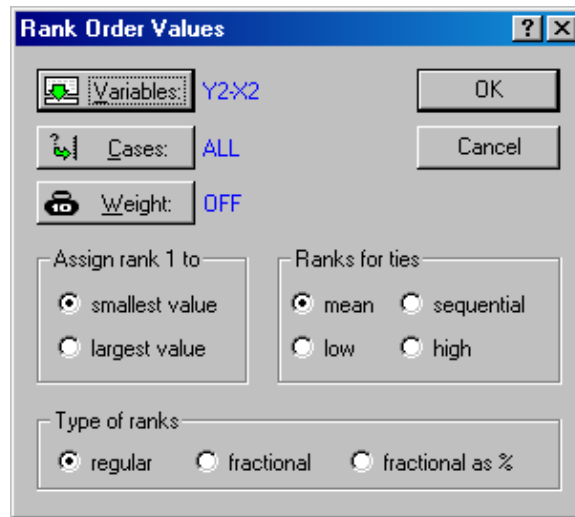


Рисунок 2.11 – Окно установок процедуры ранжирования

- где: Присвоить ранг 1 - *Assign rank 1 to*
наименьшему значению - *smallest value*
наибольшему значению - *largest value*
Ранги для совпадающих значений - *Ranks for ties*
средний - *mean*
наименьший - *low*
наибольший - *high*
последовательный - *sequential*
Тип рангов - *Type of rank*
обычный - *regular*
дробный - *fractional*
дробный в % - *fractional as %*

Шаг 3. В главном меню наберем *Statistics*→*Nonparametrics*→*Correlations* (*Spearman, Kendall tau, gamma*) (Вычисления → Непараметрические → Корреляция). В появившемся окне выберем закладку *Advanced* (Расширенные) что позволит рассчитать три коэффициента: корреляцию Спирмена R, статистику Тау Кендалла и статистику Гамма.

Шаг 3. Выберем переменную, по которой необходимо провести расчет, для этого нажмем кнопку *Variables* (в качестве переменных выберем Y2 и X2). Соответственно переменная Y2 будет выведена первой (List 1), а переменная X2 будет указана второй (List 2).

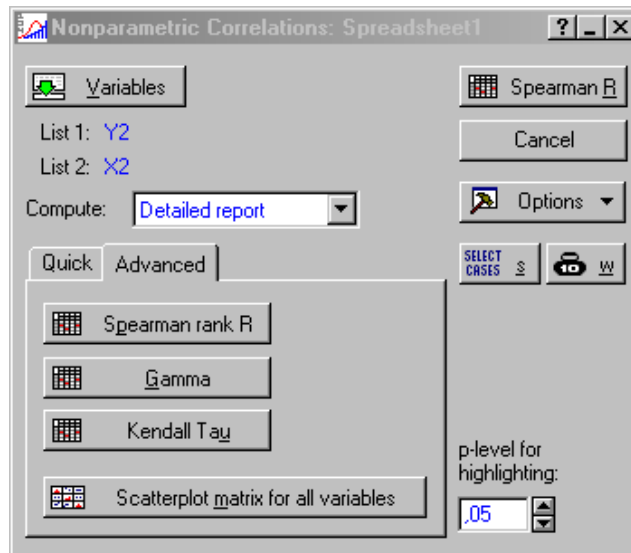


Рисунок 2.12 – Окно выбора расчета непараметрических показателей корреляции

Шаг 4. Рассчитаем ранговый коэффициент корреляции Спирмена, для этого выберем кнопки *Spearman R* (коэффициент Спирмена), получим следующие результаты.

Таблица 2.3 – Результаты оценивания значения коэффициента Спирмена

	<i>Valid N</i> (Число набл.)	<i>Spearman R</i> (Спирмена R)	$t(N-2)$	<i>p-level</i> (p-уров.)
Y2 & X2	25	0,978	22,314	0,000

Согласно данным, приведенным в таблице 2.3, коэффициент Спирмена получен равным 0,978, т.е. наблюдается прямая сильная корреляционная взаимосвязь между Y и X.

2.7 Расчет парного линейного коэффициента корреляции Пирсона

Для расчета значения парного линейного коэффициента корреляции необходимо в главном меню выбрать *Statistics*→*Basic Statistics/Tables* (Статистики→Основные статистики и таблицы). В появившемся окне *Basic Statistics and Tables* (Основные статистики и таблицы) необходимо указать переменные, по которым будет проведен расчет, для этого выберем кнопку *One variable list* (Один список), далее *Summary: Correlation matrix* (Вычислить: Корреляционная матрица).

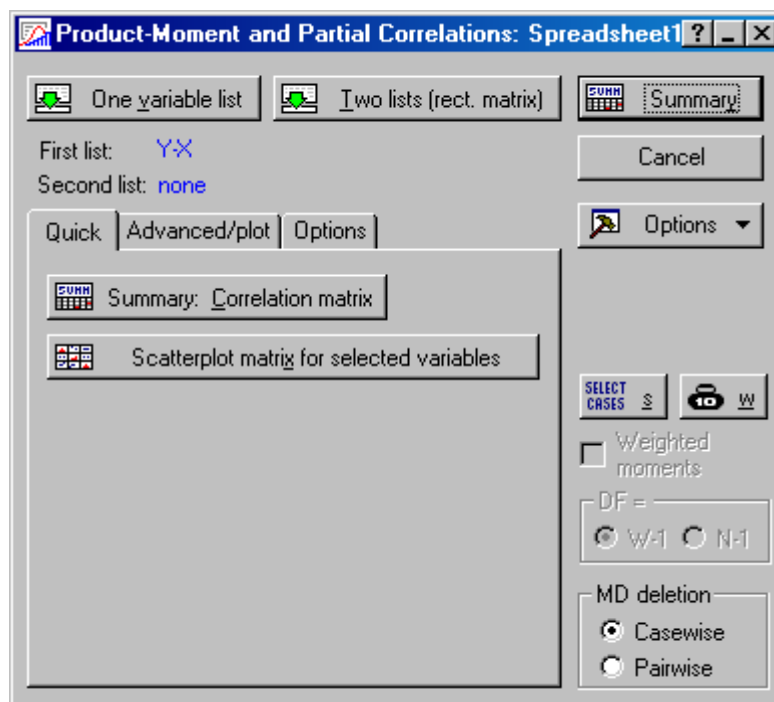


Рисунок 2.13 – Окно установок для расчета парных коэффициентов корреляции

Таблица 2.4 – Матрица парных коэффициентов корреляции

	Y	X
Y	1,000	0,977
X	0,977	1,000

Согласно полученным результатам, между исследуемыми показателями существует тесная линейная связь (коэффициент корреляции равен 0,977).

2.8 Тесты для самоконтроля

1 Какой из перечисленных методов не направлен на выявление корреляционной зависимости между переменными?

- построение корреляционного поля;
- построение аналитической группировки;
- сопоставление двух параллельных рядов;
- смыкание рядов динамики.

2 Предложенная формула $1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$ является:

- парным коэффициентом корреляции;
- коэффициентом корреляции рангов Спирмена;
- коэффициентом детерминации.

- 3 Приведенная формула $\frac{ad - bc}{\sqrt{(a+b)(b+d)(a+c)(c+d)}}$ используется для расчета:
- коэффициента ассоциации Д.Юла;
 - коэффициента контингенции К. Пирсона;
 - коэффициента взаимной сопряженности К. Пирсона.
- 4 Приведенная формула $\sqrt{\frac{\varphi^2}{1-\varphi^2}}$ необходима для расчета:
- коэффициента ассоциации Д.Юла;
 - коэффициента контингенции К. Пирсона;
 - коэффициента взаимной сопряженности К. Пирсона.
- 5 Основоположником теории корреляции являются:
- Ф. Гальтон;
 - Р. Фриш;
 - К. Пирсон;
 - В. Кетли.
- 6 Какой из перечисленных методов не направлен на выявление корреляционной зависимости между переменными?
- построение корреляционного поля;
 - построение аналитической группировки;
 - сопоставление двух параллельных рядов;
 - смыкание рядов динамики.
- 7 Парный линейный коэффициент корреляции принимает значения в интервале:
- от -2 до 0;
 - от 0 до 1;
 - от -1 до +1.
- 8 Парный коэффициент детерминации принимает значения в интервале:
- от -2 до +2;
 - от 0 до 1;
 - от -1 до +1.
- 9 Отрицательный знак парного коэффициента корреляции указывает на:
- отсутствие зависимости x и y ;
 - обратную зависимость между x и y ;
 - прямую зависимость между x и y .
- 10 Предложенная формула $\eta = \sqrt{\frac{\delta^2}{\sigma_y^2}}$ является:
- коэффициентом корреляции;
 - коэффициентом Фихнера;
 - корреляционным отношением.

2.9 Задания для самостоятельного выполнения

Задания для самостоятельной работы составлены в пяти вариантах, номер варианта выбирается в соответствии с последней цифрой зачетной книжки студента:

Таблица 2.5 – Задания для самостоятельной работы по теме «Выявление и измерение корреляционной связи»

Последняя цифра номера зачетной книжки	1 и 6	2 и 7	3 и 8	4 и 9	5 и 0
Вариант №	1	2	3	4	5

Используя данные соответствующего варианта (приложение 5), необходимо выполнить следующие этапы самостоятельной работы:

- 1 Ввести данные в STATISTICA (либо используя табличный редактор, либо вводя данные непосредственно в поле пакета программ, либо непосредственно скопировать из соответствующего приложения);
- 2 Провести сопоставление двух параллельных рядов;
- 3 Построить аналитическую группировку;
- 4 Построить поле корреляции;
- 5 Оценить коэффициент ранговой корреляции Кендала Тау;
- 6 Оценить парный коэффициент корреляции Пирсона.

3 Построение классической линейной регрессии

3.1 Цели и задачи лабораторной работы

В данной лабораторной работе на практическом примере рассмотрим этапы построения уравнения классической линейной регрессии, при этом будут решаться следующие задачи:

- 1 Рассчитать описательные статистики, характеризующие изучаемые данные;
- 2 Определить парные коэффициенты корреляции и на их основе выявить факторы, оказывающие наибольшее влияние на результативный показатель;
- 3 Оценить регрессионное уравнение имеющимися факторами. Проанализировать множественные коэффициенты корреляции и детерминации, по полученной модели;
- 4 Оценить качество модели на основе t -статистики Стьюдента и F -статистики Фишера.

3.2 Понятие классической линейной регрессии

В данной главе остановимся на рассмотрении понятия классической линейной регрессии, при этом рассматриваются два возможных случая:

Множественная регрессия представляет собой модель результативного признака с двумя и большим числом факторов, т. е. модель вида:

$$\tilde{y}_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_m x_{mi} + \varepsilon_i \quad (3.1)$$

Парная линейная регрессия представляет собой частный случай множественной регрессии и есть модель между двумя переменными - y и x , т.е. имеем:

$$\tilde{y}_i = a_0 + a_1 x_i + \varepsilon_i \quad (3.2)$$

где: $i=1, 2, \dots, n$

n – объем изучаемой совокупности;

\tilde{y} - данные полученные в результате построения модели (теоретические уровни, модельные данные)

y – зависимая переменная;

x – независимая переменная;

a_0, a_1 – искомые параметры уравнения;

ε_i – случайная величина (возмущение, остатки, отклонения).

Основным методом решения задачи нахождения параметров a_0 и a_1 уравнения связи является метод наименьших квадратов (МНК). Он состоит в минимизации суммы квадратов отклонений фактических значений от значений, вычисленных по уравнению связи.

Основным параметром парного уравнения регрессии является параметр a_1 (в случае множественной регрессии a_j где $j = 1, 2, \dots, m$) который

характеризует силу связи между вариацией факторного признака x и вариацией результативного признака y ;

Иногда в эконометрических исследованиях возникают ситуации, в которых использование параметров a_j не дает желаемого результата, так как коэффициент имеет размерность совпадающую с анализируемым показателем и не пригоден для выявления наибольшего (наименьшего) влияния той или иной независимой переменной. В этом случае используют β - коэффициент или коэффициент эластичности.

β - коэффициент (стандартизованный коэффициент регрессии) показывает, на сколько среднеквадратических отклонений (β) изменится результативный признак, если величина факторного признака изменятся на одно среднеквадратическое отклонение.

$$\beta_{ji} = a_j \cdot \frac{\sigma_{xj}}{\sigma_y} \quad (3.3)$$

Коэффициенты условно-чистой регрессии полезно выразить в виде относительных сравниваемых показателей связи, коэффициентов эластичности:

$$\varepsilon_j = a_j \cdot \frac{\bar{x}_j}{\bar{y}} \quad (3.4)$$

Значение коэффициента определяет, на сколько процентов в среднем изменится значение зависимой переменной y если независимая переменная x изменится на 1%.

В большинстве случаев при построении модели приходится пользоваться выборочными данными, поэтому прежде чем приступить к использованию модели необходимо убедиться ее адекватности фактическим данным (анализируемому явлению). Для этих целей используют t-критерий Стьюдента и F-критерий Фишера.

3.3 Рекомендуемая литература

Для лучшего понимания материала изложенного в данной главе необходимо дополнительно проанализировать следующие источники литературы (см. список используемой литературы):

- 1 Афанасьев В.Н., Гуляева Т.И., Юзбашев М.М. Эконометрика: Учебник / Под общ. редакцией М.М. Юзбашева. – М.: Финансы и статистика, 2004. (см. стр. 30-61, 77-91)
- 2 Боровиков, В.П. Программа STATISTICA для студентов и инженеров. - 2-е изд. – М.: КомпьютерПресс, 2001. – 301с. - ISBN 5-89959-080-7. (см. стр. 125-161)
- 3 Бородич, С.А. Эконометрика: учеб. пособие / С.А. Бородич. – 3 –е изд., стер. – Мн.: Новое знание, 2006. – 408 с. ISBN 985-475-206-2 (см. стр. 98-154)
- 4 Кремер Н, Путко Б. Эконометрика: Учебник для вузов. – М.: ЮНИТИ-ДАНА. 2003. - 311с. (см. стр. 50-108)

- 5 Практикум по курсу «Статистика» (в системе STATISTICA). Салин В.Н., Чурилова Э.Ю. – М.: «Издательский Дом» Социальные отношения», Издательство «Перспектива». 2002. – 188с. (см. стр. 96-122)
- 6 Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. 2-е изд., испр. - Т. 2: Айвазян С.А. Основы эконометрики. – М.: ЮНИТИ-ДАНА, 2001. -432с. (см. стр. 49-53)
- 7 Эконометрика: Учебник / Под ред. И.И. Елиссеовой. 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2005. – 576 с. (см. стр. 43-155)

3.4 Расчет описательных (дескриптивных) статик

В пакете STATISTICA 6.0 существует возможность расчета огромного числа дескриптивных (описательных, элементарных) статистик (максимальные, минимальные, средние, показатели распределения и эксцесса и т.д.).

Прежде чем приступить к выполнению данного пункта, необходимо внести данные содержащиеся в приложении 6 в рабочую таблицу пакета и сохранить их на жестком диске (см. лабораторная работа 1).

Для расчета описательных статистик необходимо:

Шаг 1. В главном меню выбирать *Statistics*→*Basic Statistics/Tables* (Вычисления →Основные статистики и таблицы).

Шаг 2. В окне *Basic Statistics and Tables* выбирать первый пункт *Descriptive statistics* (Описательные статистики).

Шаг 3. В окне *Descriptive statistics* выбирать вкладку *Advanced* (Расширенные). Данное окно разделено на три группы показателей (рисунок 3.1):

1) *Location, valid N* (Объем совокупности) – содержатся структурные и степенные средние величины: *Valid N* (Число наблюдений N), *Mean* (Средняя), *Sum* (Сумма), *Median* (Медиана), *Mode* (Мода), *Geom. Mean* (Геометрическая средняя), *Harm. Mean* (Гармоническая средняя).

2) *Variation, moments* (Вариация, момент) – содержатся показатели относящиеся к вариации изучаемого признака и отражающие распределение переменной: *Standard Deviation* (Стандартное Отклонение), *Variance* (Вариация), *Std. err. of mean* (Стандартная ошибка среднего), *Conf. limits for means* (Доверительная граница для среднего), *Skewness* (Ассиметрия), *Std. err., Skewness* (Стандартная ошибка Ассиметрии), *Kurtosis* (Эксцесс), *Std. err., Kurtosis* (Стандартная ошибка Эксцесса);

3) *Percentiles, ranges* (Перцентели, ранги) – в группе собраны следующие показатели *Minimum & Maximum* (Максимум и минимум), *Lower & upper quartiles* (Нижний и верхний квартили), *Percentile boundaries* (), *Range* (Ранг), *Quartile range* (Ранг квартиля).

Шаг 4. Выберем для анализа следующие показатели (рисунок 3.1): *Valid N* (Число наблюдений N), *Mean* (Средняя арифметическая), *Standard Deviation* (Стандартное Отклонение), *Skewness* (Ассиметрия), *Kurtosis* (Эксцесс), *Minimum & Maximum* (Максимум и минимум).

После нажатия кнопки *Summary* (Вычислить) получаем следующие результаты:

Таблица 3.1 - Описательная статистика

	<i>Valid N</i> (N набл.)	<i>Mean</i> (Среднее)	<i>Std.Dev.</i> (Стд.откл.)	<i>Skewness</i> (Асимметрия)	<i>Kurtosis</i> (Эксцесс)
<i>Y</i>	15	37,667	11,684	0,266	-1,588
<i>X1</i>	15	32,400	13,809	0,225	0,675
<i>X2</i>	15	47,533	6,621	-0,442	-1,125

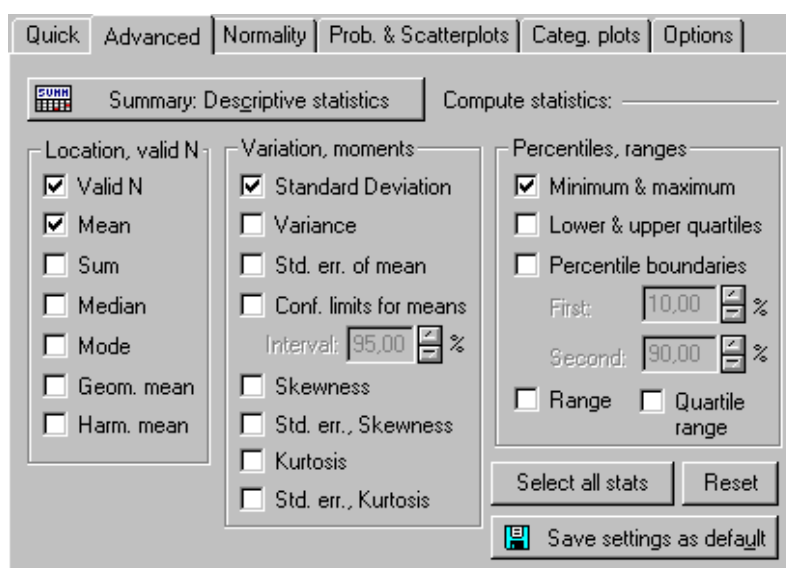


Рисунок 3.1 – Окно выбора (установок) описательных статистик (приведена часть исходного окна)

Для симметричного распределения, в частности для нормального распределения, асимметрия (*Skewness*) равна нулю. Если асимметрия больше трех, то распределение имеет более «длинный правый хвост». Если асимметрия меньше трех, то распределение имеет более «длинный левый хвост».

В нашем примере для всех переменных значение асимметрии близко к нулю. Это указывает на то, что распределения переменных *Y*, *X1* и *X2* близки к симметричным.

Если эксцесс (*Kurtosis*) больше нуля, то распределение островершинное относительно нормального. Если эксцесс меньше нуля, то распределение «туповершинное» относительно нормального. В нашем случае распределение переменных *Y* и *X2* туповершинное, а переменной *X1* – островершинное.

Более точный ответ о нормальности распределения можно получить, если обратится к вкладке *Normality* (Нормальность) в окне *Descriptive statistics* (рисунок 3.2).

В пакете программ для выявления нормальности распределения исследуемых показателей используются следующие критерии:

Kolmogorov-Smirnov & Lillifors test for normality – (Критерий Колмогорова-Смирного и Критерий Лиллиефорса) согласно этому критерию, если вычисленная *D*-статистика значима, то гипотеза о том, что данные имеют нормальный закон распределения, должна быть отвергнута. Иначе, гипотеза о нормальном распределении не отвергается

Shapiro-Wilk's W test – (*W* Критерий Шапиро-Уилкса) согласно этому критерию, если рассчитанная по данным наблюдений *W*-статистика значима, то гипотеза о том, что данные имеют нормальный закон распределения, должна быть отвергнута.

Таким образом, если вероятность отклонения гипотезы о значимости *D*-статистики имеет значения большие выбранного уровня значимости α (обычно $\alpha = 0,01, 0,05$ или $0,1$), то гипотеза о нормальном законе распределения данных принимается с вероятностью $(1 - \alpha)$.

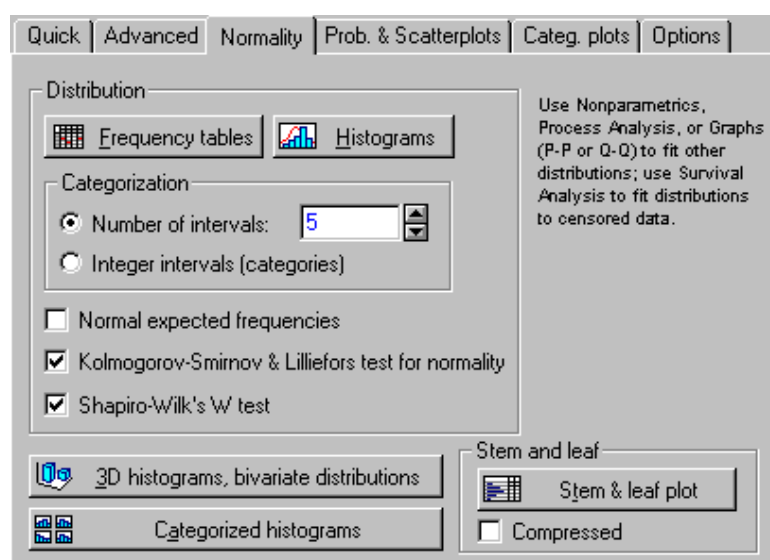


Рисунок 3.2 – Окно установок вычисления характеристики нормальности распределения (приведена часть исходного окна)

Для расчета перечисленных статистик установим, галочки как показано на рисунке 3.2, и выберем кнопку *Frequency tables* (Таблицы частот). В рабочей книге (*Workbook*) будут выведены три таблицы (в соответствии с количеством анализируемых переменных), рассмотрим результаты расчета по переменной *Y*.

Frequency table: Y (Лаб 3)						
K-S d=,17088, p> .20; Lilliefors p> .20						
Shapiro-Wilk W=,89040, p=,06801						
Category	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
20,00000<x<=25,00000	2	2	13,333	13,333	13,333	13,333
25,00000<x<=30,00000	4	6	26,667	40,000	26,667	40,000
30,00000<x<=35,00000	2	8	13,333	53,333	13,333	53,333
35,00000<x<=40,00000	0	8	0,000	53,333	0,000	53,333
40,00000<x<=45,00000	3	11	20,000	73,333	20,000	73,333
45,00000<x<=50,00000	1	12	6,667	80,000	6,667	80,000
50,00000<x<=55,00000	2	14	13,333	93,333	13,333	93,333
55,00000<x<=60,00000	1	15	6,667	100,000	6,667	100,000
Missing	0	15	0,000		0,000	100,000

Рисунок 3.3 – Результаты оценки критериев нормальности распределения переменных (приведена часть исходно окна)

В верхней части окна приведены значения показателей, в данном случае и критерий Колмогорова-Смирнова и Шапиро-Уилкса получены незначимыми и соответственно нельзя считать распределение переменной Y нормальным.

3.5 Построение классической линейной регрессии

Для построения линейной регрессии (парной и множественной), а также для оценки параметров линейных и нелинейных трендов в пакете STATISTICA 6.0 используется модуль *Multiple Regression* (Множественная регрессия).

Проведем построение уравнения регрессии зависимости фондоотдачи от среднечасовой производительности труда и удельного веса активной части ОПФ (приложение E).

Шаг 1. В главном меню выберем: *Statistics*→*Multiple Regression* (Статистика→ Множественная регрессия).

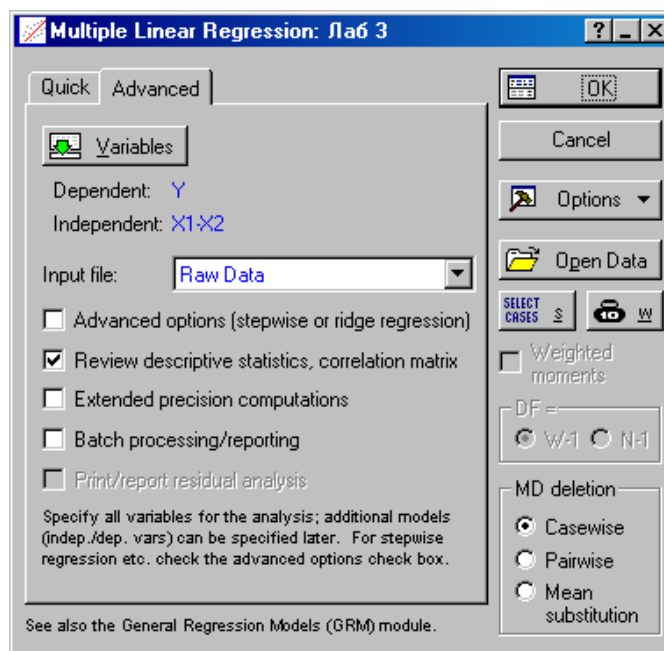


Рисунок 3.4 – Окно *Multiple Linear Regression* (Множественная линейная регрессия)

Шаг 2. В активном окне инициируем кнопку *Variables* (Переменные) и укажем зависимую и не зависимую переменную. В качестве зависимой переменной (*Dependent var.*) необходимо указать производительность труда – Y , в качестве не зависимых переменных (*Independent var.*) будут выступать $X1$ и $X2$.

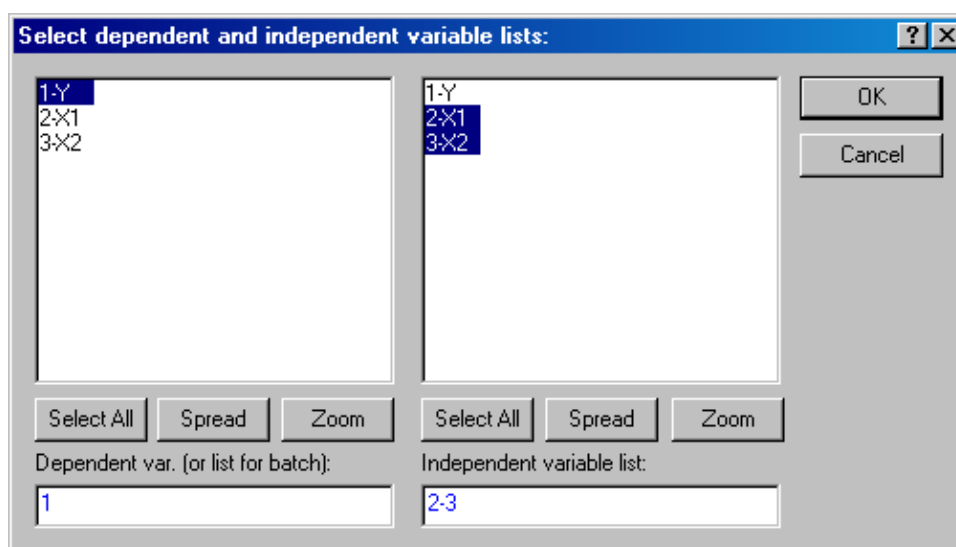


Рисунок 3.5 – Окно выбора зависимой и не зависимой переменных

Шаг 3. Выбираем опцию *Review descriptive statistics, correlation matrix* (Описательные статистик и матрица корреляции) и нажмем кнопку *OK*.

Шаг 4. В появившемся окне *Review Descriptive Statistic* необходимо выбрать вкладку *Advanced*.

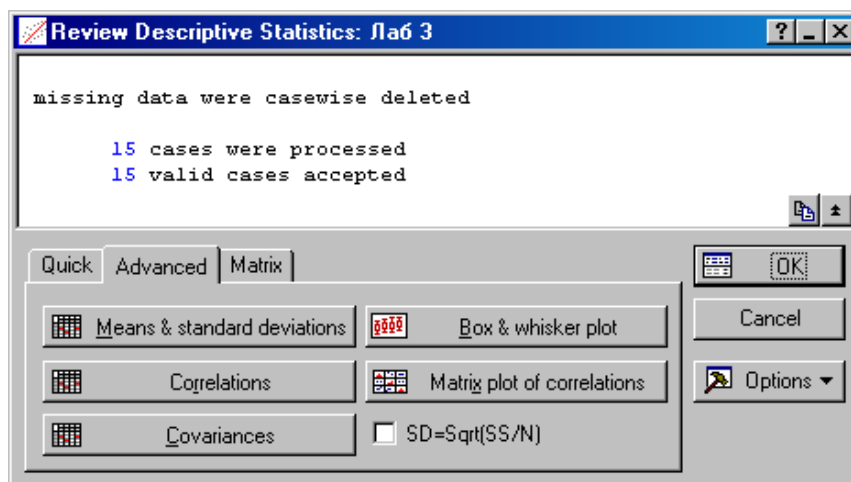


Рисунок 3.6 – Окно установок описательных статистик и корреляции

После чего становятся доступными следующие таблицы: *Means & Standard Deviation* (Средняя и стандартное отклонение), *Correlations* (Корреляция), *Covariances* (Ковариация), *Box & whisker plot* (), *Matrix plot of correlations* (Матрица диаграмм рассеяния).

Шаг 5. Выберем кнопку *Correlations*, в результате получим матрицу, содержащую значения парных коэффициентов корреляции (таблица 3.2).

Значения, представленные таблице показывают, что фактор X_2 оказывает сильное положительное влияние на зависимую переменную Y (т.к. значение на пересечении соответствующего столбца и строки равно 0,868), фактор X_1 оказывает слабое отрицательное влияние. Между переменными X_1 и X_2 связь практически отсутствует (значение коэффициента корреляции -0,117 близко к нулю).

Таблица 3.2 - Матрица парных коэффициентов корреляции

	X_1	X_2	Y
X_1	1,000	-0,117	-0,351
X_2	-0,117	1,000	0,868
Y	-0,351	0,868	1,000

Также можно представить полученные результаты в графическом виде, для этого выбираем кнопку *Matrix plot of correlations* (рисунок 3.6), полученный результат представлен на рисунке 3.7.

Рисунок 3.8 – Окно с результатами оценивания регрессии

где: *Quick* (Быстрый) – данная вкладка предназначена для неопытных пользователей так как в ней доступна только одна кнопка *Summary: Regression results* (Итоговая таблица регрессии). После ее инициализации в рабочую книгу выводятся две таблицы: таблица с коэффициентами и критериями, характеризующими качество уравнения регрессии; таблица с параметрами уравнения регрессии.

Advanced (Расширенные статистики) - вкладка предназначена для опытных исследователей, содержит дополнительные инструменты тестирования оцененной регрессионной модели.

Summary: Regression results (Итоговая таблица регрессии)

ANOVA (Overall goodness of fit) (Дисперсионный анализ)

Covariance of coefficients (Ковариации коэффициентов)

Current sweep matrix (Текущая матрица выметания)

Partial correlations (Частные корреляции)

Redundancy (Избыточность)

Stepwise regression summary (Итоги по шагам)

ANOVA adjusted for mean (Скорректированный на среднее)

Residuals/assumptions/prediction (Остатки/предсказанные/наблюдаемые значения) – вкладка содержит алгоритмы анализа отклонений построенной модели, дескриптивные статистики, а также возможность рассчитывать прогнозные значения зависимой переменной (особенности данной вкладки будут рассмотрены ниже).

Шаг 7. Выбрав кнопку *Summary: Regression results* (Вычислить: Результаты построения регрессии) перейдем в *Workbook* (Рабочая книга) где будут представлены две таблицы содержащие оцененные параметры модели и основные показатели адекватности построения регрессии.

Таблица 3.3 – Показатели адекватности множественного уравнения регрессии

	<i>Value</i> (Значение)
<i>Multiple R</i> (Множест. R)	0,903
<i>Multiple R?</i> ((Множест. R2)	0,816
<i>Adjusted R?</i> (Скорр. R2)	0,785
<i>F(2,12)</i>	26,616
<i>p</i>	0,000
<i>Std.Err. of Estimate</i> (Стд. Ош. Оценки)	5,413

Multiple R - Множественный коэффициент корреляции. Данный показатель является обобщением коэффициента линейной парной корреляции

и отражает тесноту связи между зависимой переменной и одновременно несколькими независимыми переменными. В отличие от парного коэффициента корреляции коэффициент множественной корреляции всегда неотрицателен и изменяется от 0 до 1. Чем ближе значение R к 1, тем большее одновременное влияние оказывают независимые переменные.

В данном случае множественный коэффициент корреляции получен равным 0,903 показывает, что связь между вариацией результативного показателя Y и вариацией факторных признаков $X1$ и $X2$ сильная.

Multiple R? - Множественный коэффициент детерминации. Показатель измеряет долю полной вариации переменной Y , объясняемую множественной регрессией. Величина R^2 изменяется от 0 до 1. Если значение R^2 равно 1, то между переменными существует точная линейная связь. Если R^2 равно нулю, то статистическая линейная связь отсутствует.

Согласно данным таблицы 3.3, $R^2 = 0,816$ свидетельствует, что 81,6% вариации переменной Y объясняется факторами $X1$, $X2$.

Adjusted R? - Скорректированный коэффициент множественной детерминации \bar{R}^2 . Важным свойством коэффициента детерминации является то, что R^2 - неубывающая функция от количества факторов, входящих в модель. Поэтому для сравнения коэффициентов детерминации разных моделей надо уравнивать количество факторов. Для сравнения моделей по коэффициенту детерминации корректируют коэффициент детерминации так, чтобы он как можно меньше зависел от количества факторов. Скорректированный коэффициент корреляции может быть использован для выбора лучшей модели.

$F(2,12)$ - F - статистика Фишера, служит для проверки модели на адекватность. Для проверки модели на адекватность с помощью F - статистики Фишера используют значение вероятности p . Если значение вероятности меньше принятого значения α , например, 0,5, то нулевая гипотеза отвергается. Так в рассматриваемом примере p практически равна нулю. Следовательно, нулевая гипотеза о равенстве нулю всех коэффициентов регрессии отвергается. К аналогичному выводу можно перейти, если сопоставить табличное значение критерия при $\alpha=0,05$ и $v_1=2$, $v_2=12$ равное 3,88 с фактическим значением $F(2,12)= 26,616$, т.е. получаем $F_{таб} < F_{факт}$ следовательно модель в целом статистически значима.

Необходимо обратить внимание на то, что F -тест является суммарным тестом. Поэтому может возникнуть ситуация когда все t -статистики являются незначимыми, а F -статистика показывает адекватность модели, что и наблюдается в нашем случае (таблицу 3.4), отсюда можно сделать предположение о наличии мультиколлинеарности (понятие будет введено в последующих лабораторных работах)

Таблица 3.4 – Результаты оценивания множественного уравнения регрессии

	<i>Beta</i>	<i>Std.Err. of</i>	<i>B</i>	<i>Std.Err.</i>	<i>t(12)</i>	<i>p-level</i>
--	-------------	--------------------	----------	-----------------	--------------	----------------

	(БЕТА)	<i>Betta</i> (Стд.Ош. БЕТА)		<i>of B</i> (Стд.Ош. <i>B</i>)		(p-уров.)
<i>Intercept</i> (Св.член)			-25,686	11,459	-2,242	0,045
<i>X1</i>	-0,253	0,125	-0,214	0,105	-2,032	0,065
<i>X2</i>	0,838	0,125	1,479	0,220	6,723	0,000

Рассмотрим результаты оценки параметров уравнения регрессии по столбцам. В первом столбце перечислены члены регрессионного уравнения, при этом *Intercept* это свободный член уравнения.

Во втором столбце содержатся β -коэффициент, являются отвлеченными (абстрактными) величинами и указывают на сколько среднеквадратических отклонений увеличится зависимая переменная при изменении соответствующего независимой переменной на 1 среднеквадратическое отклонение. На практике данный показатель используется для выявления фактора оказывающего наибольшее влияние на зависимую переменную. В нашем случае наибольшее (положительное) влияние оказывает показатель *X2* ($\beta_2=0,838$).

В четвертом столбце содержатся значения параметров a_j оцененного уравнения вида 3.1, т.е. в данном случае получаем следующую регрессионную модель:

$$\tilde{Y}_{ij} = -25,686 - 0,214 \cdot X1_{ij} + 1,479 \cdot X2_{ij}$$

Полученные значения параметров уравнения можно проинтерпретировать следующим образом. Если при прочих равных условиях ($a_1 = -0,214$) среднечасовая производительность увеличится на 1 ед., то фондоотдача уменьшится на 0,214 руб./чел.

Если при прочих равных условиях удельный вес активной части ОПФ ($a_2 = 1,479$) увеличится на 1 процентный пункт, то фондоотдача увеличится на 1,479 руб./чел..

Std. Error (Standart error) указаны стандартные ошибки коэффициентов уравнения. Стандартные ошибки показывают статистическую надежность коэффициента. Если стандартные ошибки имеют нормальное распределение, то примерно в 2 случаях из 3 истинный коэффициент регрессора находится в пределах одной стандартной ошибки соответствующего коэффициента, и примерно в 95 случаях из 100 в пределах двух стандартных ошибок. Значение стандартных ошибок используем для построения доверительных интервалов.

$t(12)$ – выводит расчетное значение t – статистики Стьюдента. Ее значение используется для проверки значимости соответствующего коэффициента.

p-level - показывает вероятность принять или отвергнуть гипотезу о равенстве нулю соответствующего коэффициента. При этом предполагается, что ошибки имеют нормальное или асимптотически нормальное

распределение. Значения вероятности, указанные в таблице известны в статистике как уровни значимости α . Если значение вероятности ниже уровня значимости α , то гипотеза H_0 отвергается и соответствующий коэффициент не равен нулю.

В рассматриваемом примере параметр a_2 при переменной X_2 значим при уровне значимости α больше, чем 0,0002. Коэффициент a_1 получен не значим при уровне $\alpha = 0,05$, т.к. значение вероятности 0,065 больше 0,05.

Так же выявить статистическую значимость параметров можно используя табличное значение t -критерия Стьюдента, в нашем случае при $\alpha=0,05$ и $df=12$ значение равно 2,1788, т.е. получаем:

$a_0 - |-2,242| > 2,1788 \Rightarrow$ параметр статистически значим;

$a_1 - |-2,032| < 2,1788 \Rightarrow$ параметр статистически не значим;

$a_2 - |6,723| > 2,1788 \Rightarrow$ параметр статистически значим;

Шаг 8. Так как оцененная множественная регрессионная модель получена, незначима по параметру при X_1 , необходимо исключить из рассмотрения фактор X_1 . Для этого в активном окне выберем кнопку *Cancel*, перейдя в стартовое окно, далее в качестве независимой переменной (*Independent var.*) укажем X_2 . Получаем следующие результаты:

Таблица 3.5 – Показатели адекватности парного уравнения регрессии

	<i>Value</i> (Значение)
<i>Multiple R</i> (Множест. R)	0,868
<i>Multiple R?</i> ((Множест. R2)	0,753
<i>Adjusted R?</i> (Скорр. R2)	0,734
<i>F(2,13)</i>	39,571
<i>p</i>	0,000
<i>Std.Err. of Estimate</i> (Стд. Ош. Оценки)	6,030

Сравнивая показатели, полученные по первой и второй моделям можно заметить, что значения по второй модели снизились, но при этом модель в общем можно считать статистически значимой.

Согласно данным, приведенным в таблице 3.5, параметры парной регрессионной модели получены статистически значимыми.

Таблица 3.6 – Результаты оценивания парного уравнения регрессии

	<i>Beta</i> (БЕТА)	<i>Std.Err. of Beta</i> (Стд.Ош. БЕТА)	<i>B</i>	<i>Std.Err. of B</i> (Стд.Ош. B)	<i>t(13)</i>	<i>p-level</i> (p-уров.)
<i>Intercept</i> (Св. член)			-35,110	11,673	-3,008	0,010

X2	0,868	0,138	1,531	0,243	6,291	0,000
----	-------	-------	-------	-------	-------	-------

Оценив вторую модель, можно утверждать, что она пригодна для практического использования, так как параметры модели статистически значимы по t -критерию Стьюдента (таблица 3.6), а уравнение в целом проходит тест по F -критерию Фишера (таблица 3.5).

3.6 Прогнозирование (имитация) неизвестных значений зависимой переменной

Воспользуемся полученным парным линейным регрессионным уравнением и проведем экстраполирование значений фондоотдачи при различных вариантах удельного веса активной части ОПФ.

Шаг 1. Для этого в окне *Multiple Regression Results* (рисунок 3.8) необходимо выбрать вкладку *Residuals/assumptions/prediction* (Отклонения/распределения/предсказания) и воспользоваться кнопкой *Predict dependent variable* (Прогнозирование зависимой переменной).

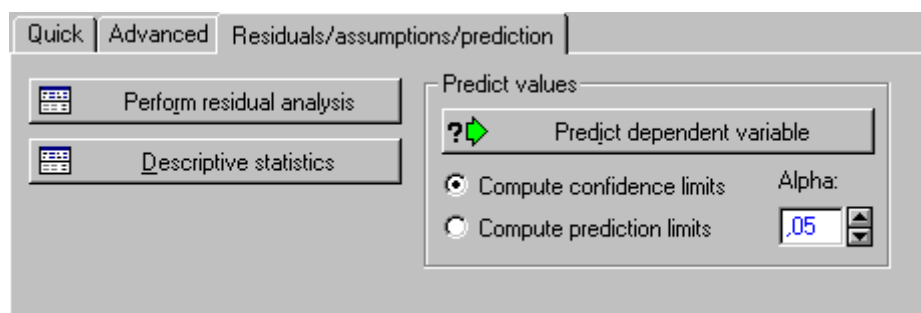


Рисунок 3.9 – Окно установок прогноза (приведена часть исходного окна)

Шаг 2. Для того чтобы определить неизвестное значение независимой переменной в пространственной модели необходимо задать значение независимой переменной. Логичным было предположить, что для увеличения фондоотдачи среднегодовая стоимость ОПФ должна была как можно выше, т.е. стремиться к максимуму в нашем случае $X2_{max} = 55$. Внесем данное значение в окно *Specify values for indep. vars* (Определение неизвестных значений для зависимой переменной).

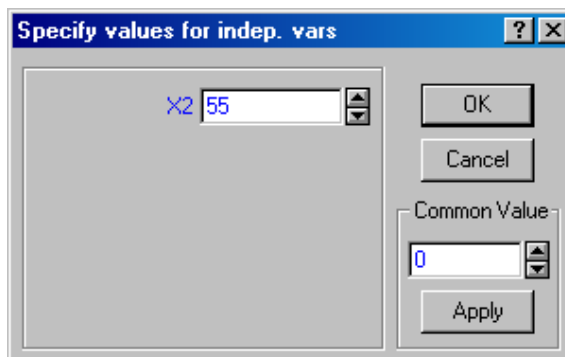


Рисунок 3.10 – Прогнозирование неизвестных значений зависимой переменной

После нажатия кнопки ОК получаем следующие результаты:

Таблица 3.7 – Прогнозные значения фондоотдачи при фиксированном значении среднегодовой стоимости ОПФ на уровне 55%

	<i>B-Weight</i> (В-Вес)	<i>Value</i> (Значение)	<i>B-Weight * Value</i> (В-Вес* Значение)
<i>X2</i>	1,531	55	84,209
<i>Intercept</i> (Св. член)			-35,110
<i>Predicted</i> (Предсказ.)			49,099
<i>-95,0%CL</i>			43,929
<i>+95,0%CL</i>			54,268

В первом столбце содержатся наименования расчетных и исходных показателей. Во втором столбце приведено значение параметра a_1 . В третьем – значение независимой переменной (или переменных) используемое для расчета прогноза. В четвертом – значение независимой переменной (с доверительным интервалом) рассчитанное в результате оценивания прогноза.

В целях сопоставления прогнозов аналогичным образом проведем прогнозирование фондоотдачи при среднем значении среднегодовой стоимости ОПФ.

Таблица 3.8 – Прогнозные значения фондоотдачи при фиксированном значении среднегодовой стоимости ОПФ на уровне 47,53%

	<i>B-Weight</i> (В-Вес)	<i>Value</i> (Значение)	<i>B-Weight * Value</i> (В-Вес* Значение)
<i>X2</i>	1,531	47,530	72,772
<i>Intercept</i> (Св. член)			-35,110
<i>Predicted</i> (Предсказ.)			37,662
<i>-95,0%CL</i>			34,298
<i>+95,0%CL</i>			41,025

Рассмотрим полученные в таблице 3.8 и 3.9 результаты. В нашем случае

прогноз фондоотдачи при значении $X_2=55\%$ находится в интервале $43,929 < 49,099 < 54,268$ руб./чел, а при среднем значении независимой переменной - $34,298 < 37,662 < 41,025$ руб./чел. т.е. наибольшее значение зависимой переменной будет получено при максимальном значении X_2 .

3.7 Тесты для самоконтроля

1 Значение параметра a_j полученное больше нуля указывает на:

- прямую связь между показателями y и x ;
- на отсутствие связи между показателями y и x ;
- на обратную связь между показателями y и x .

2 Параметр a_1 уравнения $\tilde{y}_i = a_0 + a_1 \cdot x_i$ показывает:

- на сколько процентов изменится в среднем результат y , если фактор x изменится на 1 %;
- на сколько в среднем изменится результат y , если фактор x изменится на натуральную единицу;
- на сколько среднеквадратических отклонений изменится результат y , если фактор x изменится на 1 среднеквадратическое отклонение.

3 Коэффициент эластичности показывает:

- на сколько процентов изменится в среднем результат y , если фактор x изменится на 1 %;
- на сколько в среднем изменится результат y , если фактор x изменится на натуральную единицу;
- на сколько среднеквадратических отклонений изменится результат y , если фактор x изменится на 1 среднеквадратическое отклонение.

4 Предположим оцениваем уравнение регрессии с двумя независимыми переменными X_1 и X_2 , при этом β -коэффициент при первом регрессоре получен равным 0,124, а при втором -0,673. Какой из регрессоров оказывает наибольшее влияние на результирующую переменную:

- фактор X_1 ;
- фактор X_2 ;
- оба фактора;
- ни один из факторов.

5 Получена следующая множественная регрессионная модель в стандартизованном виде:

$$\tilde{y}_i = -0,371 \cdot X_1 + 0,780 \cdot X_2$$

какой из факторов оказывает наибольшее влияние на результирующую переменную:

- фактор X_1 ;
- фактор X_2 ;
- не возможно сравнивать влияние этих факторов;
- ни один из факторов.

6 Получена следующая множественная регрессионная модель в

стандартизированном виде:

$$\tilde{y}_i = -0,971 \cdot X_1 + 0,780 \cdot X_2$$

какой из факторов оказывает наибольшее влияние на результирующую переменную:

- фактор X_1 ;
- фактор X_2 ;
- не возможно сравнивать влияние этих факторов;
- ни один из факторов.

7 Значение параметра a_1 получено равным 12,4 среднеквадратическая ошибка равна 2,34, будет ли статистически значим данный параметр если табличное значение t -критерия Стьюдента для данной выборки равно 2,20.

- параметр будет не значим
- параметр будет значим
- не представляется возможным вычислить

8 Фактическое значение t -критерия Стьюдента равно 2,34 при условии, что табличное значение равно 2,21 свидетельствует о:

- статистической значимости соответствующего параметра a_j
- статистической не значимости соответствующего параметра a_j
- статистической значимости уравнения регрессии в целом

9 Проверка значимости уравнения регрессии проводится с помощью:

- t -критерия Стьюдента
- F -критерия Фишера
- коэффициента Фихнера

10 Нулевой гипотезой при проверке значимости регрессионного уравнения является:

- объясненная дисперсия < остаточной дисперсии
- объясненная дисперсия = остаточной дисперсии
- объясненная дисперсия > остаточной дисперсии

3.8 Задания для самостоятельного выполнения

Задания для самостоятельной работы составлены в пяти вариантах, номер варианта выбирается в соответствии с последней цифрой зачетной книжки студента:

Таблица 3.9 – Задания для самостоятельной работы по теме «Построение классической линейной регрессии»

Последняя цифра номера зачетной книжки	1 и 6	2 и 7	3 и 8	4 и 9	5 и 0
Вариант №	1	2	3	4	5

Используя данные (Y , X_1 , X_2 , X_3 и X_4) соответствующего варианта (приложение 7) необходимо выполнить следующие этапы самостоятельной работы:

- 1 Ввести данные в пакет STATISTICA и сохраните их на диске (или ином носителе);
- 2 Рассчитать описательные статистики и сделайте предположение о нормальности распределения показателей;
- 3 Рассчитайте матрицу парных коэффициентов корреляции и сделайте предположение о факторах оказывающих наибольшее влияние на зависимую переменную;
- 4 Оценить множественную линейную регрессию и проверьте ее на статистическую значимость с помощью t -критерия Стьюдента и F -критерия Фишера;
- 5 Провести экстраполирование неизвестных значений зависимой переменной при разных значениях независимых переменных (максимальное, минимальное и среднее).

4 Выявление и устранение мультиколлениарности и гетероскедестичности

4.1 Цели и задачи лабораторной работы

В данной лабораторной работе рассмотрим основные алгоритмы выявления и устранения нарушений условий Гаусса-Маркова, при этом будут решаться следующие задачи:

- 1 Построение регрессионной модели на основе выборочной совокупности;
- 2 Тестирование наличия мультиколлениарности и построение статистически значимой модели;
- 3 Тестирование наличия гетероскедестичности и построение статистически значимой модели.

4.2 Понятие мультиколлениарности и гетероскедестичности, методы выявления и устранения

Для того чтобы регрессионный анализ, основанный на обычном методе наименьших квадратов, давал наилучшие из всех возможных результаты, случайный член должен удовлетворять четырем условиям, известным как условия Гаусса-Маркова.

- 1 Математическое ожидание случайного члена в любом наблюдении должно быть равно нулю.
- 2 Дисперсия случайного члена должна быть постоянна для всех наблюдений.
- 3 Отсутствие систематической связи между значениями случайного члена в любых двух наблюдениях.
- 4 Случайный член должен быть распределен независимо от объясняющих переменных.
- 5 Зависимая переменная y_i (или ε_i) есть нормально распределенная величина.

4.2.1 Мультиколлениарность, выявление и устранение

Мультиколлениарность – это понятие, которое используется для описания проблемы, когда нестрогая линейная зависимость между объясняющими переменными приводит к получению ненадежных оценок регрессии.

Выделим некоторые наиболее характерные признаки мультиколлениарности.

- 1 В первую очередь анализируют матрицу R парных коэффициентов корреляции, точнее, ту ее часть, которая относится к объясняющим

переменным. Считается, что наличие значений коэффициентов корреляции, по абсолютной величине превосходящих 0,75-0,80, свидетельствует о присутствии мультиколлинеарности.

- 2 Анализ корреляционной матрицы R позволяет лишь в первом приближении судить о наличии или отсутствии мультиколлинеарности в исходных данных. Более внимательное изучение этого вопроса достигается с помощью расчета значений коэффициентов детерминации R^2 каждой из объясняющих переменных x_i по всем остальным предикторам $X = (x_1, \dots, x_n)$
- 3 Небольшое изменение исходных данных (например, добавление новых наблюдений) приводит к существенному изменению оценок коэффициентов модели.
- 4 Оценки имеют большие стандартные ошибки, малую значимость, в то время как модель в целом является значимой (высокое значение коэффициента детерминации R^2 и соответствующей F -статистики Фишера).
- 5 Оценки коэффициентов имеют неправильные с точки зрения теории знаки или неоправданно большие значения.

Существует несколько способов борьбы с мультиколлинеарностью:



Рисунок 4.1 – Методы снижения проблемы мультиколлинеарности

4.2.2 Гетероскедестичность, выявление и устранение

Одной из ключевых предпосылок МНК является условие постоянства дисперсий случайных отклонений (дисперсия случайных отклонений ε_i постоянная $D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2$ для любых наблюдений i и j). Выполнение данной предпосылки называется гомоскедестичностью. Не выполнимость данной предпосылки называется гетероскедестичностью.

При наличии гетероскедестичности все выводы, полученные на основе соответствующих t - и F -статистик, а также интервальные оценки будут ненадежными. Следовательно, статистические выводы, полученные при

стандартных проверках качества оценок, могут быть ошибочными и приводить к неверным заключениям по построенной модели.

В ряде случаев, значения характер данных, появление проблемы гетероскедстичности можно предвидеть и попытаться устранить этот недостаток еще на этапе спецификации.

В качестве наиболее популярных из имеющихся алгоритмов выявления гетероскедстичности можно назвать следующие:

- 1 графический анализ отклонений
- 2 тест ранговой корреляции Спирмена
- 3 тест Парка
- 4 тест Глейзера
- 5 тест Голдфреда-Квандта

Для борьбы с гетероскедстичностью используют обобщенный, доступный обобщенный или взвешенный метод наименьших квадратов.

4.3 Рекомендуемая литература

Для лучшего понимания материала изложенного в данной главе необходимо дополнительно проанализировать следующие источники литературы:

- 1 Бородич, С.А. Эконометрика: учеб. пособие / С.А. Бородич. – 3 –е изд., стер. – Мн.: Новое знание, 2006. – 408 с. ISBN 985-475-206-2 (см. стр. 230-250, 271-285)
- 2 Доугерти К. Введение в эконометрику: Пер. с англ. – М.: ИНФРА-М. 2001. - 402с. (см. стр. 155-159, 200-217)
- 3 Кремер Н, Путко Б. Эконометрика: Учебник для вузов. – М.: ЮНИТИ-ДАНА. 2003. - 311с. (см. стр. 108-111, 155-157)
- 4 Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. 2-е изд., испр. - Т. 2: Айвазян С.А. Основы эконометрики. – М.: ЮНИТИ-ДАНА, 2001. -432с. (см. стр. 74-89, 102-111)

4.4 Мультиколлениарность, выявление и устранение

Используя данные, приведенные в приложении 8, рассмотрим методы выявления и устранения мультиколлениарности

4.4.1 Методы выявления мультиколлениарности

На практике, для идентификации мультиколлениарности, наиболее часто прибегают к анализу матрицы парных коэффициентов корреляции. В пакете STATISTICA данную процедуру можно реализовать двумя способами:

Первый способ:

Шаг 1. Выберем в главном меню *Statistics* → *Basic Statistics/Tables* (Статистика → Основные статистики и таблицы).

Шаг 2. В окне *Basic Statistics and Tables* выберем пункт *Correlation matrices* (Корреляционная матрица) и нажмем кнопку ОК.

Шаг 3. В окне *Moment and Partial Correlations* выберем кнопку *One variables list* и выделим переменные для анализа Y , $X1$ - $X5$, далее нажмем *Summary* (Итоги).

Таблица 4.1 – Матрица коэффициентов корреляции (первый способ)

	Y	$X1$	$X2$	$X3$	$X4$	$X5$
Y	1,000	-0,194	0,737	-0,070	0,792	-0,165
$X1$	-0,194	1,000	-0,123	0,666	-0,154	0,566
$X2$	0,737	-0,123	1,000	-0,189	0,678	-0,264
$X3$	-0,070	0,666	-0,189	1,000	-0,152	0,313
$X4$	0,792	-0,154	0,678	-0,152	1,000	-0,160
$X5$	-0,165	0,566	-0,264	0,313	-0,160	1,000

Второй способ:

Шаг 1. В главном меню выберем *Statistics* → *Multiple Regression* в окне *Multiple Linear Regressions* нажмем кнопку *Variables* (Зависимая переменная – Y , не зависимые – $X1$, $X2$, $X3$, $X4$, $X5$)

Шаг 2. Установим флажок напротив опции *Review descriptive statistics, correlation matrix* и нажмем кнопку ОК.

Шаг 3. В окне *Review descriptive statistics* (во вкладке *Advanced*) выберем кнопку *Correlations* (Корреляция).

Согласно данным, приведенным в таблице 4.2 (таблица 4.1), между переменными $X2$ и $X4$, а также $X1$ и $X3$, $X5$ наблюдается сильная взаимосвязь (0,678, 0,666 и 0,566 соответственно), что свидетельствует о наличии мультиколлениарности. В связи с этим при оценке модели с переменными $X2$ и $X4$ (оказывают сильное воздействие на Y) невозможно разделить влияние данных переменных на зависимую переменную, т.е. мы не можем одновременно включить переменную в модель.

Таблица 4.2 - Матрица коэффициентов корреляции (второй способ)

	$X1$	$X2$	$X3$	$X4$	$X5$	Y
$X1$	1,000	-0,123	0,666	-0,154	0,566	-0,194
$X2$	-0,123	1,000	-0,189	0,678	-0,264	0,737
$X3$	0,666	-0,189	1,000	-0,152	0,313	-0,070
$X4$	-0,154	0,678	-0,152	1,000	-0,160	0,792
$X5$	0,566	-0,264	0,313	-0,160	1,000	-0,165
Y	-0,194	0,737	-0,070	0,792	-0,165	1,000

Еще один распространенным способом выявления наличия мультиколлениарности является расчет показателей детерминации, для этого

(относительно рассматриваемого примера) последовательно необходимо оценить пять уравнений регрессии.

В этом случае зависимая переменная Y исключается из рассмотрения, и уравнение принимает вид:

$$\tilde{X}_1 = a_0 + a_1X_2 + a_2X_3 + a_3X_4 + a_4X_5$$

$$\tilde{X}_2 = a_0 + a_1X_1 + a_2X_3 + a_3X_4 + a_4X_5$$

$$\tilde{X}_3 = a_0 + a_1X_1 + a_2X_2 + a_3X_4 + a_4X_5$$

$$\tilde{X}_4 = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_5$$

$$\tilde{X}_5 = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4$$

Воспользуемся модулем *Multiple Regression* получим следующие результаты:

Таблица 4.3 - Показатели адекватности множественного уравнения регрессии влияния независимых переменных на фактор X_1

	<i>Value</i> (Значение)
<i>Multiple R</i> (Множест. R)	0,776
<i>Multiple R²</i> ((Множест. R ²))	0,602
<i>Adjusted R²</i> (Скорр. R ²)	0,549
<i>F(4,30)</i>	11,329
<i>p</i>	0,000
<i>Std.Err. of Estimate</i> (Стд. Ош. Оценки)	1984,209

Согласно данным, приведенным в таблице 4.3, получаем значение коэффициента детерминации $R^2_{X_1 / X_2, X_3, X_4, X_5}$ равное 0,602.

Оценивая оставшиеся четыре регрессионных уравнения, получаем следующие результаты:

$$R^2_{X_1 / X_2, X_3, X_4, X_5} = 0,602 (F(4,30) = 11,329 \text{ при } p=0,000)$$

$$R^2_{X_2 / X_1, X_3, X_4, X_5} = 0,506 (F(4,30) = 7,676 \text{ при } p=0,000)$$

$$R^2_{X_3 / X_1, X_2, X_4, X_5} = 0,469 (F(4,30) = 6,619 \text{ при } p=0,001)$$

$$R^2_{X_4 / X_1, X_2, X_3, X_5} = 0,472 (F(4,30) = 6,694 \text{ при } p=0,001)$$

$$R^2_{X_5 / X_1, X_2, X_3, X_4} = 0,380 (F(4,30) = 4,589 \text{ при } p=0,005)$$

Анализируя коэффициенты можно сделать вывод, что значимая связь наблюдается между всеми показателями (кроме X_5) и остальными независимыми показателями, т.е. в очередной раз подтверждается наличие в имеющихся данных мультиколлениарности.

4.4.2 Устранение мультиколлениарности

Воспользуемся модулем *Multiple Regression* и проиллюстрируем ход выполнения метода пошагового исключения переменных в пакете STATISTICA, для этого необходимо:

Шаг 1. В главном меню выбрать *Statistics* → *Multiple Regression*.

Шаг 2. В стартовом окне *Multiple Regression* (Множественная регрессия) установить флажок в поле *Advanced options (stepwise or ridge regression)* (Расширенные опции (пошаговая и ридж-регрессии)), далее нажмем ОК.

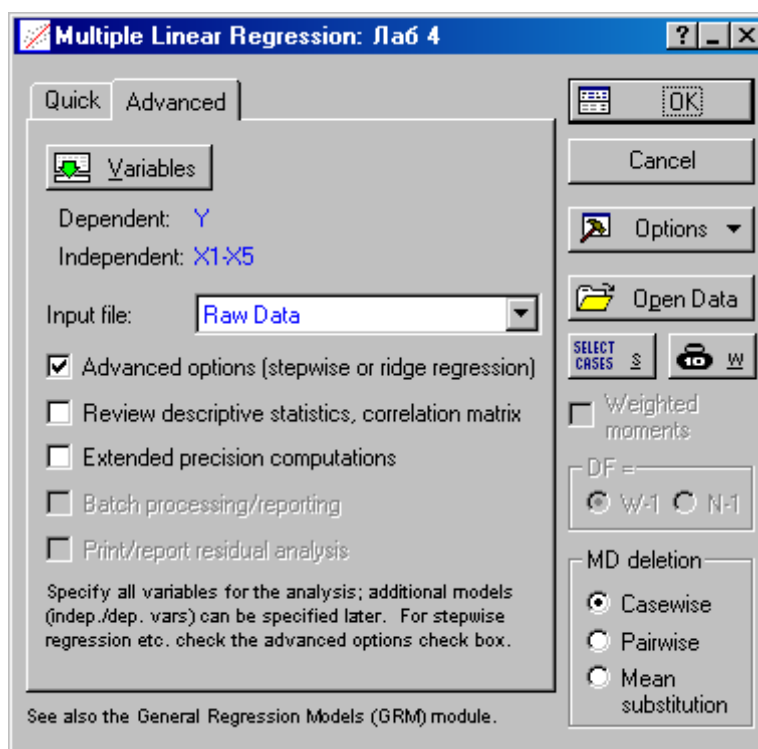


Рисунок 4.1 – Окно установки расширенных опций

Шаг 3. В окне *Model Definition* (рисунок 4.2) можно выбрать метод пошагового построения регрессии, и метод построения ридж-регрессии. Причем метод пошаговой регрессии реализован в двух вариантах, как метод пошагового исключения (*Backward stepwise*) и как метод пошагового включения (*Forward stepwise*) переменных в модель. Выберем первый способ и нажмем ОК.

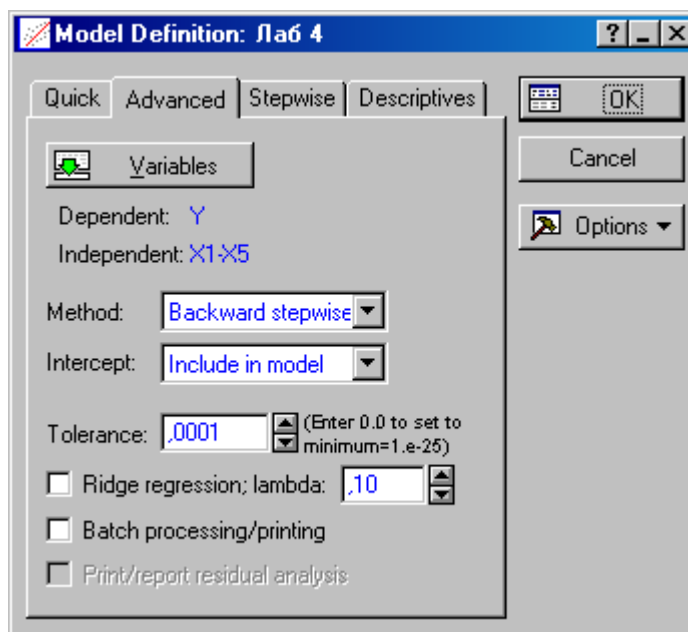


Рисунок 4.2 – Выбор метода построения регрессии в окне *Model Definition*

В результате проведения процедуры пошагового исключения получаем модель парной линейной регрессии, в качестве независимой переменной выступает X_4 :

Таблица 4.4 – Показатели адекватности регрессионной модели

	<i>Value</i> (Значение)
<i>Multiple R</i> (Множест. R)	0,792
<i>Multiple R?</i> ((Множест. R ²))	0,628
<i>Adjusted R?</i> (Скорр. R ²)	0,617
<i>F</i> (1,33)	55,701
<i>p</i>	0,000
<i>Std.Err. of Estimate</i> (Стд. Ош. Оценки)	76,895

Согласно результатам, представленным в таблице 4.4, 62,8% вариации Y описывается включенным в уравнение фактором X_4 (см. строку *Multiple R?*). При этом модель получена, статистически значима по F -критерию Фишера.

Согласно скорректированному коэффициенту детерминации полученную модель можно признать более удачной, так как значение этого показателя 0,628 больше чем значения этого же показателя (0,549) по модели со всеми независимыми переменными.

Таблица 4.5 – Результаты оценивания регрессионной модели

	<i>Beta</i> (БЕТА)	<i>Std.Err. of Beta</i> (Стд.Ош.)	<i>B</i>	<i>Std.Err. of B</i> (Стд.Ош.)	<i>t(33)</i>	<i>p-level</i> (p-уров.)
<i>Intercept</i> (Св.член)			54,518	19,769	2,758	0,009
<i>X4</i>	0,792	0,106	9,387	1,258	7,463	0,000

Также, необходимо отметить, что параметры уравнения статистически значимы по *t*-критерию Стьюдента.

Рассмотрим один из подходов построения хороших оценок коэффициентов регрессии в условиях мультиколлениарности, а именно построение ридж-регрессии (или гребневая регрессия). Для реализации метода гребневой регрессии в окне установок расширенных опций (рисунок 4.2) установим флажок в поле *Ridge regression; lambda:* (Ридж-регрессия; лямбда:) при этом возьмем значение лямбда равное 0,1, получим следующие результаты.

Таблица 4.6 – Показатели адекватности модели ридж-регрессии

	<i>Value</i> (Значение)
<i>Multiple R</i> (Множест. R)	0,756
<i>Multiple R?</i> ((Множест. R2)	0,571
<i>Adjusted R?</i> (Скорр. R2)	0,558
<i>F(1,33)</i>	43,901
<i>p</i>	0,000
<i>Std.Err. of Estimate</i> (Стд. Ош. Оценки)	82,584

Согласно скорректированному значению коэффициента детерминации модель полученная методом ридж-регрессии хуже нежели предыдущая модель.

Таблица 4.7 – Результаты оценивания модели ридж-регрессии

	<i>Beta</i> (БЕТА)	<i>Std.Err. of Beta</i> (Стд.Ош.)	<i>B</i>	<i>Std.Err. of B</i> (Стд.Ош.)	<i>t(33)</i>	<i>p-level</i> (p-уров.)
<i>Intercept</i> (Св.член)			64,624	20,676	3,126	0,004
<i>X4</i>	0,720	0,109	8,534	1,288	6,626	0,000

В результате изучения имеющихся данных было выявлено, что в рассматриваемых данных присутствует мультиколлениарность. После

применения соответствующих процедур было установлено, что наилучшей моделью является регрессионное уравнение с независимой переменной X_4 . Следующим этапом анализа является выявление наличия гетероскедстичности в оцененной модели.

4.5 Гетероскедстичность выявление и устранение

Рассмотрим реализацию алгоритмов выявления гетероскедстичности в пакете STATISTICA.

4.5.1 Графический анализ отклонений

По оси абсцисс откладываются значения (x_i) объясняющей переменной, а по оси ординат либо отклонения ε_i , либо их квадраты ε_i^2

В случае множественной регрессии по оси абсцисс откладывают либо каждую из объясняющих переменных либо \tilde{y}_i .

Для реализации данного метода используем статически значимую регрессионную модель с независимой переменной X_4 .

Шаг 1. Для построения графика отклонений необходимо в окне результатов построения регрессионной модели *Multiple Regression Results* выбрать вкладку *Residuals / assumptions / prediction* (Остатки / распределения / предсказания) и нажмем кнопку *Perform residual analysis* (Анализ остатков)

Шаг 2. В окне *Residual Analysis* (Анализ остатков) выберем вкладку *Residuals* (Остатки) и кнопку *Residuals vs. independent var.* (Остатки и независимая переменная).

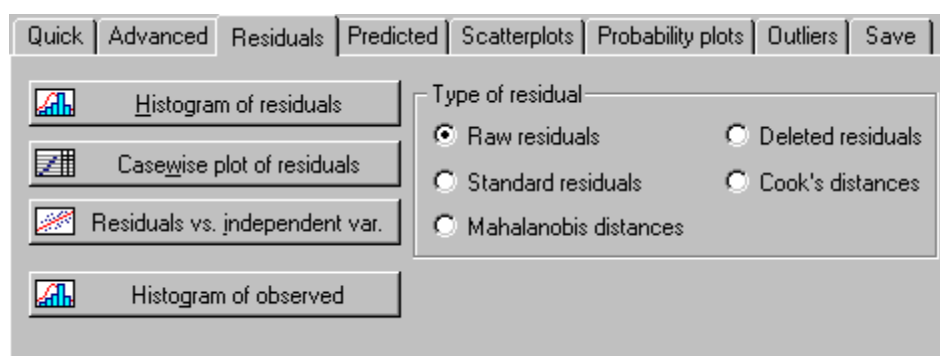


Рисунок 4.3 – Анализ остатков регрессионной модели (приведена часть исходного окна)

Шаг 3. В появившемся окне *Select variable for scatter plot* (Выбор переменных для построения поля корреляции) выберем независимую переменную, в нашем случае это X_4 , и нажать ОК.

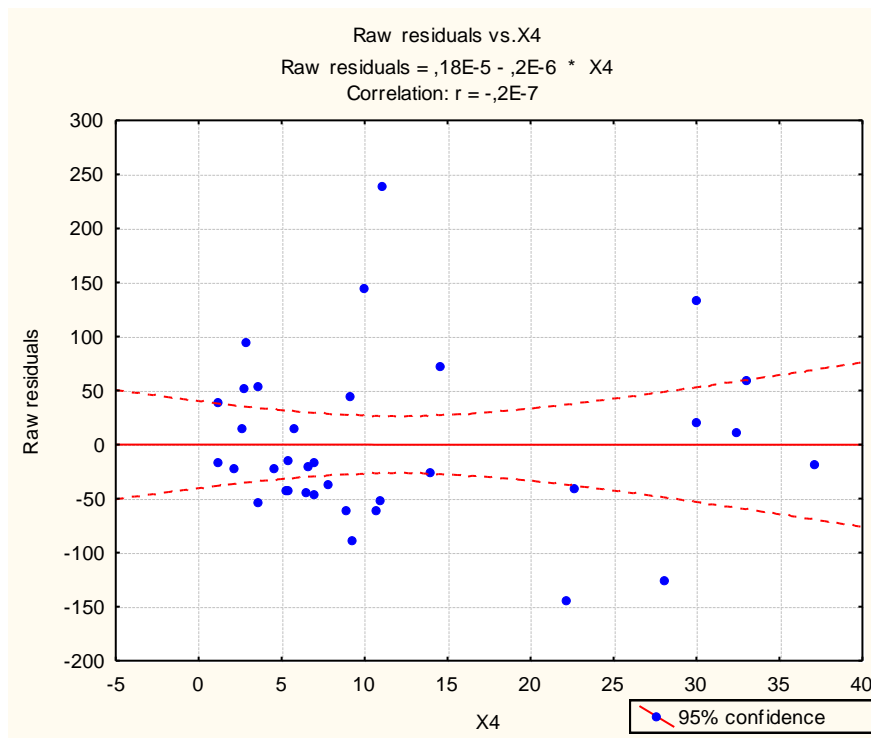


Рисунок 4.4 – Зависимость значений остатков от X_4

Согласно данным рисунка 4.4, значения распределены случайным образом (выходят за доверительные границы), поэтому можно предположить отсутствие гетероскедестичности.

4.5.2 Тест ранговой корреляции Спирмена

Для проведения теста Спирмена на наличие гетероскедестичности необходимо предварительно провести несколько процедур, а именно:

Шаг 1. Сохраним результаты построения регрессионной модели, для этого в окне *Residual Analysis* (Анализ остатков) выберем вкладку *Save* (Сохранить) и нажмем единственную доступную кнопку *Save residuals & predicted* (Сохранить остатки и факторы).

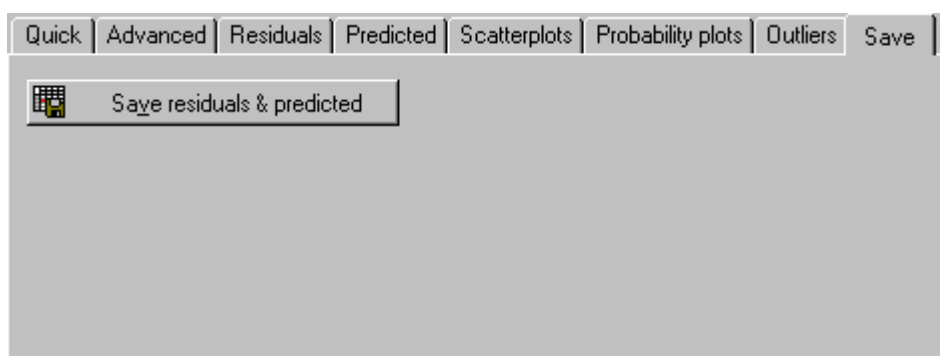


Рисунок 4.5 – Сохранение результатов построения регрессии (приведена часть исходного окна)

Шаг 2. В появившемся окне *Select variables to save with predicted/residual scor...* (Выбор переменной для сохранения с регрессорами, остатками и т.д.) выберем независимую переменную *X4*. После чего переходим к таблице результатов (рисунок 4.6), которая содержит независимую переменную, остатки стандартные отклонения и т.д. В дальнейшем тест Спирмена будет проводиться на ее основе.

	1 X4	2 Predicted	3 Residuals	4 StandardPredicted	5 StandardResidual
1	5,8	108,96	14,04	-0,58	0,18
2	4,5	96,76	-22,76	-0,70	-0,30
3	10	148,39	144,61	-0,18	1,88
4	1,2	65,78	39,22	-1,02	0,51
5	13,9	185,00	-25,00	0,20	-0,33
6	2,7	79,86	51,14	-0,87	0,67
7	32,4	358,65	11,35	1,96	0,15
8	11	157,77	239,23	-0,08	3,11
9	30	336,12	20,88	1,73	0,27
10	3,6	88,31	-53,31	-0,79	-0,69

Рисунок 4.6 – Результаты оценки регрессионной модели с независимой переменной *X4* (приведена часть исходного окна)

Шаг 3. Проранжируем сохраненные показатели *X4* и *Residuals*, для этого в главном меню необходимо указать *Date* → *Rank* → *Variables* (Данные → Ранг → Переменные) и выбрать необходимые переменные (*X4* и *Residuals*).

Шаг 4. В главном меню набрать *Statistics* → *Nonparametrics* → *Correlations (Spearman, Kendall tau, gamma)* (Статистики → Непараметрические → Корреляция).

Шаг 5. В появившемся окне нажмем кнопку *Variables* и сделаем установки как показано на рисунке 4.8.

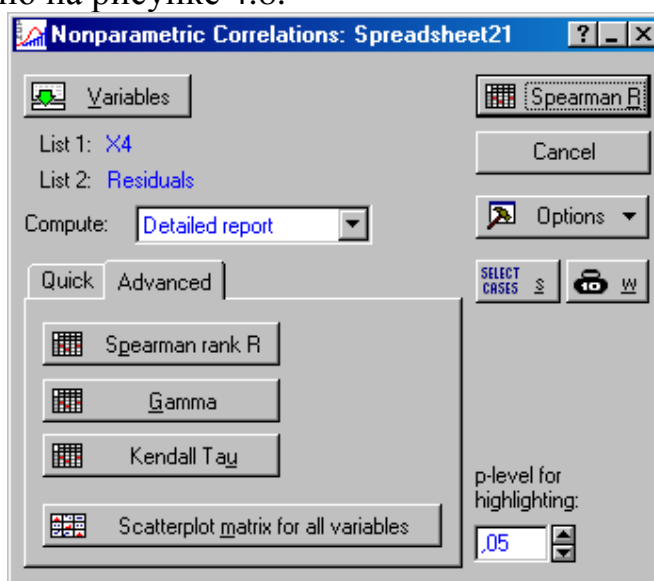


Рисунок 4.7 – Окно выбора расчета непараметрических показателей корреляции

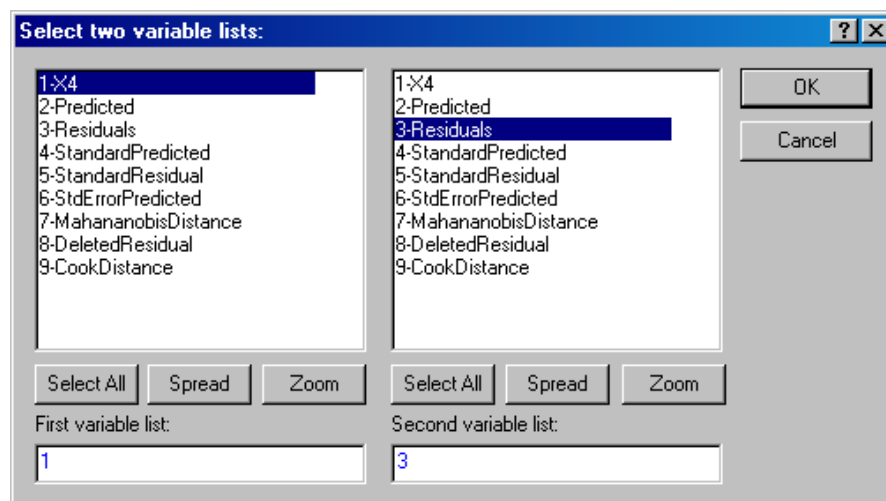


Рисунок 4.8 – Установки переменных для оценки коэффициента Спирмена

После нажатия кнопки *Spearman R* (коэффициент Спирмена) получим следующие результаты.

Таблица 4.8 – Результаты оценивания значения коэффициента Спирмена

	<i>Valid N</i> (Число)	<i>Spearman R</i> (Спирмена)	$t(N-2)$	<i>p-level</i> (p-уров.)
<i>X4 & Residuals</i> (X4 и Остатки)	35	-0,040	-0,231	0,819

Согласно данным, приведенным в таблице 4.8, полученный коэффициент корреляции знаков Спирмена статистически незначим, отсюда можно сделать вывод об отсутствии взаимосвязи между отклонениями и переменной *X4*, т.е. гетероскедстичности отсутствует.

4.5.3 Тест Гольфельда-Квандта

Шаг 1. Для проведения данного теста в системе STATISTICA необходимо воспользоваться исходными данными (Приложение 8) и упорядочить совокупность по величине исследуемой независимой переменной. Для этого в главном меню выберем *Date* → *Sort...* (Данные → Сортировка...).

В появившемся окне *Sort Options* (Установки сортировки) необходимо указать переменную, по которой проводится сортировка, в нашем случае это *X4* и нажать ОК.

Шаг 2. Воспользуемся модулем *Multiple Regression*, в стартовом окне *Multiple Linear Regression* в качестве зависимой переменной укажем *Y* в качестве не зависимой – *X4*. Также (в этом же окне) необходимо выбрать кнопку *Select Cases* (Выбор значений).

Шаг 3. В появившемся окне *Analysis/Graph cases Selection conditions* необходимо установить флажок в поле *Enable Selection Conditions* → *Specific, selected by*:

После активации поля *By Expression* введем $v0 \leq 11$, т.е. регрессия будет строиться по значениям с 1 – 11 (так как в нашем случае $n \approx 30$ соответственно $k=11$).

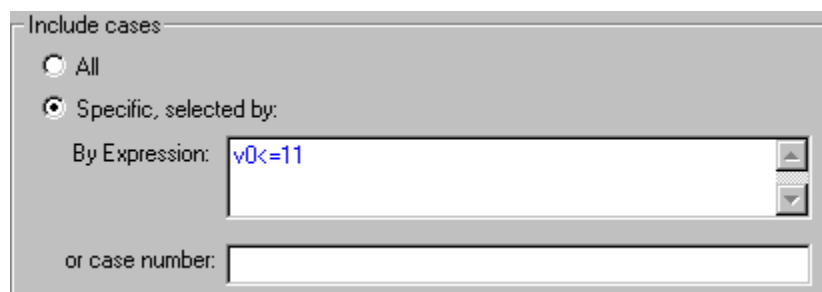


Рисунок 4.9 – Окно выбора подмножества для оценки первой регрессии (приведена часть исходного окна)

Шаг 4. Нажмем ОК→ОК. В окне *Multiple Regression Results* (Результаты множественной регрессии) во вкладке *Advanced* (Расширенные) выберем кнопку *ANOVA* (*Overall goodness of fit*) получим следующую таблицу:

Таблица 4.9 – Результаты дисперсионного анализа для модели по первым 11 наблюдениям

	<i>Sums of Squares</i> (Сумма)	<i>df</i>	<i>Mean Squares</i> (Средн.)	<i>F</i>	<i>p-level</i> (p-уров.)
<i>Regress.</i> (Регресс.)	546,68	1	546,684	0,252	0,628
<i>Residual</i> (Остатки)	19529,86	9	2169,985		
<i>Total</i> (Итого)	20076,55				

Шаг 5. Повторяем процедуру для последних 11 значений при этом в поле *By Expression* необходимо указать $v0 > 24$. В результате оценивания модели получаем следующие оценки.

Таблица 4.10 – Результаты дисперсионного анализа для модели по последним 11 наблюдениям

	<i>Sums of Squares</i> (Сумма)	<i>df</i>	<i>Mean Squares</i> (Средн.)	<i>F</i>	<i>p-level</i> (p-уров.)
<i>Regress.</i> (Регресс.)	27948,1	1	27948,07	2,252	0,168
<i>Residual</i> (Остатки)	111706,5	9	12411,83		

Total (Итого)	139654,5				
---------------	----------	--	--	--	--

Шаг 4. Рассчитываем фактическое значение F -статистики Гольфельда-Квандта, для этого возьмем соответствующие значения, стоящие на пересечении строки *Residual* и столбца *Sums of Squares*, т.е. $19529,86/111706,5 = 0,175$. Табличное значение F -критерия Фишера при степенях свободы $\nu_1=8$ и $\nu_2=8$ равно 3,44, т.е. получаем $F_{\text{факт}} < F_{\text{табл}}$ значит, гипотеза об отсутствии гетероскедостичности принимается.

4.5.4 Методы устранения гетероскедостичности

Построенные тесты и критерии указывают на отсутствие гетероскедостичности в анализируемых данных вследствие чего нет необходимости применять методы борьбы с гетероскедостичностью.

Если же в результате исследования критерии подтверждают наличие гетероскедостичности необходимо обратиться к одному из методов устранения влияния гетероскедостичности.

Обычно исследователю не известны значения дисперсий отклонений в этом случае предполагают, что дисперсии пропорциональны x_i или x_i^2 .

Рассмотрим реализацию первого алгоритма в пакете STATISTICA.

Шаг 1. Образует новые переменные $y_i/\sqrt{x_i}$, $1/\sqrt{x_i}$ и $\sqrt{x_i}$ для того чтобы оценить следующие уравнение:

$$\frac{y_i}{\sqrt{x_i}} = a_0 \frac{1}{\sqrt{x_i}} + a_1 \sqrt{x_i} + u_i \quad (4.1)$$

Для того чтобы образовать переменную $y_i/\sqrt{x_i}$ необходимо в главном меню выбрать *Insert* → *Add Variables ...* (Вставка → Добавить переменную).

Шаг 2. В появившемся окне (рисунок 10) в поле *Name*: введем обозначение переменной – $Y/X4$, в поле *Long name (label or formula with Functions)*: введем выражение $=Y/Sqrt(X4)$, что означает взятие корня (*Sqrt*) переменной $X4$ и деление на полученное значение переменной Y .

Аналогичным образом образуем переменные:

$1/\sqrt{x_i}$ - вводим обозначение $1/X4$ и формулу $=1/Sqrt(X4)$

$\sqrt{x_i}$ - вводим обозначение $SQRT_X4$ и формулу $=Sqrt(X4)$

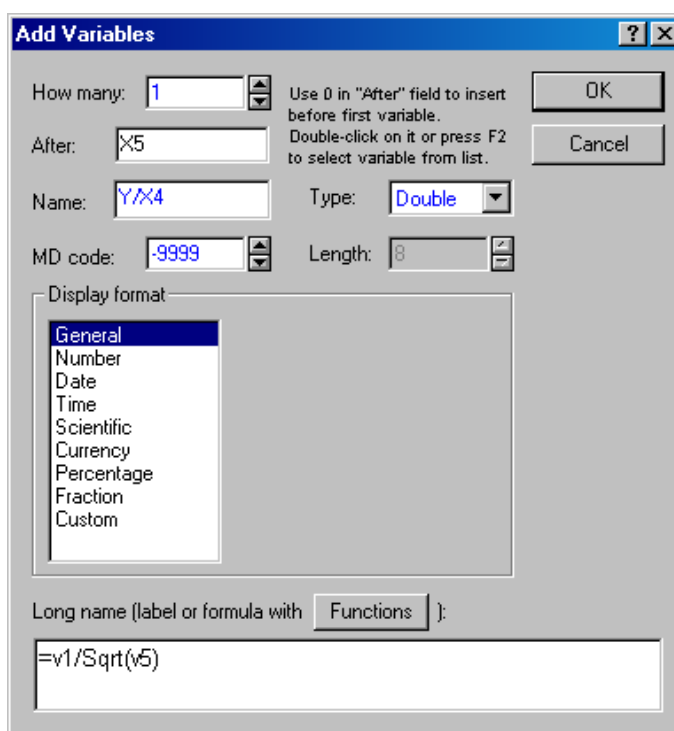


Рисунок 4.10 – Образование новой переменной $y_i / \sqrt{x_i}$

Шаг 3. В главном меню выбираем *Statistics* → *Multiple Regressions* устанавливаем галочку в опции *Advanced options (stepwise or ridge regression)*, далее в окне *Model Definition (Установки модели)* в поле *Intercept (Свободный член)* сделаем установку *Set to zero (Установить на ноль)* и нажмем кнопку ОК.

В результате получаем следующую модель:

Таблица 4.11 – Результаты оценивания модели с исключением гетероскедестичности

	<i>Beta</i> (БЕТА)	<i>Std.Err. of</i> <i>Betta</i> (Стд.Ош. БЕТА)	<i>B</i>	<i>Std.Err.</i> <i>of B</i> (Стд.Ош. B)	<i>t(33)</i>	<i>p-level</i> (p-уров.)
<i>I/X4</i>	0,480	0,099	63,886	13,233	4,828	0,000
<i>SQRT_X4</i>	0,511	0,099	8,596	1,672	5,142	0,000

Согласно полученным данным (таблица 4.11) параметры модели получены статистически значимыми.

При этом коэффициент детерминации получен равным 0,820, т.е. 82% вариации результирующего признака описывается вариацией независимых

переменных. Фактическое значение F -статистики Фишера равно 74,935, что выше табличного значения 3,32 (при $\alpha=0,05$ и степенях свободы $\nu_1=2$; $\nu_2=33$) и указывает на статистическую значимость полученной модели.

4.6 Анализ нормальности распределения случайного члена уравнения

Убедившись в том, что построенная модель статически значима и удовлетворяет условию гомоскедестичности (модель с фактором X_4), проверим нормальность распределения случайной величины, для этого в окне результатов построения регрессионной модели выберем вкладку *Residuals/assumptions/prediction* (Остатки / распределения / предсказания) и нажмем кнопку *Perform residual analysis* (анализ остатков).

В появившемся окне во вкладке *Quick* (Быстрые) выберем *Normal plot of residuals* (График отклонений на нормальной вероятностной бумаге).

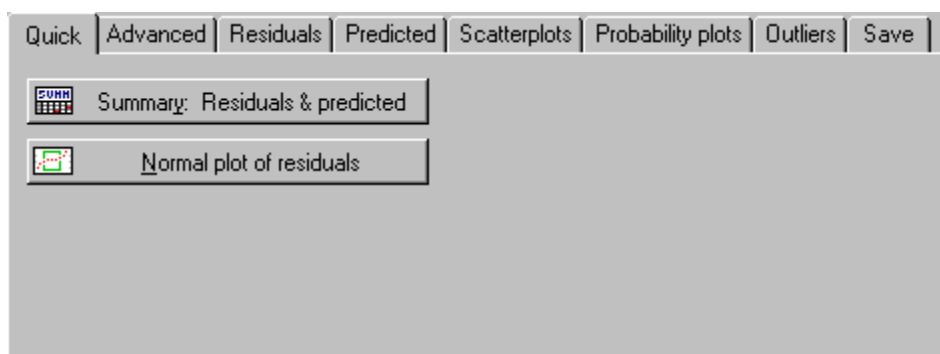


Рисунок 4.11 – *Residual Analysis* (Анализ остатков) вкладка *Quick* (Быстрые) (приведена часть исходного окна)

Рассмотрим рисунок 4.12, на котором приведен вероятностный график. Если остатки ($\varepsilon_i = \tilde{y}_i - y_i$) имеют распределение, отличное от нормального, то точки на графике также отклоняются от прямой.

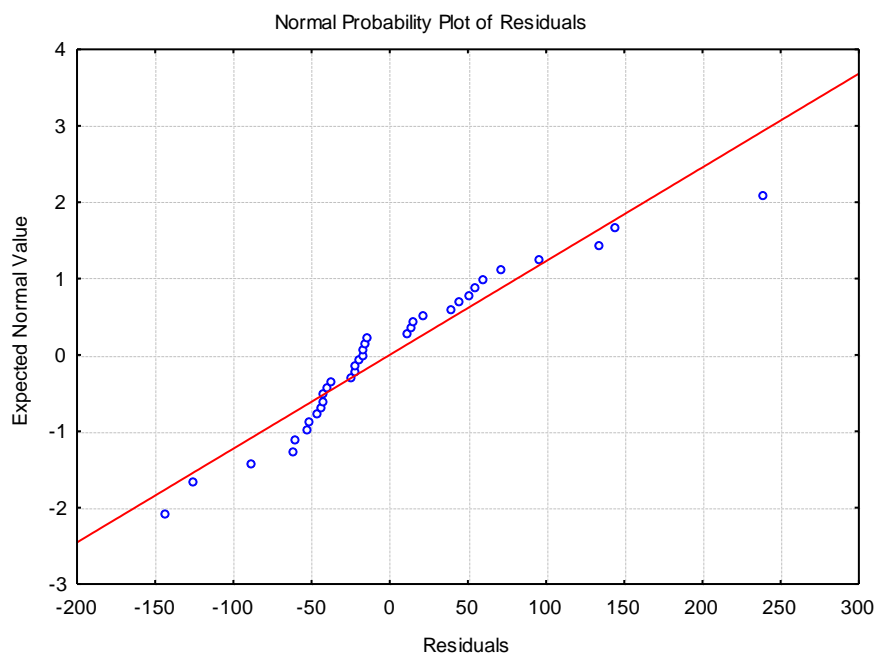


Рисунок 4.12 - Остатки на нормальной вероятностной бумаге

Из рисунка 4.12 видно, что остатки (отклонения) практически лежат на прямой (за исключением выброса), это свидетельствует о том, что они распределены нормально. То есть построенная динамическая модель значима и пригодна для проведения дальнейшего анализа.

Далее выберем вкладку *Residual* (Остатки) нажмем первую кнопку *Histogram of residuals* (Гистограмма распределения остатков), получим следующие результаты (рисунок 4.13).

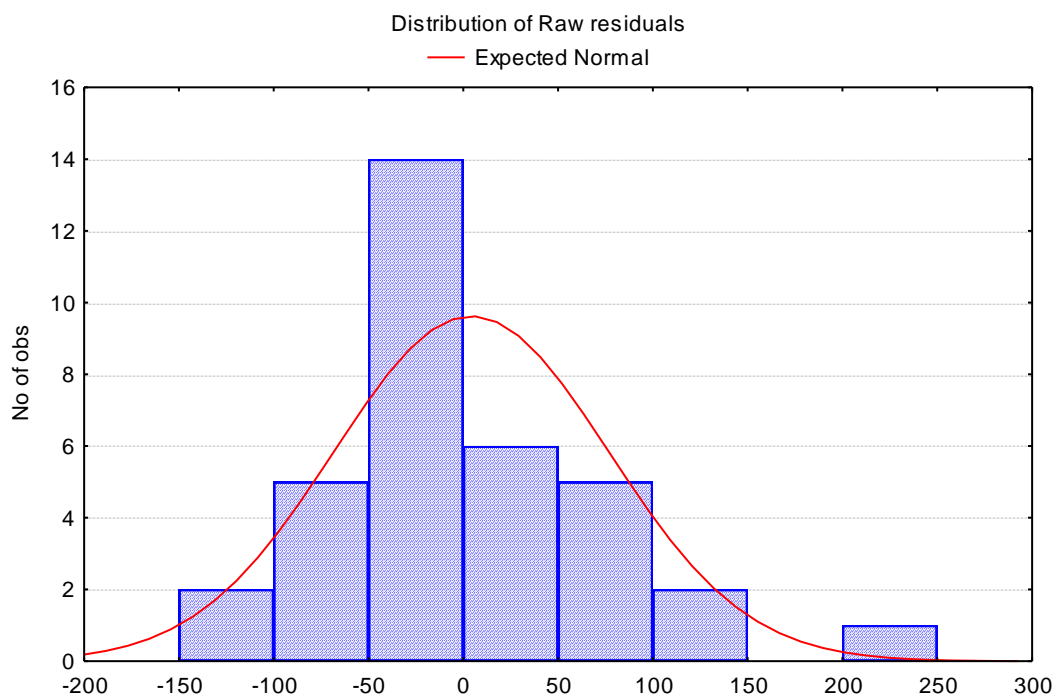


Рисунок 4.13 – Гистограмма распределения регрессионных остатков

Согласно приведенному рисунку остатки подчиняются нормальному закону распределения, т.е. построенная модель адекватна фактическим данным.

4.7 Тесты для самоконтроля

1 Какой из перечисленных тестов используются для обнаружения автокорреляции?

- Тест Парка;
- Тест Дарбина-Уотсона;
- Тест Льюинга-Бокса;
- Тест Глейзера.

2 Гетероскедастичность можно обнаружить с помощью:

- Теста Вальда;
- Теста Глейзера;
- Теста Голфелда-Квандта.

3 Если трендовая модель имеет следующие характеристики: $\tilde{y}_t = 12,032 + 0,29t$, $t_1 = 23,312$, $t_2 = 1,425$, $t_{маб} = 2,179$, $F_{факт} = 29,427$, $F_{маб} = 4,60$, $DW = 2,092$, $d_1 = 1,08$ и $d_2 = 1,36$, можно ли использовать данную модель для прогнозирования?

- да – модель удовлетворяет всем тестам и критериям;
- нет – наблюдается автокорреляция;
- нет – параметры уравнения не значимы по t-критерию Стьюдента;
- нет – параметры уравнения не значимы по F-критерию Фишера.

4 Гомоскедастичность означает:

- «одинаковый разброс»;
- «неодинаковый разброс»;
- «разное среднее значение».

5 Для оценки модели с гетероскедастичностью применяют:

- метод максимального правдоподобия;
- метод взвешенной суммы наименьших квадратов;
- метод исключения переменных.

6 Мультиколлинеарность -

- нестрогая линейная зависимость между объясняющими переменными приводит к получению ненадежных оценок;
- отсутствие корреляции переменных;
- наличие сильной связи между зависимой переменной и независимыми;
- оценка тесноты связи.

7 При оценке теста серий Бреуша-Годфри получим следующее уравнение:

$$\tilde{\varepsilon}_t = 0,322\varepsilon_{t-1} + u_t$$

Если фактическое значение t-критерия равно 1,18, а табличное при $\alpha = 0,05$ и $\nu = 14$ равно 2,145, то можно предположить:

- наличие автокорреляции;

- отсутствие автокорреляции;
 - тест не предназначен для выявления автокорреляции.
- 8 Оценка гетероскедастичной модели МНК-методом является:
- несостоятельной, но эффективной;
 - состоятельной, но неэффективной;
 - несостоятельной и неэффективной;
 - несмещенными и линейными.
- 9 Предпосылки использования метода наименьших квадратов называются:
- теоремой Байеса;
 - теоремой Крамера;
 - теоремой Бернулли;
 - теоремой Гаусса – Маркова.
- 10 Какой из перечисленных методов не способен исключить проблему мультиколлениарности:
- исключение из регрессионных модели незначимых переменных;
 - переход к смещенным методам оценивания;
 - получение дополнительных данных или новой выборки;
 - использование метода взвешенных наименьших квадратов.

4.8 Задание для самостоятельного выполнения

Задания для самостоятельной работы составлены в пяти вариантах, номер варианта выбирается в соответствии с последней цифрой зачетной книжки студента:

Таблица 4.12 – Задания для самостоятельной работы по теме «Выявление и устранение мультиколлениарности и гетероскедастичности»

Последняя цифра номера зачетной книжки	1 и 6	2 и 7	3 и 8	4 и 9	5 и 0
Вариант №	1	2	3	4	5

Используя данные ($Y, X1-X10$) соответствующего варианта (приложение 7) необходимо выполнить следующие этапы самостоятельной работы:

- 1 Построить матрицу парных коэффициентов и проанализировать взаимосвязи между независимыми переменными на наличие мультиколлениарности.
- 2 Устранить (если имеется) влияние мультиколлениарности и построить статистически значимую модель.
- 3 Проанализировать, полученную в результате выполнения пункта 2, модель на наличие гетероскедастичности
- 4 Устранить (если имеется) влияние гетероскедастичности.

5 Проверить нормальность распределения остатков конечной модели и сделать вывод о ее пригодности к использованию

5 Выявление тенденции во временных рядах

5.1 Цели и задачи лабораторной работы

В дано лабораторной работе остановимся на построение статистически значимой динамической модели социально-экономических явлений и прогнозирование неизвестных уровней ряда, при этом выделим следующие задачи (приложение 9):

- 1 с помощью соответствующих критериев выявить наличие тенденции в исследуемом ряду;
- 2 провести выравнивание уровней ряда с помощью скользящей средней;
- 3 построить модель линейного тренда с помощью аналитического выравнивания;
- 4 провести прогнозирование уровней ряда на перспективу;
- 5 выявить и устранить наличие автокорреляции.

5.2 Понятие временных рядов и методы выявления тренд составляющей ряда

Цель анализа экономических временных рядов сводится к прогнозированию будущих (прошлых) значений ряда по уже имеющимся уровням.

В соответствии с поставленной целью анализа можно выделить несколько задач.

Генезис наблюдений, образующих временной ряд (механизм порождения данных). Речь идет о структуре и классификации основных факторов, под воздействием которых формируются значения временного ряда. Как правило, выделяются 4 типа таких факторов.

- *Долговременные*, формирующие общую (в длительной перспективе) тенденцию в изменении анализируемого признака x_t . Обычно эта тенденция описывается с помощью той или иной неслучайной функции $f_{тр}(t)$ (аргументом которой является время), как правило, монотонной. Эту функцию называют функцией тренда или просто – трендом.
- *Сезонные*, формирующие периодически повторяющиеся в определенное время года колебания анализируемого признака. Поскольку эта функция $\varphi(e)$ должна быть периодической (с периодами, кратными «сезонам»), в ее аналитическом выражении участвуют гармоники (тригонометрические функции), периодичность которых, как правило, обусловлена содержательной сущностью задачи.

- *Циклические (конъюнктурные)*, формирующие изменения анализируемого признака, обусловленные действием долговременных циклов экономической или демографической природы (волны Кондратьева, демографические «ямы» и т.п.) Результат действия циклических факторов будем обозначать с помощью неслучайной функции $\psi(t)$.
- *Случайные (нерегулярные)*, не поддающиеся учету и регистрации. Их воздействие на формирование значений временного ряда как раз и обуславливает стохастическую природу элементов x_t , а, следовательно, и необходимость интерпретации x_1, \dots, x_T как наблюдений, произведенных над случайными величинами ξ_1, \dots, ξ_T . Будем обозначать результат воздействия случайных факторов с помощью случайных величин («остатков», «ошибок») ε_t .

Основные задачи анализа временных рядов. Базисная цель статистического анализа временного ряда заключается в том, чтобы по имеющейся траектории этого ряда:

- 1 определить, какие из неслучайных функций присутствуют во временном ряду;
- 2 построить «хорошие» оценки для тех неслучайных функций, которые идентифицированы во временном ряду;
- 3 подобрать модель, адекватно описывающую поведение случайных остатков ε_t , и статистически оценить параметры этой модели.

В данной работе пойдет речь о методах выявления и описания тренд-составляющей, поэтому подробнее остановимся на видах трендов.

В практике эконометрических исследований выделяют несколько разновидностей трендов.

Первым и самым очевидным типом тренда представляется тренд среднего, когда временной ряд выглядит как колебания около медленно возрастающей или убывающей величины.

Второй тип трендов - это тренд дисперсии. В этом случае во времени меняется амплитуда колебаний переменной. Иными словами, процесс гетероскедастичен. Часто экономические процессы с возрастающим средним имеют и возрастающую дисперсию.

Третий и более тонкий тип тренда, визуально не всегда наблюдаемый - изменение значимости одной из компонент временного ряда (например, уменьшение величины сезонных колебаний), или, скажем, изменение величины корреляции между текущим и предшествующим значениями ряда, т.е. тренд автоковариации и автокорреляции.

Принято выделять четыре основных способа аппроксимации временных рядов и соответственно четыре вида трендов.

- 1 Полиномиальный тренд.
- 2 Экспоненциальный тренд.
- 3 Гармонический тренд.
- 4 Тренд, выражаемый логистической функцией.

Наиболее распространенной при анализе тренд составляющей в экономической практике является построение прямой. Конечно, данный вид тренда легок в построении, но при этом во многих случаях сильно искажает реальную тенденцию исследуемого ряда.

В годы перехода экономики России к рыночным отношениям при анализе экономических рядов часто приходится прибегать к криволинейным трендам (парабола второго порядка, гипербола, степенная, экспоненциальная функция и др.). Построение данного типа моделей связано с определенными трудностями, как выбор правильной формы кривой или оценка параметров, но при этом во многих случаях кривые дают наилучший результат в сравнении с прямой.

Существуют несколько способов выбора тренда:

- 1 графический способ, при этом строится график динамики ряда и сравнивается с возможными линейными и нелинейными функциям.
- 2 на основе следующих, выработанных в результате изучения эмпирических данных, правил:

Таблица 5.1 – Функции используемые для аналитического выравнивания временных рядов

Функция	Формула	Рекомендации
Линейная	$\tilde{y}_t = a_0 + a_1 t_t$	используется в том случае, если первые разности уровней (абсолютные приросты) более или менее постоянны
Парабола второго порядка	$\tilde{y}_t = a_0 + a_1 t_t + a_2 t_t^2$	используется в том случае, если вторые разности уровней (ускорения) более или менее постоянны
Показательная	$\tilde{y}_t = a_0 + a_1^{t_t}$	используется в том случае, если цепные коэффициенты роста примерно постоянны
Гипербола	$\tilde{y}_t = a_0 + a_1 \frac{1}{t_t}$	используется в том случае, если обнаружено замедленное снижение (рост) уровней ряда, которые по логике не могут снизиться до нуля (превысить какое-либо значение)

Процедуру выявления и анализа тренд составляющей (как прямолинейной, так и криволинейной) можно представить в виде следующей схемы.

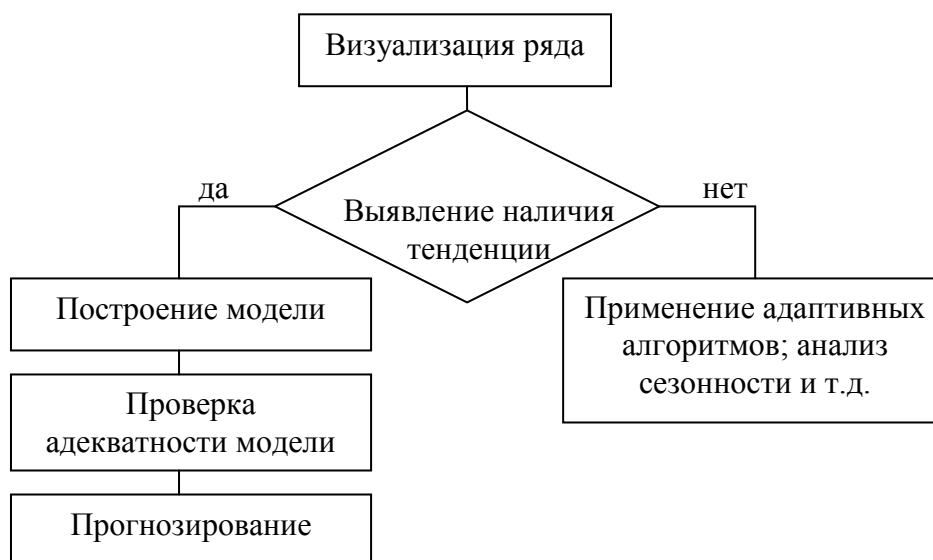


Рисунок 5.1 – Методика выявления и анализа тренд составляющей ряда

5.3 Рекомендуемая литература

Для лучшего понимания материала изложенного в данной главе необходимо дополнительно проанализировать следующие источники литературы (см. список использованных источников):

- 1 Афанасьев, В.Н. Анализ временных рядов и прогнозирование: учеб. для вузов / В.Н. Афанасьев, М.М. Юзбашев. - М. : Финансы и статистика, 2001. - 228с. : ил. - ISBN 5-279-02419-8. (см. стр. 115-121, 225-232)
- 2 Вуколов, Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: Учебное пособие. – М.: ФОРУМ: ИНФРА-М, 2004. – 464с. - ISBN 5-8199-0141-X, 5-16-002003-9. (см. стр. 240-255)
- 3 Доугерти, К. Введение в эконометрику: Учеб. для вузов / К. Доугерти.- 2-е изд. - М. : ИНФРА-М, 2004. - 432 с. - (Университетский учебник). - Библиогр.: с. 407. - ISBN 5-16-001463-2. (см. стр. 115-121, 225-232)
- 4 Кремер, Н.Ш. Эконометрика: учеб. для вузов / Н.Ш. Кремер, Б.А. Путко. - М. : ЮНИТИ-ДАНА, 2003. - 311 с. - ISBN 5-238-00333-1. (см. стр. 133-150)
- 5 Эконометрика: учебник для вузов / под ред. И.И. Елисеевой. - М. : Финансы и статистика, 2005. - 576 с. : ил. - ISBN 5-279-02786-3. (см. стр. 296-335)

5.3 Выявление наличия тенденции в динамическом ряду

В качестве данных для проведения анализа выберем временной ряд динамики ВВП России в текущих ценах (трлн. р.) с 1 квартала 1999 года до 4 квартала 2006 года (приложение 9), при этом в анализе будем использовать данные до 4 квартала 2005г., а информацию за 1-2 квартала 2006г. применим для оценки прогнозов.

В качестве метода выявления наличия тенденции во временном ряду рассмотрим метод сравнения средних уровней. Данный алгоритм предполагает,

что исходный временной ряд разбивается на две приблизительно равные части по числу членов ряда, каждая из которых рассматривается как самостоятельная, независимая выборочная совокупность, имеющая нормальное распределение.

Если временной ряд имеет тенденцию, то средние, вычисленные для каждой совокупности в отдельности, должны существенно различаться между собой. Если же расхождение незначимо и носит случайный характер, то временной ряд не имеет тенденции средней. Таким образом, проверка нулевой гипотезы (H_0) о наличии тенденции в исследуемом ряду сводится к проверке гипотезы о равенстве средних двух нормально распределенных совокупностей, то есть:

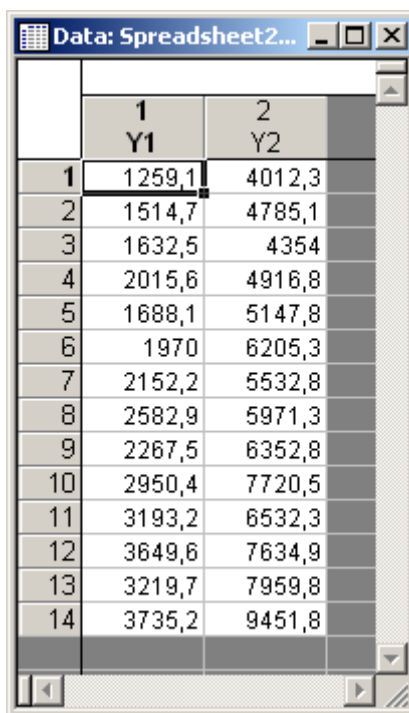
$$H_0: \bar{y}_1 = \bar{y}_2$$

$$H_1: \bar{y}_1 \neq \bar{y}_2$$

Если $t_{\text{факт}} > t_{\text{кр}}$, то гипотеза о равенстве средних уровней двух нормально распределенных совокупностей отвергается, следовательно расхождение между вычисленными средними значимо, существенно и носит неслучайный характер, и, следовательно, во временном ряду существует тенденция средней и существует тренд.

Для проведения данного теста, применительно к ряду динамики ВВП России, воспользуемся модулем *Basic Statistics/Tables*.

Шаг 1. Преобразуем исходный динамический ряд (приложение 9), разбив его на две одинаковые совокупности. Для этого образуем новую таблицу размером 2×14 (28 уровней / 2 = 14 уровней). В главном меню выбираем *File* → *New...*, в появившемся окне в поле *Number of variables* укажем число 2 в поле *Number of cases* – 14. Затем копируем данные из исходной таблицы (рисунок 5.1).



	1 Y1	2 Y2
1	1259,1	4012,3
2	1514,7	4785,1
3	1632,5	4354
4	2015,6	4916,8
5	1688,1	5147,8
6	1970	6205,3
7	2152,2	5532,8
8	2582,9	5971,3
9	2267,5	6352,8
10	2950,4	7720,5
11	3193,2	6532,3
12	3649,6	7634,9
13	3219,7	7959,8
14	3735,2	9451,8

Рисунок 5.1 – Исходная таблица для проведения теста на наличие тенденции

Шаг 2. Тесты на сравнение средних в выборках существуют в двух вариантах: для независимых групп наблюдений (*t-test, independent, by groups* и *t-test, independent variables*) и для зависимых групп наблюдений (*t-test, dependent samples*).

Так как в нашем случае выборка (временной ряд) был разбит на две части и сравниваются показатели в первой и второй группе, то имеем дело с независимыми группами.

В главном меню выбираем *Statistics* → *Basic Statistics* → *t-test for independent variables* (Расчеты → Основные статистики → *t*-тест для независимых выборок). Результаты расчетов представим в таблице 5.1.

Таблица 5.1 - Результаты сравнения двух средних на основе ряда индекса ВВП России (приведена часть таблицы)

	<i>Mean Group 2</i>	<i>Mean</i>	<i>t-value</i>	<i>df</i>	<i>p</i>	<i>Std.Dev. Group 1</i>	<i>Std.Dev. Group 2</i>	<i>F-ratio</i>	<i>p</i>
<i>Y1 vs. Y2</i>	2416,48	6184,11	-8,01	26	0,00	811,97	1559,56	3,69	0,03

Так как *t*-статистика получена, значима, можно утверждать, что во временном ряду существует тенденция средней и существует тренд.

5.4 Сглаживание уровней ряда

Прежде чем приступить к выполнению процедуры необходимо активировать исходную таблицу ряда динамики ВВП (приложение 9).

Одним из распространенных методов анализ тенденций временного ряда является использование скользящих средних. В пакете STATISTICA существует возможность проведения анализа ряда с использованием данного метода, для этого в главном меню необходимо выбрать *Statistics* → *Advanced Linear/Nonlinear Models* → *Time series/Frication* (Статистика → Выбор линейных/нелинейных моделей → Временные ряды и прогнозирование).

В появившемся окне *Time Series Analysis* (Анализ временных рядов) необходимо нажать кнопку *OK* (*transformations, autocorrelations, crosscorrelations, plots*).

Шаг 1. В окне *Transformations of Variables* (Преобразование переменной) выберем вкладку *Smoothing* (Сглаживание).

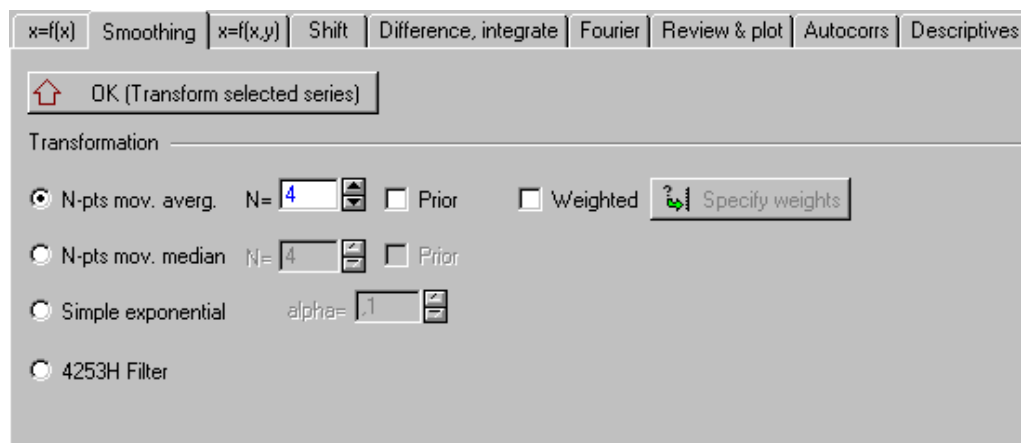


Рисунок 5.2 – Установки для сглаживания уровней динамического ряда (приведена часть исходного окна)

где: *N-pts mov. averg.* – Скользящая средняя с окном N
N-pts mov. median – Скользящая медиана с окном N
Simple exponential – Экспоненциальное сглаживание
4253H – фильтр 4253H

Шаг 2. Выберем сглаживание скользящей средней, для этого установим флажок в опции *N-pts mov. averg.*

В связи с тем, что рассматриваются поквартальные данные для снижения уровня колебаний, было бы логичным выбрать значение сглаживающего окна (интервала) равным 4.

Шаг 3. После соответствующих установок выберем кнопку *OK (Transform selected series)*. Получаем график со сглаженными уровнями (рисунок 5.3), но можно заметить, что информация на рисунке не несет ни какой информации.

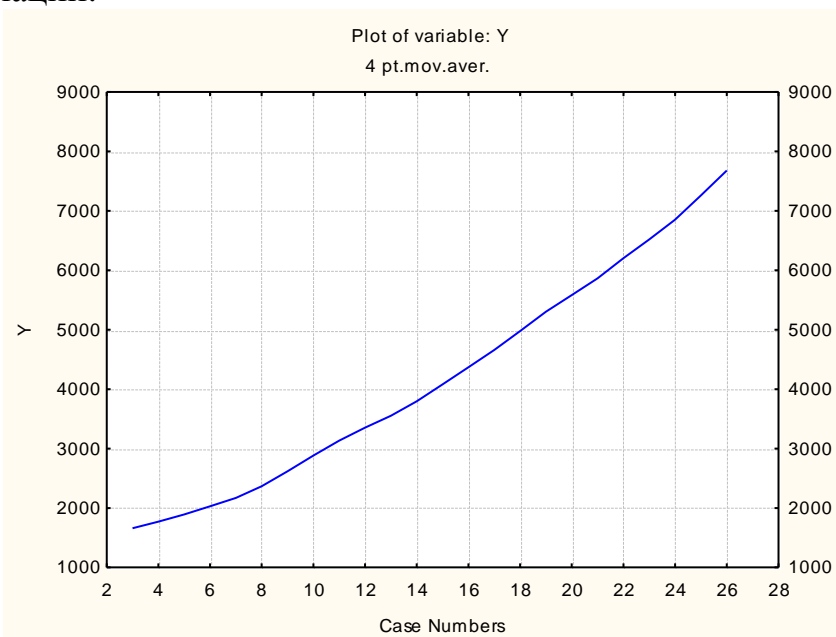


Рисунок 5.3 – Сглаженные уровни ВВП России методом скользящей средней

Для целей увеличения информативности графика в окне *Transformations of Variables* (Преобразование переменных) выберем кнопку *Save variables* (Сохранить переменные), вследствие чего будет выведена таблица с исходным временным рядом плюс уровни, сглаженные на основе скользящей средней.

Шаг 4. Воспользуемся таблицей полученной в ходе выполнения предыдущего шага, и в главном меню выберем *Graphs* → *2D Graphs* → *Line Plots (Variables)*... (Графики → 2D Графики → Линейный график) укажем опцию *Multiple* (Множественный). Полученные результаты представим на рисунке 5.4.

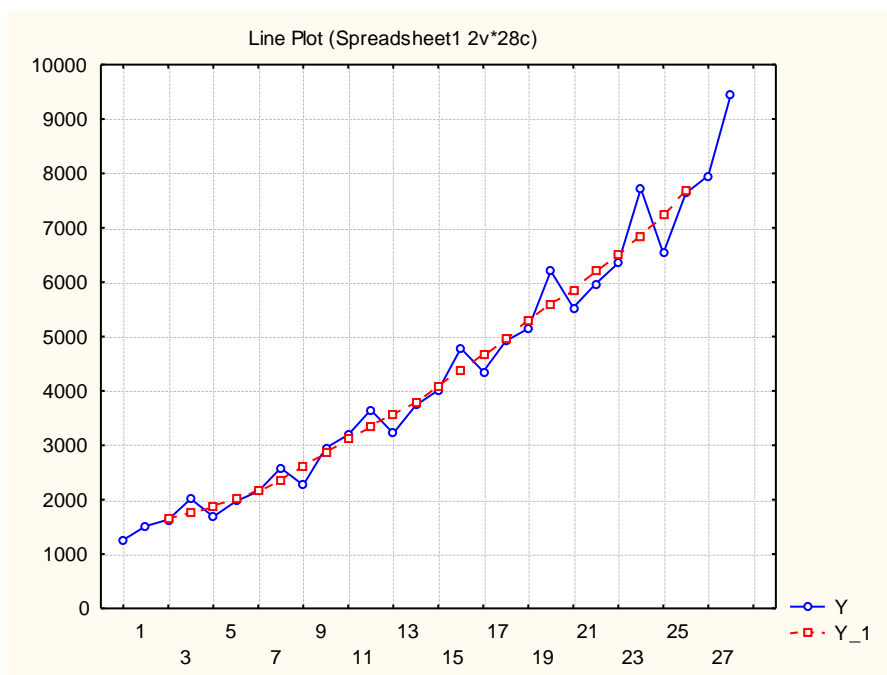


Рисунок 5.4 – Динамика фактических и выровненных уровней ряда ВВП России

Согласно данным рисунка, после сглаживания уровней тенденция показателя к росту проявляется более четко.

5.5 Построение полиномиального тренда

В пакете STATISTICA для построения линейного тренда можно воспользоваться как минимум двумя способами:

- 1 Графическим способом – с помощью опции построения графиков динамического ряда;
- 2 Аналитическим выравниванием – используются средства модуля *Multiple Regression*.

5.5.1 Построение графиков динамических рядов с учетом тренда

Покажем реализацию первого способа в пакете. Данный вариант наилучшим образом подходит для разведочного (предварительного) анализа динамики.

Шаг 1. В строке главного меню необходимо выбрать *Graphs* → *2D Graphs* → *Line Plots (Variables...)* (Графики → Двухмерные графики → Линейный рисунок для переменных).

Шаг 2. В появившемся окне (рисунок 5.5) необходимо указать вкладку *Advanced* (Расширенные) и определить переменную, на базе которой будет построен график. Для этого нажмем кнопку *Variables* (Переменные) и в появившемся окне выбрать необходимую переменную.

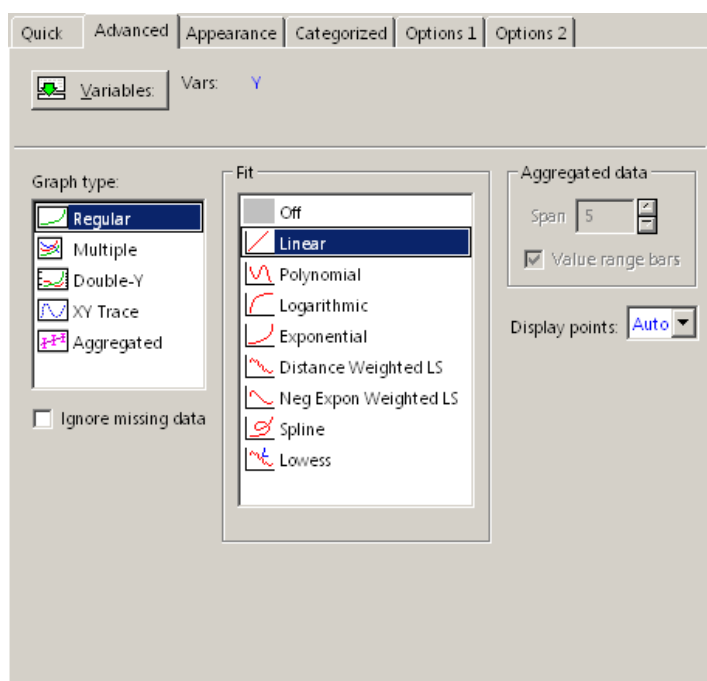


Рисунок 5.5 – Диалоговое окно *2D Line Plots – Variables* (приведена часть исходного окна)

Шаг 3. Укажем в поле *Fit* (Функции) вид желаемого уравнения, в данном случае *Linear* (Линейная функция) и нажать ОК. Получим график динамики ряда ВВП и соответствующий линейный тренд (рисунок 5.6).

Согласно данным, приведенным на рисунке 5.6, в верхней части графика выводится уравнение линейного тренда. Но как было сказано выше, рассмотренный способ не предоставляет информации о статистической значимости полученной модели и пригоден лишь в разведочном анализе.

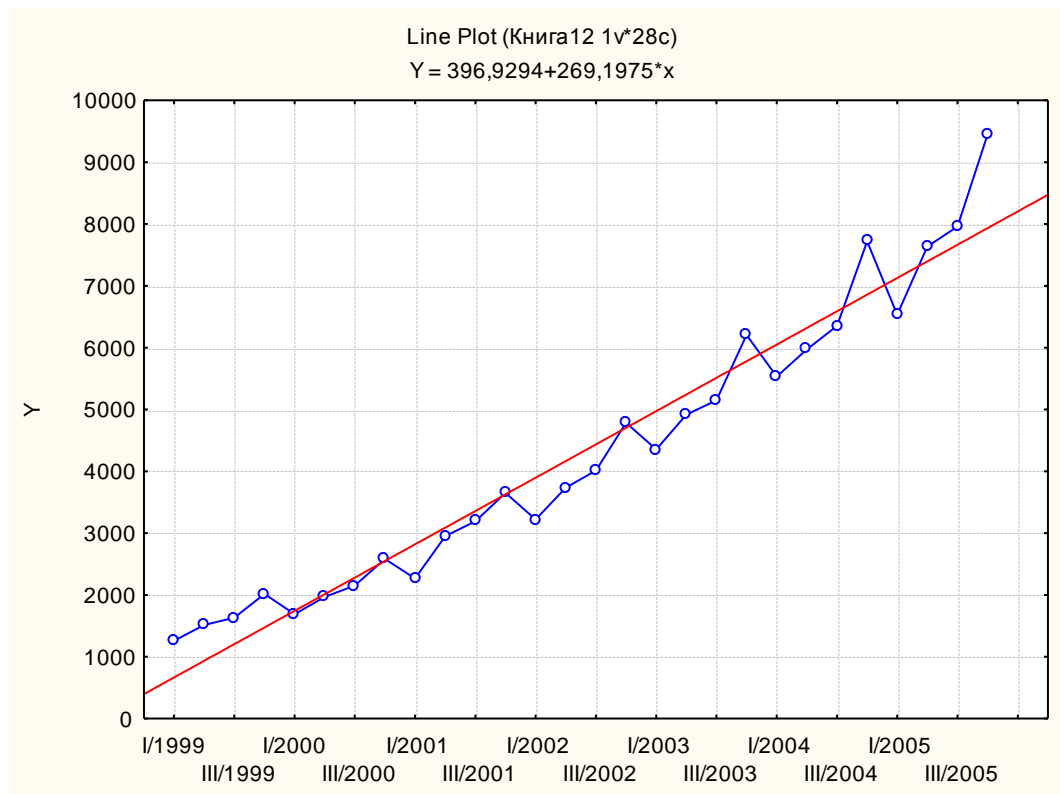


Рисунок 5.6 – Динамика ряда ВВП России, млрд. р.

5.5.2 Построение тренда с помощью аналитического выравнивания

Прежде чем приступить к построению линейного тренда данным способом, необходимо сделать замечание, что в практике экономических исследований моменты (периоды) времени t можно расставлять двояко, от начала ряда (метод используется при машинном счете) и от центра ряда (метод используется при ручном счете). При этом формулы расчета параметров уравнения и их интерпретация будут различаться.

В случае когда моменты времени выставляются от середины ряда параметр a_0 будет являться средней арифметической простой уровней ряда. В случае когда моменты времени выставляются от начала ряда a_0 указывает на точку пересечения тренда с осью OY . Параметр a_1 в обоих случаях будет интерпретироваться одинаково.

Таблица 5.2 – Методы расчета параметров линейного тренда

Моменты (периоды) времени расставлены от <u>середины</u> ряда	Моменты (периоды) времени расставлены от <u>начала</u> ряда
$a_0 = \frac{\sum y_t}{n}$	$a_0 = \bar{y} - a_1 \bar{t}$
$a_1 = \frac{\sum y_t t_t}{\sum t_t^2}$	$a_1 = \frac{\overline{yt} - \bar{y} \cdot \bar{t}}{t^2 - (\bar{t})^2}$

Образуем, новые переменные $t1$ и $t2$, для этого используем функцию вставки *Insert* → *Add Variables...* Далее вручную или с помощью «протяжки» (подобная функция используется в табличном редакторе MS Excel) заполняем новые переменные. При этом применительно к анализируемым данным расстановка моментов времени будет выглядеть следующим образом:

	1 Y	2 t	3 t2
I/1999	1259,1	1	1
II/1999	1514,7	2	4
III/1999	1632,5	3	9
IV/1999	2015,6	4	16
I/2000	1688,1	5	25
II/2000	1970	6	36
III/2000	2152,2	7	49
IV/2000	2582,9	8	64
I/2001	2267,5	9	81
II/2001	2950,4	10	100
III/2001	3193,2	11	121
IV/2001	3649,6	12	144

Рисунок 5.7 – Динамика ВВП России в текущих ценах, трлн. р. (приведена часть исходного окна)

Для построения тренда в форме полинома первой степени (прямая) в пакете STATISTICA выполним следующие действия:

Шаг 1. В главном меню выберем *Statistics* → *Multiple Regression* (Статистические методы → Множественная регрессия). В появившемся диалоговом окне *Multiple Linear Regression* (Множественная линейная регрессия) укажем вкладку *Advanced* (Расширенные) и нажмем кнопку *Variables* (Переменные).

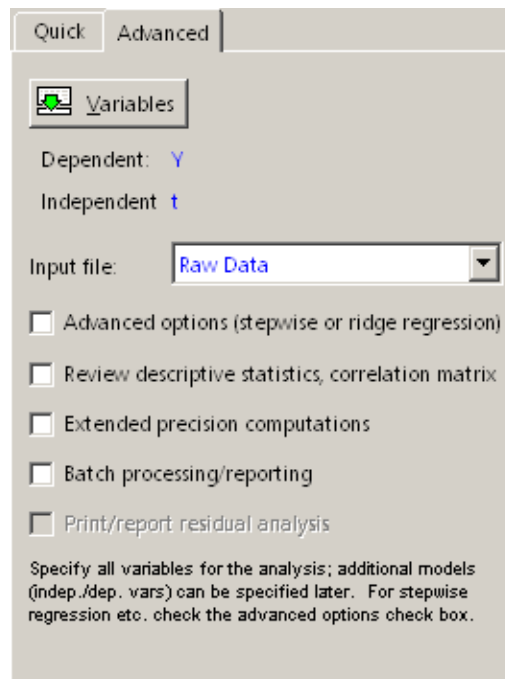


Рисунок 5.8 – Диалоговое окно *Multiple Linear Regression* (Множественная линейная регрессия) (приведена часть исходного окна)

Шаг 2. В окне *Select dependent and independent variable lists*: (Выбор зависимой и независимой переменной) необходимо в качестве зависимой переменной указать ВВП, для этого в поле *Dependent var. (or list for batch)*: (Зависимая переменная) необходимо указать GDP. В качестве независимой переменной выберем момент времени от начала ряда t_1 , для этого в поле *Independent variable lists* (Независимая переменная) необходимо указать t_1 . Далее нажать кнопку ОК.

Шаг 3. После определения зависимой и независимой переменной переходим в следующее окно *Multiple Regression Results* (Результаты построения множественной регрессии). Воспользуемся вкладкой *Quick* (Быстрые) и нажмем кнопку *Summary: Regression results* (Итоги: Результаты построения регрессии) перейдем в *Workbook* (Рабочая книга) где будут представлены две таблицы содержащие оцененные параметры модели и основные показатели адекватности построения регрессии.

Шаг 4. В рабочей книге представлены две таблицы с результатами построения регрессионной модели.

В нашем примере множественный коэффициент детерминации получен равным 0,949, т.е. 94,9% колебаний уровней ВВП описывается построенной моделью.

Оценку общей значимости регрессии проводят на основе *F*-критерия Фишера, в данном случае он получен равным 480,951, что на много выше табличного значения равного 4,22 при $\alpha=0,05$ и степенях свободы $\nu_1=1$, $\nu_2=26$.

Таблица 5.3 – Показатели адекватности динамической модели с фиктивной переменной $t1$

	<i>Value</i> (Значение)
<i>Multiple R</i> (Множест. R)	0,974
<i>Multiple R²</i> ((Множест. R2)	0,949
<i>Adjusted R²</i> (Скорр. R2)	0,947
<i>F</i> (2,26)	480,951
<i>p</i>	0,000
<i>Std.Err. of Estimate</i> (Стд. Ош. Оценки)	524,674

Таблица 5.4 – Результаты оценивания регрессионной динамической модели с фиктивной переменной $t1$

	<i>Beta</i> (БЕТА)	<i>Std.Err. of Beta</i> (Стд.Ош. БЕТА)	<i>B</i>	<i>Std.Err. of B</i> (Стд.Ош. B)	<i>t</i> (26)	<i>p-level</i> (p-уров.)
<i>Intercept</i> (Св.член)			396,929	203,742	1,948	0,062
$t1$	0,974	0,044	269,197	12,275	21,931	0,000

В таблице 5.4, во втором столбце приведены β -коэффициенты, в третьем столбце указаны средняя ошибка β -коэффициентов. В четвертом столбце расположены параметры регрессионного уравнение, в пятом средняя ошибка параметров уравнения.

Получаем следующий линейный тренд:

$$\tilde{y}_t = 396,929 + 269,197 \cdot t_{1t}$$

где: $t_1=0$ в 4 квартал 1998 года

Согласно *p-level* значимым является лишь параметр a_1 , т.е. модель не пригодна для использования в целях прогнозирования.

Повторим процедуру построения модели и в качестве независимой переменной выберем $t2$, получаем следующие результаты.

Таблица 5.5 – Показатели адекватности динамической модели с фиктивной переменной $t2$

	<i>Value</i> (Значение)
<i>Multiple R</i> (Множест. R)	0,974
<i>Multiple R²</i> ((Множест. R2)	0,949
<i>Adjusted R²</i> (Скорр. R2)	0,947
<i>F</i> (2,26)	480,951
<i>p</i>	0,000
<i>Std.Err. of Estimate</i> (Стд. Ош. Оценки)	524,674

Как видим результаты представленные в таблице 5.5 полностью совпадают с данными таблицы 5.3.

Таблица 5.6 – Результаты оценивания регрессионной динамической модели с фиктивной переменной t_2

	<i>Beta</i> (БЕТА)	<i>Std.Err. of Beta</i> (Стд.Ош. БЕТА)	<i>B</i>	<i>Std.Err. of B</i> (Стд.Ош. B)	<i>t</i> (26)	<i>p-level</i> (p-уров.)
<i>Intercept</i> (Св.член)			4300,293	99,154	43,370	0,000
t_2	0,9740	0,044	269,197	12,275	21,931	0,000

Значения таблицы 5.6 отличаются от значений таблицы 4 только по строке *Intercept* (Свободный член), при этом он получен статистически значим по t -критерию Стьюдента, соответственно в дальнейшем рассмотрении будем использовать именно эту модель.

5.6 Автокорреляция, выявление и устранение

Автокорреляция имеет место, если нарушено третье условие Гаусса-Маркова.

Последствия автокорреляции в некоторой степени сходны с последствиями гетероскедастичности, перечислим их:

- 1 оценки коэффициентов регрессии будут неэффективны;
- 2 стандартные ошибки коэффициентов будут оценены неправильно, чаще всего занижены, иногда настолько, что нет возможности воспользоваться для проверки гипотез соответствующими точечными критериями;
- 3 прогнозы по модели получаются неэффективными.

Для выявления наличия автокорреляции наиболее часто используют ряд критериев:

- 1 Графический метод;
- 2 Тест Дарбина-Уотсона
- 3 Тест серий (Бреуша-Годфри)
- 4 Q-тест Льюинга-Бокса

5.6.1 Графический метод выявления автокорреляции

Для реализации этого метода в окне результатов оценки динамической модели *Multiple Regression Results* выберем вкладку *Residuals / assumptions / prediction*, нажмем кнопку *Perform residual analysis*, деле в окне *Residual Analysis* выберем вкладку *Residuals* (Остатки) и кнопку *Residuals vs. independent var.*

В окне *Select variable for scatter plot* укажем переменную *t1*, получаем следующий график:

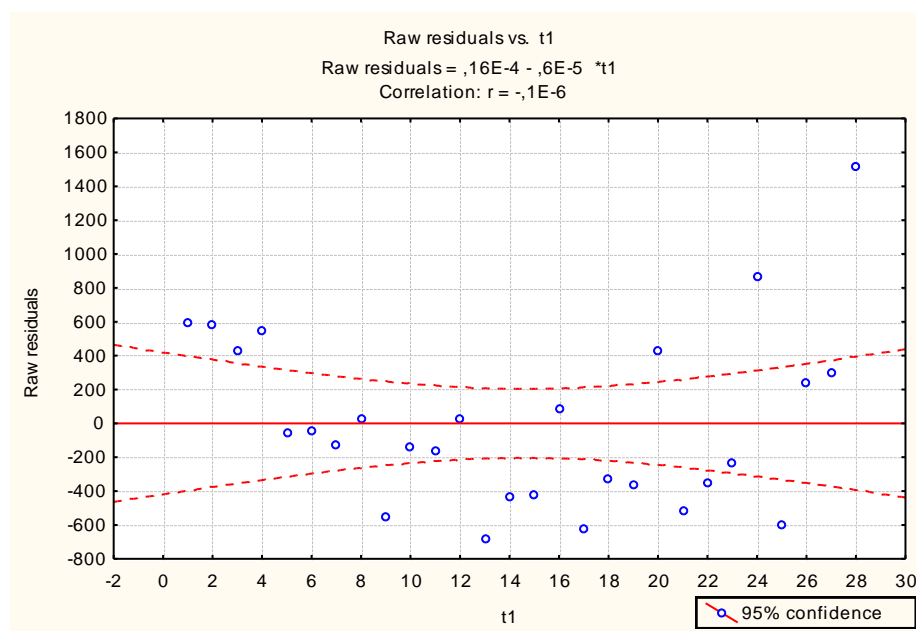


Рисунок 5.9 – Динамики случайного члена временного ряда

Согласно данным, представленным на рисунке 5.10, не прослеживается тренда в отклонениях, соответственно можно предположить отсутствие автокорреляции.

5.6.2 Тест Дарбина-Уотсона

Для реализации теста Дарбина-Уотсона в пакете STATISTICA в окне результатов *Residual Analysis* выбрать вкладку *Advanced* (Расширенные) и нажать кнопку *Durbin-Watson statistic* (Критерий Дарбина-Уотсона).

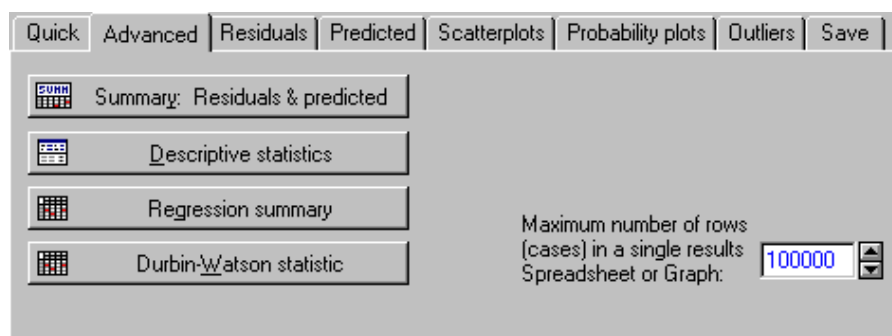


Рисунок 5.10 – Выбор критерия Дарбина-Уотсона (приведена часть исходного окна)

В результате оценки получаем следующую таблицу 5.7:

Таблица 5.7 – Значение критерия Дарбина-Уотсона

	<i>Durbin- Watson</i> (Дарбина-Уотсона)	<i>Serial Corr.</i> (Сериальная корреляция)
<i>Estimate</i> (Оценки)	1,317	0,230

Значение критерия отлично от двух (приложение 23), при этом согласно таблице критических значений данной статистики при $n=25$ и $k=2$ нижняя граница $d_n=1,21$ и верхняя граница $d_b=1,55$. Отсюда получаем, что фактическое значение попадает в зону неопределенности $d_n < DW < d_b$, соответственно невозможно точно сделать вывод о наличии (отсутствии) автокорреляции (см. рисунок 5.11).

Есть положительная автокорреляция остатков. H_0 отклоняется с вероятностью $P=(1-\alpha)$ принимается H_1	Зона неопределенности	Нет оснований отклонять H_0 (автокорреляция остатков отсутствует)	Зона неопределенности	Есть отрицательная автокорреляция остатков. H_0 отклоняется с вероятностью $P=(1-\alpha)$ принимается H_1
0	d_n	d_b	$4-d_n$	$4-d_b$
		2		4

Рисунок 5.11 - Механизм проверки гипотезы о наличии автокорреляции остатков

5.6.3 Тест серий Бреуша-Годфри

Для реализации данного теста в пакете STATISTICA, необходимо выполнить следующие шаги:

Шаг 1. Воспользуемся результатами оценки тренда - $\tilde{y}_t = a_0 + a_1 t^2$, найдем значения случайного члена ε_t . Для этого в окне *Residual Analysis* выберем вкладку *Save* и нажмем единственную доступную кнопку *Save residuals & predicted*, при этом появится *окно Select variables to save with predicted/residual scor... не выбирая* переменных сразу нажмем кнопку *OK*. (см. глава 4).

Шаг 2. В связи с тем, что для реализации теста необходимо оценить уравнение $\varepsilon_t = \rho \varepsilon_{t-1} + u_t$ образуем новую переменную ε_{t-1} . Для этого в главном меню выберем *Insert* → *Add Variables...* в появившемся окне *Add Variables* в поле *Name:* (имя) внесем название переменной – *Residuals t-1*.

В результате образования новой переменной получаем таблицу представленную на рисунке 5.12.

	1 Predicted	2 Residuals	3 Residuals t-1	4 StandardPredicted	5 StandardResidual
I/1999	666,13	592,97	592,97	-1,64	1,13
II/1999	935,32	579,38	579,38	-1,52	1,10
III/1999	1204,52	427,98	427,98	-1,40	0,82
IV/1999	1473,72	541,88	541,88	-1,28	1,03
I/2000	1742,92	-54,82	-54,82	-1,15	-0,10
II/2000	2012,11	-42,11	-42,11	-1,03	-0,08
III/2000	2281,31	-129,11	-129,11	-0,91	-0,25
IV/2000	2550,51	32,39	32,39	-0,79	0,06
I/2001	2819,71	-552,21	-552,21	-0,67	-1,05
II/2001	3088,90	-138,50	-138,50	-0,55	-0,26
III/2001	3358,10	-164,90	-164,90	-0,43	-0,31
IV/2001	3627,30	22,30	22,30	-0,30	0,04
III/2002	3896,50	-676,80	-676,80	-0,18	-1,29
III/2002	4165,69	-430,49	-430,49	-0,06	-0,82

Рисунок 5.12 – Исходная таблица для построения уравнения вида - $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ (приведена часть исходного окна)

Шаг 3. Далее произведем сдвиг на один уровень вперед, для этого выберем *Date* → *Shift (Lag)* (Данные → Выделить лаг). В появившемся окне *Shift Variables* (рисунок 5.13) в поле *Lag* укажем значение 1.

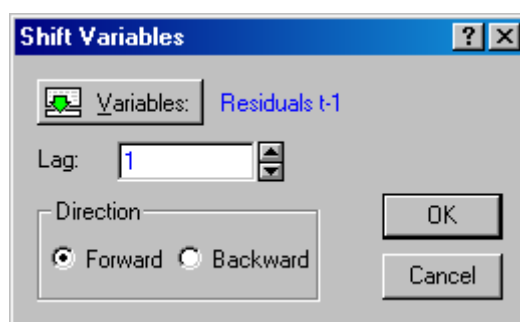


Рисунок 5.13 – Создание запаздывающей переменной

Шаг 4. На данном шаге запустим модуль *Multiple Regressions* и в качестве зависимой переменной укажем *Residuals* а в качестве независимой переменной *Residuals t-1*.

Шаг 5. Выставим галочку возле опции *Advanced options (stepwise or ridge regressions)* в появившемся окне *Model Definition* в прокрутке *Intercept* выберем *Set to zero* (т.е. регрессионное уравнение будет оценено без свободного члена). После оценки получим следующие результаты:

Таблица 5.8 – Результаты оценки модели вида - $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$

	<i>Beta</i> (БЕТА)	<i>Std.Err. of</i> <i>Betta</i> (Стд.Ош. БЕТА)	<i>B</i>	<i>Std.Err.</i> <i>of B</i> (Стд.Ош. <i>B</i>)	<i>t</i> (26)	<i>p-level</i> (<i>p</i> -уров.)

<i>Residuals t-1</i>	0,194	0,192	0,230	0,228	1,009	0,322
----------------------	-------	-------	-------	-------	-------	-------

Согласно результатам построения модели $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, представленным в таблице 5.8, параметр ρ получен статистически не значим, т.е. можно утверждать об отсутствии корреляция между соседними наблюдениями на лаге 1.

5.7 Исключение влияния автокорреляции

Для устранения корреляции во времени чаще прибегают к изменению спецификации модели (исключают или добавляют регрессоры, включают лаги переменных), так как в значительном числе случаев именно неверная спецификация и является источником автокорреляции.

Хотя полученная нами в ходе исследования динамическая модель не обнаруживает присутствие в данных автокорреляции, опишем процедуру исключения данной проблемы изменив спецификацию модели, а именно добавим в рассмотрение лаговое значение переменной ВВП.

Шаг 1. Переходим к исходной таблице и образуем новую переменную *GDP t-1* (процедура описана выше).

Шаг 2. Активизируем модуль *Multiple Regression* и в качестве зависимой переменной укажем *Y* в качестве не зависимых - *t2* и *Yt-1*. Получаем следующие результаты:

Таблица 5.9 – Результаты построения модели вида $\tilde{y}_t = a_0 + a_1t_2 + a_2y_{t-1}$

	<i>Beta</i> (БЕТА)	<i>Std.Err. of Betta</i> (Стд.Ош. БЕТА)	<i>B</i>	<i>Std.Err. of B</i> (Стд.Ош. B)	<i>t(24)</i>	<i>p-level</i> (p-уров.)
<i>Intercept</i> (св. челн)			3221,338	977,606	3,295	0,003
<i>Y t-1</i>	0,246	0,227	0,265	0,244	1,084	0,289
<i>t2</i>	0,733	0,227	206,404	63,881	3,231	0,004

Согласно результатам таблицы 5.9, параметр при лаговой переменной получен статистически не значим (чего и следовало ожидать) и модель не пригодна для прогнозирования. Вместе с тем табличное значение статистики Дарбина-Уотсона получено близким к 2 (таблица 5.10), что свидетельствует об отсутствии автокорреляции.

Таблица 5.10 - критерия Дарбина-Уотсона

	<i>Durbin- Watson</i> (Дарбина-Уотсона)	<i>Serial Corr.</i> (Сериальная корреляция)
<i>Estimate</i>	1,946	-0,186

(Оценки)		
----------	--	--

5.8 Прогнозирование уровней исследуемого показателя

Прогнозирование на основе динамической модели заключается в подстановке в исходную модель значений моментов времени соответствующих прогнозным периодам (моментам) времени и расчет доверительных границ прогноза.

В пакете программ данный алгоритм реализован в рамках модуля *Multiple Regression* (Множественная регрессия). При этом в качестве прогнозной модели будем использовать тренд оцененный в пункте 5.5.2.

Шаг 1. Для построения прогноза необходимо воспользоваться вкладкой *Residuals/assumptions/prediction* (Отклонения/распределения/предсказания) которая доступна в окне *Multiple Regression Results*. Воспользоваться кнопкой *Predict dependent variable* (Прогнозирование зависимой переменной).

Шаг 2. Чтобы определить неизвестное значение ВВП в 1 квартале 2006 года необходимо в окне *Specify values for indep. vars* (Определение неизвестных значений для зависимой переменной) задать момент времени t_2 соответствующее прогнозному периоду – в данном случае 14,5.

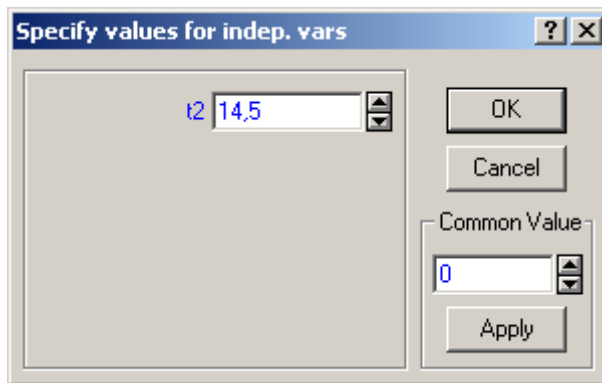


Рисунок 5.14 – Окно выбора прогнозного периода

После нажатия кнопки ОК, получаем следующие результаты.

Таблица 5.10 – Точечный прогноз ВВП России на 1 квартал 2006г.

	<i>B-Weight</i> (В-Вес)	<i>Value</i> (Значение)	<i>B-Weight * Value</i> (В-Вес* Значение)
t_2	269,197	14,500	3903,363
<i>Intercept</i> (Св.член)			4300,293
<i>Predicted</i> (Предсказ.)			8203,656
-95,0%CL			7784,858
+95,0%CL			8622,455

Т.е. уровень ВВП России в 1 квартале 2006 года будет лежать в интервале $7784,858 < 8203,656 < 8622,455$ трлн. р.

Шаг 3. Аналогичным образом рассчитаем значения для 3, 4 и 5 кварталов 2006 года ($t_2 = 15,5, 16,5$ и $17,5$ соответственно). Результаты расчетов представим в таблице 5.11.

Таблица 5.11 - Прогнозные значения ВВП России на 2006г.

Квартал/год	Фактические данные	Прогноз по линейному тренду	Нижняя доверительная граница	Верхняя доверительная граница
I/2006	7873	8203,656	7784,858	8622,455
II/2006	9575,8	8472,854	8031,842	8913,865
III/2006	-	8742,051	8278,518	9205,585
IV/2006	-	9011,249	8524,927	9497,570

Согласно полученным результатам, прогнозные значения согласуются с фактическими данными лишь в 1 квартале 2006 года. Это происходит из-за того, что построенная модель не учитывает сезонную составляющую, которая присутствует в анализируемых данных.

5.9 Тесты для самоконтроля

- 1 По какой формуле вычисляется средний темп роста:
 - средней арифметической;
 - средней геометрической;
 - средней гармонической.
- 2 Если из единицы (или 100%) вычесть коэффициент колеблемости получим:
 - коэффициент детерминации;
 - коэффициент корреляции;
 - коэффициент устойчивости.
- 3 Ряд динамики состоит из:
 - частот;
 - частостей;
 - уровней;
 - вариантов.
- 4 Предложенная формула $y_i - y_{i-1}$ является:
 - абсолютным цепным приростом;
 - абсолютным цепным темпом роста;
 - абсолютным базисным темпом роста.
- 5 По какой формуле вычисляется средняя в моментном динамическом ряду:
 - средней геометрической;
 - средней арифметической;

- средняя хронологическая.

6 Какая из предложенных формул является формулой средне квадратического отклонения:

$$- \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}};$$

$$- \sqrt{\frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}};$$

$$- \frac{\sum_{i=1}^n x_i}{n}.$$

7 Если при построении прямолинейного тренда моменты времени t были расставлены от начала ряда, параметр a_0 будет показывать:

- ничего не показывает;
- средний уровень ряда;
- точку пересечения прямой с осью OY .

8 Тест Дарбина-Уотсона применяется для:

- обнаружения недостающих регрессоров;
- выявления порядка автокорреляции;
- выявления автокорреляции в модели.

9 Какие тесты не используются для обнаружения автокорреляции?

- Тест серий (Бреуша-Годфри)
- Тест Дарбина-Уотсона
- Тест Глейзера

10 Аддитивная модель:

- представляет собой сумму компонент;
- представляет собой произведение компонент;
- представляет собой сумму и произведение соответствующих компонент;
- представляет собой отношение компонент.

5.10 Задание для самостоятельного выполнения

Задания для самостоятельной работы составлены в пяти вариантах, номер варианта выбирается в соответствии с последней цифрой зачетной книжки студента:

Таблица 5.12 – Задания для самостоятельной работы по теме «Выявление тенденции во временных рядах»

Последняя цифра номера зачетной книжки	1 и 6	2 и 7	3 и 8	4 и 9	5 и 0
Номер варианта	1	2	3	4	5

Используя данные (приложение 10) соответствующего варианта необходимо выполнить следующие этапы самостоятельной работы:

- 1 Введите данные в пакете STATISTICA (либо используя табличный редактор, либо вводя данные непосредственно в поле пакета программ);
- 2 Проведите процедуру оценки наличия тренд составляющей в изучаемом динамическом ряду;
- 3 Постройте линейный тренд двумя способами;
- 4 Сделайте выводы об адекватности полученной модели;
- 5 Провести анализ наличия автокорреляции в исследуемых рядах;
- 6 Проведите прогнозирование исследуемого показателя на 3 года вперед.

6 Экспоненциальное сглаживание временных рядов

6.1 Цели и задачи лабораторной работы

В данной лабораторной работе проведем построение и прогнозирование динамики среднедушевых денежных доходов населения за период 1999-2006гг. (р. в мес.) с помощью метода экспоненциального сглаживания, при этом будут решаться следующие задачи:

- 1 Построим адаптивную модель с помощью простого экспоненциального сглаживания;
- 2 Построим экспоненциальную модель с учетом тренда;
- 3 Построим экспоненциальную модель с учетом тренд-сезонной составляющей.

Проводя расчеты, воспользуемся информацией с 1 квартала 1999г. по 4 квартал 2005г., имеющиеся данные за 2006г. будем использовать для сравнения с выровненными уровнями (приложение 11).

6.2. Понятие экспоненциального сглаживания временных рядов

Разработку метода экспоненциального сглаживания приписывают Р. Г. Брауну, данный метод характерен тем, что при его применении изменяют каждый уровень y_t , учитывая сам этот уровень и предшествующие ему, причем «вклад» предшествующего уровня тем меньше, чем «старше» этот уровень. Веса уровней снижаются экспоненциально в степени, зависящей от принятой величины параметра сглаживания α , значение которого находится в пределах 0-1.

При обсуждении вопроса о возможности применения метода экспоненциального сглаживания для прогнозирования экономических процессов остановимся на следующих важных моментах.

- 1 Метод экспоненциального сглаживания, разработанный для анализа временных рядов, состоящих из большого числа наблюдений, при изучении экономических временных рядов нередко не «срабатывает». Это обусловлено тем, что экономические временные ряды бывают слишком короткими (15-20 наблюдений) и в случае, когда темпы роста и прироста велики, то, как показывает практика, метод не «успевает» отразить все изменения.
- 2 Для нахождения оценок коэффициентов сглаживающего полинома используется рекуррентная процедура, позволяющая при конечном числе наблюдений получить приближенное решение задачи, причем приближение тем точнее, чем больше число наблюдений.
- 3 Проблема выбора начальных условий принципиально, сводится к оценке погрешности метода, а вопрос выбора оптимального значения параметра сглаживания α для своего решения требует, прежде всего, четкой постановки задачи.

Рассмотрим содержание процедуры экспоненциального сглаживания, а также ее модификации, разработанные с учетом различной структуры временного ряда - наличие тренда и сезонных изменений.

6.3 Рекомендуемая литература

Для лучшего понимания материала изложенного в данной главе необходимо дополнительно проанализировать следующие источники литературы (см. список использованных источников):

1. Боровиков, В.П. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере : учеб. пособие / В.П. Боровиков, Г.И. Ивченко. – 2-е изд., перераб. И доп. - М. : Финансы и статистика, 2006. - 368 с: ил. ISBN 5-279-03059-7. (см. стр. 115-121)
2. Кильдишев, Г.С., Френкель, А.А. Анализ временных рядов и прогнозирование. – М.: Статистика. 1973. (см. стр. 53-66)
3. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов: учеб. пособие. – М.: Финансы и статистика, 2003. – 416 с. ISBN 5-279-02740-5. (см. стр. 17-47, 76-88)
4. Френкель, А.А. Прогнозирования производительности труда: методы и модели. – М.: Экономика, 1989. – 214 с. (см. стр. 60-86)
5. Халафян А.А. STATISTICA 6. Статистический анализ данных. 3-е изд. Учебник – М.: ООО «Бином-Пресс», 2007 г. – 512 с. ISBN 978-5-9518-0215-6. (см. стр. 464-469)

6.4 Простое экспоненциальное сглаживание (метод Брауна)

Проиллюстрируем метод простого экспоненциального сглаживания используя следующие сглаживающие константы: $\alpha = 0,1$; $\alpha = 0,5$; $\alpha = 0,9$.

Прежде чем приступить к построению данной модели необходимо сказать, что все методы анализа временных рядов (в том числе и экспоненциальное сглаживание) в пакете собраны в модуле *Time series & Frication* (Временные ряды и прогнозирование).

Для запуска метода экспоненциального сглаживания в пакете STATISTICA необходимо:

Шаг 1. В главном меню программы выбрать *Statistics* → *Advanced Linear/Nonlinear Models* → *Time series/Frication* (Статистика → Выбор линейных/нелинейных моделей → Временные ряды и прогнозирование).

Шаг 2. В появившемся окне *Time Series Analysis* (рисунок 6.1) необходимо выбрать кнопку *Exponential smoothing & forecasting* (Экспоненциальное сглаживание и прогнозирование), после чего перейдем в окно *Seasonal and Non-Seasonal Exponential smoothing* (Сезонное и не сезонное экспоненциальное сглаживание).

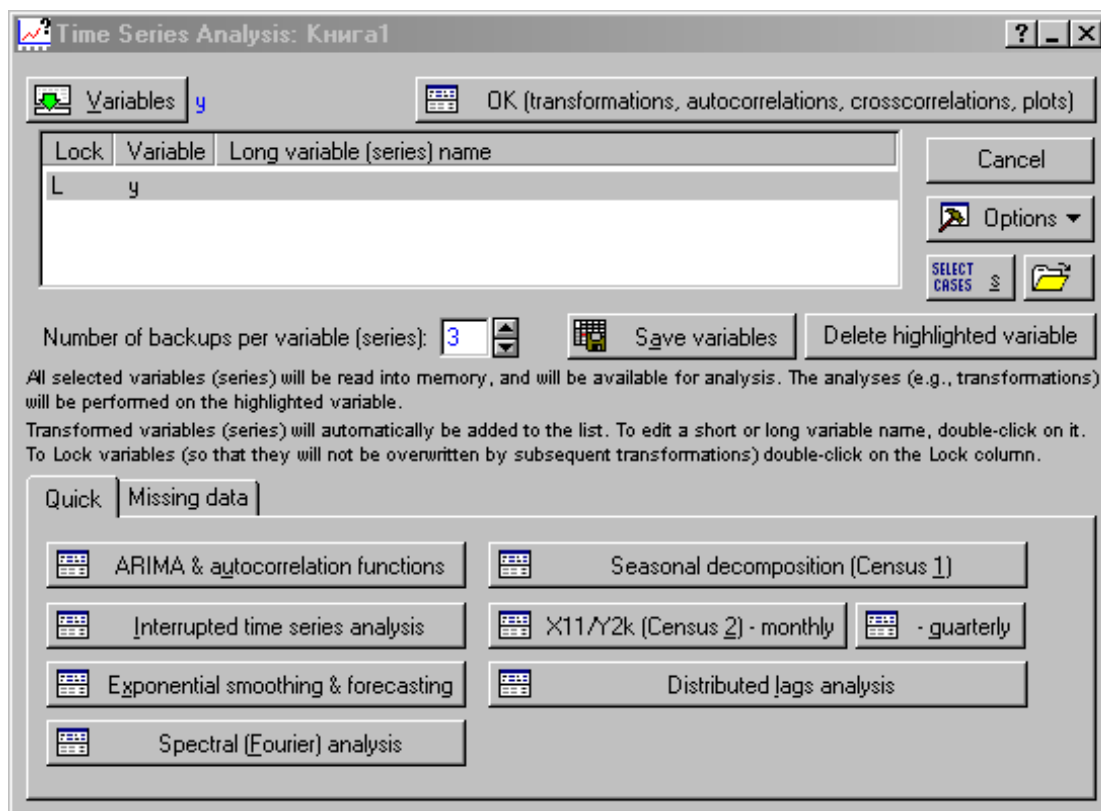


Рисунок 6.1 – Окно выбора методов анализа временных рядов

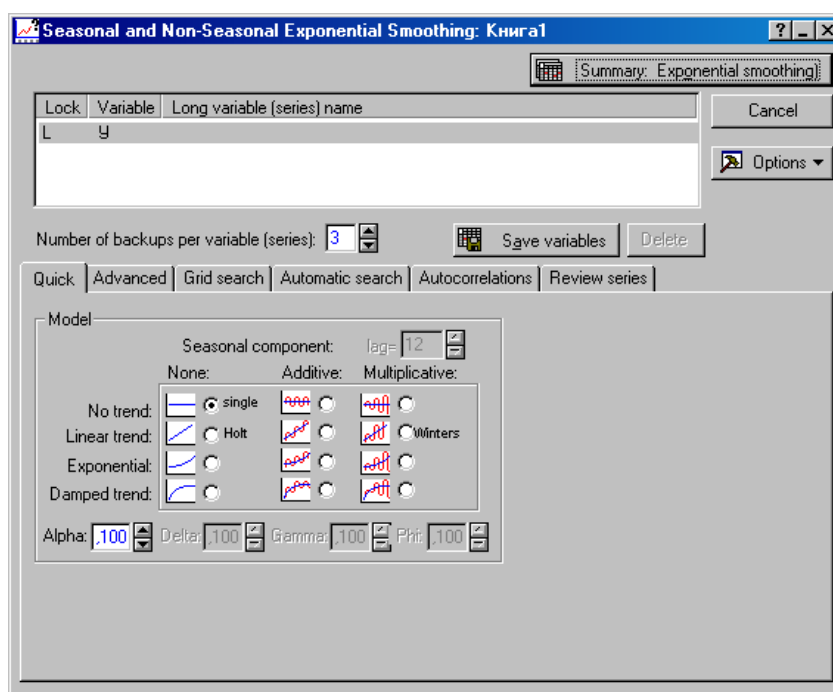


Рисунок 6.2 – Окно настроек моделей экспоненциального сглаживания

где: *Seasonal component* – опция задает лаг (задержку) сезонной компоненты;
None – в данном столбце собраны модели не учитывающие сезонную составляющую;
Additive – в столбце отображены модели содержащие аддитивную

сезонную составляющую;

Multiplicative – в столбце отображены модели содержащие мультипликативную сезонную составляющую;

No trend – в данной строке представлены модели не содержащие;

Linear trend – модели содержащие линейный тренд;

Exponential – модель содержащие экспоненциальный тренд;

Damped trend – затухающий тренд

Alpha – константа простого экспоненциального сглаживания;

Delta – константа отвечающая за сезонность;

Gamma – константа отвечающая за тренд;

Phi – константа

Шаг 3. Для построения модели с параметрам $\alpha=0,1$, необходимо установить флажок на пересечении опций *None* и *No trend*. В поле *Alpha* указать 0,1, далее нажать кнопку *Summary: Exponential smoothing* (Провести экспоненциальное сглаживание). Результаты оценивания модели становятся доступны в новом окне *Workbook* (Рисунок 6.3).

В верхней части графика выводится сообщение об уровне, с которого начинается сглаживание, в данном случае программа автоматически выбрала $S_0=4300$. Далее указывается, что модель построена без учета тренд и сезонной составляющей (*No trend, no season*), при этом сглаживающая константа равна 0,1 ($Alpha=0,100$).

На графике отображены две оси *OY*, по левой отложены значения исследуемой переменной, по правой даны значения отклонений (*Residuals*) фактических уровней от выровненных. Пунктирной линией на графике отображены сглаженные значения (*Smoothed Series*) изучаемого ряда, точечной линией – отклонения (*Resids*).

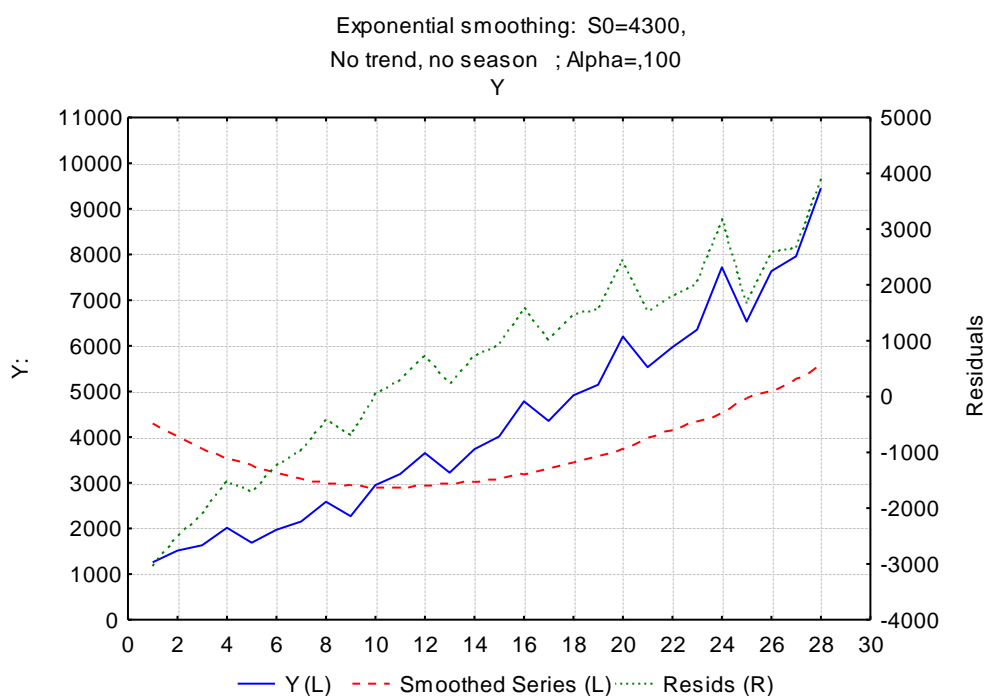


Рисунок 6.3 - Графики сглаженного временного ряда среднедушевых денежных доходов населения при $\alpha=0,1$.

Аналогичные вычисления проведем для $\alpha=0,5$ и $\alpha=0,9$.

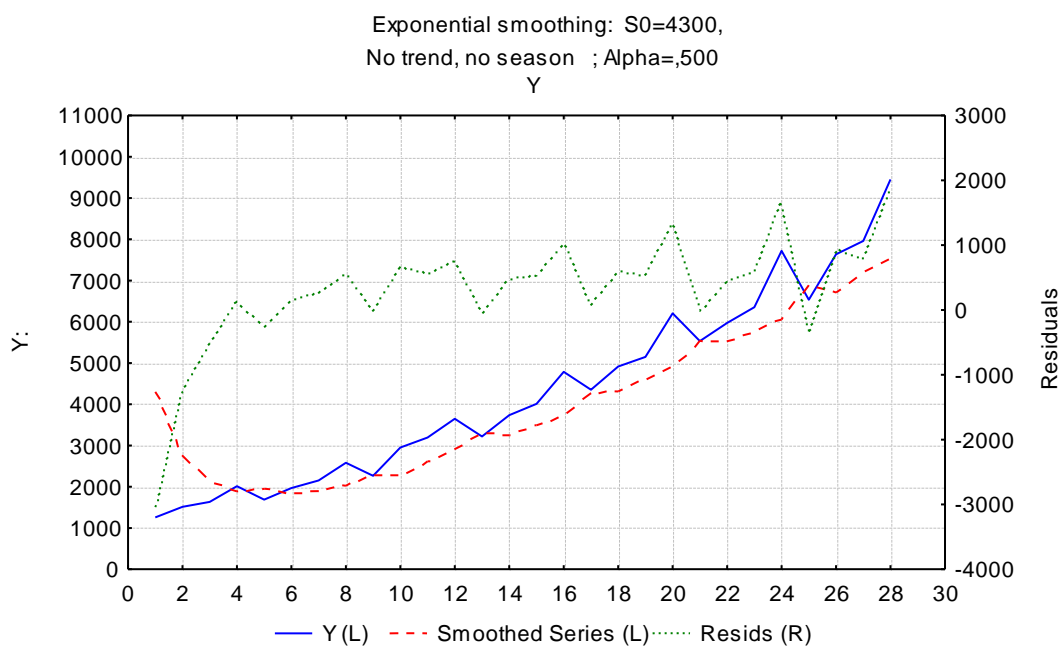


Рисунок 6.4 - Графики сглаженного временного ряда среднедушевых денежных доходов населения при $\alpha=0,5$

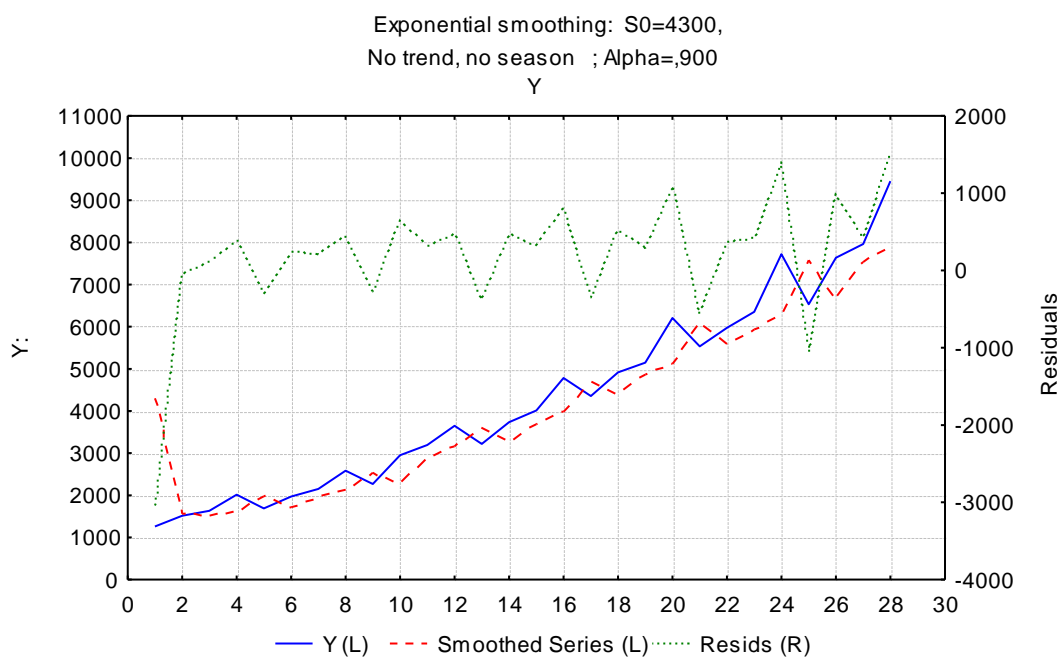


Рисунок 6.5 - Графики сглаженного временного ряда среднедушевых денежных доходов населения при $\alpha=0,9$

Анализируя приведенные графики нетрудно заметить, что чем меньше

значение сглаживающей константы, тем меньше варьируют сглаженные значения. При малой величине $\alpha = 0,1$ сглаженные значения сильно расходятся с фактическими уровнями исходного временного ряда. В общем случае сглаживание при малых α слабо реагирует на подобные скачки или поворотные точки. Большая константа $\alpha = 0,9$ дает гораздо меньший сглаживающий эффект, однако сглаженные значения в большей степени следуют фактическим значениям по всей длине исходного временного ряда.

Константа $\alpha = 0,5$ дает промежуточный эффект между первыми двумя вариантами. Это подтверждает тот факт, что использовать большие константы следует для рядов, содержащих незначительную нерегулярную компоненту.

Заметим также, что анализируемый показатель имеет тенденцию к росту, а также сезонную составляющую с максимумом в 4 квартале каждого года.

Основная цель экспоненциального сглаживания, как и любого другого метода исследования временных рядов, является прогнозирование уровней ряда на k -уровней вперед (не стоит исключать случаи прогнозирования на k -уровней назад).

Для построения прогноза на основе модели экспоненциальной средней в СПП STATISTICA необходимо в окне *Seasonal and Non-Seasonal Exponential smoothing* выбрать вкладку *Advanced* (Расширенные).

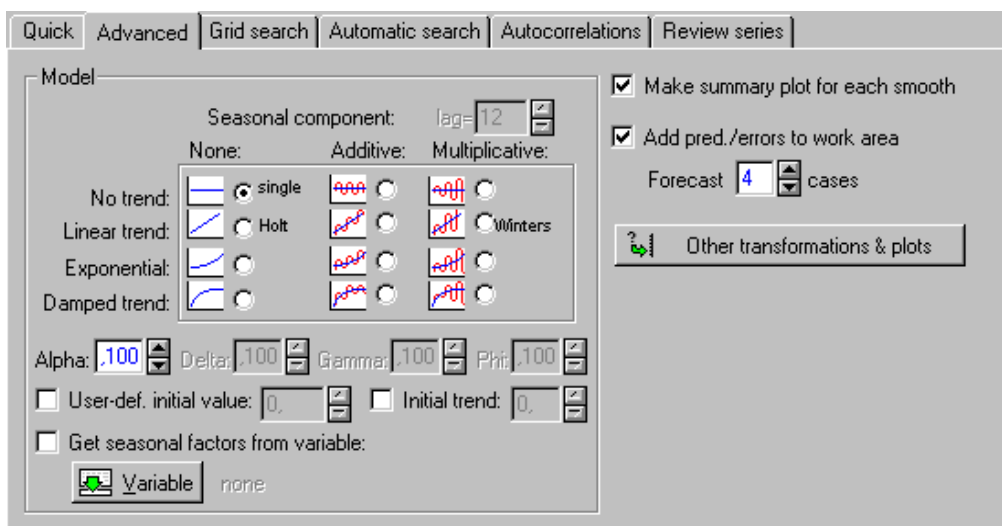


Рисунок 6.6 – Окно установок для прогнозирования (приведена часть исходного окна)

Для установки периода упреждения необходимо в поле *Forecast ___ cases* (Прогнозировать на k значений) установить необходимое значение, в нашем случае установим число 4 (что соответствует четырем кварталам 2006г.). Все результаты оценивания модели представим в таблице 6.1.

Согласно полученным прогнозным значениям, представленным в таблице 6.1, варианты прогноза по модели с параметром 0,5 и 0,9 практически не отличаются друг от друга, также среднеквадратическое отклонение в данных

случаях ниже, чем в первом. Отсюда можно сделать вывод, что наилучшие прогнозы, по-видимому, будут получены по моделям с высокими значениями сглаживающей константы. Но вместе с тем необходимо указать на то, что рассчитанные значения показателей значительно расходятся с фактическими значениями. Объясняется данное явление тем, что в анализируемом ряду содержится тренд и сезонная составляющая.

Таблица 6.1 – Результаты прогнозирования уровней ряда среднедушевых денежных доходов населения с помощью экспоненциального сглаживания

Квартал/год	Модели			
	Фактические значения	$\alpha=0,1$	$\alpha=0,5$	$\alpha=0,9$
I/2006	7873,0	5938,40	8508,93	9298,35
II/2006	9575,8	5938,40	8508,93	9298,35
III/2006	9988,2	5938,40	8508,93	9298,35
IV/2006	-	5938,40	8508,93	9298,35
Среднеквадратическое отклонение	-	3445955,49	907204,43	739004,78

6.5 Экспоненциальное сглаживание с учетом тренда (метод Хольта)

Простое экспоненциальное сглаживание временных рядов, содержащих устойчивый тренд, приводит к систематической ошибке, связанной с отставанием сглаженных значений от фактических уровней временного ряда. Для учета тренда в нестационарных рядах применяется специальное двухпараметрическое линейное экспоненциальное сглаживание. В отличие от простого экспоненциального сглаживания с одной сглаживающей константой (параметром) данная процедура сглаживает одновременно случайные возмущения и тренд с использованием двух различных констант (параметров).

Процедура прогнозирования начинается с того, что сглаженная величина S_1 полагается равной y_1 . Возникает проблема определения начального значения тренда a_1 . Рассмотрим два способа оценки a_1 .

Способ 1. Положим $a_1=0$. Такой подход хорошо работает в случае длинного исходного временного ряда. Тогда сглаженный тренд за небольшое число периодов приблизится к фактическому значению тренда.

Способ 2. Можно получить более точную оценку a_1 используя первые, пять (или более) наблюдений временного ряда. На их основе по методу наименьших квадратов решается уравнение $\tilde{y}_t = a_0 + a_1 \cdot t_t$. Величина a_1 берется в качестве начального значения тренда.

Для решения проблемы первым способом необходимо в окне *Seasonal and Non-Seasonal Exponential smoothing* установить галочку в опции *User-def.*

initial value (Установка значение пользователем) и указать значение первого уровня ряда $y_1=1259,1$. Далее выделим *Initial trend* (Установки тренда) и в появившемся поле указать $a_1=0$.

Для реализации второго способа необходимо воспользоваться возможностями модуля *Multiple Regression* (Множественная регрессия).

Шаг 1. Образуем дополнительную переменную t (см. главу 5).

Шаг 2. Проведем оценку тренда по первым 8 уровням (1999-2000гг.) для этого в модуле в окне *Multiple Linear Regression* (Множественная линейная регрессия) выбрать кнопку *Select cases* (Выбор значений).

Шаг 3. В появившемся окне (рисунок 6.7) *Analysis/Graph Case Selection Conditions* (Выбор наблюдений для анализа и построения графиков), установим галочку в опции *Enable Selection Conditions* → *Specific, selected by:* в поле *By Expression:* необходимо указать какие значения используются для оценки модели (в нашем случае 1-8 наблюдение).

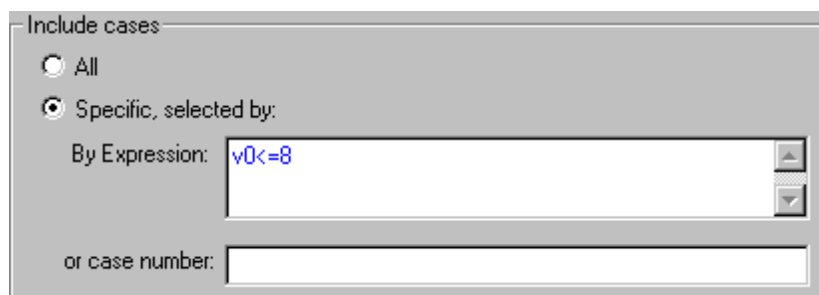


Рисунок 6.7 – Выбор наблюдений для оценки линейного тренда (приведена часть исходного окна)

После проведения процедуры получаем следующие результаты:

Таблица 6.2 – Результаты оценивания линейного тренда ряда среднедушевых денежных доходов населения

	<i>Beta</i> (БЕТА)	<i>Std.Err. of Beta</i> (Стд.Ош. БЕТА)	<i>B</i>	<i>Std.Err. of B</i> (Стд.Ош. B)	<i>t(6)</i>	<i>p-level</i> (p-уров.)
<i>Intercept</i> (Св.член)	-	-	1148,007	134,123	8,559	0,000
<i>t</i>	0,923	0,157	156,418	26,560	5,889	0,001

Определив значение параметра a_1 линейного тренда вернемся в модуль *Exponential smoothing & forecasting* и внесем начальный уровень, т.е. выберем опцию *Initial trend* и укажем значение параметра $a_1=156,418$.

Помимо озвученной проблемы остается острой проблема выбора оптимальных констант, при этом задача усложняется в виду увеличение числа

констант до двух. Решить данную проблему, призвана процедура *Perform grid search* (Поиск на сетке решений), позволяющая на основе перебора всех возможных вариантов выбрать модель с оптимальными константами.

Для запуска данной процедуры в окне *Seasonal and Non-Seasonal Exponential smoothing* выберем вкладку *Grid search* (Поиск решений) (рисунок 6.7). При этом выполним данную процедуру в двух вариантах, при $a_1=0$ и при $a_1=156,418$. полученные результаты представим в таблице 6.3 и 6.4.

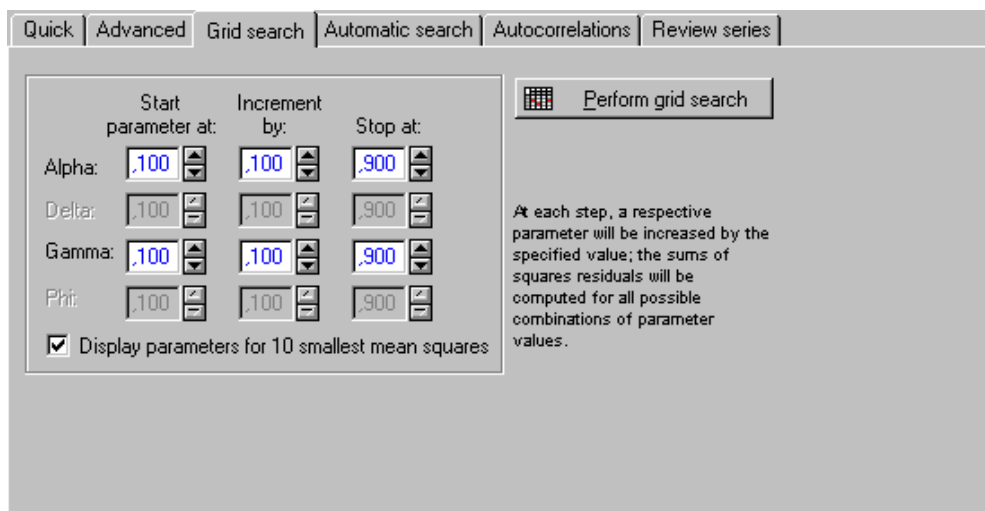


Рисунок 6.7 – Установки процедуры поиска оптимальных вариантов констант (приведена часть исходного окна)

Таблица 6.3 - Варианты поиска для модели Хольта ряда среднедушевых денежных доходов населения на сетке решений при $a_1 = 0$ и $y_1=1259,1$ (приведена часть исходной таблицы)

	<i>Alpha</i>	<i>Gamma</i>	<i>Mean Error</i>	<i>Mean Abs Error</i>	<i>Sums of Squares</i>	<i>Mean Squares</i>	<i>Mean % Error</i>	<i>Mean Abs Error</i>
9	0,1	0,9	161,153	378,617	6980228	232674,3	5,294	9,348
18	0,2	0,9	80,756	354,800	7149853	238328,4	2,368	8,133

Таблица 6.4 - Варианты поиска для модели Хольта ряда среднедушевых денежных доходов населения на сетке решений при $a_1 = 156,418$ и $y_1=1259,1$ (приведена часть исходной таблицы)

	<i>Alpha</i>	<i>Gamma</i>	<i>Mean Error</i>	<i>Mean Abs Error</i>	<i>Sums of Squares</i>	<i>Mean Squares</i>	<i>Mean % Error</i>	<i>Mean Abs Error</i>
9	0,1	0,9	91,834	344,725	6420119	214004,0	1,188	7,711
8	0,1	0,8	100,869	344,618	6454859	215162,0	1,370	7,752
16	0,2	0,7	65,909	345,908	6913086	230436,2	0,584	7,556
17	0,2	0,8	58,892	348,233	6984981	232832,7	0,451	7,619

где: *Mean Error* - Средняя ошибка

Mean Abs Error - Абсолютная средняя ошибка

Sums of Squares - Сумма квадратов отклонения

Mean Squares - Средние квадраты

Mean % Error - Процент средней ошибки

Mean Abs Error - Процент абсолютная средняя ошибка

Опишем некоторые особенности приведенных показателей:

Средняя ошибка (*Mean error*) вычисляется простым усреднением ошибок на каждом шаге.

$$\bar{\varepsilon} = \frac{\sum \varepsilon}{n} \quad (6.1)$$

Очевидным недостатком этой меры является то, что положительные и отрицательные ошибки аннулируют друг друга, поэтому она не является хорошим индикатором качества прогноза.

Средняя абсолютная ошибка (*Mean absolute error*) вычисляется как среднее *абсолютных* ошибок.

$$|\bar{\varepsilon}| = \frac{\sum |\varepsilon|}{n} \quad (6.2)$$

Если она равна 0 (нулю), то имеем совершенную подгонку (прогноз). В сравнении со средней *квадратической* ошибкой, эта мера «не придает слишком большого значения» выбросам.

Сумма квадратов ошибок (*Sum of squared error (SSE), Mean squared error.*), среднеквадратическая ошибка. Эти величины вычисляются как сумма (или среднее) квадратов ошибок.

Это наиболее часто используемые индексы качества подгонки.

Относительная ошибка (*Percentage error (PE)*). Во всех предыдущих мерах использовались действительные значения ошибок. Представляется естественным выразить индексы качества подгонки в терминах *относительных* ошибок. Например, при прогнозе месячных продаж, которые могут сильно флуктуировать (например, по сезонам) из месяца в месяц, вы можете быть вполне удовлетворены прогнозом, если он имеет точность 10%. Иными словами, при прогнозировании абсолютная ошибка может быть не так интересна как относительная. Чтобы учесть относительную ошибку, было предложено несколько различных индексов. В первом относительная ошибка вычисляется как:

$$OO_t = \frac{100 \times (y_t - \tilde{y}_t)}{y_t} \quad (6.3)$$

где: y_t - наблюдаемое значение в момент времени t ,

\tilde{y}_t - прогноз (сглаженное значение).

Средняя относительная ошибка (*Mean percentage error (MPE)*). Это значение вычисляется как среднее относительных ошибок

$$\bar{\varepsilon}_t = \frac{1}{n} \sum \frac{(y_t - \tilde{y}_t)}{y_t} \times 100 \quad (6.4)$$

Средняя абсолютная относительная ошибка (*Mean absolute percentage error (MAPE)*). Как и в случае с обычной средней ошибкой отрицательные и положительные относительные ошибки будут подавлять друг друга. Поэтому для оценки качества подгонки в целом (для всего ряда) лучше использовать среднюю *абсолютную* относительную ошибку.

$$\bar{\varepsilon} = \frac{1}{n} \sum \frac{|y_t - \tilde{y}_t|}{y_t} \times 100 \quad (6.5)$$

Данный показатель является относительным показателем точности прогноза и не отражает размерность изучаемых признаков, выражается в процентах и на практике используется для сравнения точности прогнозов полученных как по различным моделям, так и по различным объектам. Интерпретация оценки точности прогноза на основе данного показателя представлена в следующей таблице:

< 10	Высокая точность
10 – 20	Хорошая точность
20 – 50	Удовлетворительная точность
> 50	Не удовлетворительная точность

Как видим из приведенных в таблице 6.3 результатов приемлемыми являются модели 9 и 18 (с параметрами $Alpha=0,1$ и $Gamma=0,9$; $Alpha=0,2$ и $Gamma=0,9$). Из таблицы 6.4, выбираем модели с номером 9, 8, 16 и 17.

Для определения наилучшей модели построим прогноз по всем выбранным моделям (6 моделей).

Таблица 6.5 – Прогнозные значения по экспоненциальным моделям среднедушевых денежных доходов населения

Варианты моделей	I/2006	II/2006	III/2006	IV/2006	% средней относительной ошибки
Фактические значения	7873,0	9575,8	9988,2	-	-
$Alpha=0,1$ $Gamma=0,9$ ($a_1 = 0$ и $y_1=1259,1$)	8718,19	9179,22	9640,26	10101,29	5,85
$Alpha=0,2$ $Gamma=0,9$ ($a_1 = 0$ и $y_1=1259,1$)	9033,69	9604,47	10175,25	10746,03	2,91
$Alpha=0,1$ $Gamma=0,9$ ($a_1 = 156,418$ и $y_1=1259,1$)	8784,57	9223,70	9662,84	10101,97	1,49
$Alpha=0,1$ $Gamma=0,8$ ($a_1 = 156,418$ и $y_1=1259,1$)	8808,39	9241,52	9674,64	10107,77	1,70
$Alpha=0,2$ $Gamma=0,7$ ($a_1 = 156,418$ и $y_1=1259,1$)	8964,61	9486,17	10007,72	10529,27	0,85
$Alpha=0,2$ $Gamma=0,8$ ($a_1 = 156,418$ и $y_1=1259,1$)	9000,21	9550,73	10101,24	10651,75	0,83

Как видим из приведенных в таблице 6.5 данных, наиболее

приближенным к фактическим данным является прогноз по модели с параметрами $Alpha=0,1$ и $Gamma=0,9$, при этом и среднеквадратическое отклонение получено минимальным.

Согласно полученным значениям, наиболее приемлемыми являются прогнозы по модели $Alpha=0,1$ и $Gamma=0,9$ ($a_1 = 0$ и $y_1=1259,1$), так как прогнозные значения ближе к фактическим данным. Но при этом процент средней относительной ошибки (таблица 6.5) получен высоким. Если выбирать прогнозную модель по на основе данного показателя, то наилучшей стоит признать модели - $Alpha=0,2$ и $Gamma=0,8$ ($a_1 = 156,418$ и $y_1=1259,1$).

6.6 Экспоненциальное сглаживание с учетом одновременно тренда и сезонности (метод Винтерса)

Метод Хольта обобщается для временных рядов, содержащих наряду с трендом ярко выраженную сезонную компоненту. Новый метод линейного и сезонного экспоненциального сглаживания - метод Винтера - является трехпараметрическим, так как включает три сглаживающие константы. Он содержит три уравнения: к двум уравнениям, сглаживающим наблюдения и тренд, добавляется уравнение для сглаживания сезонных изменений.

6.5.1 Метод Винтерса – первый способ

Построение модели Винтера первым способом, не содержит в себе не каких сложностей и во многом схож с алгоритмом построения модели Хольта.

Шаг 1. Установим галочку на пересечении строки *Linear trend* (Линейный тренд) и столбца *Additive* (Аддитивно).

Шаг 2. В опции *Season components Lag* ___ (Ляг сезонной компоненты) укажем значение равное 4 (т.к. длина волны составляет 4 квартала).

Шаг 3. Проведем поиск параметров модели на сетке решений, для этого обратимся к вкладке *Grid search* нажмем кнопку вкладку *Perform grid search*. Результаты оценки представим в таблице 6.6.

Таблица 6.6 - Варианты поиска для модели Винтерса ряда среднедушевых денежных доходов населения на сетке решений при $a_1 = 0$ и $y_1=0$ индекс сезонности равен 1 (приведена часть исходной таблицы)

	<i>Alpha</i>	<i>Delta</i>	<i>Gamma</i>	<i>Mean Error</i>	<i>Mean Abs Error</i>	<i>Sums of Squares</i>	<i>Mean Squares</i>	<i>Mean % Error</i>	<i>Mean Abs Error</i>
319	0,4	0,9	0,4	36,712	197,157	1687713	60275,46	-0,104	6,132
310	0,4	0,8	0,4	37,788	197,938	1741775	62206,26	-0,092	6,072
321	0,4	0,9	0,6	30,107	205,065	1742592	62235,44	-0,001	6,469

Согласно полученным результатам с установленными параметрами наилучшими признаны 4 модели.

Воспользуемся оцененными параметрами по моделям и рассчитаем

прогнозы на 2006г. (результаты представим в таблице 6.7).

Таблица 6.7 – Прогнозные значения по экспоненциальным моделям среднедушевых денежных доходов населения метод Винтерса (первый способ)

Варианты моделей	I/2006	II/2006	III/2006	IV/2006	% средней относительной ошибки
Фактические значения	7873,0	9575,8	9988,2	-	-
$Alpha=0,4$ $Delta=0,9$ $Gamma=0,4$ ($a_1 = 0$ и $y_1=0$)	8337,92	9366,02	9730,30	11140,20	6,13
$Alpha=0,4$ $Delta=0,8$ $Gamma=0,4$ ($a_1 = 0$ и $y_1=0$)	8368,49	9366,86	9754,64	11136,33	6,07
$Alpha=0,4$ $Delta=0,9$ $Gamma=0,6$ ($a_1 = 0$ и $y_1=0$)	8393,16	9461,03	9850,54	11295,20	6,47

6.5.2. Метод Винтерса – второй способ

Шаг 1. Воспользуемся установками из пункта 6.5.1 и добавим в поле *Initial trend* значение $a_1 = 156,418$.

Шаг 2. Установим галочку в опции *User-def. initial value* (Установка значение пользователем) и укажем значение первого уровня ряда $y_1=1259,1$.

Шаг 3. Найдем на сетке решения оптимальные значения сглаживающих констант для этого обратимся к вкладке *Grid search* нажмем кнопку вкладки *Perform grid search*. Результаты оценки представим в таблице 6.8.

Таблица 6.8 - Варианты поиска для модели Винтерса ряда среднедушевых денежных доходов населения на сетке решений при $a_1 = 156,418$ и $y_1=1259,1$ индекс сезонности равен 1 (приведена часть исходной таблицы)

	<i>Alpha</i>	<i>Delta</i>	<i>Gamma</i>	<i>Mean Error</i>	<i>Mean Abs</i>	<i>Sums of Squares</i>	<i>Mean Squares</i>	<i>Mean %</i>	<i>Mean Abs</i>
--	--------------	--------------	--------------	-------------------	-----------------	------------------------	---------------------	---------------	-----------------

					<i>Error</i>			<i>Error</i>	<i>Error</i>
239	0,3	0,9	0,5	64,752	182,265	1381981	49356,45	1,223	5,106
238	0,3	0,9	0,4	74,993	178,374	1398731	49954,69	1,402	4,926
242	0,3	0,9	0,8	47,060	189,521	1430175	51077,67	0,908	5,509

Прогнозы на четыре шага вперед по трем моделям представим в таблице 6.9.

Таблица 6.9 – Прогнозные значения по экспоненциальным моделям среднедушевых денежных доходов населения метод Винтерса (второй способ)

Варианты моделей	I/2006	II/2006	III/2006	IV/2006	% средней относительной ошибки
Фактические значения	7873,0	9575,8	9988,2	-	-
<i>Alpha</i> =0,3 <i>Delta</i> =0,9 <i>Gamma</i> =0,5 ($a_1 = 156,418$ и $y_1=1259,1$)	8284,16	9344,05	9705,14	11145,25	5,11
<i>Alpha</i> =0,3 <i>Delta</i> =0,9 <i>Gamma</i> =0,4 ($a_1 = 156,418$ и $y_1=1259,1$)	8251,06	9290,60	9637,92	11063,27	4,93
<i>Alpha</i> =0,3 <i>Delta</i> =0,9 <i>Gamma</i> =0,8 ($a_1 = 156,418$ и $y_1=1259,1$)	8360,62	9466,90	9855,04	11327,90	5,51

Сравнивая результаты построения моделей Винтерса можно сделать вывод, что второй способ дает лучшие результаты, т.к. процент средней относительной ошибки получен значительно ниже.

6.7 Тесты для самоконтроля

1 Для описания динамического ряда с помощью экспоненциального сглаживания содержащего тренд составляющую используют:

- метод Брауна
- метод Хольта
- метод Винтера

2 Для описания динамического ряда с помощью экспоненциального сглаживания содержащего тренд и сезонную составляющую используют:

- метод Брауна
- метод Хольта
- метод Винтера

3 Для описания динамического ряда с помощью экспоненциального сглаживания не содержащего тренд и сезонную составляющую используют:

- метод Брауна
- метод Хольта
- метод Винтера

4 Следующее уравнение $\tilde{y}_{t+m} = (S_t + ma_t) \cdot F_{t+m-L}$ используется для прогноз на m периодов вперед при:

- простом экспоненциальном сглаживании
- экспоненциальном сглаживании с учетом тренда и сезонности
- экспоненциальном сглаживании с учетом тренда

5 Следующее уравнение $\tilde{y}_{t+m} = S_t + ma_t$ используется для прогноз на m периодов вперед при:

- простом экспоненциальном сглаживании
- экспоненциальном сглаживании с учетом тренда и сезонности
- экспоненциальном сглаживании с учетом тренда

6) Приведенные выражения $S_t = \alpha y_t + (1-\alpha)(S_{t-1} + a_{t-1})$, $a_t = \gamma(S_t - S_{t-1}) + (1-\gamma)a_{t-1}$ используются при:

- однопараметрический методе экспоненциального сглаживания
- двухпараметрический методе экспоненциального сглаживания
- трехпараметрический методе экспоненциального сглаживания

7 Приведенные выражения $S_t = \alpha(y_t/F_{t-L}) + (1-\alpha)(S_{t-1} + a_{t-1})$, $F_t = \delta(y_t/S_t) + (1-\delta)F_{t-L}$, $a_t = \gamma(S_t - S_{t-1}) + (1-\gamma)a_{t-1}$ используются при:

- однопараметрический методе экспоненциального сглаживания
- двухпараметрический методе экспоненциального сглаживания
- трехпараметрический методе экспоненциального сглаживания

8 Приведенное выражение $S_t = \tilde{y}_t + \alpha(y_t - \tilde{y}_t)$ используется при:

- однопараметрический методе экспоненциального сглаживания
- двухпараметрический методе экспоненциального сглаживания
- трехпараметрический методе экспоненциального сглаживания

6.8 Задания для самостоятельного выполнения

Задания для самостоятельной работы составлены в пяти вариантах, номер варианта выбирается в соответствии с последней цифрой зачетной книжки студента:

Таблица 6.10 – Задания для самостоятельной работы по теме «Экспоненциальное сглаживание временных рядов»

Последняя цифра номера зачетной книжки	1 и 6	2 и 7	3 и 8	4 и 9	5 и 0
Номер варианта	1	2	3	4	5

Используя данные (приложение 13) соответствующего варианта необходимо выполнить следующие этапы самостоятельной работы:

- 1 построить адаптивную модель с помощью простого экспоненциального сглаживания при значениях константы α равной 0,1, 0,5 и 0,9;
- 2 построить экспоненциальную модель с учетом тренда;
- 3 построить экспоненциальную модель с учетом тренд-сезонной составляющей.

7 Анализ и прогнозирования временных рядов на основе моделей ARMA

7.1 Цели и задачи лабораторной работы

В данной лабораторной работе, на основе динамического ряда денежного агрегата M0 (приложение 12) 1 квартал 1999г. по 4 квартал 2005г., рассмотрим способы идентификации и построения ARMA модели, при этом выделим следующие задачи:

- 1 на основе фактических данных построить авторегрессионную и частную авторегрессионную функцию и сделать предположение о стационарности ряда;
- 2 с помощью вычисления разностей определить порядок ARMA модели;
- 3 построить наиболее адекватную имеющимся данным ARMA модель и сделать прогноз на 1-4 квартала 2006г. (сравним полученные результаты с фактическими данными).

7.2 Теоретические аспекты адаптивных методов прогнозирования

Основоположниками разработки адаптивных методов можно назвать Дж. Бокса и Г. Дженкенса. Авторы рассматривали так называемые ARMA модели, применение которых основывалось на стационарности исследуемого ряда. В настоящее время круг данных моделей расширен до ARIMA моделей (Авто Регрессия и Проинтегрированное Скользящие Средние) что позволило применять их к более широкому кругу временных рядов из-за снижения требований предъявляемых к исходному динамическому ряду.

В отечественной литературе проблемами построения авторегрессионных моделей посвящены работы Лукашина Ю.П., Конторовича Г.Г., Носко В.П. и Давнис В.В.. В приведенных работах освещены вопросы, касающиеся описания как одномерных, так и многомерных временных рядов с применением авторегрессионных моделей. Приведены модели развивающие подход Бокса и Дженкенса, допускающие наличие детерминированного или стохастического тренда, структурных изменений и т.д.

Согласно перечисленным работам процесс построения ARIMA моделей можно разбить на несколько этапов:

1 этап. С помощью соответствующих тестов проводят анализ динамического ряда на стационарность. Если в рассматриваемый временной ряд она присутствует, то переходят к третьему этапу построения.

Ряд, для которого выполнены указанные три условия:

- $E(y_t) \equiv \mu$ - если процесс строго стационарный, то его математическое ожидание не зависит от времени;

- $D(y_t) \equiv \sigma^2$ – дисперсия строго стационарного процесса в каждый момент времени одинакова;
- $Cov(y_t, y_{t+\tau}) = \gamma(\tau)$ для любых t и τ - автокорреляционная функция стационарного временного ряда зависит только от разности моментов времени $(t_1 - t_2)$.

называют стационарным в широком смысле (слабо стационарным, стационарным второго порядка или ковариационно стационарным).

2 этап. Если временной ряд не стационарный, то его с помощью соответствующих преобразований можно привести к стационарному.

Во-первых, можно применить d раз взятие разностей. Считается, что с помощью этой процедуры можно исключить тренд составляющую ряда.

Во-вторых, можно оценить линейный (или не линейный) тренд и вычесть его из исходных данных, т.е. $\varepsilon_t = y_t - \tilde{y}_t$ (где $\tilde{y}_t = a_0 + a_1 t_t$ или другая функция).

3 этап. Если мы предполагаем, что некоторый наблюдаемый временной ряд y_1, y_2, \dots, y_T порождается моделью ARMA, то при этом возникает проблема подбора конкретной модели из этого класса, решение которой предусматривает три этапа:

- 1 идентификация модели;
- 2 оценивание модели;
- 3 диагностика модели.

На этапе идентификации производится выбор некоторой частной модели из всего класса ARMA, т.е. выбор значений p и q . при этом можно руководствоваться поведением автокорреляционных и частных автокорреляционных функций (АКФ и ЧАКФ). Опираясь на свойства АКФ и ЧАКФ за годы практической работы с авторегрессионными моделями, исследователями были сформулированы некие правила выбора наилучшей модели, приведенные в работе [дуброва] и представленные таблицей 7.1.

На втором этапе производится уточнение оценок коэффициентов модели с использованием эффективных статистических методов. Для оцененных коэффициентов вычисляются приближенные стандартные ошибки, дающие возможность, при дополнительных предположениях о распределениях случайных величин y_1, y_2, \dots , строить доверительные интервалы для этих коэффициентов и проверять гипотезы об их истинных значениях с целью уточнения спецификации модели.

На третьем этапе применяются различные диагностические процедуры проверки адекватности выбранной модели имеющимся данным.

Неадекватности, обнаруженные в процессе такой проверки, могут указать на необходимую корректировку модели, после чего производится новый цикл подбора, и т.д. до тех пор, пока не будет получена удовлетворительная модель.

Таблица 7.1 – Свойство автокорреляционных и частных автокорреляционных функций

Функции	АКФ	ЧАКФ
ARMA(1,0)	Экспоненциально затухает (монотонно или знакопеременно)	Выброс (пик) на лаге 1
ARMA(2,0)	Экспоненциально затухает или имеет форму синусоидальной волны	Выбросы (пики) на лагах 1 и 2
ARMA(0,1)	Выброс (пик) на лаге 1	Экспоненциально затухает (монотонно или знакопеременно)
ARMA(0,2)	Выбросы (пики) на лагах 1 и 2	Экспоненциально затухает или имеет форму синусоидальной волны
ARMA(1,1)	Экспоненциально затухает от значения $\rho(1)$ (монотонно или знакопеременно)	Экспоненциально убывает от значения $\rho_c(1)$ (монотонно или знакопеременно)

7.3 Рекомендуемая литература

Для лучшего понимания материала изложенного в данной главе необходимо дополнительно проанализировать следующие источники литературы (см. список использованных источников):

- 1 Андерсон, Т. Введение в многомерный статистический анализ. - М.: Физ.-мат. лит., 1963. – 245 с. (см. стр. 189-283)
- 2 Боровиков, В.П. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере : учеб. пособие / В.П. Боровиков, Г.И. Ивченко. – 2-е изд., перераб. И доп. - М. : Финансы и статистика, 2006. - 368 с: ил. ISBN 5-279-03059-7. (см. стр. 146-217)
- 3 Дуброва, Т.А. Статистические методы прогнозирования: учеб. пособие для вузов. – М.: ЮНИТИ-ДАНА, 2003. – 206 с. ISBN 5-238-00497-4. (см. стр. 141-166)
- 4 Прикладная статистика. Основа эконометрики: учебник для вузов: В 2 т. 2-е изд., испр. Т. 2: Айвазян, С.А. Основы эконометрики. - М.: ЮНИТИ-ДАНА, 2001. – 432 с. ISBN 5-238-00305-6. (см. стр. 244-331)
- 5 Эконометрика: учебник для вузов / под ред. И.И. Елисеевой. - М. : Финансы и статистика, 2005. - 576 с. : ил. ISBN 5-279-02786-3. (см. стр. 388-427)

7.4 Определение стационарности модели

Не стационарность ряда можно определить с помощью простых процедур:

Во-первых, проводят визуализацию ряда, в случае если прослеживается тенденция или сезонная составляющая, то динамический ряд можно признать нестационарным.

В главном меню выбираем *Graphs*→*2D Graphs* →*Line Plots (Variables)*... В результате получаем следующий график (рисунок 7.1). Согласно полученным данным прослеживается устойчивая тенденция к росту, соответственно можно предположить, что ряд является не стационарным.

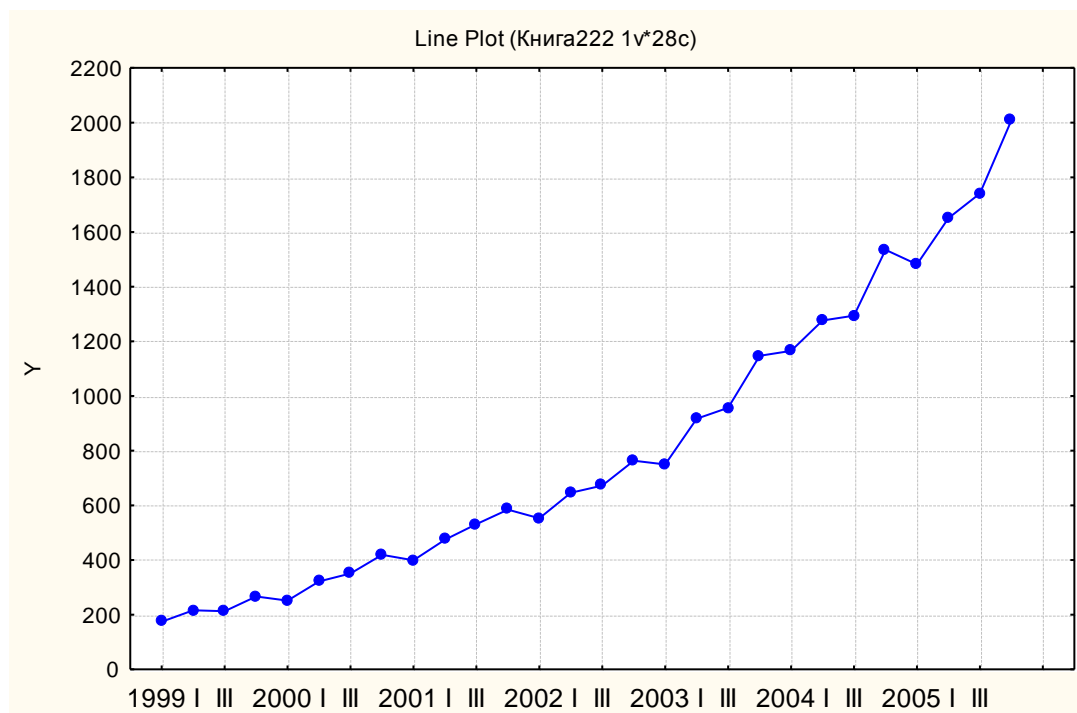


Рисунок 7.1 - Динамика денежного агрегата M0

Во-вторых, можно использовать метод выявления тенденции (подробное описание теста приведен в главе 5). В нашем случае получаем следующие результаты:

Таблица 7.1 - Результаты сравнения двух средних на основе ряда индекса ВВП России

	<i>Mean</i>	<i>Std.Dv.</i>	<i>N</i>	<i>Diff.</i>	<i>Std.Dv.Diff.</i>	<i>t</i>	<i>df</i>	<i>p</i>
<i>Y1</i>	385,836	153,489						
<i>Y2</i>	1239,929	406,564	14	-854,093	256,692	-12,450	13	0,000

Так как *t*-статистика получена, значима, можно утверждать, что во временном ряду существует тенденция средней и существует тренд, соответственно предположение о не стационарности ряда подтверждается.

7.5 Приведение динамического ряда к стационарному виду

Так как исследуемый ряд является не стационарным, необходимо попытаться его преобразовать.

Во-первых, воспользуемся методом взятия первых разностей, при этом возможна реализация данного метода в ручном режиме и в автоматическом режиме.

Для реализации данного алгоритма в программе ручным способом необходимо:

Шаг 1. В исходной таблице добавляем с помощью команды *Insert* → *Add Variables* новую переменную Y_{t-1} , далее переносим (копируем) данные из первого столбца.

Шаг 2. В главном меню выбираем *Date* → *Shift (Lag)* и проводим сдвиг вновь образованной переменной Y_{t-1} на 1 единицу вперед (*Forward*).

Шаг 3. Образуем переменную e и в поле *Long name* вводим выражение $=v1-v2$, тем самым получаем ряд первых разностей. Представим его графически, для этого *Graphs* → *2D Graphs* → *Line Plots (Variables)*... (рисунок 7.2).

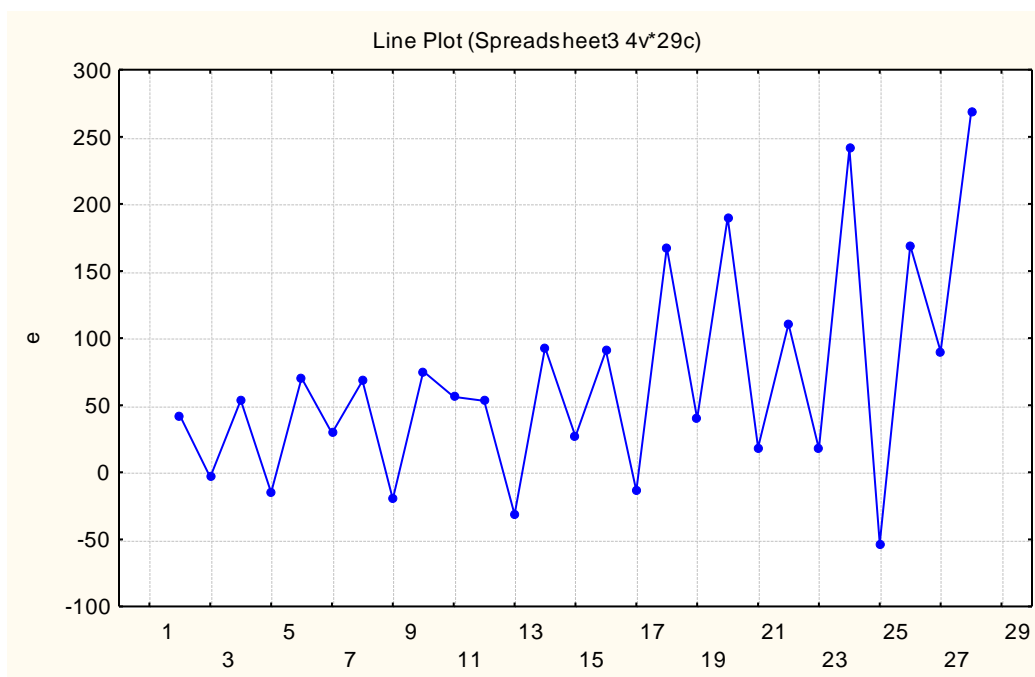


Рисунок 7.2 – Ряд первых разностей денежного агрегата M0

Для реализации данного алгоритма в программе автоматическим способом необходимо:

Шаг 1. Выберем *Statistics* → *Advanced Linear/Nonlinear Models* → *Time series/Frication* (Статистика → Выбор линейных/нелинейных моделей → Временные ряды и прогнозирование).

Шаг 2. В окне *Time Series Analysis* (Анализ временных рядов) необходимо нажать кнопку *OK* (*transformations, autocorrelations, crosscorrelations, plots*).

Шаг 3. В окне *Transformations of Variables* (Преобразование переменной) выберем вкладку $x=f(x)$.

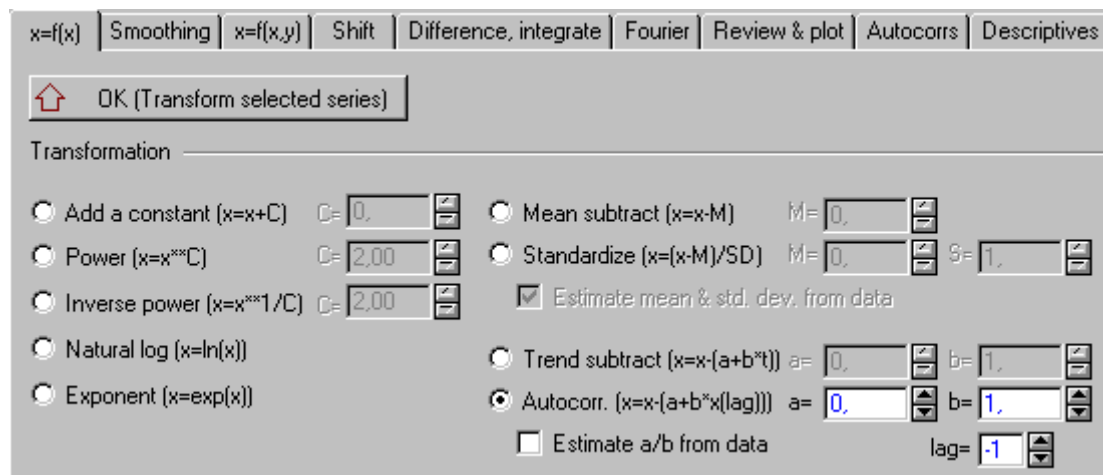


Рисунок 7.3 – Окно выбора алгоритмов преобразования переменной (приведена часть исходного окна)

Необходимо выбрать *Autocorr.* ($x=x-(a+b*x(lag))$) и сделать установки как показано на рисунке 7.3. Далее нажать кнопку *OK* (*Transform selected series*), в результате чего получаем рисунок аналогичный 7.2.

Во-вторых, можно прибегнуть к выделению тренд составляющей, для этого:

Шаг 1. В окне *Time Series Analysis* (Анализ временных рядов) необходимо нажать кнопку *OK* (*transformations, autocorrelations, crosscorrelations, plots*).

Шаг 2. Во вкладке $x=f(x)$ выделим опцию *Trend subtract* ($x=x-(a+b*t)$) в результате чего получаем следующий график:

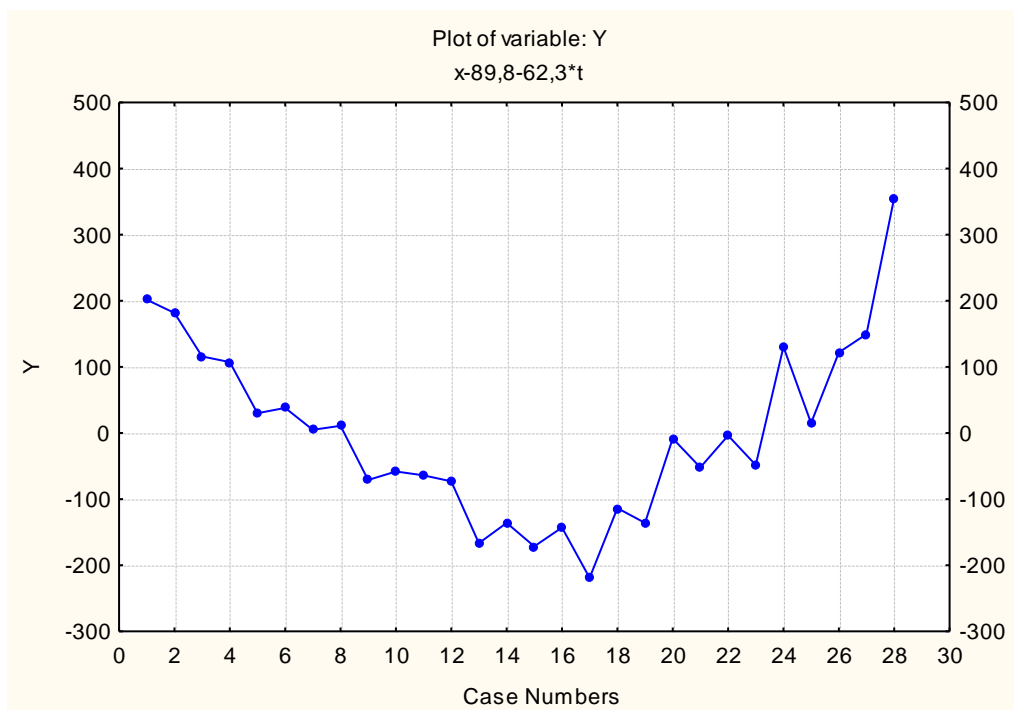


Рисунок 7.4 – Ряд денежного агрегата М0 без тренд составляющей

Согласно полученным рисункам, можно предположить что выполненные нами действия не смогли привести к желаемому результату (стационарности ряда). Так в первом случае (рисунок 7.2) имеем тенденцию к росту, при этом тренд является статистически значимым $\tilde{y}_t = 1,253 + 4,765 \cdot t$ ($F(1,25) = 7,01$). Во втором случае (рисунок 7.4) до 16 уровня (4 квартал 2002г.) имеем тенденцию к снижению, а после этого периода к росту.

7.6 Идентификация порядка разности модели

На практике для выявления значений параметров p и q АРПСС модели используются автокорреляционные и частные автокорреляционные функции.

Шаг 1. Воспользуемся рядом первых разностей (рисунок 7.2) и в окне *Transformations of Variables* (Преобразование переменной) выберем вкладку *Autocorr* (Автокорреляция).

Шаг 2. Для построения автокорреляционной функции необходимо выбрать кнопку *Autocorrelations* (Автокорреляция), при этом можно регулировать уровень значимости, указывая соответствующее значение в окне *p-level for highlighting*. В результате проведения процедуры получаем график представленный на рисунке 7.6.

Для расчета частной автокорреляционной функции выберем кнопку *Partial autocorrelations*, результаты представим на рисунке 7.7.

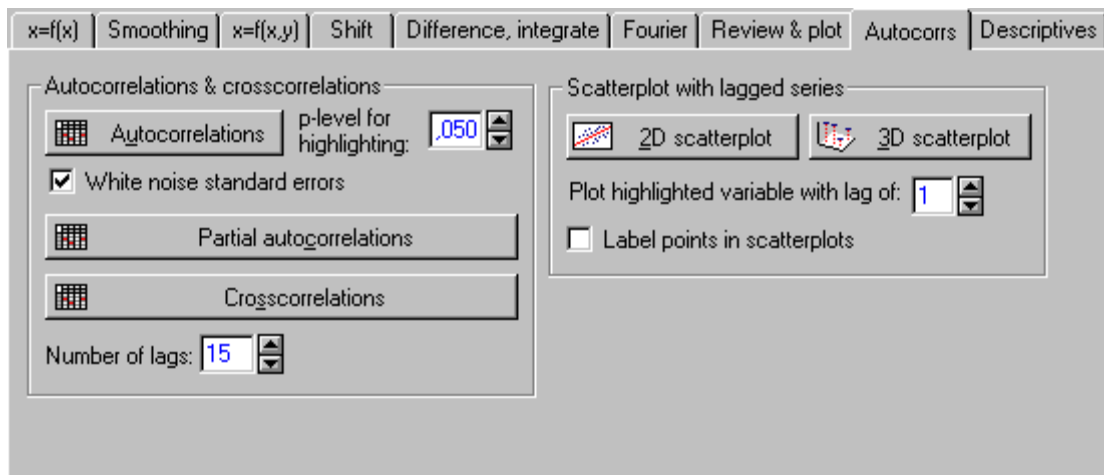


Рисунок 7.5 – Окно выбора расчета автокорреляционной, частной автокорреляционной функции и кросскорреляции (приведена часть исходного окна)

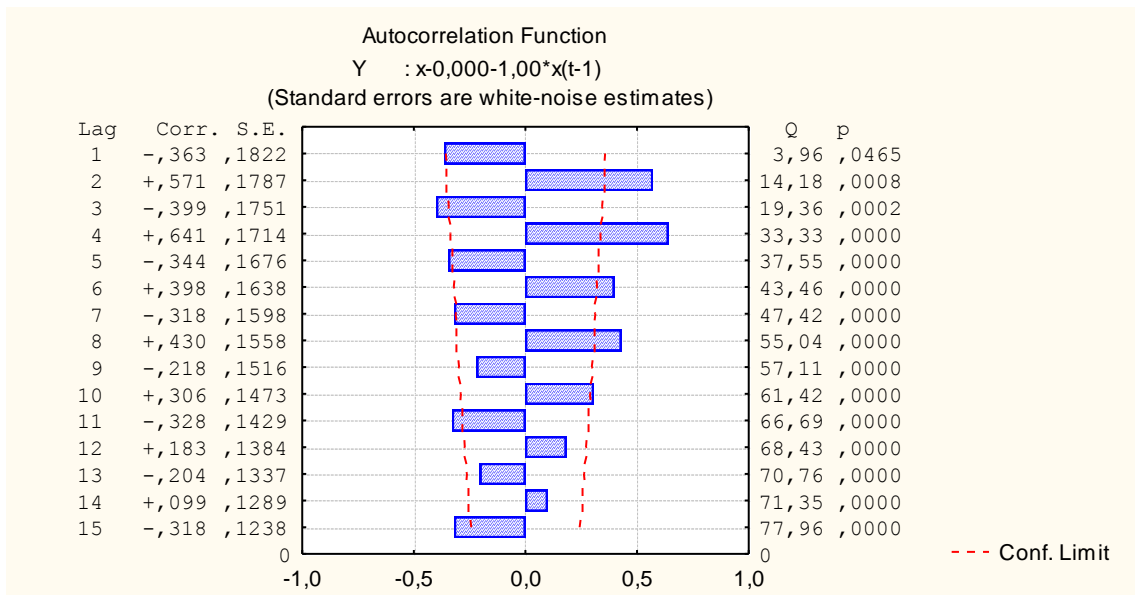


Рисунок 7.6 – Автокорреляционная функция

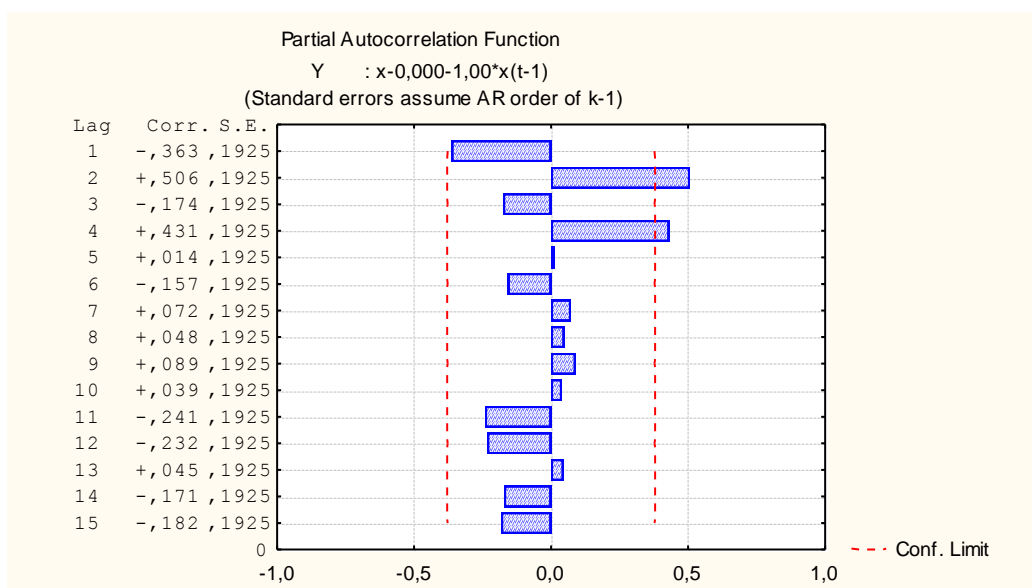


Рисунок 7.7 – Частная автокорреляционная функция

Обратимся к таблице 7.1 и попытаемся идентифицировать ARMA модель. Поведение автокорреляционной функции (экспоненциально затухает) согласуется с моделью ARMA (1,0) или ARMA (2,0). При этом поведение частной автокорреляционной функции (согласно рисунку 7.7, имеем выбросы на втором и четвертом лаге) не согласуется с данными моделями. Поэтому необходимо рассчитаем вторые разности, для этого во вкладке $x=f(x)$ выбрать *Autocorr.* ($x=x-(a+b*x(lag))$), в качестве исходной переменной (в верхней части окна) выберем $x-0,000-1,00*x(t-1)$

Lock	Variable	Long variable (series) name
L	Y	
Y		x-0,000-1,00*x(t-1)

Рисунок 7.8– Окно выбора исходной переменной для расчета (приведена часть исходного окна)

В результате получаем следующие автокорреляционные и частные автокорреляционные функции:

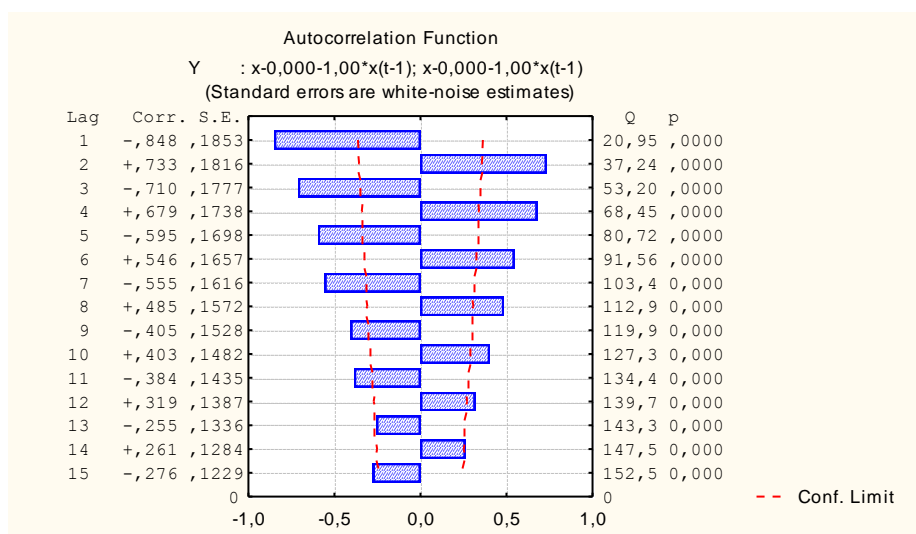


Рисунок 7.9 – Автокорреляционная функция

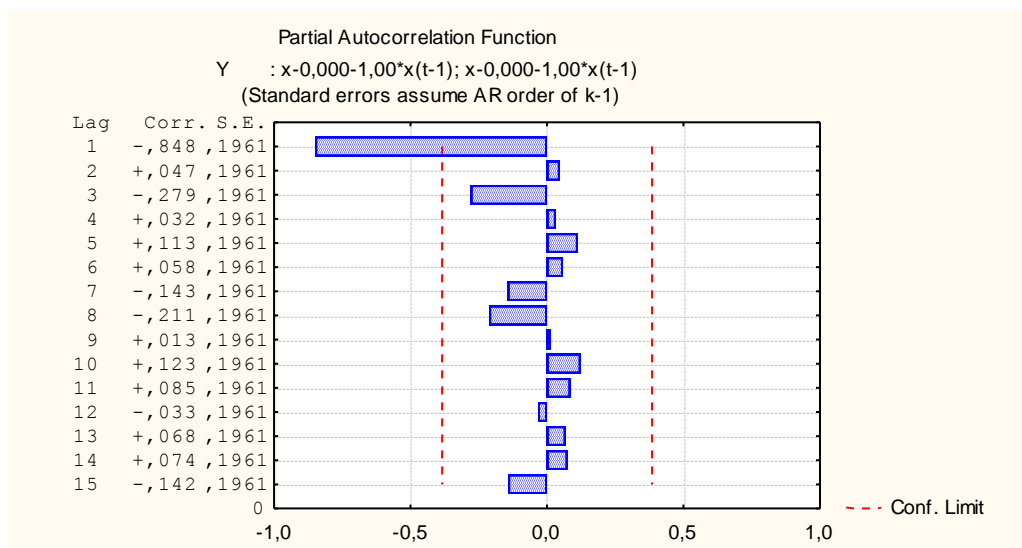


Рисунок 7.10 – Частная автокорреляционная функция

Полученные данные соответствуют модели ARMA(1,0), так как АКФ - экспоненциально затухает, ЧАКФ - выброс (пик) на лаге 1. После идентификации моделей перейдем к непосредственной оценки ее параметров.

7.7 Построение ARMA модели

В рассматриваемом пакете программ существует два метода построения авторегрессионных моделей:

- 1 Построение ARMA модели в модуле *Multiple regression*
- 2 Построение ARMA модели в модуле *ARIMA & Autocorrelation function*

7.7.1 Построение ARMA модели в модуле *Multiple regression*

Оценить параметры ARMA модели можно с помощью обычного МНК (но при этом оценки будут смещены). Для реализации этого метода в пакете необходимо:

Шаг 1. Сохранить результаты расчета вторых разностей, для этого в окне *Transformations of Variables* нажать кнопку *Save variables* получаем новую таблицу, содержащую три переменные:

Y – исходный ряд денежного агрегата $M0$;

Y_1 – ряд первых разностей;

Y_2 – ряд вторых разностей.

Шаг 2. Проведем замену обозначений выведенных переменных: Y_1 засеменим на D' , Y_2 на D'' .

Также необходимо ввести переменную $D''t-1$, для этого в главном меню выбираем *Insert* → *Add Variables*. В появившемся окне *Add Variables* в поле *Name* заменим *NewVar* на $D''t-1$, в поле *Long name* введем выражение $=v3$.

Шаг 3. Выбираем в главном меню *Date* → *Shift (Lag)* и проводим сдвиг переменной D''_{t-1} на один шаг вперед (*Forward*).


Шаг 4. В главном меню выбираем *Statistics* → *Multiple Regression*. В окне *Select dependent and independent variable lists* (запускается после нажатия кнопки *Variables*) в качестве зависимой переменной указываем D'' в качестве не зависимой - D''_{t-1} .

Шаг 5. Установим галочку возле опции *Advanced options (stepwise or ridge regression)* в появившемся окне *Model Definition* в прокрутке *Intercept* выберем *Set to zero* (т.е. регрессионное уравнение будет оценено без свободного члена). После оценки получим следующие результаты:

Таблица 7.2 – Результаты оценки модели вида $\tilde{\Delta}_t'' = \alpha \cdot \Delta_{t-1}''$

	<i>Beta</i> (БЕТА)	<i>Std.Err. of Beta</i> (Стд.Ош. БЕТА)	<i>B</i>	<i>Std.Err. of B</i> (Стд.Ош. B)	<i>t(12)</i>	<i>p-level</i> (p-уров.)
D''_{t-1}	-0,878	0,098	-0,911	0,101	-8,992	0,000

Согласно результатам построения авторегрессионной модели вида ARMA(1,0), параметр α получен статистически значим. Также необходимо отметить, что модель статистически значима по *F*-критерию Фишера ($F_{\text{факт}}(1,24)=80,86$ против $F_{\text{табл}}(1,24)=4,26$).

Шаг 6. В окне итогов построения регрессионной модели (*Multiple Regression Results*) необходимо выбрать вкладку *Residuals / assumptions / prediction*, далее кнопку .

В появившемся окне *Residual Analysis* выбираем вкладку *Save* → *Save residuals & predicted*, в промежуточном окне *Select variables to save with predicted/residual sc...* выделим переменную D'' и нажмем кнопку ОК. Получаем следующую таблицу с результатами:

	1 D''	2 Predicted	3 Residuals	4 Standard Predicted
1	57,400	41,82	15,58	0,37
2	-68,900	-52,29	-16,61	-0,44
3	85,400	62,77	22,63	0,56
4	-41,100	-77,80	36,70	-0,65
5	39,100	37,44	1,66	0,34
6	-88,200	-35,62	-52,58	-0,29
7	95,200	80,35	14,85	0,71

Рисунок 7.11 – Таблица результат оценки ARMA модели

Шаг 7. На основе полученной таблицы построим график. Выберем в главном меню *Graphs* → *2D Graphs* → *Line Plots (Variables...)*. Нажмем кнопку *Variables* (Переменные) и в появившемся окне укажем переменные D'' и *Predicted*. Также в поле *Graph type* (окно *2D Line Plots –Variables*) выделим

Multiple (таким образом, на график будет выведены обе переменные одновременно).

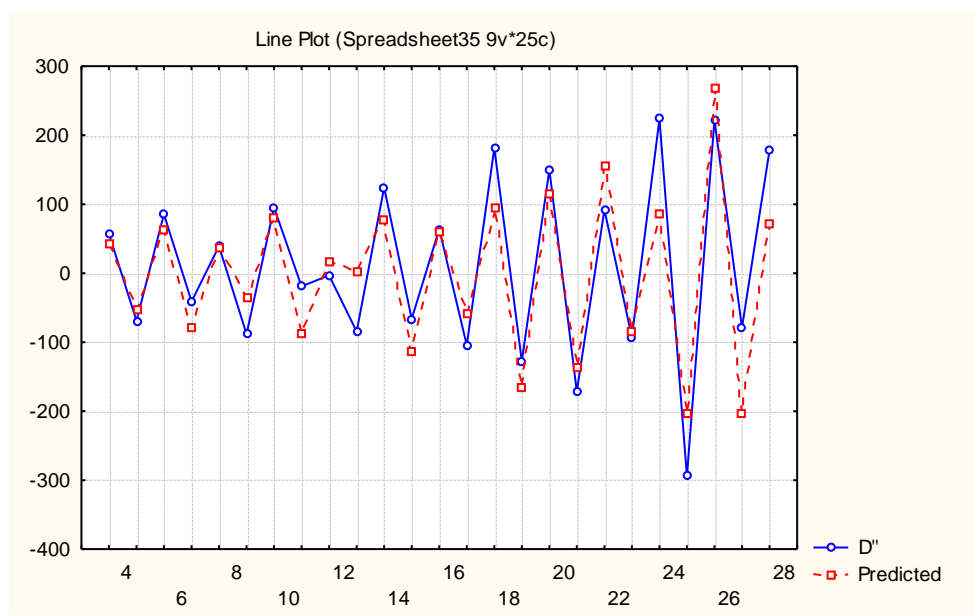


Рисунок 7.12 - Ряд вторых разностей денежного агрегата М0 и модельные данные

Согласно приведенному графику используемая для описания динамики первых разностей авторегрессионная модель ARMA (1,0) очень хорошо описывает анализируемый ряд.

Шаг 8. Снова обратимся к вкладке *Residuals/assumptions/prediction* нажмем кнопку *Predict dependent variable* для прогнозирования уровней исследуемого ряда.

Шаг 9. В окне *Specify values for indep. vars* укажем последнее значение переменной $D''t-1$ равное -79, получим следующий результат:

Таблица 7.3 – Точечный прогноз вторых разностей денежного агрегата М0 на 1 квартал 2006г.

	<i>B-Weight</i> (В-Вес)	<i>Value</i> (Значение)	<i>B-Weight * Value</i> (В-Вес* Значение)
$D''t-1$	-0,911	-79,000	71,969
<i>Predicted</i> (Предсказ.)			71,969
-95,0%CL			55,451
+95,0%CL			88,488

Аналогичным образом рассчитаем значения для 2-4 квартала 2006г. при этом в качестве прогнозного значения будем использовать значение точечного прогноза предыдущего квартала:

Таблица 7.4 - Прогнозные значения вторых разностей денежного агрегата М0 на 2006г.

Квартал/год	Значения показателя для прогноза	Прогноз модели ARMA (1,0)	Нижняя доверительная граница	Верхняя доверительная граница
I/2006	-79,000	71,969	55,451	88,488
II/2006	71,969	-65,564	-80,612	-50,516
III/2006	-65,564	59,729	46,020	73,438
IV/2006	59,729	-54,413	-66,902	-41,924

7.7.2 Построение ARMA модели в модуле *ARIMA & Autocorrelation function*

Для построения семейства ARMA моделей в пакете STATISTICA существует специальный модуль:

Шаг 1. В главном меню выберем *Statistics* → *Advanced Linear/Nonlinear Models* → *Time Series Analysis* → *ARIMA & Autocorrelation function* (Расчеты → Основные линейные и нелинейные модели → Анализ временных рядов → ARIMA & Автокорреляционные функции).

Шаг 2. В появившемся окне *Single Series ARIMA* (АРПСС модель для одномерного ряда) выберем вкладку *Quick* (Быстрые). В верхней части окна выберем ряд, на основе которого будет проведена оценка модели, в нашем случае это – $x - 0,000 - 1,00 * x(t-1)$; $x - 0,000 - 1,00 * x(t-1)$ (ряд вторых разностей).

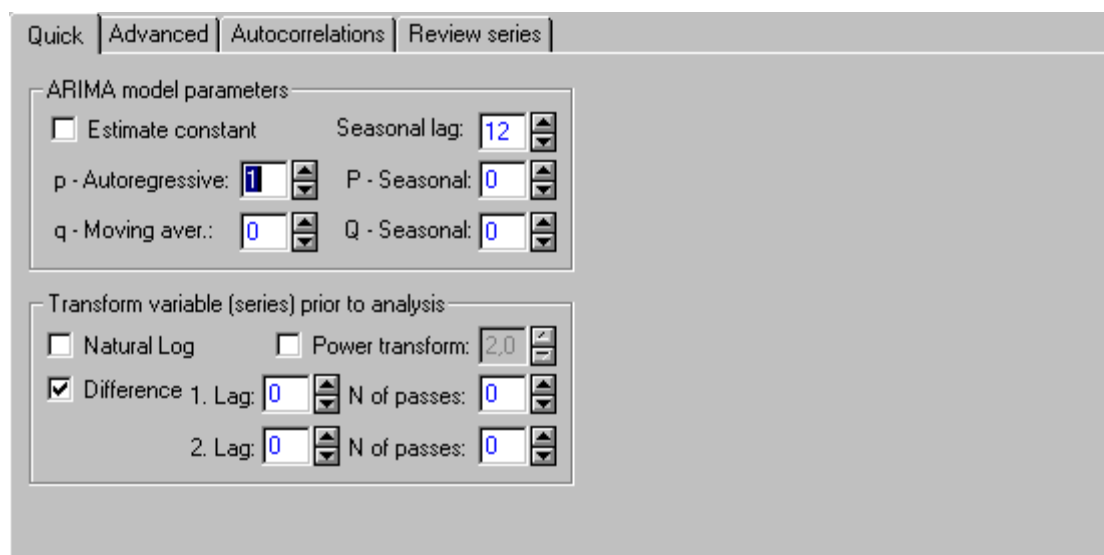



Рисунок 7.13 – Установки ARMA модели (приведено часть исходного окна)

Шаг 3. После нажатия кнопки  будет выведено окно *Single Series ARIMA Results* (Результаты построения АРПСС модели для одномерного ряда).

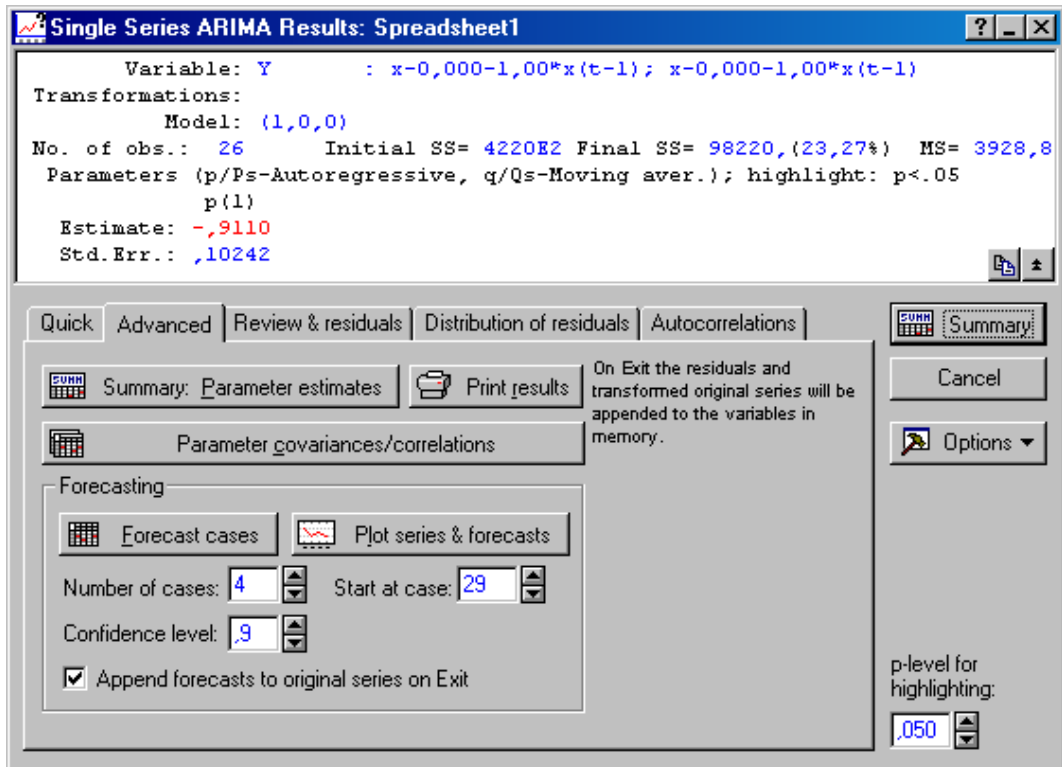


Рисунок 7.14 – Окно результатов построения АРПСС модели

Приведенное окно (рисунок 7.14) характерно для рассматриваемого пакета программ, оно разделено на две половины: в верхней части приведены результаты расчетов ARMA модели, в нижней части приведены вкладки, позволяющие подробнее посмотреть результаты построения модели и оценить прогноз.

Шаг 4. Выберем вкладку *Advanced* и нажмем кнопку *Summary: Parameter estimates* результаты приведем в таблице 7.5.

Таблица 7.5 – Результаты построения ARMA(1,0) модели

	<i>Param.</i>	<i>Asympt. Std. Err.</i>	<i>Asympt. t(25)</i>	<i>p</i>	<i>Lower 95% Conf</i>	<i>Upper 95% Conf</i>
<i>p(1)</i>	-0,911	0,102	-8,895	0,000	-1,122	-0,700

Во втором столбце расположены параметры уравнения в данном случае модель имеет вид:

В третьем столбце расположены асимптотические ошибки.

В четвертом асимптотический *t*-критерий.

Далее *p*-уровень значимости и 95% нижняя (верхняя) границы.

Судя по полученному значению асимптотического *t*-критерия полученная модель статистически значима.

Шаг 5. Для прогнозирования значений вторых разностей в данном модуле необходимо во вкладке *Advanced* обратиться к группе опций *Forecasting*

(Прогнозирование). Допустим необходимо построить прогноз на 4 уровня вперед, для этого установим цифру 4 в опции *Number of cases* (Номер наблюдения). И нажмем кнопку *Forecast cases* (Прогноз значения), в результате получим таблицу 7.6 и рисунок 7.15.

Таблица 7.6 - Прогнозные значения вторых разностей денежного агрегата М0 на 2006г.

	<i>Forecast</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Std.Err.</i>
29	-162,614	-291,707	-33,522	62,680
30	148,142	-26,487	322,772	84,791
31	-134,958	-339,833	69,917	99,476
32	122,947	-103,989	349,884	110,188

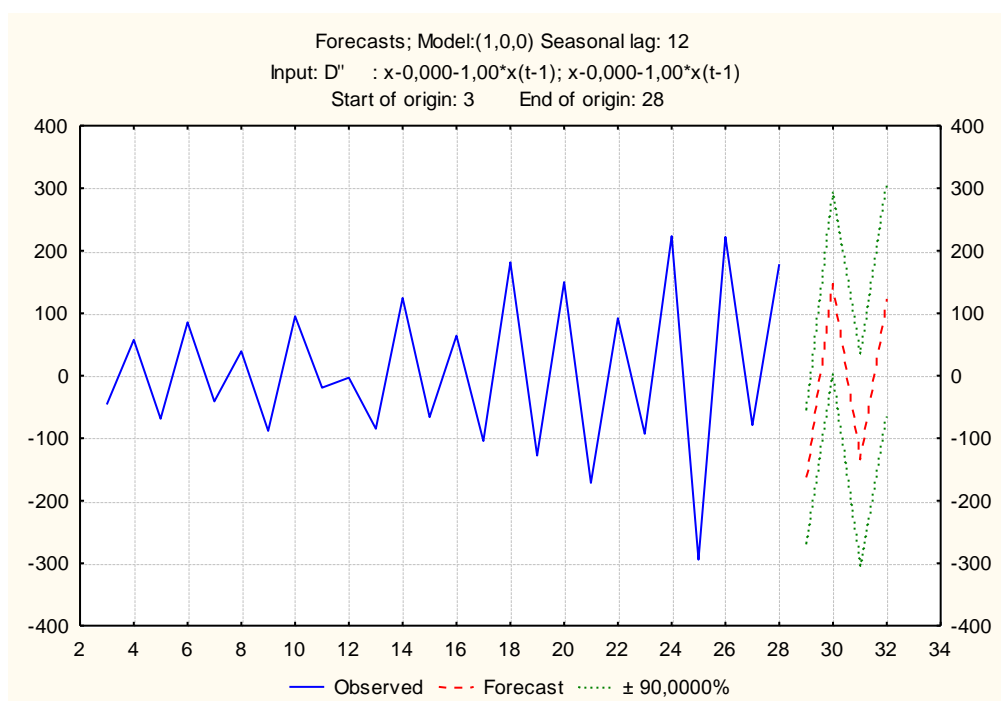


Рисунок 7.15 – Прогноз вторых разностей по модели ARMA (1,0)

7.8 Тесты для самоконтроля

1 Если автокорреляционная функция (АКФ) выказывает выброс на первом лаге, а частная автокорреляционная функция (ЧАКФ) экспоненциально затухает, то можно предположить что ряд наиболее адекватно опишет модель в форме:

- ARMA(0,2)
- ARMA(0,1)
- ARMA(1,0)

2 Если АКФ экспоненциально затухает, а ЧАКФ обнаруживает выброс на первом лаге, то можно предположить что ряд наиболее адекватно опишет модель в форме:

- ARMA(0,2)
- ARMA(0,1)
- ARMA(1,0)

3 Если АКФ экспоненциально затухает, а ЧАКФ обнаруживает выброс на первом и втором лаге, то можно предположить что ряд наиболее адекватно опишет модель в форме:

- ARMA(2,0)
- ARMA(0,1)
- ARMA(1,0)

4 Если АКФ обнаруживает выбросы (пики) на лагах 1 и 2, а ЧАКФ экспоненциально затухает, то можно предположить что ряд наиболее адекватно опишет модель в форме:

- ARMA(0,2)
- ARMA(0,1)
- ARMA(1,0)

5 Приведенная модель $y_t = \alpha y_{t-1} + \varepsilon_t$ является:

- AR(1)
- AR(2)
- MA(1)
- MA(2)

6 Приведенная модель $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t$ является:

- AR(1)
- AR(2)
- MA(1)
- MA(2)

7 Приведенная модель $y_t = \varepsilon_t + \theta \varepsilon_{t-1}$ является:

- AR(1)
- AR(2)
- MA(1)
- MA(2)

8 Приведенная модель $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$ является:

- AR(1)
- AR(2)
- MA(1)
- MA(2)

9 Приведенная модель $y_t = \alpha_1 y_t + \varepsilon_t + \theta \varepsilon_{t-1}$ является:

- ARIMA (1, 1, 1)
- ARMA (1, 1)
- ARMA (2, 2)

10 Марковским процессом называют модель вида:

- $y_t = \alpha y_{t-1} + \varepsilon_t$
- $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t$
- $y_t = \alpha_1 y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$

7.9 Задания для самостоятельного выполнения

Задания для самостоятельной работы составлены в пяти вариантах, номер варианта выбирается в соответствии с последней цифрой зачетной книжки студента:

Таблица 7.6 – Задания для самостоятельной работы по теме «Анализ и прогнозирование временных рядов на основе моделей ARMA»

Последняя цифра номера зачетной книжки	1 и 6	2 и 7	3 и 8	4 и 9	5 и 0
Номер варианта	1	2	3	4	5

Используя данные соответствующего варианта (приложение 13) необходимо выполнить следующие этапы самостоятельной работы:

- 1 с помощью соответствующих тестов определить стационарность исследуемого ряда;
- 2 привести анализируемый временной ряд к стационарному виду используя метод взятия разностей или переход к отклонениям от тренда;
- 3 определить параметры p и q используя автокорреляционную и частную автокорреляционную функцию;
- 4 оценить ARMA модель построить прогноз.

8 Выявление и анализ сезонности во временных рядах

8.1 Цели и задачи лабораторной работы

Используя фактические данные динамического ряда среднедушевых денежных доходов населения (приложение 14) выявим сезонную составляющую, при этом выделим следующие задачи:

- 1 Для выявления сезонности необходимо провести визуализацию ряда;
- 2 Проведем выявление сезонности методом *Seasonal decomposition 1*;
- 3 Проведем выявление сезонности методом *X11/Y2k (Census 2) – monthly*

8.2 Понятие сезонности и методы ее выявления

Иногда в экономических исследованиях возникает потребность в сезонной корректировке ряда с целью выявления основной тенденции и построения адекватной модели. Так как если демонический ряд содержит сезонную составляющую, то коэффициент вариации будет значителен и модель, построенная на основе этого ряда, не будет адекватно отражать направление развития изучаемого явления или процесса.

В качестве еще одного случая, в котором необходимо изучать сезонность является потребность в моделировании сезонной составляющей с целью прогнозирования выделенной волны. Этот случай возникает при изучении таких показателей как надой молока или привес животных в течение года и т.д.

В экономике чаще всего встречаются два вида сезонной составляющей динамического ряда, это детерминированная сезонная волна и эволюционирующая сезонная волна. Как правило, первый тип встречается в стабильно развивающихся экономиках, а второй характерен для экономики переходного периода, т.к. в результате трансформации экономических механизмов, как правило, происходят изменения в механизме генерации динамического ряда.

Методы выделения сезонной составляющей в соответствии с типами сезонной волны можно также разделить на две группы. Для выявления детерминированной сезонной волны разработано большое количество алгоритмов, самыми распространенными из которых являются:

- исчисление индексов сезонности;
- сезонная декомпозиция временного ряда;
- фиктивные переменные;
- преобразование Фурье;

В качестве алгоритмов выявления эволюционирующей сезонной волны можно назвать следующие методики:

- процедура X-12-ARIMA является расширенной версией процедуры сезонной корректировке X-11 разработанной Бюро переписей США. Эта процедура содержит множество усовершенствований, к ее достоинствам относятся: возможность учитывать различное влияние дней недели, отдельных месяцев, улавливать сложные календарные составляющих, возможность корректировки экстремальных наблюдений, более широкий выбор способов вычисления скользящих средних и другие.
- процедура ES (Extract Seasons) рассматриваемая Губановым В.А. в его работе. Суть предложенной методики заключается в применении непараметрического алгоритма сезонной корректировки временных рядов основанному на использовании вариационных принципов.
- процедура TRAMO/SEATS. Данная процедура основывается на ARIMA моделях, была разработана Maravall и Gomez и реализована в программе Burman (1980).
- BV4 - метод сезонной корректировки основанный на скользящем фильтре методом регрессии. Первоначально был разработан в Берлинском техническом университете. В настоящее время является официальным методом сезонной корректировки Центрального Статистического Офиса Германии. Данный метод способен выделить такие составляющее временного ряда как: тренд-циклическую, сезонную, календарную, нерегулярную.

Обобщая накопленный опыт схему выявления сезонной составляющей в динамическом ряду в общем виде можно представить следующим образом (рисунок 8.1). В пакете STATISTICA существует несколько вариантов выявления сезонной составляющей.

Во-первых, это сезонная корректировка в *Seasonal decomposition 1* (Сезонная декомпозиция 1).

Во-вторых, это использование возможностей модуля *X11 Seasonal decomposition* (X11 Сезонная декомпозиция).

В-третьих, использование фиктивных переменных для выявления сезонности в модуле *Multiple Regression* (Множественная Регрессия) (метод будет рассмотрен в главе 9).

В-четвертых, описание сезонности с использованием адаптивных методов – экспоненциального сглаживания и АРПСС моделей (методы рассмотрены в главе 6 и 7).



Рисунок 8.1 – Схема выявления сезонной составляющей в динамическом ряду

8.3 Рекомендуемая литература

Для лучшего понимания материала изложенного в данной главе необходимо дополнительно проанализировать следующие источники литературы (см. список использованных источников):

1. Афанасьев, В.Н. Анализ временных рядов и прогнозирование: учеб. для вузов / В.Н. Афанасьев, М.М. Юзбашев. - М. : Финансы и статистика, 2001. - 228с. : ил. - ISBN 5-279-02419-8. (см. стр. 69-77)
2. Вуколов, Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: Учебное пособие. – М.: ФОРУМ: ИНФРА-М, 2004. – 464с. - ISBN 5-8199-0141-X, 5-16-002003-9. (см. стр. 255-283)
3. Дуброва, Т.А. Статистические методы прогнозирования: учеб. пособие для вузов. – М.: ЮНИТИ-ДАНА, 2003. – 206 с. ISBN 5-238-00497-4. (см. стр. 84-113)
4. Халафян А.А. STATISTICA 6. Статистический анализ данных. 3-е изд. Учебник – М.: ООО «Бином-Пресс», 2007 г. – 512 с. ISBN 978-5-9518-0215-6. (см. стр. 469-497)
5. Эконометрика: учебник для вузов / под ред. И.И. Елисеевой. - М. : Финансы и статистика, 2005. - 576 с. : ил. - ISBN 5-279-02786-3. (см. стр. 311-312)

8.4 Сезонная корректировка в модуле *Seasonal decomposition 1*

В качестве исходной информации для иллюстрации метода *Seasonal decomposition 1* (сезонная корректировка) воспользуемся временным рядом среднедушевых денежных доходов населения (р. в мес.) за период с I квартала 1993г. по VI квартал 2005г. (приложение 14).

Для начала представим данные в графическом виде, для этого в главном меню выберем *Graphs* → *2D Graphs* → *Line Plots (Variables)*... в результате получим следующий рисунок.

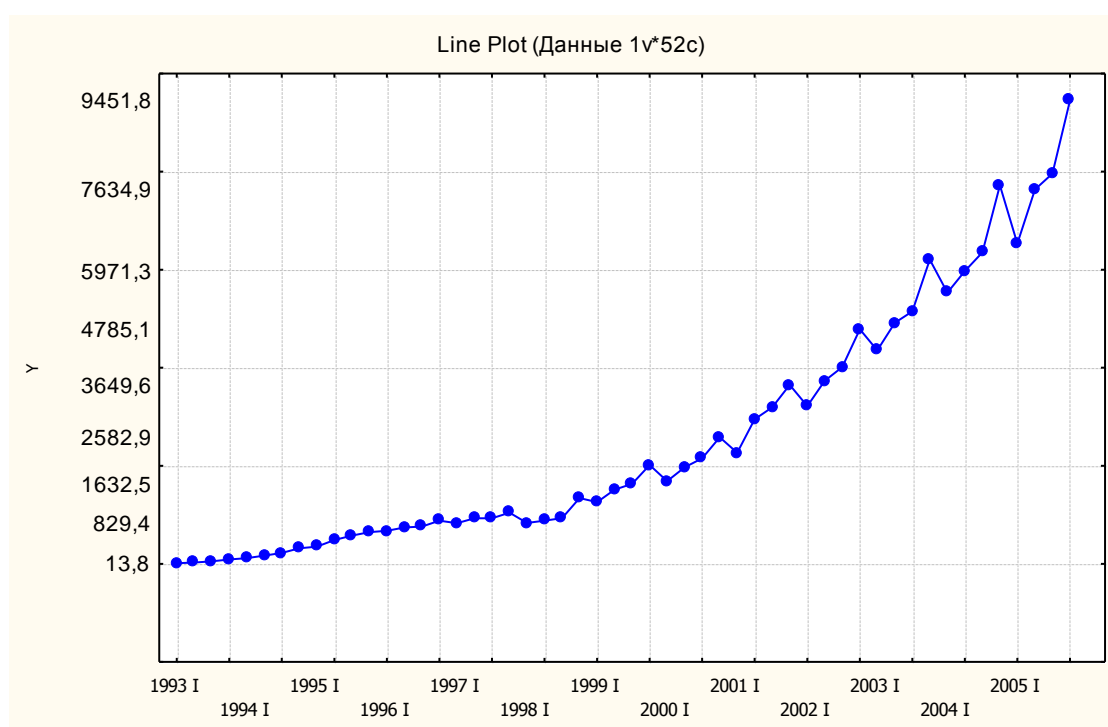


Рисунок 8.2 – Динамика среднедушевых денежных доходов населения, р. в мес.

Согласно приведенному рисунку в рассматриваемом динамическом ряду присутствует сезонная и трендовая составляющая. Также можно сделать предположение, что существуют структурные изменения в механизме генерирования ряда, так как до 1998 года не прослеживается сезонности, начиная с 1999 г. имеет место пик в четвертом квартале каждого года. Поэтому для дальнейшего анализа будем использовать данные за период I кв. 1999г. – VI кв. 2005г.

Шаг 1. В главном меню выберем *Statistics* → *Advanced Linear/Nonlinear Models* → *Time series/Frication*.

Шаг 2. В окне *Time Series Analysis* необходимо выбрать кнопку *Seasonal decomposition 1 (Census 1)*, после чего перейдем в окно *Ratios-To-Moving Averages Classical Seasonal Decomposition (Census Method)* (Классическая сезонная корректировка).

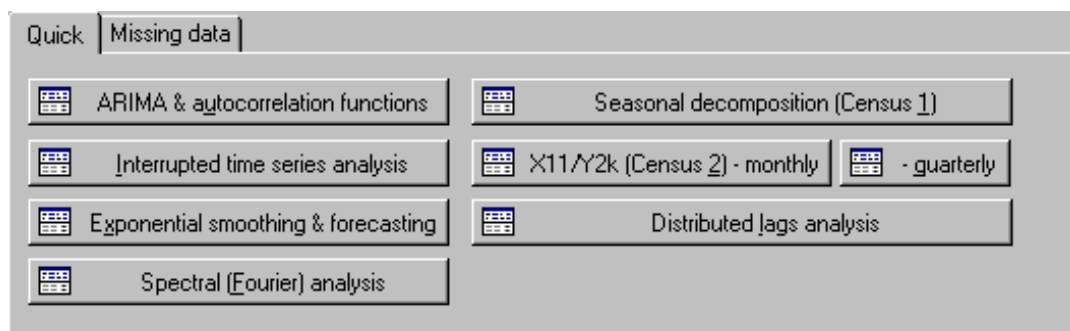


Рисунок 8.3 – Окно выбора методов анализа временных рядов (приведена часть исходного окна)

В данном окне необходимо выбрать вкладку *Advanced* (Расширенные), появится доступ к следующим опциям (рисунок 8.4):

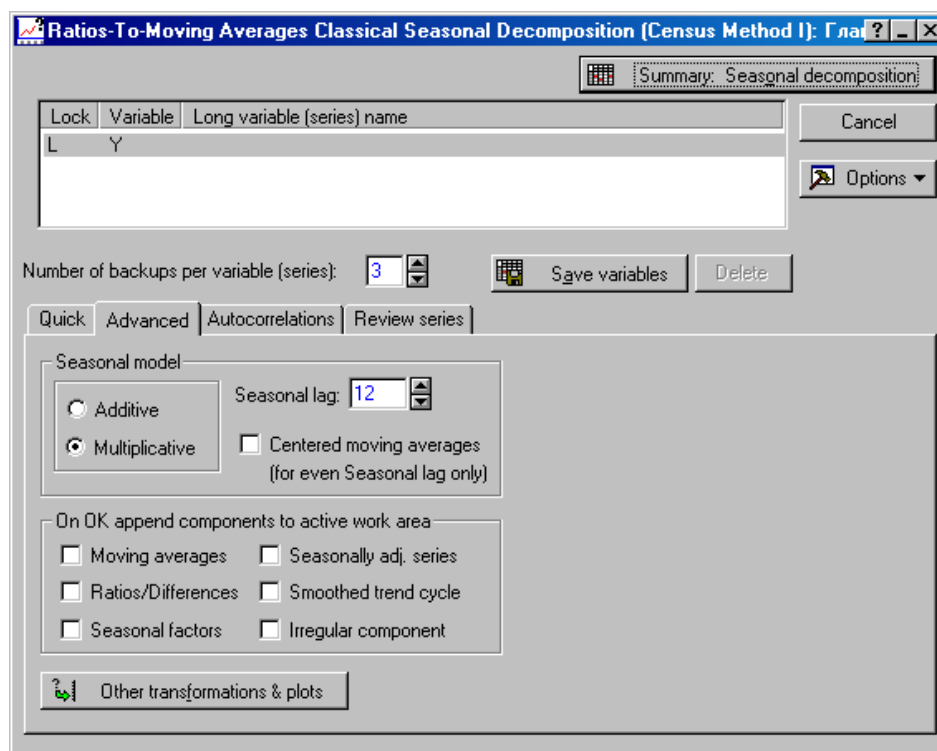


Рисунок 8.4 – Установки для проведения процедуры сезонной корректировки

Группа *Seasonal model* (Сезонная модель) позволяет представить модель в *Additive* (Аддитивная) или *Multiplicative* (Мультипликативная) форме.

Опция *Seasonal lag* (Сезонный лаг) позволяет задавать длину сезонного лага (12 месяцев или 4 квартала).

Если поставить галочку в опции *Centered moving averages (for even Seasonal lag only)* (Центрированная скользящая средняя (только для Сезонного лага))

Установив флажок около одной из опций *On OK append components to active work area* сможем в дальнейшем сохранить и использовать соответствующие компоненты динамического ряда.

Moving averages – Скользящая средняя

Ratios/Differences – Отношения/Различия

Seasonal factor – Сезонная составляющая

Seasonally adj. series – Ряд с исключенной сезонной составляющей

Smoothed trend cycle – Сглаженная тренд-циклическая компонента

Irregular component – Нерегулярная составляющая

Опция *Other transformations & plots* (Другие преобразования и графики) позволяет провести предварительный анализ исследуемого ряда, например, выделить из ряда тренд составляющую, или среднюю, или провести процедуру сглаживания ряда с помощью различных скользящих средних и т.д.

Шаг 3. Для анализа временного ряда среднедушевых денежных доходов населения выберем мультипликативную (*Multiplicative*) модель.

В мультипликативной модели исходный временной ряд Y представляется в виде [боровиков]:

$$Y(t) = S(t) \cdot TC(t) \cdot I(t), \quad t = 0, 1, 2, \dots$$

Последовательность вычислений в классической сезонной декомпозиции следующая:

$Y \rightarrow$ сглаженный $Y - Y_I \rightarrow$ отношение $Y/Y_I \rightarrow$ оценка сезонной компоненты - ES (усредненное на сезоне отношение Y/Y_I) \rightarrow поправка на сезонную компоненту исходного ряда: $Y/ES \rightarrow$ оценка тренд-циклической компоненты ETC (сглаживание Y/ES) \rightarrow выделение нерегулярной составляющей: $I = (Y/ES)/ETC$.

Для аддитивной модели выполняется та же последовательность действий с заменой умножения на сложение, деления на вычитание.

На данном шаге необходимо в опции *Seasonal lag* указать число 4, установить флажки возле опций *Seasonal factor*, *Smoothed trend cycle* и нажать кнопку *Summary: Seasonal decomposition* (Провести процедуру сезонной корректировки).

В результате проведения данной процедуры в пакете STATISTICA получаем таблицу с результатами (рисунок 8.5).

Case	Y	Moving Averages	Ratios	Seasonal Factors	Adjusted Series	Smoothed Trend-c.	Irreg. Compon.
1	13,8			92,185	14,970	17,740	0,844
2	26,1			99,544	26,220	30,686	0,854
3	50,0	44,800	111,607	98,290	50,870	56,580	0,899
4	89,3	70,575	126,532	109,981	81,196	88,316	0,919
5	116,9	106,275	109,998	92,185	126,810	128,902	0,984
6	168,9	149,575	112,920	99,544	169,673	176,188	0,963
7	223,2	206,350	108,166	98,290	227,083	233,798	0,971
8	316,4	265,025	119,385	109,981	287,687	304,148	0,946
9	351,6	344,150	102,165	92,185	381,406	389,726	0,979
10	485,4	432,250	112,296	99,544	487,623	474,754	1,027
11	575,6	514,875	111,794	98,290	585,613	557,999	1,049
12	646,9	595,700	108,595	109,981	588,195	626,007	0,940
13	674,9	660,100	102,242	92,185	732,113	691,766	1,058

Рисунок 8.5 – Результаты проведения сезонной корректировки (приведена часть исходного окна)

Рассмотрим результаты корректировки по столбцам: Y – приведены значения исходного динамического ряда; *Moving averages* (Сглаженные данные); *Ratios*; *Seasonal factor* (Сезонная компонента); *Adjusted Series*; *Smoothed trend cycle* (Сглаженная тренд-циклическая составляющая); *Irregular component* (Не регулярная составляющая).

Шаг 4. Вернемся в исходное окно сезонной корректировки (рисунок 8.4) и нажмем кнопку *Save variables* (Сохранить переменные), получим следующую таблицу:

	1	2	3
	Y	Y_2	Y_1
1	1259,1	91,035	1457,802
2	1514,7	99,044	1523,200
3	1632,5	98,510	1653,996
4	2015,6	111,411	1774,320
5	1688,1	91,035	1889,035
6	1970,0	99,044	2019,193
7	2152,2	98,510	2168,237

Рисунок 8.6 – Таблица с сохраненными результатами сезонной корректировки (приведена часть исходного окна)

Соответственно: Y – исходный временной ряд, Y_2 – сезонная составляющая динамического ряда, Y_1 – тренд-циклическая составляющая.

Шаг 5. Представим результаты корректировки графически, для этого выберем *Graphs* → *2D Graphs* → *Line Plots (Variables)*... Полученный рисунок является не информативным, так как Y_2 в данном представлении является прямой линией.

Шаг 6. Для улучшения наглядности графика щелкнем правой кнопкой мышки на данной составляющей, в появившемся меню (рисунок 8.7) выберем пункт *Plot properties* ...

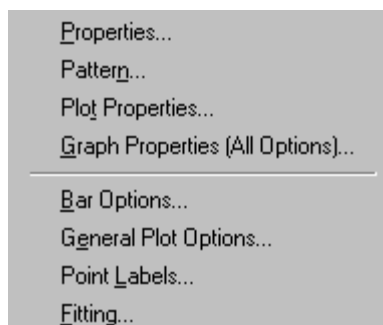


Рисунок 8.7 – Меню установок графика

Шаг 7. В окне *Plot Layout* в группе опций *Assignment of axis* установим флажок возле опции *Right* (Правая) тем самым поместим отображение данной компоненты на дополнительную (правую) часть оси *OY*.

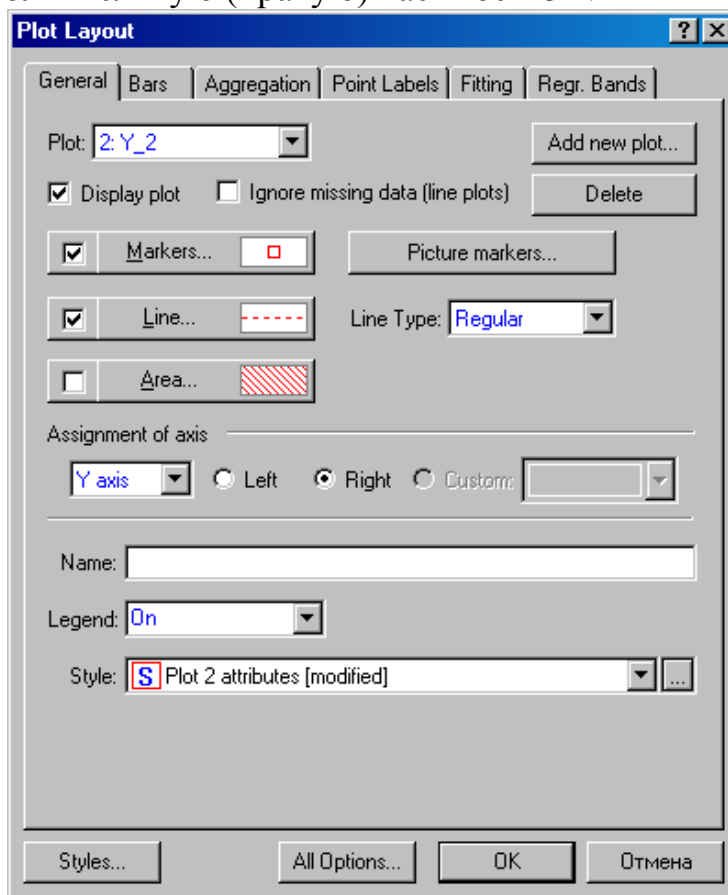


Рисунок 8.8 – Окно настроек графика

После нажатия кнопки ОК получаем следующие результаты:

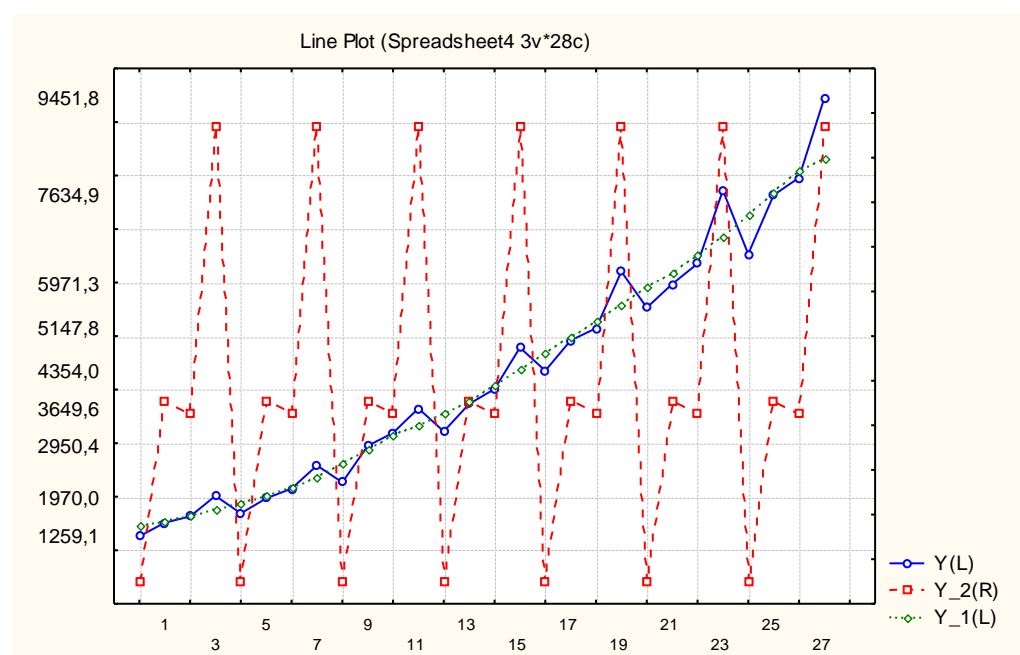


Рисунок 8.9 – Результаты сезонной корректировки ряда среднедушевых доходов методом сезонной декомпозиции

Как видим на рисунке 8.9, рассмотренная процедура выделила детерминированную сезонную волну с максимумом в четвертом квартале каждого года и минимумом в первом.

8.5 Сезонная корректировка в модуле *X11/Y2k (Census 2) – monthly*

Шаг 1. В данной работе рассматривается динамический ряд, состоящий из квартальных данных, поэтому в окне *Time Series Analysis* (рисунок 8.3) вместо кнопки *X11/Y2k (Census 2) – monthly* (которая запускает месячную сезонную корректировку) необходимо выбрать кнопку *quarterly* (запускающая квартальную сезонную корректировку).

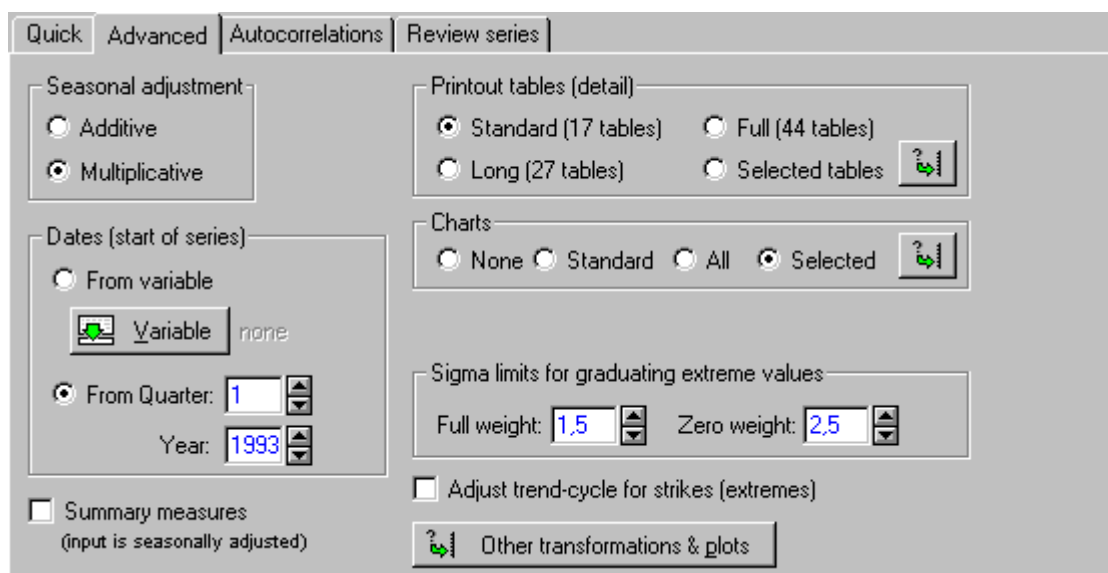


Рисунок 8.10 – Окно выбора варианта итоговых таблиц и графиков сезонной корректировки (приведена часть исходного окна)

Согласно приведенному рисунку стартовое окно модуля содержит множество опций и вкладок.


Укажем тип сезонной волны (*Seasonal adjustment*), при этом доступны два варианта аддитивная (*Additive*) и мультипликативная (*Multiplicative*), в нашем случае выберем второй тип.

В следующей группе опций – *Dates (start of series)* (Дата начала временного ряда) предлагается указать начало отсчета временного ряда, в нашем случае 1 квартал 1993 года.

В соответствии со стандартом метода *X-11*, принятым Бюро переписи США, предусмотрены три степени подробности вывода (*Printout tables (detail)*): *Standard (17 to 27)* (Стандартный (17 - 27 таблиц)), *Long (27 to 39)* (Длинный (27

- 39 таблиц)), *Full (44 to 59)* (Полный (44 - 59 таблиц)), а также *Selected tables* (Выбор таблиц).

Также в данном окне можно выбрать количество выводимых графиков, для этого в предназначена группа опций *Charts* (Графики): *None* (Отсутствуют), *Standard* (Стандартное количество), *All* (Все), *Selected* (Выборочные).

Шаг 2. Если необходимо выбрать необходимые графики (например, только итоговые или только характеризующие сезонность) то необходимо нажать кнопку , появится следующее окно (рисунок 8.11).

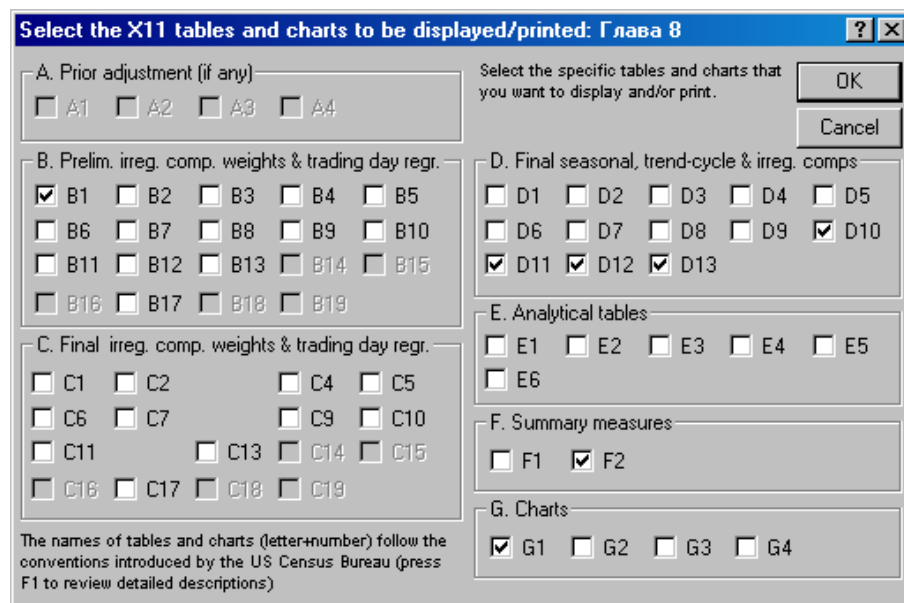


Рисунок 8.11 – Варианты вывода итоговых таблиц и графиков

Процедура корректировки разбивается на семь этапов, которые обычно обозначаются буквами А - G. [см. справочник]

А. Априорная корректировка (помесячная сезонная корректировка). Перед тем, как к временному ряду, содержащему ежемесячные значения, будет применяться какая-либо сезонная корректировка, могут быть произведены различные корректировки, заданные пользователем. Можно ввести еще один временной ряд, содержащий априорные корректирующие факторы; значения этого ряда будут вычитаться из исходного ряда (аддитивная модель), или же значения исходного ряда будут поделены на значения корректирующего ряда (мультипликативная модель). В случае мультипликативной модели пользователь может также определить свои собственные поправочные коэффициенты (веса) на число рабочих дней. Эти веса будут использоваться для корректировки ежемесячных наблюдений, так чтобы учитывалось число рабочих дней в этом месяце.

В. Предварительное оценивание вариации числа рабочих дней (месячный вариант X-11) и весов. На следующем шаге вычисляются

предварительные поправочные коэффициенты на число рабочих дней (только в месячном варианте X-11) и веса, позволяющие уменьшить эффект выбросов.

С. Окончательное оценивание вариации числа рабочих дней и нерегулярных весов (месячный вариант X-11). Поправки и веса, вычисленные в пункте В, используются для построения улучшенных оценок тренд-циклической и сезонной компонент. Эти улучшенные оценки используются для окончательного вычисления факторов числа рабочих дней (в месячном варианте X-11) и весов.

Д. Окончательное оценивание сезонных факторов, тренд-циклической, нерегулярной и сезонно скорректированной компонент ряда. Окончательные значения факторов рабочих дней и весов, вычисленные в пункте С, используются для вычисления окончательных оценок для компонент ряда.

Е. Модифицированные ряды: исходный, сезонно скорректированный и нерегулярный. Исходный и окончательный сезонно скорректированный ряды, а также нерегулярная компонента модифицируются путем сглаживания выбросов. Полученные в результате этого, модифицированные ряды позволяют пользователю проверить устойчивость сезонной корректировки.

Ф. Месяц (квартал) циклического доминирования (МЦД, КЦД), скользящее среднее и сводные показатели. На этом этапе вычислений рассчитываются различные сводные характеристики, позволяющие пользователю исследовать относительную важность разных компонент, среднюю флуктуацию от месяца к месяцу (от квартала к кварталу), среднее число идущих подряд изменений в одну сторону и др.

Г. Графики. Данная группа опций позволяет строить различные графики итоговых результатов. Например, можно построить окончательно скорректированный ряд в хронологическом порядке или по месяцам.

В нашем случае ограничимся графиками представленными в пункте Г.

В результате проведения процедуры получаем следующее:

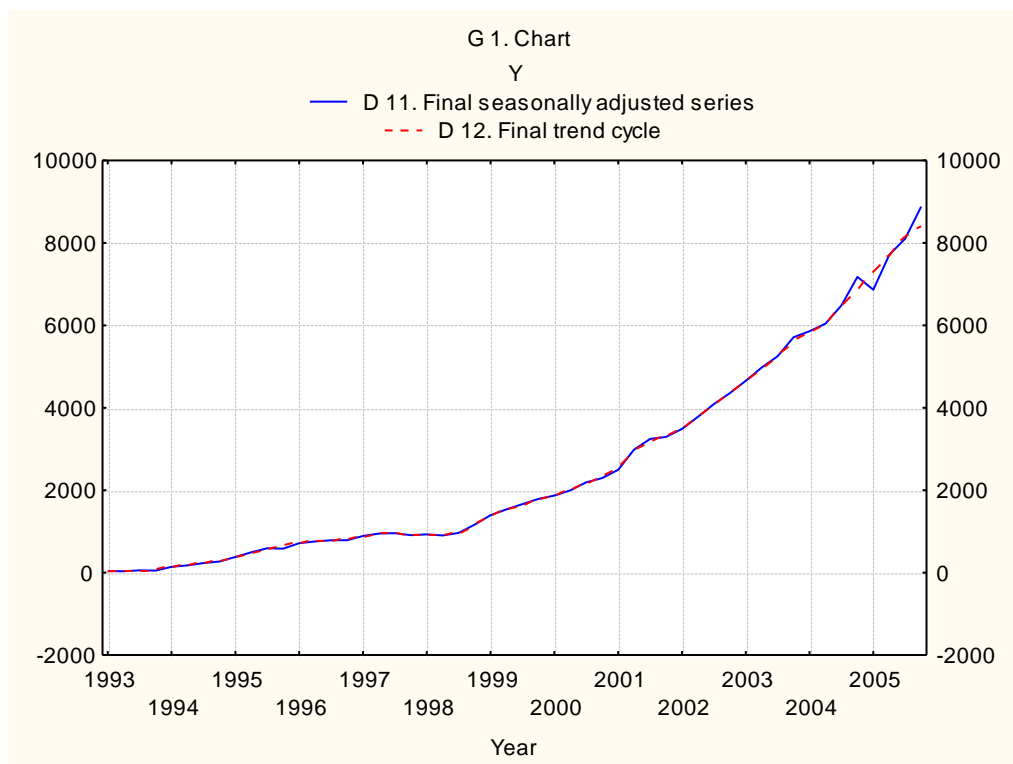


Рисунок 8.12 – Скорректированный ряд и тренд-циклическая составляющая

Согласно приведенному рисунку ряд среднедушевых денежных доходов до 1999 года развивается с одной скоростью (незначительный наклон тренда), а после данной точки скорость роста увеличивается.

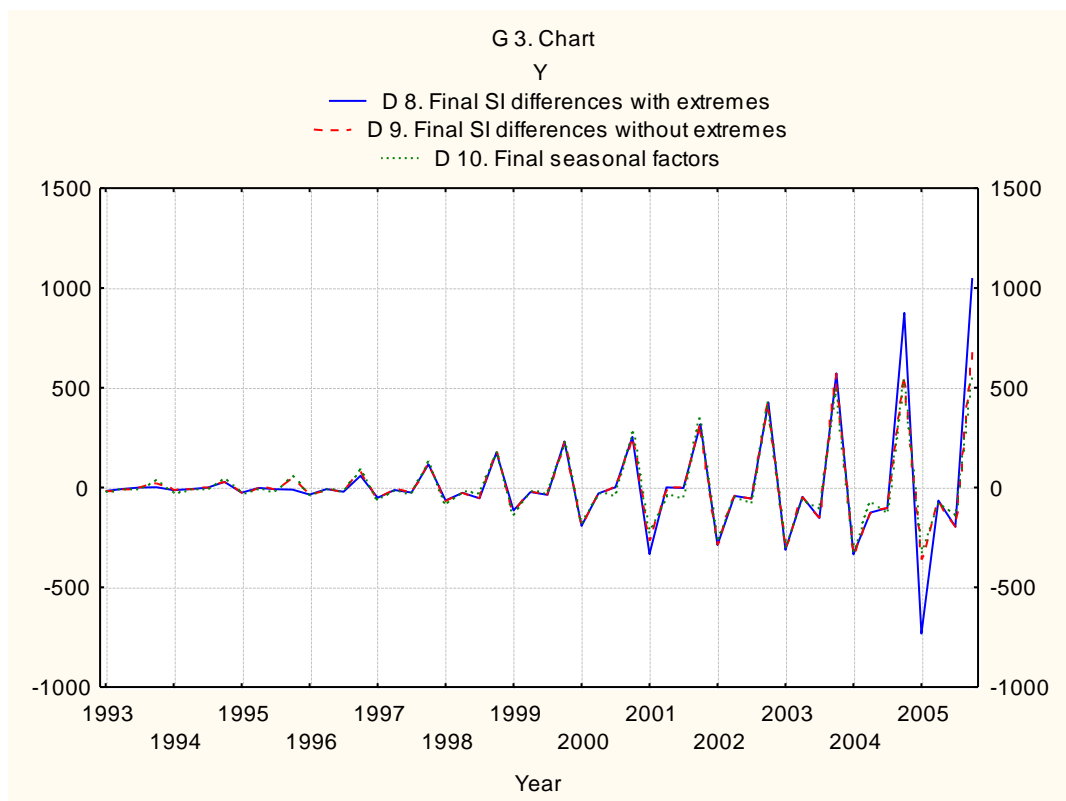


Рисунок 8.13 – Выделенная сезонная составляющая

Согласно графику выделенной сезонной волны (составляющей динамического ряда) с развитием наблюдаемого ряда прослеживается усиление сезонной составляющей, т.е. можно сделать вывод о присутствии во временном ряду эволюционирующей сезонной составляющей.

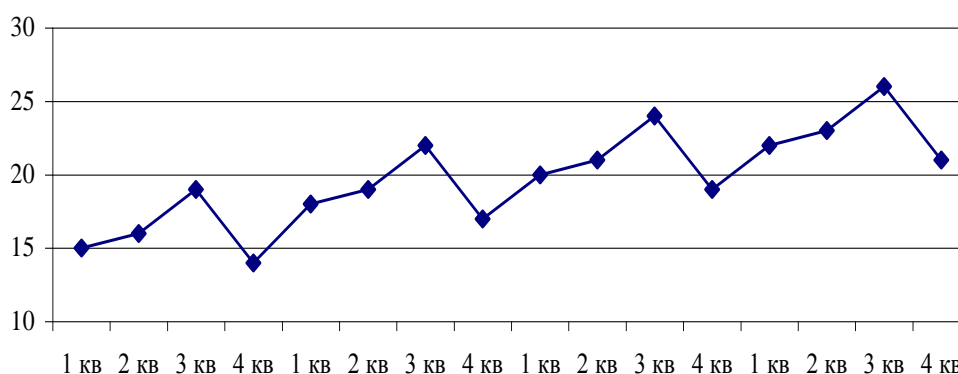
8.6 Тесты для самоконтроля

1 Чему равно значение индекса сезонности в приведенной таблице:

	1 квартал	2 квартал	3 квартал	4 квартал
Индекс сезонности	?	1,20	0,95	0,79

- 1,09;
- 1,06;
- 1,25;
- 0,95.

2 Если график динамического ряда имеет приведенный вид, то это указывает на присутствие:



- аддитивную сезонность;
- мультипликативную сезонность;
- кратная сезонность;
- отсутствие сезонности.

3 Если график динамического ряда имеет вид, приведенный на рисунке, то это указывает на присутствие:



- детерминирующей сезонной волны;
- эволюционирующей сезонной волны;
- отсутствие сезонной волны.

4 Метод Х12 используется для выявления:

- детерминирующей сезонной составляющей;
- эволюционирующей сезонной составляющей;
- тренд составляющей ряда.

5 С помощью метода фиктивных переменных можно выявить:

- детерминирующей сезонной составляющей;
- эволюционирующей сезонной составляющей;
- случайную составляющую.

6 Приведенная формула $\tilde{y}_t = a + \sum_{k=1}^m b_{1k} \cos kt_i + b_{2k} \sin kt_i$ необходима для расчета:

- тренда;
- описания сезонности гармониками Фурье;
- оценки сезонности с помощью фиктивных переменных.

7 Приведенная формула $\frac{\bar{y}_t}{\bar{y}}$ необходима для расчета?

- индекс сезонности;
- индекс колеблемости;
- индекс цен.

8.7 Задания для самостоятельного выполнения

Задания для самостоятельной работы составлены в пяти вариантах, номер варианта выбирается в соответствии с последней цифрой зачетной книжки студента:

Таблица 8.1 – Задания для самостоятельной работы по теме «Выявление и анализ сезонности во временных рядах»

Последняя цифра номера зачетной книжки	1 и 6	2 и 7	3 и 8	4 и 9	5 и 0
Номер варианта	1	2	3	4	5

Используя данные соответствующего варианта (приложение 15) необходимо выполнить следующие этапы самостоятельной работы:

- 1 Провести визуализацию ряда если прослеживается сезонность, то оценить данную составляющую используя возможности модуля *Seasonal decomposition 1*;
- 2 Выделить сезонную составляющую используя возможности модуля *X11/Y2k (Census 2) – monthly*.

9 Построение регрессионных моделей с фиктивными переменными

9.1 Цели и задачи лабораторной работы

В данной лабораторной работе, на основе фактического материала, рассмотрим способы применения фиктивных переменных в регрессионных моделях, при этом выделим следующие задачи:

- 1 на основе данных о ценах на квартиры, оценить регрессионную модель с фиктивными переменными;
- 2 рассмотреть способ применения фиктивных переменных для описания сезонности;
- 3 рассмотреть использование фиктивных переменных для описания структурных сдвигов в макроэкономических рядах.

9.2 Понятие фиктивных переменных и их применение в пространственных и временных моделях

Факторы, применяемые в регрессионных задачах, обычно могут принимать значения из какого-либо непрерывного интервала. Но часто случается так, что отдельные факторы, которые необходимо (или хотелось бы) ввести в регрессионную модель, являются качественными по своей природе и, следовательно, не измеряются в числовой шкале. В качестве примеров можно привести следующие случаи:

- 1 Исследуется зависимость между продолжительностью полученного образования и доходом, и в выборке представлены лица как мужского, так и женского пола. Нужно выяснить, обуславливает ли пол различие в результатах.
- 2 Исследуется зависимость между доходом и потреблением в Бельгии, и выборка включает как франкоговорящие семьи, так и семьи, говорящие по-фламандски. Нужно выяснить, имеет ли существенное значение это этническое различие.
- 3 Исследуются факторы, определяющие инфляцию, и в некоторые годы периода наблюдений правительство проводило политику регулирования доходов. Нужно проверить, оказало ли это какое-либо влияние на исследуемую зависимость.

В каждом из этих примеров одним из возможных решений было бы оценивание отдельных регрессий для двух указанных категорий с последующим выяснением, различаются ли полученные коэффициенты. Другой возможный подход к решению состоит в оценивании единой регрессии с использованием всей совокупности наблюдений и измерением степени влияния качественного фактора посредством введения так называемой фиктивной переменной (манекены, дамми переменные, в отечественной литературе

структурные переменные). Второй подход обладает двумя важными преимуществами: во-первых, имеется простой способ проверки, является ли воздействие качественного фактора значимым; во-вторых, при условии выполнения определенных предположений регрессионные оценки оказываются более эффективными.

В качестве фиктивных переменных обычно используются дихотомические (бинарные, булевы) переменные которые принимают всего два значения 0 и 1. Обычно фиктивные переменные обозначаются как Z_i (или D_i).

Следует отметить не совсем удачный перевод на русский язык термина *dummy variables* как «фиктивные» переменные. Во-первых, в модели регрессионного анализа уже имеет фиктивная переменная x_0 при α_0 , всегда равная единицы. Во-вторых все процедуры регрессионного анализа проводятся при включении фиктивных переменных так же как и обычных количественных переменных. «Фиктивность» же переменных Z_i состоит только в том, что они количественным образом описывают качественный признак.

Вводя в уравнение регрессии фиктивную переменную и найдя параметры уравнения следующим шагом проверяют H_0 о равенстве коэффициента при $Z_i=0$, этим устанавливается существенность влияния фактора отражаемого фиктивной переменной на Y .

Если рассматриваемый качественный признак имеет несколько (k) уровней (градаций), то в принципе можно ввести в регрессионное уравнение дискретную переменную, принимающую такое же количественное значение. Однако так не поступают из-за трудности содержательной интерпретации соответствующих коэффициентов регрессии, а вводят ($k-1$) бинарных переменных.

Вводить бинарную переменную Z_k нельзя, так как при этом для любой i -го наблюдения $Z_1+Z_2+\dots+Z_k=1$, т.е. при суммировании элементов столбцов общей матрицы, соответствующих фиктивным переменным Z_1, Z_2, \dots, Z_k получим столбец, состоящий из одних единиц. А так как в матрице такой столбец из единиц уже есть, то это означает линейную зависимость значений (столбцов) общей матрицы X , т.е. нарушилось одна из предпосылок МНК. Таким образом, возникает проблема мультиколлинеарности и как следствие – невозможности получения не смещенных (достоверных) оценок МНК. Такая ситуация, когда сумма значений несколько переменных, включенных в регрессию, равна постоянному числу (единице) получила название «*dummy trap*» или «ловушка». Чтобы избежать такие ловушки, число вводимых переменных должно быть на единицу меньше числа уровней (градации) качественного признака.

9.3 Рекомендуемая литература

Для лучшего понимания материала изложенного в данной главе необходимо дополнительно проанализировать следующие источники литературы (см. список использованных источников):

- 1 Бородич, С.А. Эконометрика: учеб. пособие / С.А. Бородич. – 3 –е изд., стер. – Мн.: Новое знание, 2006. – 408 с. ISBN 985-475-206-2 (см. стр. 285-310)
- 2 Доугерти, К. Введение в эконометрику: Учеб. для вузов / К. Доугерти.- 2-е изд. - М. : ИНФРА-М, 2004. - 432 с. - (Университетский учебник). - Библиогр.: с. 407. - ISBN 5-16-001463-2. (см. стр. 262-288)
- 3 Кремер, Н.Ш. Эконометрика: учеб. для вузов / Н.Ш. Кремер, Б.А. Путко. - М. : ЮНИТИ-ДАНА, 2003. - 311 с. - ISBN 5-238-00333-1. (см. стр. 115-124)
- 4 Магнус, Я. Р. Эконометрика. Начальный курс: учебник для вузов / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий.- 7-е изд., испр. - М. : Дело, 2005. - 504 с. - ISBN 5-7749-0055-X. (см. стр. 283-316)
- 5 Эконометрика: учебник для вузов / под ред. И.И. Елисеевой. - М. : Финансы и статистика, 2005. - 576 с. : ил. - ISBN 5-279-02786-3. (см. стр. 233-246)

9.4 Использование фиктивных переменных при построении классической регрессии

Построение регрессионной модели с фиктивными переменными не чем не отличается от построения множественной регрессионной модели

Для иллюстрации использования фиктивных переменных в пространственных моделях обратимся к приложению 16, при этом представленный материал содержит две фиктивные переменные:

$D1$ – характеризует этаж квартире, при этом 0 присваивается квартире с первым или последним этажом, 1 в противном случае;

$D2$ – характеризует категорию дома, при этом если дом кирпичный то объекту присваивается 1, цифра 0 в противном случае.

Шаг 1. Запускаем модуль *Multiple regression* далее в окне *Multiple Linear Regression* установим галочку в опции *Review descriptive statistics, correlation matrix*. В окне *Review Descriptive Statistic* необходимо выбрать вкладку *Advanced* и нажать кнопку *Correlations*, в результате чего получаем:

Таблица 9.1 – Матрица парных коэффициентов корреляции

	$X1$	$X2$	$X3$	$X4$	$X5$	$D1$	$D2$	Y
$X1$	1,000	-0,212	-0,203	0,045	-0,344	0,108	-0,677	-0,440
$X2$	-0,212	1,000	0,080	0,010	0,012	-0,281	0,104	-0,012
$X3$	-0,203	0,080	1,000	0,620	0,712	-0,064	0,312	0,743
$X4$	0,045	0,010	0,620	1,000	0,265	0,049	0,005	0,518
$X5$	-0,344	0,012	0,712	0,265	1,000	0,107	0,282	0,562
$D1$	0,108	-0,281	-0,064	0,049	0,107	1,000	-0,075	0,141
$D2$	-0,677	0,104	0,312	0,005	0,282	-0,075	1,000	0,599
Y	-0,440	-0,012	0,743	0,518	0,562	0,141	0,599	1,000

Согласно приведенной таблице получаем, что наибольшее влияние на зависимую переменную Y (см. столбец Y) оказывают показатели $X3$

($r_{X_2Y}=0,743$), X_4 ($r_{X_4Y}=0,518$), X_5 ($r_{X_5Y}=0,562$) и D_2 ($r_{D_2Y}=0,599$). При этом необходимо указать на присутствие мультиколлинеарности в данных.

Шаг 2. Переходим в стартовое окно модуля и устанавливаем галочку в опции *Advanced options*. Выбираем кнопку *Variables* в качестве зависимой переменной указываем Y в качестве независимых указываем X_3 , X_4 , X_5 и D_2 .

Шаг 3. В окне *Model Definition* в прокрутке *Method* выберем *Backward stepwise* (Метод пошагового исключения) и нажмем ОК. Получаем следующие результаты:

Таблица 9.2 – Показатели адекватности множественного уравнения регрессии с фиктивными переменными

	Value (Значение)
<i>Multiple R</i> (Множест. R)	0,837
<i>Multiple R²</i> ((Множест. R ²))	0,701
<i>Adjusted R²</i> (Скорр. R ²)	0,692
<i>F</i> (2,66)	77,444
<i>p</i>	0,000
<i>Std.Err. of Estimate</i> (Стд. Ош. Оценки)	5,573

Таблица 9.3 – Результаты оценки множественной линейной регрессии с фиктивными переменными

	<i>Beta</i> (БЕТА)	<i>Std.Err. of Beta</i> (Стд.Ош. БЕТА)	<i>B</i>	<i>Std.Err. of B</i> (Стд.Ош. B)	<i>t</i> (66)	<i>p-level</i> (p-уров.)
<i>Intercept</i> (Св.член)			-5,480	4,817	-1,138	0,259
X_3	0,616	0,071	1,202	0,138	8,697	0,000
D_2	0,407	0,071	8,944	1,556	5,748	0,000

Согласно данным, приведенным в таблицах 9.2 и 9.3, оцененная модель статистически значима по F -критерию Фишера, при этом $R^2 = 0,701$ и указывает на высокую адекватность модели.

Согласно параметрам уравнения получаем, что при увеличении общей площади квартиры на 1 м^2 цен увеличивается на 1,202 тыс. USD.

Проинтерпретировать параметр при фиктивной переменной можно следующим образом – цена за 1 м^2 в кирпичных домах по сравнению с остальными в среднем выше на 8,94 тыс. USD. Т.е. можно сделать вывод о том, что категория дома оказывает достаточно сильное влияние на стоимость квартиры.

Отообразим на графике линии регрессии квартир в кирпичных домах и

остальных, для этого в исходной таблице образуем две новых переменных $Y1$ и $Y2$. При этом в поле Long name вносим следующие выражения:

- для $Y1$ внесем $=-5,48+1,202 \times v3$ (выровненные значения для квартир не в кирпичных домах)
- для $Y2$ внесем $=-5,48+1,202 \times v3+8,944$ (выровненные значения для квартир в кирпичных домах)

Далее в главном меню *Graphs* → *2D Graphs* → *Scatterplots* в появившемся окне *2D Scatterplots* выберем кнопку *Variables* и укажем в поле X: - X3, а в поле Y: - Y1- Y2. Также в этом окне группе *Graph Type* укажем *Multiple*, получаем следующий результат:

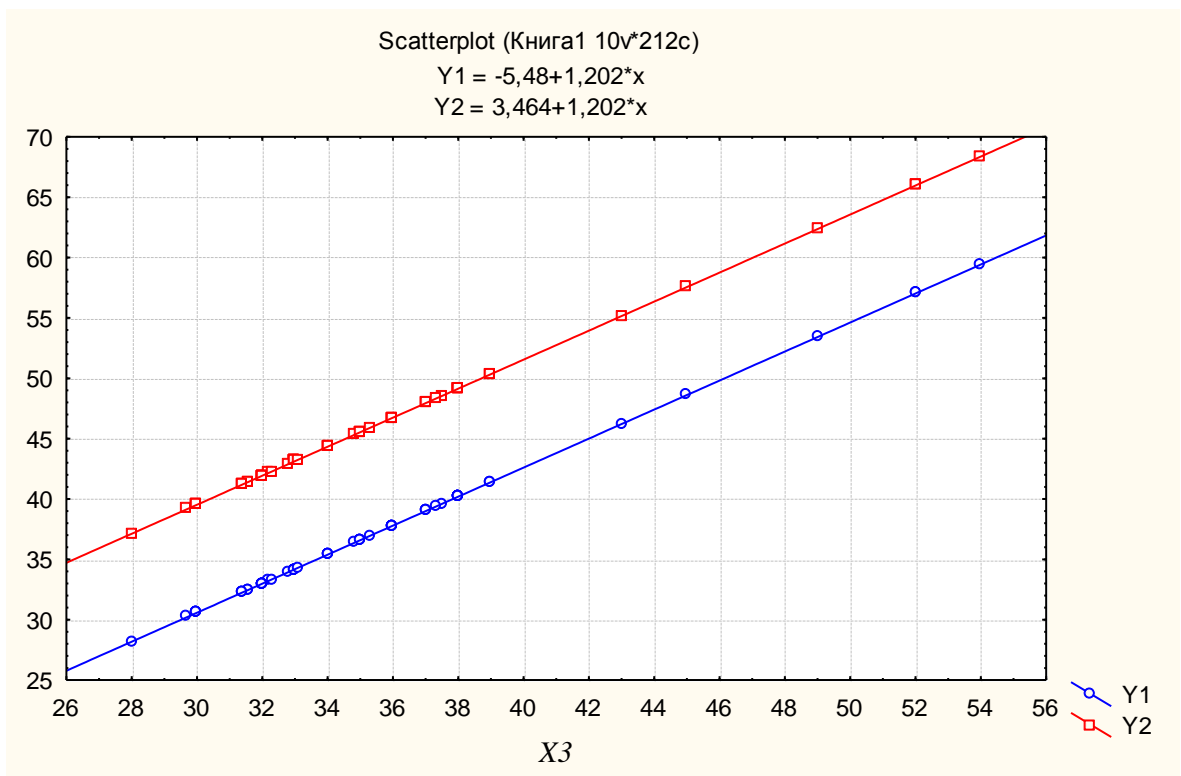


Рисунок 9.1. – Линии регрессии для моделей зависимости цены квартиры от типа дома

Как видим, приведенные уравнения отличаются друг от друга только свободным членом, а линии регрессии параллельны.

9.5 Выявление сезонности с использованием сезонных фиктивных переменных в модуле *Multiple regression*

Для выявления описания сезонных колебаний на практике используют фиктивные переменные. При этом модель имеет следующий вид:

$$\tilde{y}_i = a_0 + a_1 t + c_2 Z_2 + c_3 Z_3 + c_4 Z_4 + \varepsilon_i \quad (9.1)$$

где:

$$Z_i = \begin{cases} 1, & \text{если наблюдение принадлежит } i\text{-му кварталу} \\ 0 & \text{в остальных случаях} \end{cases}$$

a_0, a_1, c_2, c_3, c_4 - коэффициенты модели;

В приведенной формуле 1-й квартал взят в качестве эталонной категории, а фиктивные переменные позволят оценить разницу в уровнях сезонности между эталонным кварталом и остальными.

Регрессионная модель, описывающая динамику уровней ряда, относящихся к эталонному 1-му кварталу, примет вид:

$$y_t = a_0 + a_1 t$$

соответственно для наблюдений

$$\text{2-го квартала} \quad y_t = a_0 + a_1 t + c_2;$$

$$\text{3-го квартала} \quad y_t = a_0 + a_1 t + c_3;$$

$$\text{4-го квартала} \quad y_t = a_0 + a_1 t + c_4;$$

Переход из одного квартала в другой будет отражаться лишь в изменении свободного члена регрессионного уравнения и не будет касаться значения параметра b , определяющего угол наклона линейного тренда и характеризующего средний абсолютный прирост уровней ряда под воздействием тенденции.

Найденные значения коэффициентов c_2, c_3, c_4 позволяют оценить «сдвиги» в уровнях за счет фактора сезонности относительно i -го, эталонного квартала. Можно усреднить четыре полученные линии регрессии:

$$\tilde{y}_t = \bar{a}_0 + \bar{a}_1 t \quad (9.2)$$

Тогда расстояние между отдельной регрессионной прямой для любого квартала и усредненной моделью, даст оценку сезонных отклонений в этом квартале. Очевидно, что для аддитивной модели сумма сезонных отклонений будет равна нулю.

Рассмотрим реализацию применения фиктивных переменных для моделирования сезонных колебаний в пакете STATISTICA.

В качестве исходных данных используем квартальный ряд динамики ВВП (приложение 17) с 1 квартала 1999 г. до 4 квартала 2004 г.

Шаг 1. Для начала проведем визуализацию ряда, для этого в главном меню программы выберем *Graphs* → *2D Graphs* → *Line Plots (Variables)*. После выбора переменной (кнопка *Variables*) на основе которой необходимо построить график (в данном случае это переменная Y), получаем следующий результат:

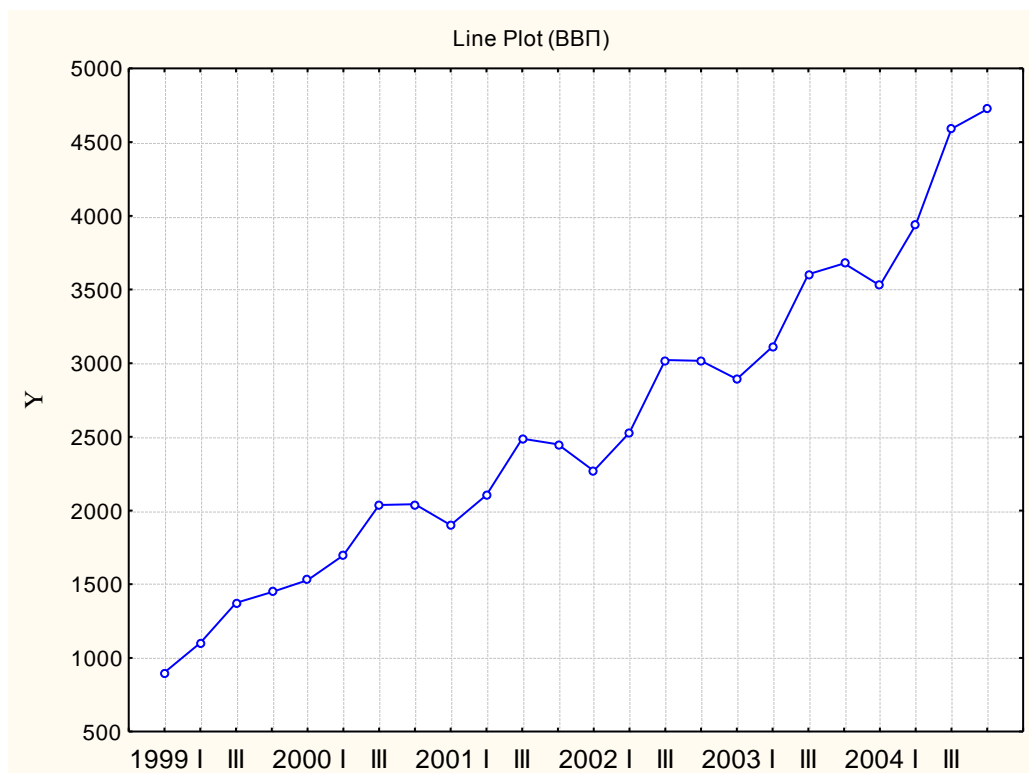


Рисунок 9.2 - Динамика ВВП России 1 квартала 1999г-4 квартал 2004г

Согласно приведенному графику наблюдается значительный рост показателя за анализируемый период, а также сезонность с пиком в каждом 3 квартале года.

Шаг 2. Для описания сезонных колебаний создадим 4 фиктивных переменных. Для этого переходим в рабочую таблицу и образуем, переменную t – характеризующую моменты (периоды) времени переменные и переменные Z_2 , Z_3 и Z_4 – характеризующие сезонность в анализируемом ряду:

	1 Y	2 t	3 Z ₂	4 Z ₃	5 Z ₄
1999 I	901,3	1	0	0	0
II	1101,5	2	1	0	0
III	1373,1	3	0	1	0
IV	1447,3	4	0	0	1
2000 I	1527,4	5	0	0	0
II	1696,6	6	1	0	0
III	2037,8	7	0	1	0
IV	2043,8	8	0	0	1

Рисунок 9.3 – Рабочая таблица с набором фиктивных переменных (приведена часть исходного окна)

Шаг 3. В главном меню выберем: *Statistics* → *Multiple Regression* (Статистика → Множественная регрессия). В появившемся окне *Multiple Linear*

Regression необходимо нажать кнопку *Variables* (Переменные) и указать в качестве зависимой переменной (*Dependent var.*) *Y*, а в качестве не зависимых (*Independent var.*) переменных - *t*, *Z2*, *Z3* и *Z4*.

Нажав кнопку ОК, перейдем в следующее окно, содержащее результаты построения модели.

Шаг 4. В появившемся окне *Multiple Regression Results* выберем кнопку *Summary: Regression results* (Итоги: Результаты построения регрессии) перейдем к двум таблицам содержащим оцененные параметры модели и основные показатели адекватности построения регрессии.

Таблица 9.4 – Показатели адекватности модели

	<i>Value</i> (Значение)
<i>Multiple R</i> (Множест. R)	0,989
<i>Multiple R?</i> ((Множест. R2)	0,977
<i>Adjusted R?</i> (Скорр. R2)	0,973
<i>F(11,144)</i>	204,893
<i>p</i>	0,000
<i>Std.Err. of Estimate</i> (Стд. Ош. Оценки)	174,096

Согласно данным, приведенным в таблице 9.5 полученная модель статистически значима по *F*-критерию Фишера, но параметр при фиктивной переменной *Z2* не проходит тест на статистическую значимость по *t*-критерию Стьюдента.

Таблица 9.5 – Результаты оценивания сезонной модели

	<i>Beta</i> (БЕТА)	<i>Std.Err. of Betta</i> (Стд.Ош. БЕТА)	<i>B</i>	<i>Std.Err. of B</i> (Стд.Ош. B)	<i>t(19)</i>	<i>p-level</i> (p-уров.)
<i>Intercept</i> (Св.член)			606,006	91,247	6,641	0,000
<i>t</i>	0,956	0,035	142,156	5,202	27,327	0,000
<i>Z2</i>	0,043	0,042	101,828	100,649	1,012	0,324
<i>Z3</i>	0,168	0,043	398,189	101,051	3,940	0,001
<i>Z4</i>	0,125	0,043	296,600	101,718	2,916	0,009

В общем, опираясь на построенную модель можно сказать, что в анализируемом ряду присутствует сезонность, с максимум в 3 квартале каждого года, т.к. β -коэффициент при *Z3* имеет наибольшее значение.

9.6 Тест (критерий) Г. Чоу

Для выявления структурных изменений в ряду динамики на практике

можно прибегнуть к проведению теста Чоу (тесты на устойчивость). Существует несколько модификаций теста Чоу, это тест на обоснованность объединения двух выборок при оценке регрессии и тест на неудачу предсказания. В первом случае тестируется предсказательная способность модели, во втором определяется, происходит ли сдвиг параметров в период предсказания.

В данном случае нас интересует первый подход, рассмотрим механизм его приведения подробнее.

Методика проведения данного теста сводится к следующему: на основе имеющихся данных оценивается уравнение регрессии сначала по всему ряду, а затем уравнения по кусочно-линейной модели.

Далее определяется фактическое значение F -статистики Фишера по формуле:

$$F_{\text{факт}} = \frac{(RSS_P - RSS_A - RSS_B)/(k_A + k_B - k_P)}{(RSS_A - RSS_B)/(T - k_A - k_B)} \quad (9.3)$$

где RSS_P – остаточная сумма квадратов модели построенной на основе всего ряда;

RSS_A – остаточная сумма квадратов первой модели построенной на основе ряда до момента t^* (предполагаемый (или реальный) момент наступления события повлекший структурные изменения ряда);

RSS_B – остаточная сумма квадратов второй модели построенной на основе ряда после момента t^* ;

k_P, k_A, k_B – число параметров в регрессии по всему ряду и в первой и второй регрессиях;

T – число уровней ряда.

Далее с помощью F -статистики тестируется гипотеза H_0 о структурной стабильности тенденции изучаемого временного ряда. Для этого найденное значение $F_{\text{факт}}$ сравнивается с табличным полученным при уровне значимости α и степенями свободы $\nu_1=m$; $\nu_2=T-k-1$. Если $F_{\text{факт}} > F_{\text{табл}}$, то гипотеза отклоняется, а влияние структурных изменений на динамику изучаемого показателя признается значимым.

Для иллюстрации описанной процедуры воспользуемся динамическим рядом ВВП России за период с 1 квартала 1994г. по 4 квартал 2004 года (приложение 17).

Шаг 1. Вначале построим уравнение регрессии на основе всей совокупности данных. Перед этим необходимо образовать переменную t_1 ($t_1=0$ в 4 квартале 1993 года). Далее запускаем процедуру *Multiple Regression*.

Шаг 2. В окне результатов оценки модели *Multiple Regression Results* необходимо выбрать вкладку *Advanced* и кнопку *ANOVA (Overall goodness of fit)*, тем самым на экран будет выведена таблица с результатами дисперсионного анализа (необходимо заметить, что модель статистически значима по F -критерию Фишера и t -критерию Стьюдента).

Таблица 9.6 – Результаты дисперсионного анализа общей регрессионной модели

	<i>Sums of Squares</i> (Сумма)	<i>df</i>	<i>Mean Squares</i> (Средн.)	<i>F</i>	<i>p-level</i> (p-уров.)
<i>Regress.</i> (Регресс.)	69023330	1	69023330	433,83	0,000
<i>Residual</i> (Остатки)	6682327	42	159103		
<i>Total</i> (Итого)	75705658				

Для оценки *F*-критерия Фишера (тест Чоу) из данной таблицы понадобится остаточная сумма квадратов, которая находится на пересечении столбца *Sums of Squares* и строки *Residual*, т.е. значение 6682327.

Шаг 3. Последовательно оценим две кусочно-линейные модели, первая до 4 квартала 1999г., вторая после данного периода.

Выбор 4 квартала 1999г. как предполагаемый момент наступления события повлекший структурные изменения ряда неслучаен, так как в 1998г. в Россию потряс финансовый кризис который и стал причиной изменения механизма генерации макроэкономических рядов, т.е. начиная с 1999г. (в результате инерционности экономики) имеем совершенно иной динамический ряд который не сопоставим с предыдущей динамикой.

Перед тем как приступить к оценке моделей необходимо ввести две переменные t_2 (равна единице в 1 квартале 1994г.) и t_3 (равна единице в 1 квартале 1999г.). При построении кусочно-линейных моделей необходимо выбрать кнопку *Select Cases* и в первом случае указать $v0 < 21$, во втором случае указать $v0 > 20$ (тем самым будут заданы диапазоны для оценки регрессионных уравнений).

В результате оценки будут получены следующие результаты:

Таблица 9.7 – Результаты дисперсионного анализа первой кусочно-линейной регрессионной модели (до 1 кв. 1999г.)

	<i>Sums of Squares</i> (Сумма)	<i>df</i>	<i>Mean Squares</i> (Средн.)	<i>F</i>	<i>p-level</i> (p-уров.)
<i>Regress.</i> (Регресс.)	681651,4	1	681651,4	200,03	0,0000
<i>Residual</i> (Остатки)	61339,9	18	3407,8		
<i>Total</i> (Итого)	742991,2				

Таблица 9.8 – Результаты дисперсионного анализа первой кусочно-линейной регрессионной модели (до 1 кв. 1999г.)

	<i>Sums of Squares</i> (Сумма)	<i>df</i>	<i>Mean Squares</i> (Средн.)	<i>F</i>	<i>p-level</i> (p-уров.)
<i>Regress.</i> (Регресс.)	24262153	1	24262153	462,40	0,000
<i>Residual</i> (Остатки)	1154337	22	52470		
<i>Total</i> (Итого)	25416490				

Шаг 4. Находим расчетное значение F -критерия Фишера:

$$F_{\text{факт}} = \frac{(RSS_P - RSS_A - RSS_B)/(k_A + k_B - k_P)}{(RSS_A - RSS_B)/(T - k_A - k_B)} = \frac{(6682327 - 61339,9 - 1154337)/(2 + 2 - 2)}{(61339,9 - 1154337)/(44 - 2 - 2)} = -2706000,12$$

По таблице находим табличное значение F -критерия Фишера при степенях значимости $v_1 = m = 2$; $v_2 = T - k - 1 = 44 - 2 - 1 = 41$, т.е. получаем 3,23.

Сравнивая расчетное значение с табличным, получаем $F_{\text{факт}} > F_{\text{табл}}$, отсюда можно сделать вывод, что подтверждается предположение о значительных изменениях в механизме генерации ряда инвестиций в основной капитал под влиянием финансового кризиса 1998г.

9.7 Тесты для самоконтроля

1 Если качественный признак, который необходимо отразить в регрессионной модели имеет четыре градации, то в уравнение включается:

- четыре фиктивные переменные
- пять фиктивных переменных
- одна фиктивная переменная
- три фиктивная переменная

2 Если качественный признак, который необходимо отразить в регрессионной модели имеет две градации, то в уравнение включается:

- одна фиктивная переменная
- две фиктивных переменных
- три фиктивная переменная
- нельзя включать качественные переменные в уравнение

3 Приведенный пример расстановки фиктивных переменных используется для:

Дата	y_t	t	t'
1998г.	y_1	1	-3
1999г.	y_2	2	-2
2000г.	y_3	3	-1
2001г.	y_4	4	0

2002г.	y_5	5	1
2003г.	y_6	6	2
2004г.	y_8	7	3

- элиминирования линейного временного тренда
- элиминирования тренда в виде параболы второго порядка
- выделения сезонной составляющей

4 Приведенный пример расстановки фиктивных переменных используется для:

Дата	y_t	t_1	t_2	t_3
1996г.	y_1	1	0	0
1997г.	y_2	2	0	0
1998г.	y_3	3	0	0
1999г.	y_4	4	0	0
2000г.	y_5	5	0	1
2001г.	y_6	5	1	1
2002г.	y_8	5	2	1
2003г.	y_9	5	3	1
2004г.	y_{10}	5	4	1

- выделения двух прямых, точка пересечения которых известна
- выделения двух прямых, точка пересечения которых не известна
- выделения линейного временного тренда

5 Приведенный пример расстановки фиктивных переменных используется для:

Дата	y_t	t_1	t_2
1996г.	y_1	-4	0
1997г.	y_2	-3	0
1998г.	y_3	-2	0
1999г.	y_4	-1	0
2000г.	y_5	0	0
2001г.	y_6	0	1
2002г.	y_8	0	2
2003г.	y_9	0	3
2004г.	y_{10}	0	4

- выделения двух прямых, точка пересечения которых известна
- выделения двух прямых, точка пересечения которых не известна
- выделения линейного временного тренда

6 Приведенное выражение $\tilde{y}_t = a_0 + a_1 t + c_2 Z_2 + c_3 Z_3 + c_4 Z_4 + \varepsilon_t$ используют при:

- построении парного линейного уравнения регрессии
- описании сезонных колебаний
- построении нелинейного уравнения

7 Приведенная формула $F_{\text{факт}} = \frac{\Delta S_{\text{ост}}}{m_1 + m_2 - m_3} / \frac{S_{\text{ост}}^{k-l}}{n - m_1 - m_2}$ используется при:

- проверке гипотезы о статистической значимости регрессионного уравнения

- проверке гипотезы о статистической значимости параметров регрессионного уравнения
- проверке гипотезы о гетероскедстичности случайных отклонений
- проверке гипотезы о адекватности линейного тренда построенного на основе всей совокупности

8 Приведенная таблица используется при:

Периоды	Число наблюдений в совокупности	Остаточная сумма квадратов	Число параметров в уравнении	Число степеней свободы остаточной дисперсии
Первое уравнение	n_1	$S^1_{ост}$	m_1	$n_1 - m_1$
Второе уравнение	n_2	$S^2_{ост}$	m_2	$n_2 - m_2$
Объединенное уравнение	n	$S^3_{ост}$	m_3	$n - m_3 = (n_1 + n_2) - m_3$

- дисперсионном анализе
- построении теста Чоу
- выводе результатов оценки параметров регрессионного уравнения

9.8 Задания для самостоятельного выполнения

Задания для самостоятельной работы составлены в пяти вариантах, номер варианта выбирается в соответствии с последней цифрой зачетной книжки студента:

Таблица 9.9 – Задания для самостоятельной работы по теме «Построение регрессионных моделей с фиктивными переменными»

Последняя цифра номера зачетной книжки	1 и 6	2 и 7	3 и 8	4 и 9	5 и 0
Номер варианта	1	2	3	4	5

Используя данные соответствующего варианта необходимо выполнить следующие этапы самостоятельной работы:

- 1 Провести визуализацию ряда (приложение 15), если прослеживается сезонность, то оценить ее влияние, используя фиктивные переменные;
- 2 Используя соответствующий динамический ряд (приложение 15), построить тест Чоу и сделать соответствующие выводы о структурной стабильности ряда.

10 Оценка нелинейных регрессионных моделей

10.1 Цели и задачи лабораторной работы

В данной лабораторной работе на основе пространственных данных и временного ряда рассмотрим методы оценки нелинейных моделей, при этом выделим следующие задачи:

- 1 На основе данных об объеме промышленного производства, стоимости основных фондов и среднегодовой численности занятых в промышленности оценить производственную функцию Кобба-Дугласа.
- 2 На основе ряда численности безработных в РФ оценить параболу второго порядка, гиперболу и логарифмическую прямую.

10.2 Определение нелинейной регрессии и основные формы нелинейных моделей

В ходе проведения экономических исследований возникают ситуации, в которых линейные модели не приносят желаемых результатов. В подобных случаях, вероятно, необходимо прибегнуть к использованию нелинейных моделей.

В эконометрических исследованиях различают два класса нелинейных регрессий:

- 1 регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам (полиномы разных степеней, равноугольная гипербола).
- 2 регрессии, нелинейные по оцениваемым параметрам (степенная, показательная, экспоненциальная функция).
 - а) нелинейные модели внутренне линейные;
 - б) нелинейные модели внутренне нелинейные.

Для оценки параметров нелинейных моделей используются два подхода.

Первый основан на линеаризации модели и заключается в том, что с помощью подходящих преобразований исходных переменных исследуемую зависимость представляют в виде линейного соотношения между преобразованными переменными.

Второй подход обычно применяется в случае, когда подобрать соответствующее линеаризующее преобразование не удастся. В этом случае применяются методы нелинейной оптимизации на основе исходных переменных.

10.3 Рекомендуемая литература

Для лучшего понимания материала изложенного в данной главе необходимо дополнительно проанализировать следующие источники литературы (см. список использованных источников):

- 1 Бородич, С.А. Эконометрика: учеб. пособие / С.А. Бородич. – 3 –е изд., стер. – Мн.: Новое знание, 2006. – 408 с. ISBN 985-475-206-2 (см. стр. 200-230)
- 2 Винн, Р., Холден, К. Введение в прикладной эконометрический анализ. - М.: Финансы и статистика, 1981. (см. стр. 34-95)
- 3 Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: Учебное пособие. – М.: ФОРУМ: ИНФРА-М, 2004. – 464с. (см. стр. 232-240)
- 4 Прикладная статистика. Основа эконометрики: учебник для вузов: В 2 т. 2-е изд., испр. Т. 2: Айвазян, С.А. Основы эконометрики. - М.: ЮНИТИ-ДАНА, 2001. – 432 с. ISBN 5-238-00305-6. (см. стр. 172-199)
- 5 Эконометрика: учебник для вузов / под ред. И.И. Елисеевой. - М. : Финансы и статистика, 2005. - 576 с. : ил. - ISBN 5-279-02786-3. (см. стр. 77-96)

10.4 Оценка нелинейной модели на основе пространственных данных

В качестве примера нелинейной зависимости рассмотрим класс производственных функции. Которые отражают зависимости, существующие между объемом произведенной продукции и основными факторами производства – трудом, капиталом и т.п. Ярким примером подобных моделей является производственная функция Кобба-Дугласа:

$$\tilde{y}_i = AK^\alpha L^\beta \quad (10.1)$$

где: Y - объем производства,
 K - затраты капитала,
 L - затраты труда.

Показатели α и β являются коэффициентами частной эластичности. Это означает, что при увеличении одних только затрат капитала (труда) на 1% объем производства увеличится на $\alpha\%$ ($\beta\%$).

Данная функция относится к подклассу 2.1, поэтому данное уравнение можно свести к линейному виду. Для этого берут логарифмы левой и правой частей модели, в результате чего получаем:

$$\ln y = \ln A + \alpha \ln K + \beta \ln L$$

Далее заменяем $y' = \ln y$, $x_1 = \ln K$, $x_2 = \ln L$, получаем

$$\tilde{y}' = a_0 + a_1 x_1 + a_2 x_2 \quad (10.2)$$

a_1 – оценка α ; a_2 – оценка β ; a_0 – оценка $\ln A$

Для построения производственной функции Кобба-Дугласа в пакете STATISTICA воспользуемся данными по 14 субъектам Приволжского федерального округа (приложение 18). В данном пакете программ существует два способа построения нелинейных моделей:

- 1 Использование модуля *Fixed Nonlinear Regression*
- 2 Использование модуля *Multiple Regression*

10.4.1 Построение нелинейной модели в модуле *Fixed Nonlinear Regression*

В пакете STATISTICA 6.0 существует специальный модуль *Fixed Nonlinear Regression* (Фиксированная нелинейная регрессия) предназначенный для оценки нелинейных моделей посредством линейризации исходных зависимостей.

Для построения производственной функции в данном модуле необходимо:

Шаг 1. В главном меню выберем *Statistics* → *Advanced Linear/Nonlinear Models* → *Fixed Nonlinear Regression* (Статистика → Выбор линейной/нелинейной модели → Фиксированная нелинейная регрессия).

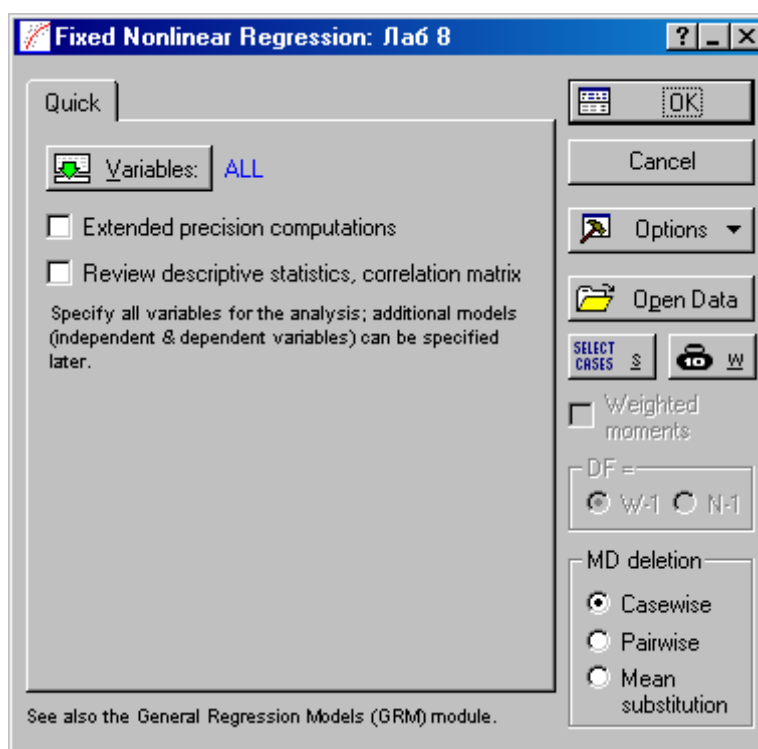


Рисунок 10.1 - Окно установок нелинейной регрессии

где: *Extended precision computations* - Вычисления с повышенной точностью
Review descriptive statistics, correlation matrix – Обзор описательных статистик, корреляционная матрица

Шаг 2. В появившемся окне *Fixed Nonlinear Regression* необходимо выбрать кнопку *Variables* и выделить переменные участвующие в расчете (в данном случае все переменные - *ALL*).

Шаг 3. В очередном окне *Non-linear Components Regression* (Компоненты не линейной регрессии) установим флажок напротив опции *LN(X)*, тем самым будут рассчитаны логарифмы для всех переменных участвующих в оценке модели.

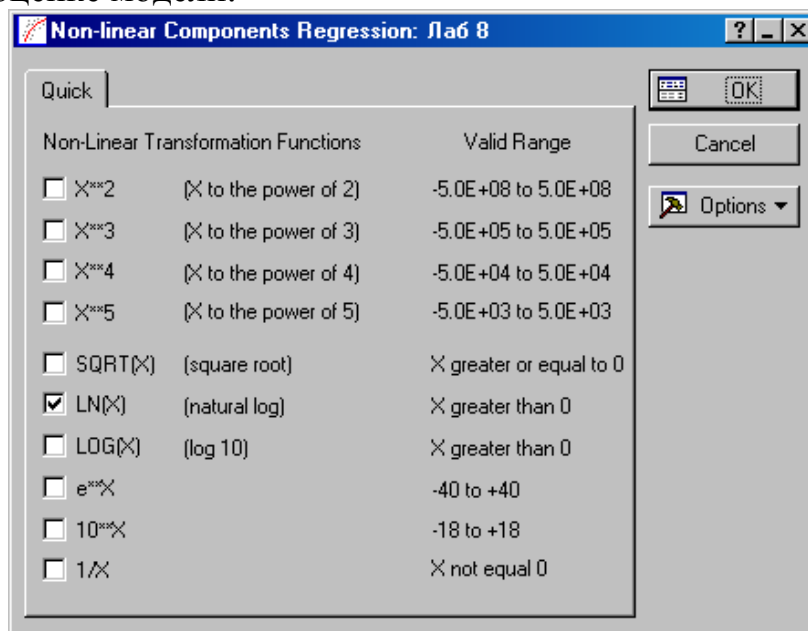


Рисунок 10.2 – Установка процедур линеаризации исходных переменных

Шаг 4. В окне *Model Definition* (Установки модели) выберем кнопку *Variables* и сделаем установки как показано на рисунке 3. В качестве зависимой переменной укажем *LN-V1* в качестве независимых переменных укажем - *LN-V2* и *LN-V3*.

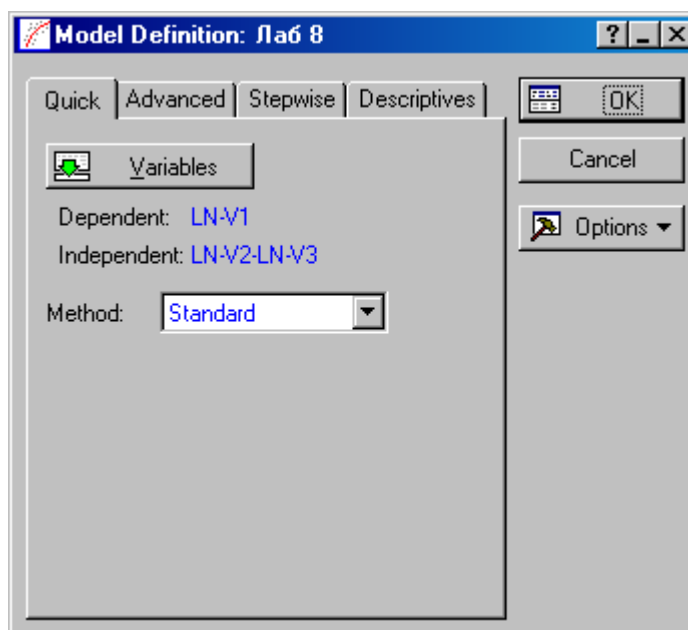


Рисунок 10.3 – Окно выбора зависимой и независимых переменных для построения производственной функции

Рассмотрим результаты оценки производственной функции Кобба-Дугласа.

Согласно данным, приведенным в таблице 2, 98,2% вариации объема промышленного производства описывается вариацией факторов производства, также можно утверждать, что наблюдается сильная связь между исследуемыми показателями.

Таблица 10.1 – Показатели адекватности производственной функции Кобба-Дугласа

	<i>Value</i> (Значение)
<i>Multiple R</i> (Множест. R)	0,991
<i>Multiple R²</i> ((Множест. R ²))	0,982
<i>Adjusted R²</i> (Скорр. R ²)	0,979
<i>F(2,11)</i>	308,655
<i>p</i>	0,000
<i>Std.Err. of Estimate</i> (Стд. Ош. Оценки)	0,135

Полученная модель статистически значима согласно *F*-критерию Фишера, параметры модели значимы согласно *t*-критерию Стьюдента.

Таблица 10.2 – Результаты оценки производственной функции Кобба-Дугласа

	<i>Beta</i> (БЕТА)	<i>Std.Err. of Beta</i> (Стд.Ош.)	<i>B</i>	<i>Std.Err. of B</i> (Стд.Ош.)	<i>t(11)</i>	<i>p-level</i> (p-уров.)
<i>Intercept</i> (Св. член)			1,505	0,424	3,548	0,005
<i>LN-V2</i>	0,602	0,094	0,556	0,087	6,423	0,000
<i>LN-V3</i>	0,413	0,094	0,657	0,149	4,400	0,001

Интерпретировать полученные показатели можно следующим образом: при изменении стоимости основных фондов на 1%, объем промышленного производства изменится на 0,66%, а при увеличении среднегодовой численности занятых в данном секторе экономики на 1% произойдет увеличение зависимой переменной на 0,56%.

10.4.2 Построение нелинейной модели в модуле *Multiple Regression*

Прежде чем приступить к оценке функции Кобба-Дугласа вторым способом предварительно преобразуем исходные переменные, а именно рассчитаем логарифмы.

Шаг 1. Образует новую переменную $LN Y$ (сразу после переменной Y), для этого в поле *Long name* введем формулу $=\log10(v1)$, далее образуем $LN K$ (после переменной K) введя формулу $=\log10(v3)$ и $LN L$ (после переменной L) - $=\log10(v5)$

Шаг 2. В главном меню выберем *Statistics* → *Multiple Regression*, при этом в качестве зависимой переменной укажем $LN Y$ в качестве не зависимых - $LN K$ и $LN L$.

Сравнивая полученные итоговые таблицы с данными приведенными в таблицах 2 и 3 можно убедиться в идентичности полученных результатов.

10.5 Оценка нелинейной модели на основе временного ряда

Наиболее часто к нелинейным моделям прибегают в случае оценки трендовых моделей на основе временных рядов. В данном пункте остановимся на двух способах оценки нелинейных моделей на основе ряда количество безработных (в среднем за период), млн. руб. (приложение 19).

При этом в анализе будем использовать данные с 1 квартала 1999г. по 4 квартал 2005г., данные за 2006г. используем для сопоставления с прогнозными значениями.

10.5.1 Графический способ построения нелинейной регрессии

Для реализации графического метода необходимо выполнить следующие шаги:

Шаг 1: В строке главного меню необходимо выбрать *Graphs* → *2D Graphs* → *Line Plots (Variables...)* (Графики → Двухмерные графики → Линейный рисунок для переменных).

Шаг 2: В появившемся окне *2D Line Plots* необходимо выбрать вкладку *Advanced* (Расширенные) и определить переменную, на базе которой будет построен график. Для этого необходимо нажать кнопку *Variables* (Переменные) и в появившемся окне выбрать необходимую переменную.

Шаг 3: В поле *Fit* (Функции) укажем вид желаемого уравнения, в данном случае *Polynomial* (Парабола второй степени) и нажать ОК. Получаем график динамики ряда численности безработных в РФ и соответствующий тренд (рисунок 10.4).

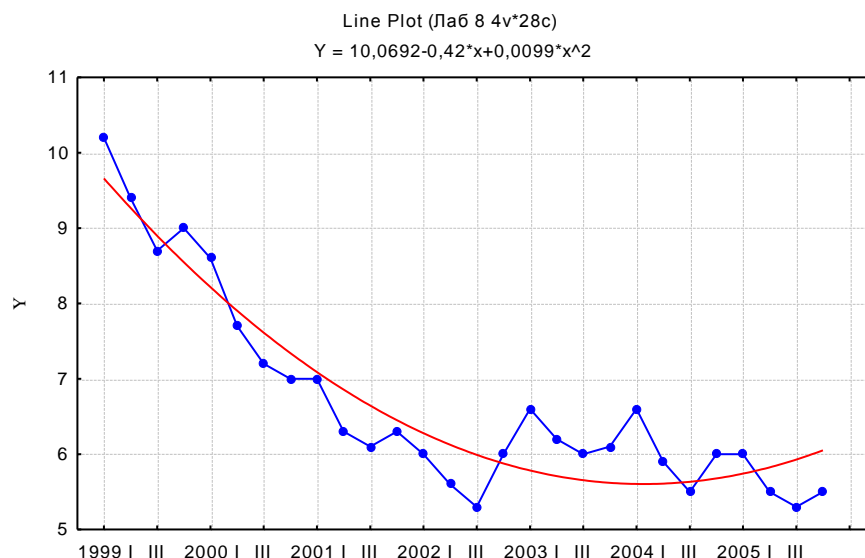


Рисунок 10.4 – Динамика численности безработных в РФ и выровненные на основе параболы второго порядка уровни

В верхней части графика выводится уравнение параболы второго порядка (рисунок 10.4). Согласно приведенным на рисунке 10.4 результатам, выбранный тип графика, довольно хорошо описывает динамику показателя.

Для сравнения результатов оценим логарифмическую прямую вида - $\tilde{y}_t = a_0 + a_1 \cdot \log_{10} t_t$, для этого в полу *Fit* укажем *Logarithmic* и получим следующие результаты:

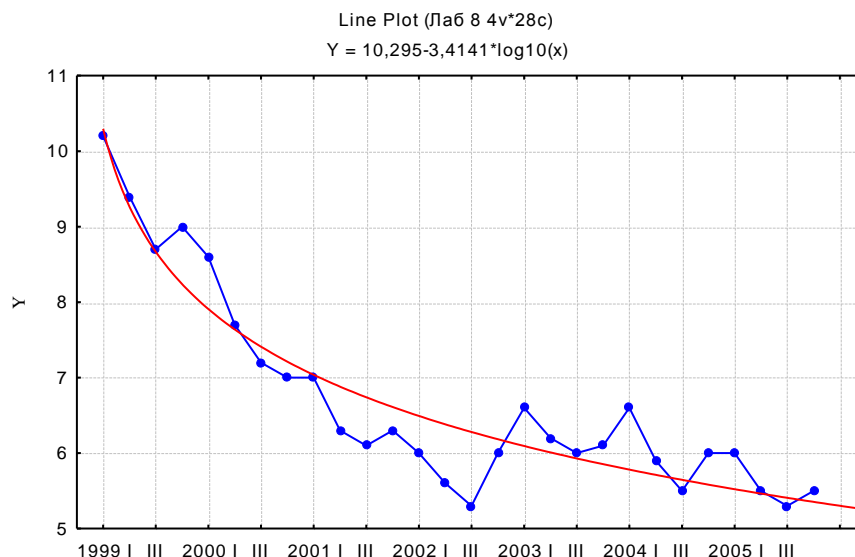


Рисунок 10.5 – Динамика численности безработных в РФ и выровненные на основе логарифмической прямой уровни

Согласно данным, приведенным на рисунке 10.5, выбранная функция (в отличие от параболы второго порядка) имеет тенденцию к снижению, что в

прогнозом периоде может привести к пересечению с осью OX . С математической точки зрения интерпретировать это случай можно как прекращение существования показателя, с социально-экономической точки зрения как полную занятость населения. Но экономическая теория утверждает, что подобного прецедента в рыночной экономике не существует. Из вышесказанного можно сделать вывод о неприемлемости подобной функции для описания динамики рассматриваемого явления.

10.5.2 Построение нелинейных трендов с помощью модуля *Multiple Regression*

Особенностью построения нелинейных моделей на основе временных данных (в отличие от пространственных данных) является то, что в качестве независимой переменной в данном случае будет выступать момент времени t . В остальном все действия аналогичны описанным в главе 5:

Шаг 1. Вводим в рассмотрение переменную t , при этом $t=0$ в 4 квартале 1998 года. Далее преобразуем данную переменную, а имен вводим следующие переменные:

- t^2 – для этого после переменной t образуем новый столбец t^2 и в поле *Long name (label or formula with Functions)*: вводим $=v2^2$;
- $\frac{1}{t}$ – после столбца t^2 добавляем переменную $1/t$ и указываем $=1/v2$;
- $\log_{10}t$ – после столбца $1/t$ добавляем переменную $\log_{10}t$ и указываем $=\log_{10}(v2)$

Шаг 2. В главном меню выбираем *Statistics* → *Multiple Regression*, для построения параболы второго порядка в качестве зависимой переменной укажем Y , не зависимых – t и t^2 .

Для оценки гиперболы необходимо в поле *Dependent var.* указать Y , а в поле *Impendent var.* - $1/t$.

Для оценки логарифмической функции необходимо в поле *Dependent var.* указать Y , в поле *Impendent var.* - $\log_{10}t$.

Результаты оценки трех кривых приведем в таблице 10.3.

Таблица 10.3 - Сравнительные характеристики динамических моделей ряда численности безработных в РФ

Модель	Парабола второго порядка	Гипербола	Логарифмическая прямая
	$\tilde{y}_t = 10,06 - 0,42t + 0,01t^2$ (34,15) (-8,96) (6,30)	$\tilde{y}_t = 5,91 + 5,66 \frac{1}{t}$ (34,76) (7,98)	$\tilde{y}_t = 10,295 - 3,41 \log_{10}t$ (38,72) (-14,29)
Коэффициент детерминации R^2	0,878	0,71	0,887

Скорректированный коэффициент детерминации AR^2	0,868	0,699	0,883
F-критерии Фишера	89,615	63,748	204,159
Стандартное отклонение	0,483	0,729	0,455

В результате построения моделей было получено, что наилучший коэффициент детерминации наблюдается у логарифмической прямой и равняется 0,89, т.е. 89% колебаний изучаемого показателя описывается данной моделью.

Корректную оценку адекватности построенной модели дает скорректированный коэффициент детерминации, так как он показывает такую оценку тесноты связи, которая не зависит от числа факторов в модели и потому может сравниваться по разным моделям с разным числом факторов. В данном случае наибольшее значение так же получено у логарифмической модели.

Значение F-критерий Фишера у всех трех моделей получены довольно большими, т.е. подтверждается статистическая значимость оцененных уравнений регрессии.

Так как полученные модели являются статистически значимыми переходим к прогнозированию уровней ряда на 2006 год. При этом по причинам описанным выше логарифмическая прямая будет исключена из рассмотрения. Поэтому для прогнозирования используем параболу второго порядка и гиперболу.

Шаг 3. Для прогнозирования неизвестных уровней по параболе необходимо в окне *Multiple Regression Results* выбрать вкладку *Residuals/assumptions/prediction* (Отклонения/распределения/предсказания) и воспользоваться кнопкой *Predict dependent variable* (Прогнозирование зависимой переменной).

В появившемся окне в окно *Specify values for indep. vars* (Определение неизвестных значений для зависимой переменной) последовательно укажем значения (результаты расчета представим в таблице 10.4):

t	t^2
29	841
30	900
31	961
32	1024

Шаг 4. Для прогнозирования неизвестных уровней по гиперболу необходимо после оценки соответствующей модели в окне *Specify values for indep. Vars* последовательно указать результаты расчета представим в таблице 5):

$1/t$
0,034
0,033
0,032

Таблица 10.4 – Прогнозируемые значения числа безработных в РФ

Годы	Фактические значения	Прогнозные значения по параболы второго порядка	Нижняя дов. граница - 95,0%	Верхняя дов. граница +95,0%	Прогнозные значения по гиперболе	Нижняя дов. граница -95,0%	Верхняя дов. граница +95,0%
2006 I	5,8	6,195	5,587	6,802	6,099	5,776	6,422
II	5,6	6,357	5,661	7,054	6,093	5,770	6,417
III	5,4	6,540	5,746	7,333	6,088	5,763	6,412
IV	-	6,742	5,843	7,640	6,082	5,757	6,407

Согласно полученным прогнозным значениям на основе параболы второго порядка, имеем рост показателя в 2006 году, при этом относительно фактических данных прогнозные значения завышены.

Прогнозы, полученные на основе гиперболы, имеют тенденцию к снижению, т.е. они более адекватно описывают фактические данные (с 1-3 квартал 2006 года наблюдается снижение числа безработных). В пользу гиперболы говорит и тот факт, что прогнозные уровни на k периодов будут стремиться к значению 5,91 (согласно свойствам гиперболы), но не пересекут это значение.

Подводя итоги проведенного анализа ряда численности безработных можно сделать вывод о том, что наилучшей моделью описывающих тенденцию показателя является гипербола, т.к. параметры модели получены статистически значимыми и прогнозы в большей степени соответствуют фактическим значениям.

10.6 Тесты для самоконтроля

1 Если в при расчете параметров параболы второго порядка получаем, что $a_1 < 0$ и $a_2 > 0$, то:

- кривая симметрична относительно высшей точки;
- кривая симметрична относительно низшей точки;
- имеем медленно повышающуюся функцию с верхней асимптотой.

2 Если в при расчете параметров параболы второго порядка получаем, что $a_1 > 0$ и $a_2 < 0$, то:

- кривая симметрична относительно высшей точки;
- кривая симметрична относительно низшей точки;
- имеем медленно повышающуюся функцию с верхней асимптотой.

3 Какое уравнение соответствует следующей системе нормальных уравнений, применяемой для определения параметров уравнения (по методу наименьших квадратов)

$$\begin{cases} \sum y = n \cdot a_0 + a_1 \cdot \sum \frac{1}{x} \\ \sum \frac{y}{x} = a_0 \cdot \sum \frac{1}{x} + a_1 \cdot \sum \frac{1}{x^2} \end{cases}$$

- $\tilde{y}_i = a_0 + a_1 \cdot \frac{1}{x_i}$;
- $\tilde{y}_i = a_0 + a_1 \cdot x_i + a_2 \cdot x_i^2$;
- $\tilde{y}_i = a_0 + a_1 \cdot x_i$.

4 Какое уравнение соответствует следующей системе нормальных уравнений, применяемой для определения параметров уравнения (по методу наименьших квадратов)

$$\begin{cases} \sum y = n \cdot a_0 + a_1 \cdot \sum x + a_2 \cdot \sum x^2 \\ \sum y \cdot x = a_0 \cdot \sum x + a_1 \cdot \sum x^2 + a_2 \cdot \sum x^3 \\ \sum y \cdot x^2 = a_0 \cdot \sum x^2 + a_1 \cdot \sum x^3 + a_2 \cdot \sum x^4 \end{cases}$$

- $\tilde{y}_i = a_0 + a_1 \cdot \frac{1}{x_i}$;
- $\tilde{y}_i = a_0 + a_1 \cdot x_i + a_2 \cdot x_i^2$;
- $\tilde{y}_i = a_0 + a_1 \cdot x_i$.

5 Приведенная модель $\tilde{y}_i = a_0 + a_1 \cdot x_i + a_2 \cdot x_i^2$ является:

- параболой второго порядка;
- гиперболой ;
- параболой первого порядка.

6 Приведенная модель $\tilde{y}_i = a_0 + a_1 \cdot x_i + a_2 \cdot x_i^2 + a_3 \cdot x_i^3$ является:

- параболой второго порядка;
- параболой третьего порядка;
- параболой первого порядка.

7 Приведенная модель $\tilde{y}_i = a_0 + a_1 \cdot \frac{1}{x_i}$ является:

- параболой второго порядка;
- равносторонней гиперболой;
- параболой первого порядка.

10.7 Задания для самостоятельного выполнения

Задания для самостоятельной работы составлены в пяти вариантах, номер варианта выбирается в соответствии с последней цифрой зачетной книжки студента:

Таблица 4.12 – Задания для самостоятельной работы по теме «Оценка нелинейных регрессионных моделей»

Последняя цифра номера зачетной книжки	1 и 6	2 и 7	3 и 8	4 и 9	5 и 0
Номер варианта	1	2	3	4	5

Используя данные соответствующего варианта необходимо выполнить следующие этапы самостоятельной работы:

- 1 Введите данные в пакете STATISTICA (либо используя табличный редактор, либо вводя данные непосредственно в поле пакета программ) и сохраните данные на диске (или ином носителе);
- 2 Оценить производственную функцию Кобба-Дугласа двумя способами (приложение 20);
- 3 На основе фактических данных оценить наилучшую нелинейную модель для соответствующего временного ряда и провести прогнозирование на 3 уровня вперед (приложение 15).

Список использованных источников

- 1 Андерсон, Т. Введение в многомерный статистический анализ. - М.: Физ.-мат. лит., 1963. – 245 с.
- 2 Афанасьев, В.Н. Анализ временных рядов и прогнозирование: учеб. для вузов / В.Н. Афанасьев, М.М. Юзбашев. - М. : Финансы и статистика, 2001. - 228с. : ил. ISBN 5-279-02419-8.
- 3 Афанасьев, В.Н. Эконометрика: учеб. для вузов / В.Н. Афанасьев, М.М. Юзбашев, Т.И. Гуляева; под ред. В.Н. Афанасьева. - М. : Финансы и статистика, 2005. - 256 с. : ил. - Библиогр.: с. 254-255. ISBN 5-279-02738-3.
- 4 Бендат, Дж., Пирсол, А. Применение корреляционного и спектрального анализа. – М.: Мир, 1979. – 331 с.
- 5 Боровиков, В.П. STATISTICS. Искусство анализа данных на компьютере : Дин профессионалов 2-е изд. (+CD) - СПб.: Питер. 2003. - 668 с.: ил. ISBN 5-272-0007В-1
- 6 Боровиков, В.П. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере : учеб. пособие / В.П. Боровиков, Г.И. Ивченко. – 2-е изд., перераб. И доп. - М. : Финансы и статистика, 2006. - 368 с: ил. ISBN 5-279-03059-7.
- 7 Боровиков, В.П. Программа STATISTICA для студентов и инженеров. - 2-е изд. – М.: КомпьютерПресс, 2001. – 301с. ISBN 5-89959-080-7
- 8 Бородич, С.А. Эконометрика: учеб. пособие / С.А. Бородич. – 3 –е изд., стер. – Мн.: Новое знание, 2006. – 408 с. ISBN 985-475-206-2
- 9 Бююль, А., Цефель, П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. Пер. с нем. - СПб.: «ДиаСофт», 2005. – 608 с. ISBN 5-93772-032-6, 3-8273-1673-1
- 10 Винн, Р., Холден, К. Введение в прикладной эконометрический анализ. - М.: Финансы и статистика, 1981.
- 11 Вуколов, Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: Учебное пособие. – М.: ФОРУМ: ИНФРА-М, 2004. – 464с. ISBN 5-8199-0141-Х, 5-16-002003-9
- 12 Джонстон, Дж. Эконометрические методы. – М.: Финансы и статистика, 1960.
- 13 Доугерти, К. Введение в эконометрику: Учеб. для вузов / К. Доугерти.- 2-е изд. - М. : ИНФРА-М, 2004. - 432 с. - (Университетский учебник). - Библиогр.: с. 407. ISBN 5-16-001463-2.
- 14 Дрейнер, Н., Смит, Г. Прикладной регрессионный анализ. В 2-х кн. Пер. с англ. – М.: Финансы статистика, 1986. – 366 с.
- 15 Дуброва, Т.А. Статистические методы прогнозирования: учеб. пособие для вузов. – М.: ЮНИТИ-ДАНА, 2003. – 206 с. ISBN 5-238-00497-4
- 16 Кендэл, М. Временные ряды / Пер. с англ. и предисловие Ю.П. Лукашина –М.: Финансы и статистика. 1981. – 199с.

- 17 Кильдишев, Г.С., Френкель, А.А. Анализ временных рядов и прогнозирование. – М.: Статистика. 1973.
- 18 Кремер, Н.Ш. Эконометрика: учеб. для вузов / Н.Ш. Кремер, Б.А. Путко. - М. : ЮНИТИ-ДАНА, 2003. - 311 с. ISBN 5-238-00333-1.
- 19 Кулаичев, А.П. Методы и средства анализа данных в среде Windows. STADIA. Изд. 4-е. – М.: Информатика и компьютеры, 2002. – 341с.
- 20 Лукашин, Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов: учеб. пособие. – М.: Финансы и статистика, 2003. – 416 с. ISBN 5-279-02740-5
- 21 Магнус, Я. Р. Эконометрика. Начальный курс: учебник для вузов / Я. Р. Магнус, П. К. Катыхов, А. А. Пересецкий. - 7-е изд., испр. - М. : Дело, 2005. - 504 с. ISBN 5-7749-0055-X.
- 22 Многомерные статистические методы и основы эконометрики: Учеб.-практ. пособие / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. - М. : МЭСИ, 1998. - 108 с. - (Система дистанционного образования).
- 23 Многомерный статистический анализ в экономике: Учеб. пособие для вузов / Л.А. Сошникова, В.Н. Тамашевич, Г. Уебе; Под ред. В.Н. Тамашевича. - М. : ЮНИТИ-ДАНА, 1999. - 598 с; 21см.
- 24 Носко, В. П. Эконометрика для начинающих: доп. главы / В. П. Носко. - М. : Ин-т экономики переход. периода, 2005. - 379 с. - Библиогр. : с. 372-373. - Предм. указ. : с. 374-378. ISBN 5-93255-173-9.
- 25 Носко, В. П. Эконометрика: элемент. методы и введ. в регрес. анализ врем. рядов / В. П. Носко; Ин-т экономики переход. периода. - М. : [Б. и.], 2004. - 501 с. - Библиогр.: с. 485-492. - Предм. указ.: с. 493-501. - ISBN 5-93255-141-0.
- 26 Плис, А.И., Сливина, Н.А. Mathcad 2000. Математический практикум. – М.: Финансы и статистика. 2003. – 656 с. ISBN 5-279-02281-0, 5-279-02550-X
- 27 Практикум по курсу «Статистика» (в системе STATISTICA). / В.Н. Салин, Э.Ю. Чурилова – М.: «Социальные отношения». 2002. – 188с. ISBN 5-94907-001-1
- 28 Прикладная статистика. Основа эконометрики: учебник для вузов: В 2 т. 2-е изд., испр. Т. 2: Айвазян, С.А. Основы эконометрики. - М.: ЮНИТИ-ДАНА, 2001. – 432 с. ISBN 5-238-00305-6
- 29 Рябушкин, Т.В., Френкель, А.А. Методологические проблемы анализа и прогноза краткосрочных процессов – М.: Статистика, 1979.
- 30 Симчера, В.М. Методы многомерного анализа статистических данных: учеб. Пособие. – М.: Финансы и статистика, 2008. – 400 с. ISBN 978-5-279-03184-9
- 31 Тюрин, Ю.Н., Макарова, А.А. Анализ данных на компьютере / Под ред. В.Э. Фигурнова. – 3-е изд., перераб. и доп. – М.: ИНФРА-М, 2003. – 544с. 5- ISBN 16-001316-4
- 32 Френкель, А.А. Прогнозирования производительности труда: методы и модели. – М.: Экономика, 1989. – 214 с.

33 Халафян, А.А. STATISTICA 6. Статистический анализ данных. 3-е изд. Учебник – М.: ООО «Бином-Пресс», 2007 г. – 512 с. ISBN 978-5-9518-0215-6

34 Эконометрика: учебник для вузов / под ред. И.И. Елисейевой. - М. : Финансы и статистика, 2005. - 576 с. : ил. ISBN 5-279-02786-3.

Таблица 1.1 - Динамика урожайности зерновых культур, ц/га

	У
1955г.	13,5
1956г.	11,6
1957г.	14,7
1958г.	16,2
1959г.	15,2
1960г.	15,1
1961г.	16,9
1962г.	16,8
1963г.	12,9
1964г.	13,8
1965г.	16,1
1966г.	20,4
1967г.	17,8
1968г.	18,3
1969г.	18,9
1970г.	22,8
1971г.	23,1
1972г.	19,6
1973г.	27,0

Таблица 1.2 – Показатели деятельности малых предприятий

	У	X1	X2
1	20	10	12
2	35	15	10
3	30	20	9
4	45	25	9
5	60	40	8
6	69	37	8
7	75	43	6
8	90	35	4
9	105	38	4
10	110	55	5
11	120	50	3
12	130	35	1
13	130	40	2
14	130	55	3
15	135	45	1
16	140	65	2

где: У – Прибыль предприятия, тыс. руб.

X1 – Среднегодовая стоимость основных производственных фондов, тыс. руб.

X2 – Численность работников предприятия, чел.

Годы	Вариант 1	Вариант 2	Вариант 3	Вариант 4	Вариант 5
	Y1	Y2	Y3	Y4	Y5
1993г.	148673,4	58,7	45,2	27,1	13,3
1994г.	148365,8	220,4	206,3	108,8	36,5
1995г.	148306,1	472,4	515,5	267,0	80,8
1996г.	147976,4	790,2	770,0	376,0	103,8
1997г.	147502,4	950,2	942,1	408,8	130,4
1998г.	147104,6	1051	1013,0	407,1	187,8
1999г.	146693,3	1523	1664,0	670,4	266,6
2000г.	145924,9	2223	2290,1	1165,2	419,3
2001г.	145173,8	3240	3078,4	1504,5	584,3
2002г.	144309,2	4360	3947,2	1758,7	763,3
2003г.	145166,7	5499	5170,4	2183,3	1147,0
2004г.	144168,2	6740	6383,3	2729,8	1534,8
2005г.	143400,0	8550	7848,4	3431,0	2009,2

Y1 – Имеются следующие данные о численность населения России (на начало года), тыс.чел

Y2 – Имеются данные о средней номинальной заработной плате в РФ, руб./мес.

Y3 – Имеются следующие данные о среднедушевых денежных доходах населения России, руб./мес.

Y4 – Имеются следующие данные о динамике инвестиций в основной капитал (текущие цены), млрд. (трлн.) рублей

Y5 – Имеются данные о динамике денежного агрегата M0 (на конец периода), млрд. (трлн.) руб.

Вариант 1

Президента компании интересует зависимость между приростом годового дохода (X) измеряемый в процентах и качеством работы (Y) коммерческих агентов измеряемое в баллах. Он выбрал 12 агентов и определил размеры дохода, приносимого компании каждым из них (в процентах от окладов), а также количество продаж, проведенных каждым агентом в течение года.

Вариант 2

Используя приведенные в таблице данные, установите, есть ли значимая зависимость между X - объем инвестиций (млн. руб.) и Y - цена за акцию (руб.).

Вариант 3

Автосервисное предприятие имеет следующие данные по стоимости ежегодного технического обслуживания автомобилей (Y , тыс. руб.) определенной марки в зависимости от времени эксплуатации (X , лет).

Вариант 4

В таблице приведены данные, характеризующие стоимость продукции (Y , тыс. руб.) в зависимости от количества исходного сырья (X , тонн).

Вариант 5

Определите регрессионную модель, описывающую зависимость нераспределенной прибыли (Y , млн. руб.) от инвестиций в основные фонды (X , млн. руб.) по результатам обследования 11 предприятий.

№ пп	Вариант 1		Вариант 2		Вариант 3		Вариант 4		Вариант 5	
	Y	X	Y	X	Y	X	Y	X	Y	X
1	64	7,8	12	108	5,3	5	10	0,25	2,3	0,3
2	73	6,9	4	4,4	5,2	4	7	0,2	3,4	0,3
3	42	6,7	5	3,5	6	5	5	0,16	4,3	0,4
4	49	6	6	3,6	5,7	6	6	0,17	5,0	0,6
5	71	6,9	13	39	6,6	7	7	0,19	6,0	1
6	46	5,2	19	68,4	6,8	8	6	0,18	2,0	0,16
7	32	6,3	8,5	7,5	8,1	10	4	0,15	3,6	0,2
8	88	8,4	5	5,5	6,9	8	3	0,14	4,2	0,3
9	53	7,2	15	375	10,3	11	8	0,21	5,8	1
10	84	10,1	6	12	4	3	9	0,21	4,7	0,6
11	85	10,8	12	51	2,5	2	10	0,24	2,7	0,11

Приложение 4

Выборочные данные по 25 предприятий одной из отраслей экономики

№	Y	X
1	166	127
2	76	69
3	112	73
4	32	29
5	49	45
6	150	128
7	120	78
8	7	8
9	53	41
10	48	43
11	57	55
12	48	43
13	109	91
14	12	14
15	86	76
16	36	36
17	67	44
18	84	69
19	69	46
20	67	58
21	179	147
22	104	74
23	155	109
24	30	33
25	28	27

где: Y - стоимость произведенной продукции, тыс. руб.

X - среднегодовая стоимость основных производственных фондов, тыс. руб.

Вариант 1. За изучаемый период имеются следующие данные о работе малых предприятий отрасли:

Y- Произведено продукции, тыс. т

X- Сумма затрат на производство продукции, тыс. руб.

Вариант 2. За отчетный период имеются следующие данные о реализации товаров и издержках обращения по предприятиям торговли района, млн. руб.

Y- Розничный товарооборот

X-Сумма издержек обращения

Вариант 3. Имеются следующие данные по 25 предприятиям одной из отраслей промышленности:

Y- Возраст оборудования, лет

X- Затраты на капитальный ремонт, млн. руб.

Вариант 4. Имеются следующие выборочные данные по 25 предприятиям

Y- Розничный товарооборот, млн. руб.

X- Издержки обращения, млн. руб.

Вариант 5. Имеются следующие выборочные данные по 25 предприятиям:

Y- Объем работ, млн. руб.

X- Накладные расходы, млн. руб.

№ п/п	Вариант 1		Вариант 2		Вариант 3		Вариант 4		Вариант 5	
	Y	X	Y	X	Y	X	Y	X	Y	X
1	3,6	330	5,1	0,30	5,6	6,7	510	30	9,0	2,7
2	4,6	396	5,6	0,34	6,8	23,0	560	33	10,3	3,0
3	5,5	460	7,0	0,46	10,6	24,2	800	46	7,0	2,5
4	4,8	430	4,6	0,30	3,9	12,0	465	31	5,2	2,2
5	2,7	243	3,3	0,15	7,0	20,0	225	16	6,4	2,5
6	2,9	170	3,9	0,25	8,4	14,8	390	25	9,5	2,7
7	7,5	618	6,4	0,42	8,0	27,0	640	39	14,0	4,0
8	6,3	540	4,4	0,26	5,8	6,9	405	26	13,0	4,0
9	4,1	369	5,0	0,16	6,4	10,0	200	15	5,0	2,0
10	4,8	425	4,2	0,34	8,5	15,0	425	34	7,4	2,6
11	7,6	646	5,0	0,37	3,9	9,3	570	37	9,3	2,6
12	6,5	598	4,2	0,28	5,2	13,0	472	28	8,0	3,2
13	11,5	858	2,0	0,18	7,5	16,7	250	19	10,2	2,3
14	10,6	820	6,6	0,39	4,0	8,0	665	38	10,0	3,0
15	9,0	810	6,5	0,36	3,5	9,5	650	36	12,0	2,6
16	6,9	566	6,2	0,36	10,2	24,5	620	35	15,0	3,0
17	5,0	450	3,8	0,25	6,2	14,1	380	24	16,0	5,0
18	11,2	858	5,5	0,38	4,3	10,9	550	38	17,0	4,3
19	8,1	656	7,5	0,44	3,5	9,0	750	44	21,0	5,0
20	7,8	640	6,6	0,37	6,0	11,0	660	36	19,0	4,8
21	4,2	399	4,5	0,27	6,2	10,2	450	27	12,5	4,0
22	6,3	518	5,6	0,35	3,0	8,0	563	34	8,0	2,0
23	12,1	920	6,0	0,40	8,9	12,6	400	26	11,0	3,0
24	9,8	780	4,0	0,24	9,0	14,0	553	38	6,5	2,0
25	8,5	696	3,5	0,24	4,0	15,0	772	45	13,0	5,0

Источник - Данные из работы Вуколова Э.А. – IV, 6

Выборочные данные по 15 промышленным предприятиям

	<i>Y</i>	<i>X1</i>	<i>X2</i>
1	26	37	39
2	33	33	40
3	24	15	35
4	29	36	48
5	42	26	53
6	24	24	42
7	52	15	54
8	56	33	54
9	26	44	45
10	45	34	53
11	27	63	46
12	54	8	55
13	34	44	43
14	48	43	55
15	45	31	51

где: *Y* - Фондоотдача, руб./чел

X1 - Среднечасовая производительность труда, ед.

X2 - Удельный вес активной части ОПФ, %

Приложение 7

Вариант 1 (1995г.)

Страны СНГ	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Россия	2 276	66,4	6712	29,5	7,4	4,6	82913	62603	1460	1570	128000
Азербайджан	313	3,6	28	27,3	25,1	4417	680	955	330	119	425
Армения	408,9	1,5	132	24,2	40,7	405,9	271	674	25	100	387
Белорусия	1009	4,4	131	31,4	17,7	11,5	4803	5469	15	377	1,527
Грузия	535	1,7	1364	14	29,8	1,3	363	700	6	157	1217
Казахстан	1040	6,6	808	23,5	12,3	61,1	5440	5326	964	1136	4765
Киргизия	330,7	1,6	1422	10,8	36,6	10,8	408,9	530,9	96,1	81	913,8
Туркменистан	592	1,7	1235	52,8	16,2	240	2084	1644	233	1165	550
Таджикистан	114	1,9	0,034	34	36,2	0,104	779	838	10	4	869
Украина	720	23,7	1437	34,4	14,5	1,47	14244	16946	257	1051	7941
Узбекистан	404	8,4	25	20	32	33	3475	3238	-24	815	2004
Молдова	387	1,7	1502	25	29,3	4,5	739	794	73	257	668

Вариант 2 (1996г.)

Страны СНГ	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Россия	2 829	66	6732	28,2	7	5,1	90563	68092	1657	1290	136100
Азербайджан	407	3,7	32	25,8	24,7	4300	789	1338	627	214	521
Армения	507	1,4	159	23,8	33	414	290	760	18	171	533
Белорусия	1399	4,4	182	34,6	16	13,3	5790	6939	105	469	991
Грузия	563	1,7	45	11,4	27	1,3	417	768	54	158	1357
Казахстан	1333	6,5	971	23,5	11,7	67,8	6292	6627	1137	1295	5807
Киргизия	398,4	1,7	1469	10,1	42,4	12,8	531,2	782,9	46,8	94,6	1301,3
Туркменистан	464	1,8	49	54,4	12,6	3546	1692	1388	108	1172	668
Таджикистан	177	1,7	46	25,7	36	0,293	770	786	18	14	867
Украина	869	23,2	1998	34	12,2	1,83	15547	19843	516	1960	8839
Узбекистан	541	8,6	28	20	26	44,7	3534	4240	90	772	2357
Молдова	443	1,7	51	23,1	27,5	4,6	823	1075	23	312	815

Вариант 3 (1997г.)

Страны СНГ	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Россия	2 903	64,7	8058	26,8	7	5,8	89008	71983	1679	1480	134600
Азербайджан	503	3,7	38	25,2	20	3983	808	1375	1115	467	602
Армения	522,2	1,4	174	23,9	30,4	490,8	234	793	52	243	679
Белорусия	1368	4,4	126	34,3	15,4	26,2	6919	8326	350	394	976
Грузия	657	1,7		12,5	35,5	1,3	494	1052	236	173	1508
Казахстан	1429	6,5	968	24	11,4	75,6	6899	7176	1320	1697	7750
Киргизия	382,1	1,7	1414	16,4	40,8	17,4	630	646	83	169,8	1711,7
Туркменистан	495	1,8		32,9	20,2	4627	743	1003	108	1285	1356
Таджикистан	187	1,8	51	19,7	27,1	0,564	746	806	18	30	1106
Украина	985	22,6	2330	34,3	14,4	1,86	15418	19623	581	2374	9634
Узбекистан	457,1	8,7	29	19	29	90,7	3695	3767	167	374	2864
Молдова	507	1,6		20,2	26	4,6	890	1237	78	366	1335

Продолжение приложения 7

Вариант 4 (1998г.)

Страны СНГ	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Россия	1802	6,8	8902	29,9	5,7	10	74444	58015	1492	7801	185700
Азербайджан	559	3,7	42	22,2	18	3869	678	1724	1023	447	708
Армения	605	1,3	134	19,9	30,8	504,9	229	806	221	298	787
Белорусия	1503	4,4	106	33,4	13,9	46,1	6172	7673	201	703	1,011
Грузия	771	1,7	295	12,3	26,7	1,4	478	1164	221	118	1636
Казахстан	1457	6,1	925	23,9	9,4	78,3	5871	6672	1143	1461	9932
Киргизия	350	1,7	1305	20,8	36,1	20,8	535	756	87	164	1480
Туркменистан	510	1,8	304	29	26,5	5500,4	614	1137	62	1379	1842
Таджикистан	216	1,8	54	18,1	19,8	0,8	586	725	25	65	1179
Украина	828	22,3	2937	25,2	11,9	2,5	13699	16283	747	761	11973
Узбекистан	449	8,8	33	14,9	26,8	131,8	3049	2938	140	533	3230
Молдова	449	1,6	167	19,5	30,5	5,4	644	1032	76	144	1451

Вариант 5 (1999г.)

Страны СНГ	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Россия	1347	64	9094	30,8	7,7	24,6	75551	39537	1102	8457	177100
Азербайджан	571	3,7	45	22,2	18,2	4120,2	1025	1433	510	673	870
Армения	590	1,3	175	21,2	27	535,1	247	721	122	305	855
Белорусия	1207	4,4	95	31,9	14,6	248,8	5646	6216	443	309	886
Грузия	524	1,7	277	13	24,7	2	477	1013	62	132	1722
Казахстан	1132	6,1	950	23,9	11,1	119,5	5989	5648	1468	1479	12081
Киргизия	266	1,8	1190	24,7	34,9	39	463	551	38	230	1647
Туркменистан	387	1,9	102	33,6	27,8	8523,9	1164	1374	125	1555	2026
Таджикистан	174	1,7	50	21,3	18,7	1,2	666	693	21	58	1233
Украина	631	21,8	2699	27,2	11,7	4,1	13189	12945	489	1046	13532
Узбекистан	340	8,9	39	14,3	29	257,2	2790	2587	121	763	4805
Молдова	306	1,5	187	18,8	27,9	10,5	475	610	154	186	1495

Условные обозначения используемых показателей:

Y - ВВП на душу населения (долл. США)

X1 - Среднегодовая численность занятых в экономике, млн. чел.

X2 - Общая численность безработных, тыс. чел.

X3 - Доля промышленности в ВВП, %

X4 - Доля сельского хозяйства в ВВП, %

X5 - Обменный курс, ед. национальной валюты за 1 доллар США

X6 - Экспорт, млн. долл. США

X7 - Импорт, млн. долл. США

X8 - Прямые иностранные инвестиции в чистой сумме, млн. долл. США

X9 - Валовые резервы (без учета золота), млн. долл. США

X10 - Внешний долг, млн. долл. США

Вариация численности безработных и факторов оказывающих на нее влияние

	Y	X1	X2	X3	X4	X5
Абдулинский	123	1635	31,2	-76	5,8	1385
Адамовский	74	10019	22,7	227070	4,5	3057
Акбулакский	293	7182	33,3	44262	10,0	2767
Александровский	105	4916	31,6	21357	1,2	2699
Асекеевский	160	7130	31,0	50534	13,9	2447
Беляевский	131	5423	36,7	27733	2,7	2888
Бугурусланский	370	5741	54,2	-10236	32,4	2457
Бузулукский	397	7252	44,8	43250	11,0	3492
Гайский	357	3283	28,6	97147	30,0	3143
Грачевский	35	5021	20,0	4935	3,6	3155
Домбаровский	91	4310	25,0	15363	5,4	3093
Илекский	93	8241	20,0	-4858	10,7	2158
Кваркенский	63	6392	22,1	37692	5,4	2731
Красногвардейский	90	6184	30,0	866	7,8	3142
Кувандыкский	424	3868	43,5	42731	33,0	2566
Курманаевский	103	6346	20,0	67055	7,0	3219
Матвеевский	72	4674	25,0	25577	6,5	2219
Новоорский	60	8057	15,6	-34445	5,2	4915
Новосергиевский	176	11268	25,7	214151	2,8	3222
Октябрьский	52	6914	21,2	36755	2,1	3329
Оренбургский	73	17532	29,3	334646	6,9	4870
Первомайский	192	8816	43,5	-21727	28,0	3607
Перволюцкий	226	6774	35,0	40542	22,6	2919
Пономаревский	105	4505	29,4	35256	10,9	2747
Сакмарский	143	5871	36,8	24859	3,6	3287
Саракташский	77	11691	27,5	24409	8,9	2765
Светлинский	53	5172	25,0	16044	9,3	3348
Северный	50	5144	27,1	-27504	1,2	3360
Соль-Илецкий	119	5948	52,2	34542	22,2	1895
Сорочинский	470	5534	57,6	38867	30,0	2464
Ташлинский	94	9176	24,3	128652	2,6	2460
Тоцкий	385	6225	43,3	22859	37,1	2972
Тюльганский	262	6589	56,0	-32454	14,5	2611
Шарлыкский	97	5897	19,0	33186	6,6	2788
Ясенский	184	1211	30,0	15054	9,1	1797

Y - Численность официально зарегистрированных безработных, тыс. чел.

X1 - Среднесписочная численность работников, чел.

X2 - Удельный вес убыточных предприятий и организаций, %

X3 - Прибыль (убыток) предприятий, организаций, тыс. руб.

X4 - Задолженность предприятий, организаций по ЗП, тыс. руб.

X5 - Среднемесячная заработная плата, руб.

Динамика ВВП России

	GDP
I/1999	1259,1
II/1999	1514,7
III/1999	1632,5
IV/1999	2015,6
I/2000	1688,1
II/2000	1970
III/2000	2152,2
IV/2000	2582,9
I/2001	2267,5
II/2001	2950,4
III/2001	3193,2
IV/2001	3649,6
I/2002	3219,7
II/2002	3735,2
III/2002	4012,3
III/2002	4785,1
I/2003	4354
II/2003	4916,8
III/2003	5147,8
IV/2003	6205,3
I/2004	5532,8
II/2004	5971,3
III/2004	6352,8
IV/2004	7720,5
I/2005	6532,3
II/2005	7634,9
III/2005	7959,8
IV/2005	9451,8
I/2006	7873
II/2006	9575,8

Динамика макроэкономических показателей

Годы	Y1	Y2	Y3	Y4	Y5
1993г.	148673,4	58,7	45,2	27,1	13,3
1994г.	148365,8	220	206,3	108,8	36,5
1995г.	148306,1	472	515,5	267,0	80,8
1996г.	147976,4	790	770,0	376,0	103,8
1997г.	147502,4	950	942,1	408,8	130,4
1998г.	147104,6	1051	1013,0	407,1	187,8
1999г.	146693,3	1523	1664,0	670,4	266,6
2000г.	145924,9	2223	2290,1	1165,2	419,3
2001г.	145173,8	3240	3078,4	1504,5	584,3
2002г.	144309,2	4360	3947,2	1758,7	763,3
2003г.	145166,7	5499	5170,4	2183,3	1147,0
2004г.	144168,2	6740	6383,3	2729,8	1534,8
2005г.	143400,0	8550	7848,4	3431,0	2009,2

- где: Y1- Численность населения (на начало года), тыс.чел
Y2 - Средняя номинальная заработная плата, руб./мес.
Y3 - Среднедушевые денежные доходы населения, руб./мес.
Y4 - Инвестиции в основной капитал текущие цены, млрд. (трлн.) рублей
Y5 - M0 (на конец периода), млрд. (трлн.) руб.

Динамика среднедушевых денежных доходов населения, руб./мес.

Квартал/Год	У
I/1999	1259,1
II/1999	1514,7
III/1999	1632,5
IV/1999	2015,6
I/2000	1688,1
II/2000	1970,0
III/2000	2152,2
IV/2000	2582,9
I/2001	2267,5
II/2001	2950,4
III/2001	3193,2
IV/2001	3649,6
I/2002	3219,7
II/2002	3735,2
III/2002	4012,3
IV/2002	4785,1
I/2003	4354
II/2003	4916,8
III/2003	5147,8
IV/2003	6205,3
I/2004	5532,8
II/2004	5971,3
III/2004	6352,8
IV/2004	7720,5
I/2005	6532,3
II/2005	7634,9
III/2005	7959,8
IV/2005	9451,8
I/2006	7873,0
II/2006	9593,1
III/2006	9979,5
IV/2006	11723,0

Динамика денежного агрегата М0 (на конец периода), млрд. (трлн.) руб.

Квартал/Год	У	Квартал/Год	У
1992 I	0,3	2000 I	251,5
II	0,5	II	321,8
III	0,9	III	351
IV	1,7	IV	419,3
1993 I	2,6	2001 I	399,4
II	5,1	II	474,7
III	8,4	III	531
IV	13,3	IV	584,3
1994 I	15,9	2002 I	552,9
II	23,8	II	645,9
III	30	III	672,6
IV	36,5	IV	763,3
1995 I	35,3	2003 I	749,5
II	55,2	II	917,1
III	68,4	III	957,1
IV	80,8	IV	1147
1996 I	86,7	2004 I	1165,5
II	104,4	II	1276,1
III	96,2	III	1293,7
IV	103,8	IV	1534,8
1997 I	105,2	2005 I	1481,7
II	136,9	II	1650,7
III	134,9	III	1740,7
IV	130,4	IV	2009,2
1998 I	119,1	2006 I	1928,8
II	129,8	II	2233,4
III	154,2	III	2400,8
IV	187,8	-	-
1999 I	174,1	-	-
II	216,4	-	-
III	212,8	-	-
IV	266,6	-	-

Квартал/Год	М2 (на конец периода), млрд. (трлн.) руб.	Международные резервы ЦБР (на конец периода), млрд. долл.	Валютные резервы ЦБР (на конец периода), млрд. долл.	Доходы федерального бюджета, млрд. (трлн.) руб.	Расходы федерального бюджета, млрд. (трлн.) руб.
	Y1	Y2	Y3	Y4	Y5
1999 I	473,8	10,8	6,7	89,1	108,3
II	567,7	12,2	8,2	137,5	163,8
III	597,4	11,2	6,6	160,7	162,1
IV	704,7	12,5	8,5	224,4	230,5
2000 I	751,4	15,5	11,5	221,8	191,5
II	892,2	21	17,7	285,9	213,9
III	992,4	25	21,5	275,5	226,8
IV	1144,3	28	24,3	344,4	321,9
2001 I	1149,5	29,709	25,942	318	268,8
II	1294,3	35,053	31,251	395,1	310
III	1414,4	37,957	34,04	395,3	350,8
IV	1602,6	36,622	32,542	482,3	396,1
2002 I	1562,4	37,295	33,179	472,4	364,4
II	1751,1	43,579	39,848	518,7	463,9
III	1846,6	45,619	41,887	586,9	503,3
IV	2119,6	47,793	44,054	624,2	714,4
2003 I	2208,5	55,525	51,79	580,9	491,6
II	2604,5	64,43	60,691	663	568,1
III	2732,1	62,073	58,33	652,4	597,7
IV	3212,7	76,938	73,175	686,8	701,2
2004 I	3430,4	83,398	79,639	688,6	412,8
II	3694	88,226	84,463	850,1	630,7
III	3727,5	95,082	91,328	886,3	652,4
IV	4363,3	124,541	120,809	1001,3	861,9
2005 I	4474,6	137,381	133,651	1196,4	671,1
II	4927,4	151,578	147,776	1190	773,1
III	5292,8	159,56	155,83	1295,3	1075,5
IV	6045,6	182,24	175,891	1443,4	992,5
2006 I	6169,4	205,881	198,676	1399,1	823,2
II	7092,3	250,561	243,175	1539	1031,5
III	7757,5	266,197	258,705	-	-

Среднедушевые денежные доходы населения, руб./мес.

Год/квартал	У	Год/квартал	У
1993 I	13,8	2000 I	1688,1
II	26,1	II	1970
III	50	III	2152,2
IV	89,3	IV	2582,9
1994 I	116,9	2001 I	2267,5
II	168,9	II	2950,4
III	223,2	III	3193,2
IV	316,4	IV	3649,6
1995 I	351,6	2002 I	3219,7
II	485,4	II	3735,2
III	575,6	III	4012,3
IV	646,9	IV	4785,1
1996 I	674,9	2003 I	4354
II	743	II	4916,8
III	761,5	III	5147,8
IV	880,7	IV	6205,3
1997 I	829,4	2004 I	5532,8
II	929,9	II	5971,3
III	929,5	III	6352,8
IV	1042,2	IV	7720,5
1998 I	835,1	2005 I	6532,3
II	881	II	7634,9
III	932,7	III	7959,8
IV	1341,4	IV	9451,8
1999 I	1259,1	2006 I	7873,0
II	1514,7	II	9593,1
III	1632,5	III	9979,5
IV	2015,6	IV	11723,0

Приложение 15

Год/ квартал	Средняя номинальная зарботная плата, руб./мес.	Среднедушевые денежные доходы населения, руб./мес.	Инвестиции в основной капитал (текущие цены), млрд. (трлн.) рублей	Ввод в действие жилых домов, млн.кв.метров	Оборот розничной торговли в текущих ценах, млрд.руб.
	Y1	Y2	Y3	Y4	Y5
1	2	3	4	5	6
1993 I	19,4	13,8	1,6	4,2	-
II	38,5	26,1	3,8	7,3	-
III	67,4	50	6,2	7,3	-
IV	109,5	89,3	15,6	23	-
1994 I	148,1	116,9	12,3	4	-
II	187,5	168,9	22,1	6,6	-
III	232,3	223,2	26,4	6,9	-
IV	298,6	316,4	48	21,7	-
1995 I	328,2	351,6	29,5	4,3	88,3
II	432,2	485,4	48,4	7,7	114,5
III	528,3	575,6	66,2	7,4	140,6
IV	643,2	646,9	122,8	21,6	168,5
1996 I	695,3	674,9	57,8	4,4	167,9
II	789,7	743	75,2	7,7	175,7
III	838,3	761,5	85,2	5,6	183,8
IV	903,8	880,7	157,8	16,6	199,7
1997 I	844	829,4	73,2	4	198,2
II	941	929,9	85,4	6,2	202,8
III	1001	929,5	108,9	6,1	211,9
IV	1074	1042,2	141,3	16,4	237
1998 I	1010	835,1	71,9	4,2	212,3
II	1069	881	83,9	5,4	211,6
III	1094	932,7	107,6	5,7	259,2
IV	1255	1341,4	143,7	15,4	350,3
1999 I	1248	1259,1	96,8	4,1	372,6
II	1511	1514,7	131,1	5,5	409,4
III	1642	1632,5	185,6	5,9	456,7
IV	1927	2015,6	256,9	16,5	528,5
2000 I	1899	1688,1	165,8	4,1	508,5
II	2148	1970	236	5,7	533
III	2336	2152,2	330,2	5,9	587
IV	2652	2582,9	433,2	14,6	678,2
2001 I	2781	2267,5	230,3	4	658,7
II	3082	2950,4	318,8	5,7	722,4
III	3393	3193,2	421,1	6,6	775,5
IV	3872	3649,6	534,5	15,4	883,3
2002 I	3836	3219,7	270,1	4,8	825,4
II	4257	3735,2	376,4	6,4	881,5
III	4547	4012,3	494,5	7,3	960,5
IV	5018	4785,1	621,4	15,3	1086,2

Продолжение приложения 15

1	2	3	4	5	6
2003 I	4800	4354	330	5,6	1010,5
II	5296	4916,8	470,6	7,1	1068,5
III	5549	5147,8	607,5	7,6	1138,8
IV	6401	6205,3	778,3	16	1297
2004 I	6173	5532,8	423,9	6	1227,5
II	6650	5971,3	603,4	7,2	1311,8
III	6930	6352,8	774	8,6	1419,1
IV	7582	7720,5	1003,5	19,2	1639,3
2005 I	7638	6532,3	526,5	6,7	1510,5
II	8234	7634,9	755,7	7,7	1674,1
III	8674	7959,8	980	8,9	1796,1
IV	9651	9451,8	1271,8	20,3	2057,6
2006 I	9397	7873	624,2	6,3	1841,5
II	10401	9619,6	958,4	9,4	2043,4
III	10908	9988,2333	1230,1	10,3	2203,8
IV	-	-	-	-	-

Стоимость однокомнатных квартир в Москве

Данные из газеты «Из рук в руки» за период с декабря 1996 г. по сентябрь 1997 г.

Была выбрана Юго-Западная часть города, в которой высок спрос на жилые площади (всего 69 наблюдений).

N	Регион	X1	X2	X3	X4	X5	D1	D2	Y
1	2	3	4	5	6	7	8	9	10
1	Фрунзенская	4	10	34	7,5	19	1	1	54
2	Ленинский проспект	5,7	7	36	10	20	0	0	35
3	Ленинский проспект	5,7	12	45	13	20	1	1	59
4	Академическая	7,6	10	35,3	10	20	1	0	35
5	Университет	8,7	6	33	5,5	22	1	0	33
6	Новые Черемушки	10,3	3	33	8,5	18	1	1	57
7	Юго-Запад	13,3	10	37	10	19	1	0	43
8	Коньково	14,8	2	38	8,5	19,1	1	0	39
9	Фрунзенская	4	15	54	9,2	27,2	1	1	70
10	Университет	8,7	15	35	6	20	0	1	43
11	Проспект Вернадского	11,4	10	31,4	5,2	21,3	1	0	33
12	Ленинский проспект	5,7	7	32	6	21	1	0	37
13	Новые Черемушки	10,3	7	38	8	19	0	0	33
14	Университет	8,7	10	31,6	8,8	14	0	0	31
15	Юго-Запад	13,3	5	32	8	17	1	0	37
16	Юго-Запад	13,3	10	37	10	19	1	0	43
17	Ленинский проспект	5,7	5	32	8	17	1	1	38
18	Академическая	7,6	10	37	8	19	1	1	51
19	Академическая	7,6	15	32,2	6,5	17	0	1	30
20	Коньково	14,8	3	33	8	19	1	0	30
21	Коньково	14,8	5	37,5	9,6	19,8	1	0	36
22	Коньково	14,8	10	33	7	19	1	0	33
23	Университет	8,7	15	32	6	21,5	1	0	35
24	Проспект Вернадского	11,4	5	29,7	6	16,1	0	0	28
25	Проспект Вернадского	11,4	15	36	8,6	18	0	0	40
26	Юго-Запад	13,3	15	36	10	19	0	0	33
27	Ленинский проспект	5,7	2	31,6	6	21,6	1	1	35
28	Ленинский проспект	5,7	5	52	12	34	1	1	75
29	Коньково	14,8	3	36	10	19	1	0	40
30	Коньково	14,8	5	33	8	18	1	0	30
31	Университет	8,7	5	32	5,5	20,1	1	0	31
32	Академическая	7,6	15	35	9,8	20	1	0	37
33	Новые Черемушки	10,3	15	38	10	19,5	1	0	40
34	Коньково	14,8	1	39	8,5	19	1	0	40
35	Фрунзенская	4	5	34	8	19	1	1	58
36	Фрунзенская	4	10	38	6,5	18	0	1	48
37	Проспект Вернадского	11,4	3	35	10	20	1	0	40
38	Юго-Запад	13,3	7	36	9	19,5	1	0	42
39	Новые Черемушки	10,3	7	34	8	18	1	1	51
40	Коньково	14,8	5	38	8,5	19	1	0	43

Продолжение приложения 16

1	2	3	4	5	6	7	8	9	10
41	Коньково	14,8	7	33	6	19	1	0	30
42	Коньково	14,8	10	32	8	17	1	0	40
43	Коньково	14,8	10	38	8,5	19,1	1	0	43
44	Академическая	7,6	5	43	8,5	25	0	1	53
45	Академическая	7,6	10	30	6	18,3	1	1	28
46	Коньково	14,8	7	34,8	7,8	17,8	0	0	29
47	Коньково	14,8	15	35	10	19,6	1	0	37
48	Коньково	14,8	3	32,8	6,5	18,5	1	0	30
49	Новые Черемушки	10,3	10	39	9	19	1	0	45
50	Университет	8,7	15	49	9	20,5	0	1	52
51	Фрунзенская	4	3	32	6,2	19	1	1	53
52	Проспект Вернадского	11,4	10	33	6,5	19	1	0	32
53	Проспект Вернадского	11,4	15	32,3	6	21,9	0	0	28
54	Юго-Запад	13,3	10	30	7	19,8	1	0	34
55	Юго-Запад	13,3	10	34	9	19	1	0	42
56	Юго-Запад	13,3	7	33	7	19	0	0	33
57	Академическая	7,6	10	30	6	18,3	1	1	28
58	Академическая	7,6	15	32	6	18	1	0	30
59	Коньково	14,8	5	33,1	7,5	18	1	0	32
60	Коньково	14,8	2	38	7,5	19	1	0	41
61	Коньково	14,8	7	38	8,6	19	1	0	43
62	Коньково	14,8	5	37,3	6,5	19	1	0	31
63	Ленинский проспект	5,7	8	31,4	5,6	21	1	0	33
64	Ленинский проспект	5,7	7	52	10	34	1	1	60
65	Новые Черемушки	10,3	15	30	6	17	1	1	37
66	Новые Черемушки	10,3	5	36	11	20	1	0	41
67	Проспект Вернадского	11,4	5	28	6,7	14,4	1	0	35
68	Проспект Вернадского	11,4	10	31,4	5,2	21,3	1	0	33
69	Юго-Запад	13,3	5	32	8	17	1	0	37

- N — номер по порядку;
 X1 — удаленность от центра, км;
 X2 — удаленность от метро, мин;
 X3 — общая площадь квартиры, м²;
 X4 — площадь кухни, м²;
 X5 — площадь комнаты, м²;
 D1 — этаж: 0 — первый/последний, 1 — нет;
 D2 — категория дома: 1 — кирпичный, 0 — нет;
 Y — цена квартиры, тыс. USD.

Объем ВВП в текущих ценах, млрд. руб.

	GDP
1994 I	88
II	130
III	168
IV	225
1995 I	235
II	324,3
III	421,1
IV	448,1
1996 I	425,3
II	468,4
III	548,9
IV	565,2
1997 I	512,4
II	555,1
III	634,2
IV	640,8
1998 I	550,9
II	602,5
III	675,5
IV	800,7
1999 I	901,3
II	1101,5
III	1373,1
IV	1447,3
2000 I	1527,4
II	1696,6
III	2037,8
IV	2043,8
2001 I	1900,9
II	2105
III	2487,9
IV	2449,8
2002 I	2269
II	2528,4
III	3021,1
IV	3015,7
2003 I	2892
II	3112,5
III	3603,8
IV	3676,8
2004 I	3527,7
II	3938,2
III	4589,6
IV	4723,3

Факторы для оценки производственной функции Кобба-Дугласа

	<i>Y</i>	<i>K</i>	<i>L</i>
Республика Башкортостан	197163	202949	396,8
Республика Марий Эл	16817	14591	71,2
Мордовия республика	28629	20170	97,7
Республика Татарстан	281847	278621	399,6
Удмуртская республика	85030	82503	221,7
Чувашская республика	39664	45766	151,9
Кировская обл.	45694	43421	194,6
Нижегородская обл.	196365	126031	526,4
Оренбургская обл.	103299	160149	202,4
Пензенская обл.	29439	26583	143,6
Пермская обл.	199365	185376	372,5
Самарская обл.	297551	337460	421,8
Саратовская обл.	95034	109757	226,9
Ульяновская обл.	47784	34628	170,8

где: *Y* - объем промышленной продукции, млн. руб.;
K - стоимости основных фондов, млн. руб.;
L - среднегодовая численность занятых, тыс. чел..

Динамика количества безработных в РФ

Год/квартал	У
1999 I	10,2
II	9,4
III	8,7
IV	9
2000 I	8,6
II	7,7
III	7,2
IV	7
2001 I	7
II	6,3
III	6,1
IV	6,3
2002 I	6
II	5,6
III	5,3
IV	6
2003 I	6,6
II	6,2
III	6
IV	6,1
2004 I	6,6
II	5,9
III	5,5
IV	6
2005 I	6
II	5,5
III	5,3
IV	5,5
2006 I	5,8
II	5,6
III	5,4
IV	-

Приложение 20

Основные показатели деятельности промышленности по Приволжскому федеральному округу в 2003г.

	Вариант 1			Вариант 2			Вариант 3		
	Сельское хозяйство			Строительство			Связь		
	У	К	L	У	К	L	У	К	L
Республика Башкортостан	50769	294,2	40649	29246,4	155,3	12558	796,2	21,8	7350
Республика Марий Эл	8011	56,5	5705	2260	16,1	806	570,4	3,9	1136
Мордовия республика	11958	70	11803	4136,5	25,3	8391	505,3	5,4	2157
Республика Татарстан	46449	218,3	43568	36571,5	141,5	18088	787,5	19,6	7504
Удмуртская республика	14999	93,2	13844	9078,6	45,2	3927	706,3	8,3	3767
Чувашская республика	11617	109,9	14342	5652,6	40,6	2302	463,1	6,7	1649
Кировская обл.	15492	93,6	17959	4282,4	32	5181	480,9	10,2	2997
Нижегородская обл.	21249	141	19793	18464,4	114,6	5577	636,6	20,4	8073
Оренбургская обл.	25557	189,3	23802	13075,5	68,9	3744	810,7	13,5	3402
Пензенская обл.	11661	119,6	13924	5491,4	52,2	1741	559,7	8,1	3123
Пермская обл.	17454	103,9	14089	16799,1	116,9	6170	1035	14,8	12420
Самарская обл.	22181	110	23248	20195,5	130,2	5959	1468,6	22,1	13406
Саратовская обл.	30537	168,1	20698	9910,8	84,6	4271	785,5	17,1	4795
Ульяновская обл.	10072	90,1	14703	3703,9	32,7	2557	878,1	7,5	2863

Основные показатели деятельности промышленности по Сибирскому федеральному округу в 2003г.

	Вариант 4			Вариант 5		
	Промышленность			Строительство		
	У	К	L	У	К	L
Респ. Алтай	814	2354	6,2	932,5	220	3,7
Респ. Бурятия	26170	26114	69,2	5211,7	2200	20,9
Респ. Тыва	1523	1328	9,1	649,5	134	4,8
Респ. Хакасия	20728	27722	55,6	2494,5	525	17,2
Алтайский край	60340	62142	202	5374,9	2040	53,1
Красноярский край	231365	237466	350,2	18953,8	4617	95,6
Иркутская обл.	139439	171634	286	10581,7	23030	72,8
Кемеровская обл.	200731	156090	396,7	14072,1	6297	74,5
Новосибирская обл.	77787	81533	218,2	9821,4	6005	68,9
Омская обл.	56964	62481	166,8	7843,8	3430	66,9
Томская обл.	57636	92207	97,1	13479,7	2859	48,8
Читинская обл.	16530	28015	66,5	6337	1662	18,1

Критические значения *t*-критерия Стьюдента на уровне значимости 0,10; 0,05; 0,01
(двухсторонний)

Число средней свободы <i>df</i>	α		
	0,10	0,05	0,01
1	6,3138	12,7060	63,6570
2	2,9200	4,3027	9,9248
3	2,3534	3,1825	5,8409
4	2,1318	2,7764	4,6041
5	2,0150	2,5706	4,0321
6	1,9432	2,4469	3,7074
7	1,8946	2,3646	3,4995
8	1,8595	2,3060	3,3554
9	1,8331	2,2622	3,2498
10	1,8125	2,2281	3,1693
11	1,7959	2,2010	3,1058
12	1,7823	2,1788	3,0545
13	1,7709	2,1604	3,0123
14	1,7613	2,1448	2,9768
15	1,7530	2,1315	2,9467
16	1,7459	2,1199	2,9208
17	1,7396	2,1098	2,8982
18	1,7341	2,1009	2,8784
19	1,7291	2,0930	2,8609
20	1,7247	2,0860	2,8453
21	1,7207	2,0796	2,8314
22	1,7171	2,0739	2,8188
23	1,7139	2,0687	2,8073
24	1,7109	2,0639	2,7969
25	1,7081	2,0595	2,7874
26	1,7056	2,0555	2,7787
27	1,7033	2,0518	2,7707
28	1,7011	2,0484	2,7633
29	1,6991	2,0452	2,7564
30	1,6973	2,0423	2,7500
40	1,6839	2,0211	2,7045
60	1,6707	2,0003	2,6603
120	1,6577	1,9799	2,6174
∞	1,6449	1,9600	2,5758

Таблица значений F -критерия Фишера на уровне значимости $\alpha=0,05$

$v_1 \backslash v_2$	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Распределение критерия Дарбина-Уотсона для положительной автокорреляции
(для 5%-го уровня значимости)

<i>n</i>	<i>V</i> =1		<i>V</i> =2		<i>V</i> =3		<i>V</i> =4		<i>V</i> =5	
	<i>d</i> ₁	<i>d</i> ₂	<i>d</i> ₁	<i>d</i> ₂	<i>d</i> ₁	<i>d</i> ₂	<i>d</i> ₁	<i>d</i> ₂	<i>d</i> ₁	<i>d</i> ₂
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,1	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,9	1,71	0,78	1,9	0,67	2,1
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,4	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,2	1,41	1,10	1,54	1	1,68	0,9	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,8	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,9	1,92
24	1,27	1,45	1,19	1,55	1,1	1,66	1,01	1,78	0,93	1,9
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,3	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,89
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,1	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,2	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,3	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,63	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,8
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,8
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,8
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,5	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,6	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,5	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,6	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,6	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78