ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ОРЕНБУРГСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Пакеты прикладных программ

Направление подготовки *39.03.02* Социальная работа

Профиль образовательной программы «Социальная работа в системе социальных служб»

Форма обучения за*очная*

СОДЕРЖАНИЕ

1. КОНСПЕКТ ЛЕКЦИЙ

Лекция № 1 Тема: «Понятие и виды пакетов прикладных программ»

1. Вопросы лекции:

1.1. Понятие пакетов прикладных программ
1.2. Виды пакетов прикладных программ
1.3. Виды статистических пакетов

2. Краткое содержание вопросов

2.1. Понятие пакетов прикладных программ

Пакеты прикладных программ (аббр. ППП, англ. Software package) — комплекс взаимосвязанных программ, предназначенных для решения задач определенного класса конкретной предметной области. Служат программным инструментарием решения функциональных задач и являются самым многочисленным классом программных продуктов. В данный класс входят программные продукты, выполняющие обработку информации различных предметных областей.

Пакеты прикладных программ (ППП) — это специальным образом организованные программные комплексы, рассчитанные на общее применение в определенной проблемной области и дополненные соответствующей технической документацией.

В зависимости от характера решаемых задач различают следующие разновидности ППП:

пакеты для решения типовых инженерных, планово-экономических, общенаучных задач;

пакеты системных программ;

пакеты для обеспечения систем автоматизированного проектирования и систем автоматизации научных исследований;

пакеты педагогических программных средств и другие.

Чтобы пользователь мог применить ППП для решения конкретной задачи, пакет должен обладать средствами настройки (иногда путём введения некоторых дополнений).

2.2. Виды пакетов прикладных программ

Выделяются следующие виды ППП:

1. проблемно-ориентированные. Используются для тех проблемных областей, в которых возможна типизация функций управления, структур данных и алгоритмов обработки. Например, это ППП автоматизации бухучета, финансовой деятельности, управления персоналом и т.д.;

2. автоматизации проектирования (или САПР). Используются в работе конструкторов и технологов, связанных с разработкой чертежей, схем, диаграмм;

3. общего назначения. Поддерживают компьютерные технологии конечных пользователей и включают текстовые и табличные процессоры, графические редакторы, системы управления базами данных (СУБД);

4. офисные. Обеспечивают организационное управление деятельностью офиса. Включают органайзеры (записные и телефонные книжки, календари, презентации и т.д.), средства распознавания текста;

5. настольные издательские системы – более функционально мощные текстовые процессоры;

6. системы искусственного интеллекта. Используют в работе некоторые принципы обработки информации, свойственные человеку. Включают информационные системы, поддерживающие диалог на естественном языке; экспертные системы, позволяющие давать рекомендации пользователю в различных ситуациях;

интеллектуальные пакеты прикладных программ, позволяющие решать прикладные задачи без программирования.

2.3. Виды статистических пакетов

Основную часть имеющихся пакетов составляют специализированные пакетов и пакеты общего назначения.

Специализированные пакеты обычно содержат методы из одного - двух разделов статистики или методы, используемые в конкретной предметной области (контроль качества промышленной продукции, расчет страховых сумм и т.д.). Чаще всего встречаются пакеты для анализа временных рядов (например, ЭВРИСТА, МИЗОЗАВР, ОЛИМП: Стат-Эксперт), регрессионного и факторного анализа. Обычно эти пакеты содержат весьма полный набор традиционных методов в своей области, а иногда включают также и оригинальные методы и алгоритмы, созданные разработчиками пакета. Как правило, пакет и его документация ориентированы на специалистов, хорошо знакомых с соответствующими методами.

Лекция № 2 Тема: «Пакеты прикладных программ используемые в экономико-статистическом исследовании»

1. Вопросы лекции:

1.1. Специализированные пакеты. Пакеты общего назначения
1.2. Свойства пакетов прикладных программ
1.3. Достоинства и недостатки пакетов прикладных программ

2. Краткое содержание вопросов

2.1. Специализированные пакеты. Пакеты общего назначения

Статистический пакет. STATGRAPHICS, SPSS, SYSTAT, BMDP,SAS, CSS, STATISTICA, S-plus, и др. STADIA, ЭВРИСТА, МЕЗОЗАВР, ОЛИМП: Стат-Эксперт, Статистик-Консультант, САНИ, КЛАСС-МАСТЕР и др. Специализированные пакеты. Пакеты общего назначения.

Основную часть имеющихся пакетов составляют специализированные пакетов и пакеты общего назначения.

Специализированные пакеты обычно содержат методы из одного - двух разделов статистики или методы, используемые в конкретной предметной области (контроль качества промышленной продукции, расчет страховых сумм и т.д.). Чаще всего встречаются пакеты для анализа временных рядов (например, ЭВРИСТА, МИЗОЗАВР, ОЛИМП: Стат-Эксперт), регрессионного и факторного анализа. Обычно эти пакеты содержат весьма полный набор традиционных методов в своей области, а иногда включают также и оригинальные методы и алгоритмы, созданные разработчиками пакета. Как правило, пакет и его документация ориентированы на специалистов, хорошо знакомых с соответствующими методами.

2.2. Свойства пакетов прикладных программ

Характерные черты (3 свойства) :

1. Содержит набор готовых алгоритмических решений доводимых до конкретной машинной реализации;
2. Содержит механизм настройки на параметры конкретного объекта применения;
3. Пакет ПП должен предусматривать возможность дополнения его программами, привязывающими к специфике конкретного объекта, а также к изменившимся во времени условиям эксплуатации.

2.3. Достоинства и недостатки пакетов прикладных программ

Достоинства ППП:
1) Сокращение затрат на разработку; (до нескольких десятков процентов, в среднем 20–30%)
2) По сравнению с элементарными средствами, более высокая комплексная увязка решений;
3) Более высокое качество документирования ПИ;
4) Более высокая функциональная надежность;
5) Наличие развитой системы сопровождения (набор сервисных услуг, которые поддерживают эксплуатацию у пользователя);
6) ППП – средство передачи и обмена опытом между разработчиками и между конечными пользователями;

Недостатки ППП:
1) Сложность освоения ППП;
2) Большое разнообразие ППП по распространенным задачам затрудняет выбор. На сегодня отсутствуют объективные методы оценки ППП;
3) Низкая степень системной увязки существующих ППП (в случае увязки нескольких конкретных программ по входам–выходам);
4) Проблема наращивания и модификации;
5) Малая функциональная полнота.

Лекция № 3 Тема: «Статистические таблицы и графики»

1. Вопросы лекции:

1.1.Таблицы: правила построения и оформления
1.2. Графический метод представления социально-экономической информации
1.3. Построение таблиц и диаграмм в MS Excel

2. Краткое содержание вопросов

2.1.Таблицы: правила построения и оформления

Табличные процессоры (типичный пример - MS Excel) позволяют обрабатывать большие объемы числовой информации (не исключая при этом обычную символьную), формируя из данных таблицы. Можно сказать, что это очень мощные калькуляторы, хранящие в своей памяти огромные числовые массивы и позволяющие выполнять над ними различные арифметические и логические операции, формировать диаграммы и делать множество других операций, полезных для решения различных задач пользователя. Аналогично пакету MS Word, табличный процессор MS Excel изучается в лабораторном практикуме по информатике.

2.2. Графический метод представления социально-экономической информации

Графические редакторы позволяют генерировать различные изобразительные объекты. Они делятся на 2 класса - растровой и векторной графики - в зависимости от того, какое внутреннее представление этих объектов в них поддерживается. Редакторы растровой графики используются для работы с фотографиями. Они кодируют фотоизображения в цифровую форму и позволяют выполнять над ними различные редактирующие операции (выделение фрагментов, перемещение, вырезание, копирование и т.д.). Примерами редакторов этого класса являются: Adobe Photoshop, Aldus Photo Styler, Picture Publisher, Photo Works Plus. Редакторы векторной графики используются для профессиональной

работы, связанной с технической и художественной иллюстрацией с последующей цветной печатью. Они занимают промежуточное место между САПР и настольными издательскими системами. Включают инструментарий для создания графического объекта; средства манипулирования объектами; средства обработки текста в части оформления и модификации параграфов, работы со шрифтами; средства вывода на печать и настройки цвета.

2.3. Построение таблиц и диаграмм в MS Excel

Примерами редакторов этого класса являются: Adobe Photoshop, Aldus Photo Styler, Picture Publisher, Photo Works Plus. Редакторы векторной графики используются для профессиональной работы, связанной с технической и художественной иллюстрацией с последующей цветной печатью. Они занимают промежуточное место между САПР и настольными издательскими системами. Включают инструментарий для создания графического объекта; средства манипулирования объектами; средства обработки текста в части оформления и модификации параграфов, работы со шрифтами; средства вывода на печать и настройки цвета.

Лекция № 4 Тема: «Описательная статистика в MS Excel и в ППП "Statistica"»

1. Вопросы лекции:

1.1. Описательная статистика: понятие, система показателей
1.2. Показатели средних величин
1.3. Показатели вариации

2. Краткое содержание вопросов

2.1. Описательная статистика: понятие, система показателей

Показатели, применяемые для изучения статистической практики и науки, подразделяют на группы по следующим признакам:

1) по сущности изучаемых явлений – это объемные, характеризующие размеры процессов, и качественные, которые выражают количественные соотношения, типичные свойства изучаемых совокупностей;

2) по степени агрегирования явлений – это индивидуальные, которые характеризуют единичные процессы, и обобщающие, отображающие совокупность в целом или ее части;

3) в зависимости от характера изучаемых явлений – интервальные и моментные. Данные, отображающие развитие явлений за определенные периоды времени, называют интервальными показателями, т. е. это статистический показатель, который характеризуют процесс изменения признаков. К моментным показателям относят показатели, которые отражают состояние явления на определенную дату (момент);

4) в зависимости от пространственной определенности различают показатели: федеральные – характеризуют изучаемый объект в целом по стране; региональные и местные – эти показатели относятся к определенной части территории или отдельному объекту;

5) в зависимости от свойств конкретных объектов и формы выражений статистические показатели делятся на относительные, абсолютные и средние, данные показатели будут рассмотрены ниже.

2.2. Показатели средних величин

Средняя величина — важнейший обобщающий показатель. Эта функция средних отражена в формуле

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n},$$

где $\bar{x}$— средняя величина для признака x; x_i — значение признака для i-й единицы совокупности; n — количество единиц совокупности.

Расчет средней величины включает две операции: суммирование данных по всем единицам (обобщение данных) и деление на число единиц (приведение обобщенной характеристики к единице совокупности).

Виды и формы средних

Значение средней зависит от того, каков порядок ее расчета. Применяются средние двух видов: простые и взвешенные.

Основные виды средних, чаще всего применяемых в практике статистических расчетов, приведены ниже в табл. 1.

Таблица 1 – Виды средних

Наименование	Простая форма	Взвешенная форма
Средняя арифметическая [x_a]	$\bar{x} = \frac{\sum x}{n}$	$\bar{x} = \frac{\sum xf}{\sum f}$
Средняя квадратическая [x_q]	$\bar{x} = \sqrt{\frac{\sum x^2}{n}}$	$\bar{x} = \sqrt{\frac{\sum x^2 f}{\sum f}}$
Средняя гармоническая [x_h]	$\bar{x} = \frac{n}{\sum 1/x}$	$\bar{x} = \frac{\sum M}{\sum M/x}$
Средняя геометрическая [x_g]	$\bar{x} = \sqrt[n]{\Pi x}$	$\bar{x} = \sqrt[\Sigma f]{\Pi x^f}$

В представленных формулах применены следующие обозначения:

x - значение признака;

$\bar{x}$ - среднее значение признака;

$\sum$ - знак суммирования;

П – знак перемножения;

f (частота) и M (произведение частоты на значения признака) – веса для расчета взвешенной средней:

N и f – численность единиц совокупности;

M – общий объем варьирующего признака.

Если средние вычислить по одним и тем же данным, то приведенные виды средних по своим численным значениям встают в следующий ряд: $x_h<x_g<x_a<x_q$, иллюстрируя так называемое правило мажорантности средних.

Одна из задач определения средней состоит в правильности выбора вида средней величины.

2.3. Показатели вариации

Показатели вариации являются числовой мерой уровня колеблемости признака. Одновременно по размеру показателя вариации делают вывод о типичности, надежности средней величины, найденной для данной совокупности, и об однородности самой совокупности.

Важнейшие виды показателей вариации:

размах вариации [R]: $R = x_{max} - x_{min}$;

среднее линейное отклонение [$\overline{d}$]: $\overline{d} = \frac{\Sigma[x - \overline{x}]f}{\Sigma f}$;

дисперсия [σ^2]: $\sigma^2 = \frac{\Sigma(x - \overline{x})^2 f}{\Sigma f}$;

среднее квадратическое отклонение [σ]: $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x - \overline{x})^2 f}{\Sigma f}}$;

коэффициент вариации [ν]: $\nu = \frac{\sigma}{\overline{x}} \cdot 100.$

Размах вариации учитывает только крайние значения признака и не учитывает все промежуточные.

Дисперсия не имеет единиц измерения.

Лекция № 5 Тема: «Статистическое изучение динамики социально-экономических явлений»

1. Вопросы лекции:

1.1. Понятие и классификация рядов динамики
1.2. Правила построения динамических рядов
1.3. Показатели, характеризующие динамику
1.4. Табличное и графическое представление динамических рядов

2. Краткое содержание вопросов

2.1. Понятие и классификация рядов динамики

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 в разложении (1), т.е. определить значения индикаторов χ_i;
2. построить «хорошие» оценки для тех неслучайных функций, которые присутствуют в разложении (1);
3. подобрать модель, адекватно описывающую поведение случайных остатков ε_t, и статистически оценить параметры этой модели.

Успешное решение перечисленных задач, обусловленных базовой целью статистического анализа временного ряда, является основой для достижения конечных прикладных целей исследования и, в первую очередь, для решения задачи кратко- и среднесрочного прогноза значений временного ряда. Приведем кратко основные элементы эконометрического анализа временных рядов.

Временные ряды отражают тенденцию изменения параметров системы во времени, поэтому входным параметром x является момент времени.

2.2. Правила построения динамических рядов

Существуют моментальные и интервальные ряды. В моментальных рядах отражаются абсолютные величины, по состоянию на определенный момент времени, а в интервальных – относительные величины (показатель за год, месяц, и т.д.). Исследование данных при помощи рядов позволяет во многих случаях более четко представить детерминированную функцию. При этом рассчитываются базисные и цепные показатели (прирост, коэффициент роста, коэффициент роста, темп роста, темп прироста, и др.). Под

базисными показателями понимают, показатели, которые соотносятся к начальному уровню ряда. Цепные показатели относятся к предыдущему уровню.

Прогноз явлений по временным рядам состоит из двух этапов:

- Прогноз детерминированной компоненты.
- Прогноз случайной компоненты.

Обе проблемы связаны с анализом результатов парных экспериментов. В отличие от аппроксимации и интерполяции анализ временных рядов включает в себя методы оценки случайных компонент. Поэтому прогнозирование при помощи временных рядов является более точным.

Исследование рядов имеет большое значение и для технических, и для экономических систем.

Одна из важнейших задач статистики - определение в рядах динамики общей тенденции развития.

2.3. Показатели, характеризующие динамику

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

4. определить, какие из неслучайных функций присутствуют в разложении (1), т.е. определить значения индикаторов χ_i;
5. построить «хорошие» оценки для тех неслучайных функций, которые присутствуют в разложении (1);
6. подобрать модель, адекватно описывающую поведение случайных остатков ε_t, и статистически оценить параметры этой модели.

Успешное решение перечисленных задач, обусловленных базовой целью статистического анализа временного ряда, является основой для достижения конечных прикладных целей исследования и, в первую очередь, для решения задачи кратко- и среднесрочного прогноза значений временного ряда. Приведем кратко основные элементы эконометрического анализа временных рядов.

Временные ряды отражают тенденцию изменения параметров системы во времени, поэтому входным параметром x является момент времени.

Выходной параметр y называется уровнем ряда. В случае отсутствия ярко выраженных изменений в течение времени, общая тенденция сохраняется. Ряд можно описать уравнением вида

$$Y_T = F(t) + E_T ,$$

где

F (t) – детерминированная функция времени.

E_T – случайная величина

Во временных рядах проводится операция анализа и сглаживания тренда, который отражает влияние некоторых факторов. Для построения тренда применяется МНК-критерий.

2.4. Табличное и графическое представление динамических рядов

С этой целью ряды динамики подвергаются обработке методами укрупнение интервалов, скользящей средней и аналитического выравнивания:

1. Метод укрупнения интервалов.

Одним из наиболее элементарных способов изучения общей тенденции в ряду динамики является укрупнение интервалов. Этот способ основан на укрупнении периодов, к которым относятся уровни ряда динамики. Например, преобразование месячных периодов в квартальные, квартальных в годовые и т.д.

2. Метод скользящей средней.

Выявление общей тенденции ряда динамики можно произвести путем сглаживания ряда динамики с помощью скользящей средней.
Скользящая средняя - подвижная динамическая средняя, которая рассчитывается по ряду при последовательном передвижении на один интервал, то есть сначала вычисляют средний уровень из определенного числа первых по порядку уровней ряда, затем - средний уровень из такого же числа членов, начиная со второго. Таким образом, средняя как бы скользит по ряду динамики от его начала к концу, каждый раз отбрасывая один уровень в начале и добавляя один следующий.

Лекция № 6 Тема: «Измерение взаимосвязей общественных явлений»

1. Вопросы лекции:

1.1. Сущность и задачи корреляционно-регрессионного анализа
1.2. Параметрические методы изучения связи
1.3. Технология решения корреляционного и регрессионного анализа в MS Excel и в ППП "Statistica"

2. Краткое содержание вопросов

2.1. Сущность и задачи корреляционно-регрессионного анализа

В многомерном статистическом анализе каждый объект описывается вектором, размерность которого произвольна (но одна и та же для всех объектов). Однако человек может непосредственно воспринимать лишь числовые данные или точки на плоскости. Анализировать скопления точек в трехмерном пространстве уже гораздо труднее. Непосредственное восприятие данных более высокой размерности невозможно. Поэтому вполне естественным является желание перейти от многомерной выборки к данным небольшой размерности, чтобы «на них можно было посмотреть».
Кроме стремления к наглядности, есть и другие мотивы для снижения размерности. Те факторы, от которых интересующая исследователя переменная не зависит, лишь мешают статистическому анализу. Во-первых, на сбор информации о них расходуются ресурсы. Во-вторых, как можно доказать, их включение в анализ ухудшает свойства статистических процедур (в частности, увеличивает дисперсию оценок параметров и характеристик распределений). Поэтому желательно избавиться от таких факторов.

2.2. Параметрические методы изучения связи

Обсудим с точки зрения снижения размерности пример использования регрессионного анализа для прогнозирования объема продаж, рассмотренный в подразделе 3.2.3. Во-первых, в этом примере удалось сократить число независимых переменных с 17 до 12. Во-вторых, удалось сконструировать новый фактор – линейную функцию от 12 упомянутых факторов, которая лучше всех иных линейных комбинаций факторов прогнозирует объем продаж. Поэтому можно сказать, что в результате размерность задачи уменьшилась с 18 до 2. А именно, остался один независимый фактор и один зависимый – объем продаж.
При анализе многомерных данных обычно рассматривают не одну, а множество задач, в частности, по-разному выбирая независимые и зависимые переменные. Поэтому рассмотрим задачу снижения размерности в следующей формулировке. Дана многомерная

выборка. Требуется перейти от нее к совокупности векторов меньшей размерности, максимально сохранив структуру исходных данных, по возможности не теряя информации, содержащихся в данных. Задача конкретизируется в рамках каждого конкретного метода снижения размерности.
Метод главных компонент является одним из наиболее часто используемых методов снижения размерности. Основная его идея состоит в последовательном выявлении направлений, в которых данные имеют наибольший разброс. Пусть выборка состоит из векторов, одинаково распределенных с вектором X = (x(1), x(2), … , x(n)).

2.3. Технология решения корреляционного и регрессионного анализа в MS Excel и в ППП "Statistica"

Для визуального анализа данных часто используют проекции исходных векторов на плоскость первых двух главных компонент. Обычно хорошо видна структура данных, выделяются компактные кластеры объектов и отдельно выделяющиеся вектора.
Метод главных компонент является одним из методов факторного анализа. Различные алгоритмы факторного анализа объединены тем, что во всех них происходит переход к новому базису в исходном n-мерном пространстве. Важным является понятие «нагрузка фактора», применяемое для описания роли исходного фактора (переменной) в формировании определенного вектора из нового базиса.
Новая идея по сравнению с методом главных компонент состоит в том, что на основе нагрузок происходит разбиение факторов на группы. В одну группу объединяются факторы, имеющие сходное влияние на элементы нового базиса. Затем из каждой группы рекомендуется оставить одного представителя. Иногда вместо выбора представителя расчетным путем формируется новый фактор, являющийся центральным для рассматриваемой группы. Снижение размерности происходит при переходе к системе факторов, являющихся представителями групп. Остальные факторы отбрасываются.

Лекция № 7 Тема: «Российские статистические пакеты прикладных программ.»

1. Вопросы лекции:

1.1. Отечественные статистические пакеты прикладных программ.
1.2. ППП "STADIA"
1.3. История создания системы ЭВРИСТА
1.4. ППП "ОЛИМП"
1.5. ППП "МЕЗОЗАВР"

2. Краткое содержание вопросов

2.1. Отечественные статистические пакеты прикладных программ

Из отечественных можно назвать такие пакеты, как STADIA, ЭВРИСТА, МЕЗОЗАВР, ОЛИМП: Стат-Эксперт, Статистик-Консультант, САНИ, КЛАСС-МАСТЕР и др.

Отечественные статистические пакеты, которые устойчиво представлены на рынке в течение последних лет, в значительной степени лишены таких **недостатков**, которые есть у западных продуктов. Они предполагают наличие широкого первоначального статистического образования, доступной литературы и консультационных служб. Поэтому они содержат мало экранных подсказок и требуют внимательного изучения документации на английском языке.

2.2. ППП "STADIA"

Пакет STADIA разработан и поддерживается НПО “Информатика и компьютеры” при активном участи ведущих специалистов МГУ им. М.В.Ломоносова. Пакет содержит широкий набор методов анализа данных из всех областей статистики и доступен широкому кругу прикладных специалистов, менеджеров и студентов. Сейчас распространяется версия 6.2 для среды Windows. Пакет может появляться в трех вариантах: study, base и prof, различающихся лишь объемами обрабатываемых массивов и ценой. Самый дешевый вариант study имеет максимальный объем матрицы данных в 400 чисел. Он предназначен главным образом для учебных заведений и задач с небольшими объемами данных. Самая дорогая версия STADIA 6.2 prof. имеет максимальный объем матрицы данных 20000 чисел и расширенные возможности статистических процедур для их обработки по сравнению с базовыми версиями. У пакета имеется бесплатная учебно-демонстрационная версия, позволяющая обрабатывать большое количество демонстрационных примеров из всех разделов статистического анализа. Эта версия также допускает ввод с клавиатуры и полную обработку данных пользователей. Однако при этом существуют ограничения на объемы вводимых данных, и отсутствует возможность сохранения введенных данных в файле. Документация пакета является одновременно детальным справочником по использованию статистических методов и может быть приобретена отдельно от пакета.

2.3. История создания системы ЭВРИСТА

Идея создания специализированного статистического пакета по анализу и прогнозированию временных рядов возникла вначале 80-х годов на кафедре математической статистики Московского государственного университета. Главным идеологом будущей программной системы выступил старший научный сотрудник кафедры, к.ф.-м.н. Ю.Г.Баласанов. Первая версия системы *ЭВРИСТА* была реализована на языке ФОРТРАН для ЭВМ БЭСМ-6 и с 1984 года началось и использование системы в учебном процессе факультета.

Первая коммерческая версия системы ЭВРИСТА для персонального компьютера появилась 1987 году и ее первым покупателем стало объединение КАМАЗ (г. Набережные Челны). Несмотря на то, что первые персональные компьютеры имели слабые (особенно с нынешних позиций) графические возможности, разработчики по максимуму старались их использовать, и в результате *ЭВРИСТА*, одна из немногих программных систем того времени, уже имела полностью графический многооконный интерфейс.

2.4. ППП "ОЛИМП"

Пакет «Олимп» предназначен для автоматизации обработки данных на основе широкого набора современных методов прикладной статистики. Он реализован в расчете на самых разнообразных пользователей – от новичков до экспертов в области статистики.

В состав пакета, кроме основных программ, входят также электронная таблица MNCALC и программное средство «Прикладные социологические исследования (ПСИ)».

Пакет «ОЛИМП» позволяет организовать полный цикл исследований по статистическому анализу и прогнозированию данных, начиная с ввода исходных данных, их проверке и визуализации и заканчивая проведением расчетов и анализом результатов.

С функциональной точки зрения пакет состоит из следующих программ (процедур): редактора средств графического отображения и утилит преобразования данных, а также программ реализации методов статистического анализа.

Редактор данных обеспечивает возможность ввода, просмотра и редактирования исходных данных (в том числе пропущенных наблюдений).

Средства графического отображения данных позволяют выводить различные виды графиков на экран, а также сохранять их на диске для дальнейшего использования.

Утилиты преобразования данных выполняют арифметические преобразования данных (унарные и бинарные), различные виды сортировки (в том числе по нескольким переменным), агрегирование (объединение по одному признаку) и фильтрование данных (отбор по одному признаку).

Программы пакета «ОЛИМП» реализуют следующие методы статистического анализа: корреляционный, регрессионный, дисперсионный, дискриминантный, факторный и компонентный, анализ таблиц сопряженности рядов и др.

2.5. ППП "МЕЗОЗАВР"

Основное назначение пакета «МЕЗОЗАВР» заключается в проведении разведочного анализа временных рядов. Это касается ситуации, когда необходимо «пощупать» имеющуюся числовую информацию, по усмотрению исследователя применяя различные методы обработки и анализируя получающиеся при этом результаты и их адекватность. Пакет позволяет осуществлять подобные исследования весьма оперативно и эффективно.

Пакет «МЕЗОЗАВР» используется для анализа временных рядов умеренной (не более нескольких тысяч наблюдений) длины. Диалог происходит по желанию пользователя на русском или английском языке. Управление осуществляется с помощью меню и клавиш быстрого доступа.

Под временным рядом понимается последовательность наблюдений за некоторой числовой характеристикой показателей, сделанных с постоянным шагом во времени (например ежегодно, ежемесячно, каждый час и т.п.). В статистике примерами подобных показателей могут служить на макроэкономическом уровне ежегодные, ежеквартальные, ежемесячные и т.п. объемы производства, поставок, перевозок, потребления; индексы цен и другие макроэкономические показатели; на уровне предприятия – объемы выпуска продукции, затраты, расход ресурсов, эволюция характеристик качества и др.

Лекция № 8 Тема: «Зарубежные статистические пакеты прикладных программ.»

1. Вопросы лекции:

1.1. SAS
1.2. MINITAB
1.3. SPSS
1.4. Statistica

2. Краткое содержание вопросов

2.1. SAS

Система SAS существует и развивается с 1976 г. и работает на самых различных платформах под управлением одной из 12-ти операционных систем (ОС). Фирма-разработчик SAS в 1995 г. занимала 13-е место в мире (и 14-е в 1994 г.) среди ведущих разработчиков разнообразных программных продуктов, имея 3200 сотрудников, поддерживающих более 3 миллионов пользователей в 120 странах.

По сути, SAS сегодня является мощным комплексом из свыше 20-ти различных программных продуктов, объединенных друг с другом «средствами доставки информации» (Information Delivery System или IDS, так что весь пакет иногда обозначается как SAS/IDS). Одной из последних версий для Windows является версия 6.11.

Если позиционировать SAS как товар на рынке статистического программного обеспечения, где одни сконцентрировались на графике, а другие на удобстве управления, то SAS прежде всего *статистическая программа.*

То есть основным «козырем» SAS является его непревзойденная мощность по набору статистических алгоритмов. Эту оценку мощности следует воспринимать лишь на фоне других *универсальных* СПП. Это не значит, например, что по богатству и качеству методов статистического анализа временных рядов соответствующий раздел SAS превосходит ряд других *специализированных* пакетов, например, широко известный отечественный пакет MESOSAUR.

2.2. MINITAB

Сейчас распространяется версия 10.0 для среды MS-Windows этой системы и уже появилась его улучшенная 32-х разрядная версия 11.0. Кроме рассматриваемых платформ, пакет также работает на Macintosh в среде MS-DOS, на рабочих станциях и других компьютерах.

Пакет развивается более 20 лет и широко известен в США, где он является одним из основных учебных пакетов. Во многом правда, это объясняется тем, что пакет в свое время захватил этот сегмент рынка, а вовсе не его исключительными свойствами.

Он хорошо продуман по разделу описательной (дескриптивной) статистики, хорошо сконструирован и управляется с помощью очень удобного меню, или, по желанию пользователя, через команды, составлять которые помогают диалоговые окна пакета. Часто используемые команды можно запускать и по их первой букве. Общее число команд превышает 200. Можно составлять и специальные макросы для выполнения последовательностей команд.

Импорт/экспорт данных из других Windows-приложений делается через стандартный буфер обмена (то есть последовательным выбором команд в меню двух пакетов типа *Cut*/*Copy — Paste to*/*From*). В пакете имеются разнообразные возможности по управлению данными.

2.3. SPSS

Пакет SPSS стал известен в научном и деловом мире, будучи реализован на больших машинах. Основными пользователями его «пакетного варианта» традиционно были ученые, работающие в академических институтах и университетах, а также в разнообразных приложениях математической статистики, например, в области контроля качества.

Как и SAS, пакет предназначен в первую очередь для статистиков-профессионалов, так как имеет достаточно мощный аппарат статистического анализа, вполне соизмеримый по мощности с SAS.

Благодаря покупке фирмой SPSS компаний BMDP и SYSTAT, а также переориентации разработчиков в последние годы на платформу Windows, программа SPSS версии 6.1 для Windows 3.1 и версий 7.0 и 7.5 для Windows 95, стала в настоящее время одним из лидеров среди универсальных статистических пакетов. В частности, версия 7.0 является призером редакции журнала PC Magazine.

Однако, как и все мощные универсальные пакеты, SPSS, «любит хорошее железо»: процессор должен быть 486DX-2 и выше, для его использования рекомендуется 16 Мб оперативной памяти, а на винчестере модули **Base** и **Professional Statistics** для управления данными и с алгоритмами классификации потребуют, как минимум, 65–80 Мб (вместе с файлами «подкачки» — swap files).

Да и цена достаточно полного комплекта системы SPSS (SPSS Base + набор из 7 модулей) впечатляет ($4290 для версии 6.1 или 7.0).

2.4. Statistica

Система *STATISTICA* производится компанией **StatSoft Inc.**, основанной в 1984 г в городе Тулса (Оклахома, США). Первые программные продукты - PsyhoStat-2,3 - были ориентированы на статистические исследования социологических данных. Первый коммерческий продукт - Statistical Supplement for Lotus 1-2-3, появился в 1985 г.

С 1985 г. начался быстрый рост фирмы. **StatSoft** выпускает первую систему статистического анализа для компьютеров Apple Macintosh под названием StatFast и статистический пакет для IBM PC под названием STATS+. В 1986 г. начинается работа над основной линией программных продуктов фирмы - интегрированных статистических пакетов для комплексной обработки данных.

В 1991 г. выходит первая версия системы *STATISTICA*/DOS, которая представляет собой новое направление развития статистического программного обеспечения. В ней реализован так называемый графически-ориентированный подход к анализу данных. Этот пакет обладал рядом существенных преимуществ перед другими статистическими пакетами (за счет оптимизации удалось добиться повышения скорости обработки более, чем в 10 раз по сравнению с другими пакетами, пакет мог работать фактически с неограниченным объемом данных). В 1992 г. вышла версия *STATISTICA* для Macintosh, которая быстро приобрела заслуженную популярность среди пользователей.

2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

Практическое занятие 1 (ПЗ - 1) Понятие и виды пакетов прикладных программ

1. Понятие пакетов прикладных программ.
2. Виды пакетов прикладных программ.
3. Виды статистических пакетов.

Практическое занятие 2 (ПЗ - 2) Пакеты прикладных программ, используемые в статистико-социологических исследованиях

1. Специализированные пакеты. Пакеты общего назначения.
2. Свойства пакетов прикладных программ.
3. Достоинства и недостатки пакетов прикладных программ

Практическое занятие 3 (ПЗ - 3) Статистические таблица и графики

1. Таблицы: правила построения и оформления.
2. Графический метод представления социально-экономической информации.
3. Построение таблиц и диаграмм в MS Excel.

Практическое занятие 4 (ПЗ-4) Описательная статистика в MS Excel и в ППП "Statistica"

1. Описательная статистика: понятие, система показателей.
2. Показатели средних величин.
3. Показатели вариации

Задача 1. Построение классической линейной регрессии

1 Цели и задачи лабораторной работы

В данной лабораторной работе на практическом примере рассмотрим этапы построения уравнения классической линейной регрессии, при этом поставим 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.

2 Понятие классической линейной регрессии

В данной глава 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:

$$\tilde{y}_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + ... + a_m x_{mi} + \varepsilon_i \qquad (2.1)$$

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

$$\tilde{y}_i = a_0 + a_1 x_i + \varepsilon_i \qquad (2.2)$$

где: i =1, 2, …, 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;

e_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 случая множественной регрессии a_j где j = 1, 2, …, 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 изменяются на одно среднеквадратическое отклонение.

$$\beta_{ji} = a_j \cdot \frac{\sigma_{xj}}{\sigma_y} \qquad (2.3)$$

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

$$Э_j = a_j \frac{\overline{x}_j}{\overline{y}} \qquad (2.4)$$

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 приступать к использованию модели необходимо убедится ее адекватности фактическим данным (анализируемому явлению). Для этих целей используют t-критерий Стьюдента и F-критерий Фишера.

3 Расчет описательных (дескриптивных) статистик

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

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 2.1):

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* (Персентели, 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: *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 2.1 - Описательная статистика

	Valid N	*Mean*	*Minimum*	*Maximum*	*Std.Dev.*	*Skewness*	*Kurtosis*
Y	15	37,667	24	56	11,684	0,266	-1,588
X1	15	32,400	8	63	13,809	0,225	0,675
X2	15	47,533	35	55	6,621	-0,442	-1,125

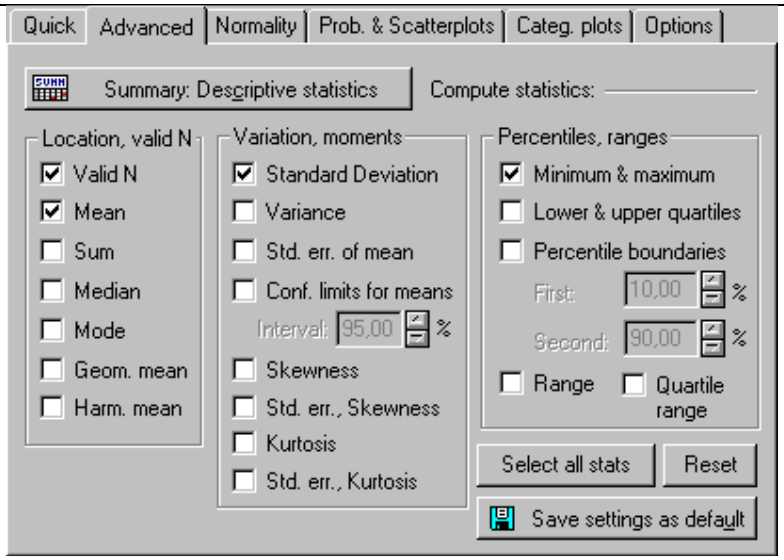

Рисунок 2.1 – Окно выбора (установок) описательных статистик (приведена часть исходного окна)

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.*

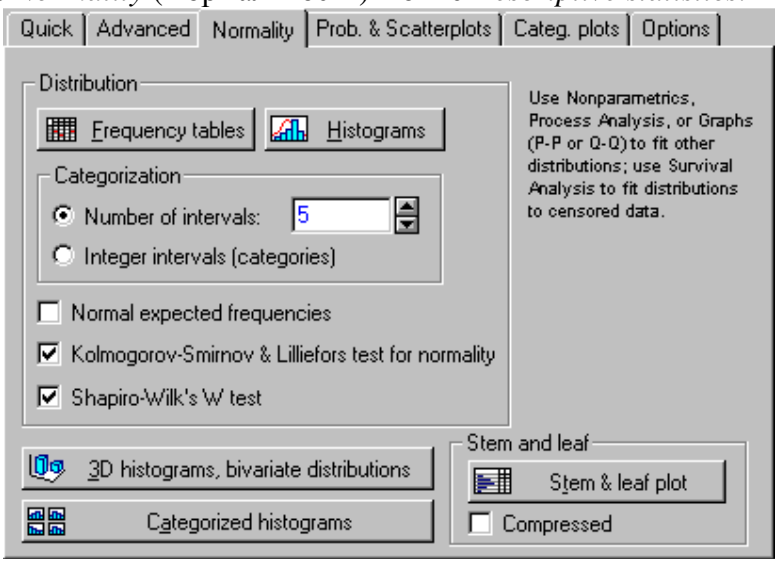

Рисунок 2.2 – Окно установок вычисления характеристики нормальности распределения (приведена часть исходного окна)

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 α = 0,01, 0,05 или 0,1), то гипотеза о нормальном законе распределения данных принимается с вероятностью (1 - α).

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 2.3 – Результаты оценки критериев нормальности распределения переменных

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.

4 Построение классической линейной регрессии

Для построения линейной регрессии (парной и множественной), а также для оценки параметров линейных и нелинейных трендов в СПП 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 3А).

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

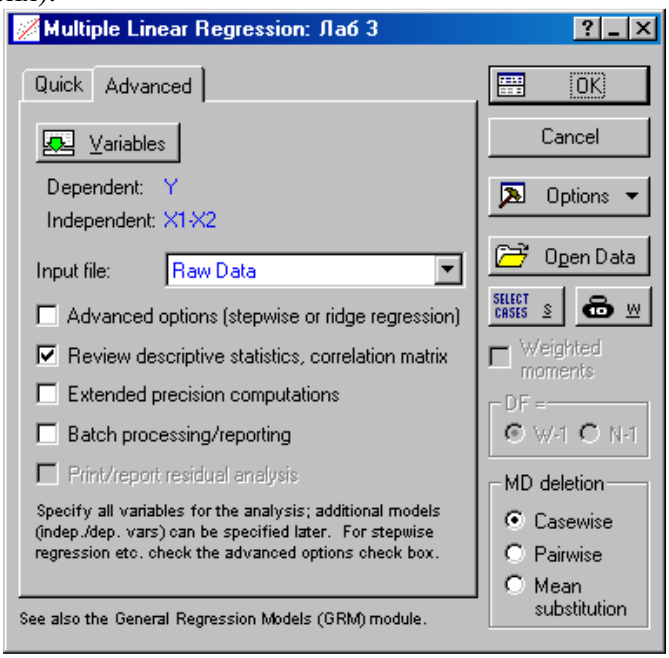

Рисунок 2.4 – Окно *Multiple Linear Regression* (Множественная линейная регрессия)

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*.

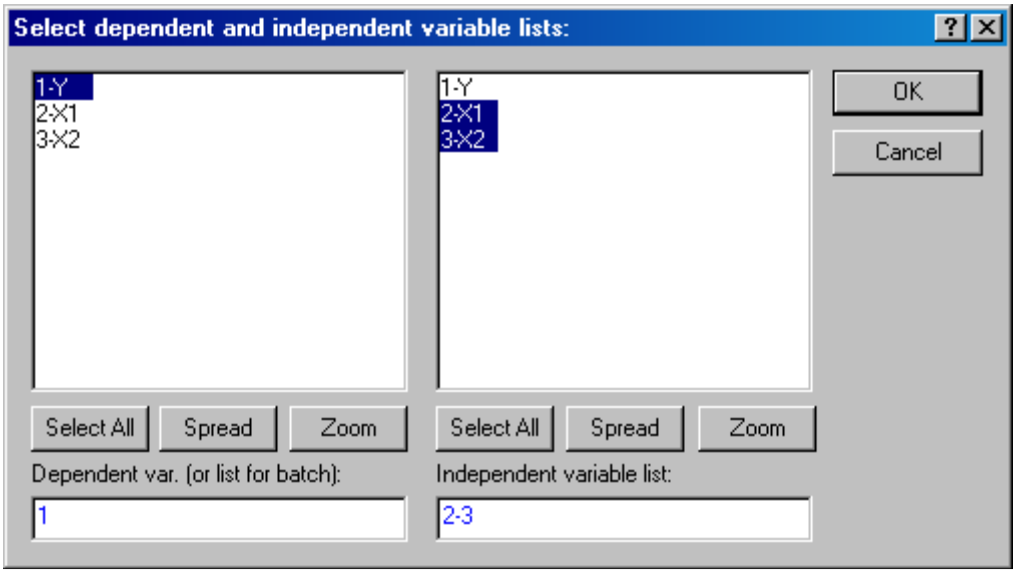

Рисунок 2.5 – Окно выбора зависимой и не зависимой переменных

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

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

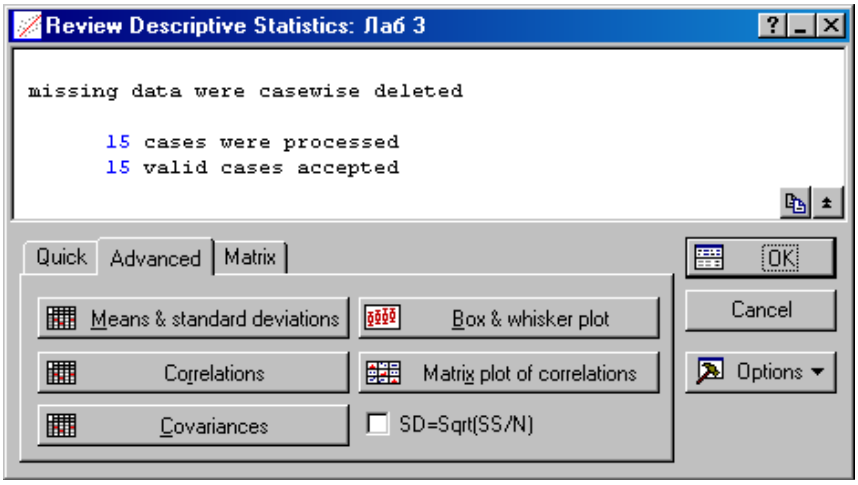

Рисунок 2.6 – Окно установок описательных статистик и корреляции

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* получим следующие данные:

Таблица 3.2 - Матрица парных коэффициентов корреляции

	X1	*X2*	*Y*
X1	1,000	-0,117	-0,351
X2	-0,117	1,000	**0,868**
Y	-0,351	**0,868**	1,000

Значения представленные таблице показывают, что фактор *X2* оказывает сильное положительное влияние на зависимую переменную *Y* (т.к. значение на пересечении соответствующего столбца и строки равно 0,868), фактор *X1* оказывает слабое отрицательное влияние. Между переменными *X1* и *X2* связь практически отсутствует (значение коэффициента корреляции -0,117 близко к нулю).

Также можно представить полученные результаты в графическом виде, для этого выбираем кнопку *Matrix plot of correlations* (рисунок 2.6), полученный результат представлен на рисунке 2.7.

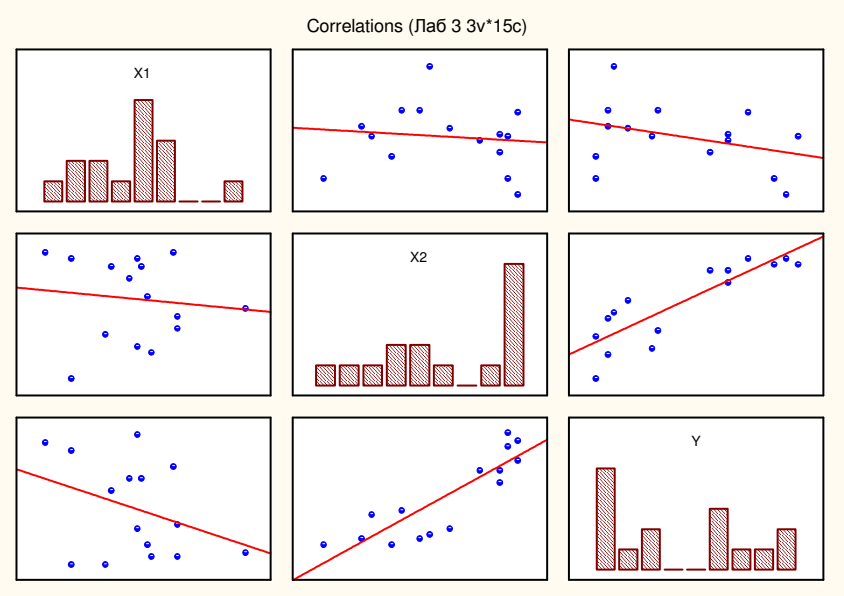

Рисунок 2.7 - Матрица диаграмм рассеяния

Интерпретация приведенного рисунка такова: чем ближе к теоретической линии регрессии сгруппированы точки, тем теснее связь между изучаемыми показателями.

Шаг 6. Вернемся в окно *Multiple Linear Regression* (рисунок 3.4), для этого в окне *Review Descriptive Statistic* выберем кнопку *Cancel* (Отмена), далее снимем флажок с опции *Review descriptive statistics, correlation matrix.*

Нажав кнопку ОК, перейдем в следующее окно, содержащее результаты построения модели.

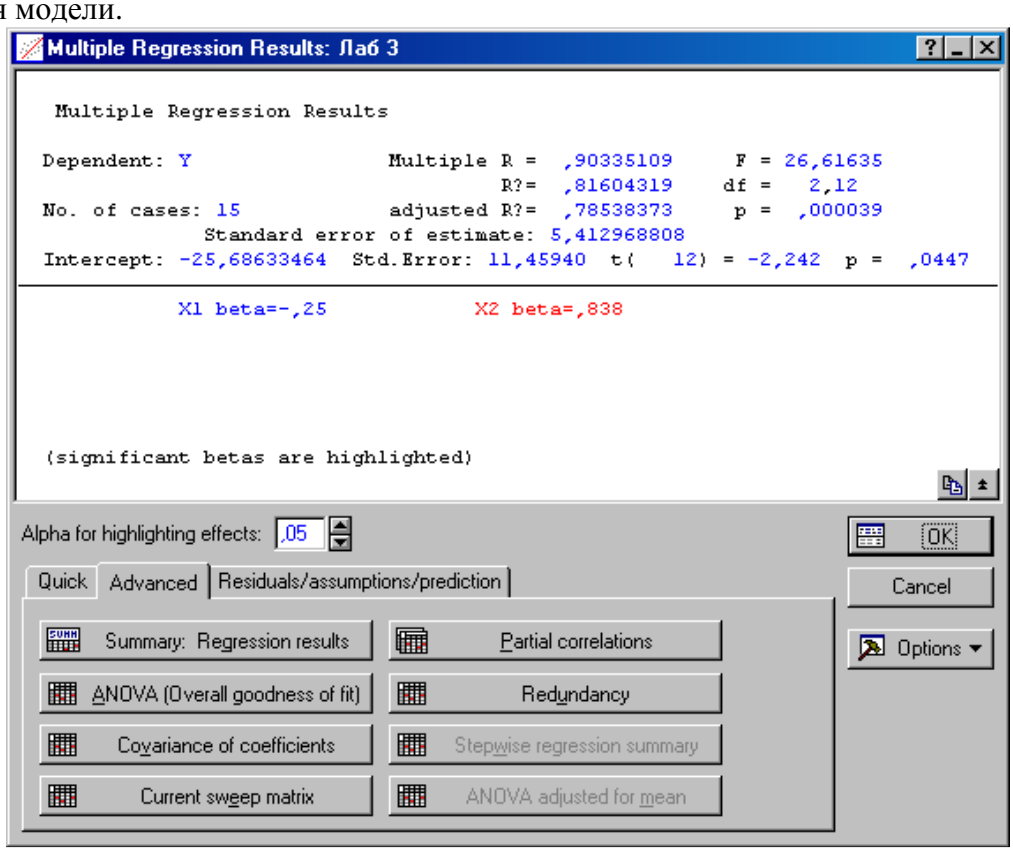

Рисунок 2.8 – Окно с результатами оценивания регрессии

Вкладки:
Quick (Быстрые статистики) – данная вкладка предназначена для неопытных пользователей так как в ней доступна только одна кнопка *Summary: Regression results*. 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 (Вычислить: Результаты построения регрессии)
ANOVA (Overall goodness of fit)
Covariance of coefficients
Current sweep matrix
Partial correlations (Частная корреляция)
Redundancy
Stepwise regression summary
ANOVA adjusted for mean
Residuals/assumptions/prediction (Отклонения/распределения/предсказания) – вкладка содержит алгоритмы анализа отклонений построенной модели, дескриптивные статистики, а также возможность рассчитывать прогнозные значения зависимой переменной.

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 2.3 – Показатели адекватности множественного уравнения регрессии

	Value
Multiple R	0,903
Multiple R?	0,816
Adjusted R?	0,785
F(2,12)	26,616
p	0,000
Std.Err. of Estimate	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 $\overline{R}^2$. 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 прейти, если сопоставить табличное значение критерия при α=0,05 и v_1=2, v_2=12 равное 3,88 с фактическим значением *F(2,12)*= 26,616, т.е. получаем Fтаб < Fфакт 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 – см. таблицу 2.4).

Таблица 2.4 – Результаты оценивания множественного уравнения регрессии

	Beta	*Std.Err. of Betta*	*B*	*Std.Err. of B*	*t(12)*	*p-level*
Intercept			-25,686	11,459	-2,242	0,045
X1	-0,253	0,125	-0,214	0,105	-2,032	0,065
X2	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* (β_2=0,838).

В четвертом столбце содержатся значения параметров a_j оцененного уравнения вида 2.1, т.е. в данном случае получаем следующую регрессионную модель:

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

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 *H0* отвергается и соответствующий коэффициент не равен нулю.

В рассматриваемом примере параметр a_2 при переменной *X2* значим при уровне значимости α больше, чем 0,0002. Коэффициент a_1 получен не значим при уровне α = 0,05, т.к. значение вероятности 0,065 больше 0,05.

Так же выявить статистическую значимость параметров можно использовав табличное значение *t*-критерия Стьюдента, в нашем случае при α=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 *X1,* необходимо исключить из рассмотрения фактор *X1*. Для этого в активном окне выберем кнопку *Cancel,* перейдя в стартовое окно, далее в качестве независимой переменной (*Independent var.*) укажем *X2*. Получаем следующие результаты:

Таблица 2.5 – Показатели адекватности парного уравнения регрессии

	Value
Multiple R	0,868
Multiple R?	0,753
Adjusted R?	0,734
F(1,13)	39,571
p	0,000
Std.Err. of Estimate	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 2.5, параметры парной регрессионной модель получены статистически значимыми.

Таблица 2.6 – Результаты оценивания парного уравнения регрессии

	Beta	*Std.Err. of Betta*	*B*	*Std.Err. of B*	*t(12)*	*p-level*
Intercept			-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 2.6), а уравнение в целом проходит тест по *F*-критерию Фишера (таблица 2.5).

5 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 2.8) необходимо выбрать вкладку *Residuals/assumptions/prediction* (Отклонения/распределения/предсказания) и воспользоваться кнопкой *Predict dependent variable* (Прогнозирование зависимой переменной).

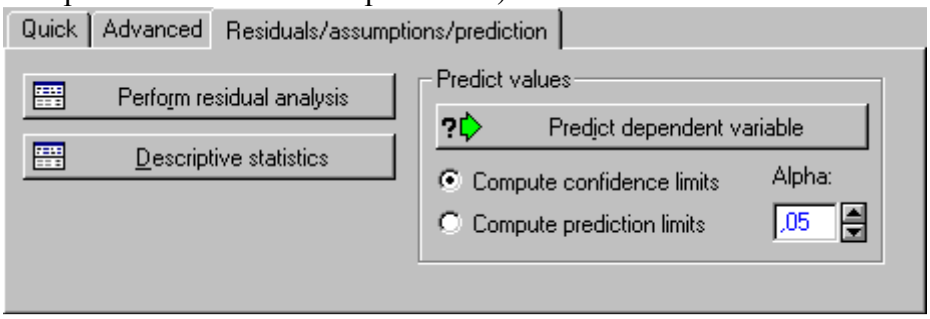

Рисунок 2.9 – Окно установок прогноза (приведена часть исходного окна)

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 *X2max* = 55. Внесем данное значение в окно *Specify values for indep. vars* (Определение неизвестных значений для зависимой переменной).

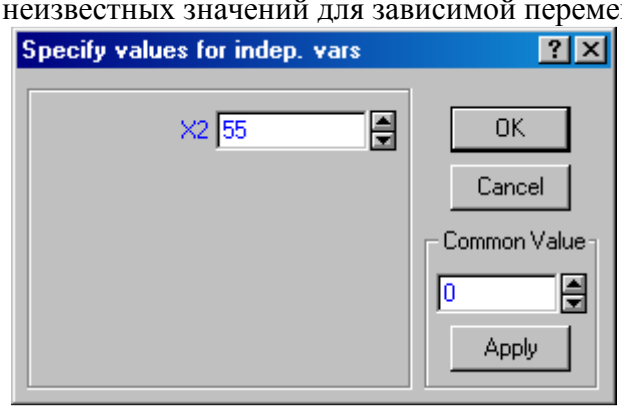

Рисунок 2.10 – Прогнозирование неизвестных значений зависимой переменной

После нажатия кнопки ОК получаем следующие результаты:

Таблица 2.7 – Прогнозные значения фондоотдачи при фиксированном значении среднегодовой стоимости ОПФ на уровне 55%

	B-Weight	*Value*	*B-Weight * Value*
X2	1,531	55	84,209
Intercept			-35,110
Predicted			49,099
-95,0%CL			43,929
+95,0%CL			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 прогнозирования фондоотдачи при среднем значении среднегодовой стоимости ОПФ.

Таблица 2.8 – Прогнозные значения фондоотдачи при фиксированном значении среднегодовой стоимости ОПФ на уровне 47,53%

	B-Weight	*Value*	*B-Weight * Value*
X2	1,531	47,530	72,772
Intercept			-35,110
Predicted			37,662
-95,0%CL			34,298
+95,0%CL			41,025

Рассмотрим полученные в таблице 2.8 и 2.9 результаты. В нашем случае прогноз фондоотдачи при значении *X2*=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 *X2*.

ЗАДАНИЕ:

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

Практическое занятие 5 (ПЗ-5) Статистическое изучение динамики социально-экономических явлений

1. Понятие и классификация рядов динамики.
2. Правила построения динамических рядов.
3. Показатели, характеризующие динамику.
4. Табличное и графическое представление динамических рядов.

Задача 2. Выявление тренда во временных рядах

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:

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 помощь аналитического выравнивания;

4) провести прогнозирование уровней ряда на перспективу;
5) выявить и устранить наличие автокорреляции.

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 обусловливает стохастическую природу элементов x_t, а, следовательно, и необходимость интерпретации $x_1,\ldots, x_T$ как наблюдений, произведенных над случайными величинами $\xi_1,\ldots, \xi_T$. 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:

7. определить, какие из неслучайных функций присутствуют во временном ряду;
8. построить «хорошие» оценки для тех неслучайных функций, которые идентифицированы во временном ряду;
9. 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 самим 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 трен.
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:

– Так, выравнивание по прямой линии (линейной функции) эффективно для рядов, уровни которых изменяются примерно в арифметической прогрессии, т.е. когда первые разности уровней (абсолютные приросты) более или менее постоянны.
– Если вторые разности уровней (ускорения) более или менее постоянны, то такое развитие хорошо описывается параболой 2-го порядка. Если постоянны *n*-е разности уровней, можно использовать параболу *n*-го порядка, позволяющую «улавливать» перегибы, изломы в кривой, смену направлений изменения уровней. Парабола 2-го порядка отражает развитие с ускоренным или замедленным изменением уровней ряда.
– Если при последовательном расположении *t* (меняющемся в арифметической прогрессии) значения уровней меняются в геометрической прогрессии, т.е. цепные коэффициенты роста примерно постоянны, то такое развитие можно отразить показательной функцией.
– Если обнаружено замедленное снижение уровней ряда, которые по логике не могут снизиться до нуля, для описания характера тренда выбирают гиперболу.

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

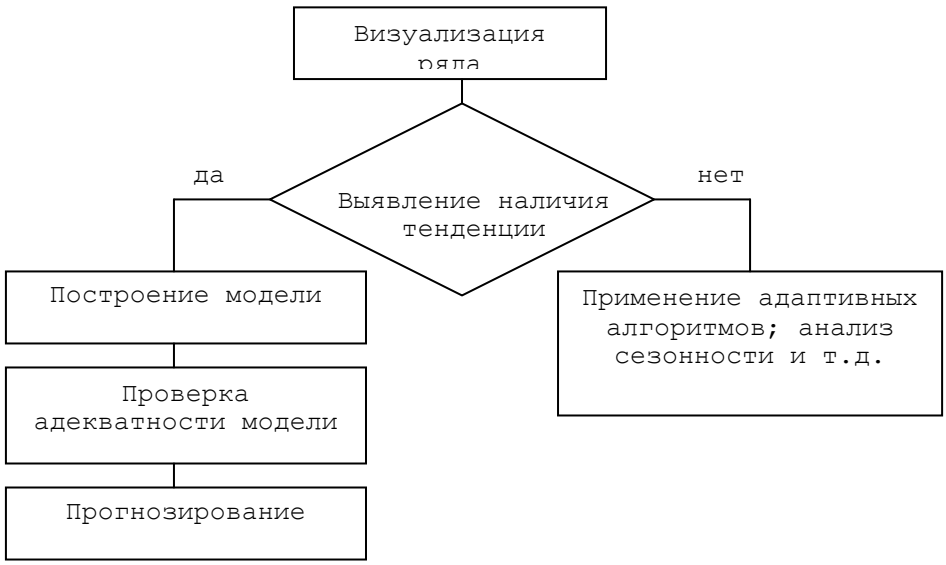

Рисунок 1 – Методика выявления и анализа тренд составляющей ряда

3. Выявление наличия тенденции в динамическом ряду

В качестве данных для проведения анализа выберем временной ряд динамики ВВП России в текущих ценах (трлн. руб.) с 1 квартала 1999 года до 4 квартала 2006 года, 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 $t_{факт} > t_{кр}$, 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 5А), 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 5.2).

	1 GDP1	2 GDP2
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 5.2 – Исходная таблица для проведения теста на наличие тенденции

Шаг 2. В главном меню выбираем *Statistics* → *Basic Statistics* → *t-test for Dependent Samples* (Расчеты → Основные статистики → *t*-тест для зависимых выборок). Результаты расчетов представим в таблице 5.1.

Таблица 5.1 - Результаты сравнения двух средних на основе ряда индекса ВВП России

	Mean	*Std.Dv.*	*N*	*Diff.*	*Std.Dv.Diff.*	*t*	*df*	*p*
GDP1	2416,479	811,969						
GDP2	6184,107	1559,563	14	-3767,63	868,668	-16,228	13	0,000

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

4. Сглаживание уровней ряда

Прежде чем приступить к выполнению процедуры необходимо активировать исходную таблицу ряда динамики ВВП (приложение 5А).

Одним из распространенных методов анализ тенденций временного ряда является использование скользящих средних. В пакете STATISTICA 6.0 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).

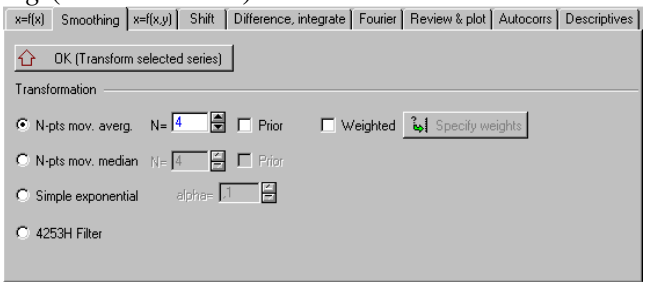

Рисунок 5.3 – Установки для сглаживания уровней динамического ряда (приведена часть исходного окна)

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 5.4), но как можно заметить информация плохо читаема. Поэтому 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 →2DГрафики → Линейный график) укажем опцию *Multiple* (Множественный). Полученные результаты представим на рисунке 5.5.

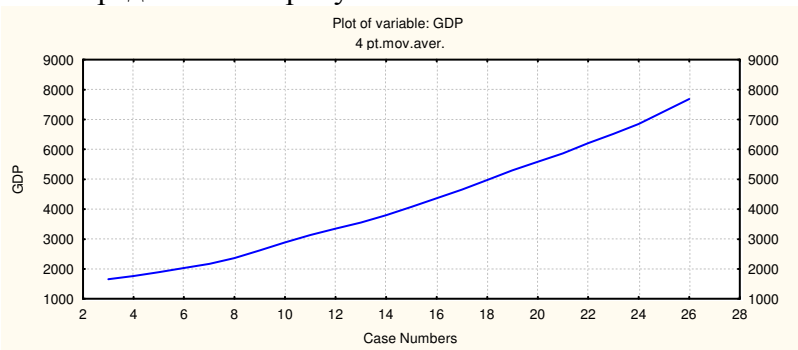

Рисунок 5.4 – Сглаженные уровни ВВП России методом скользящей средней

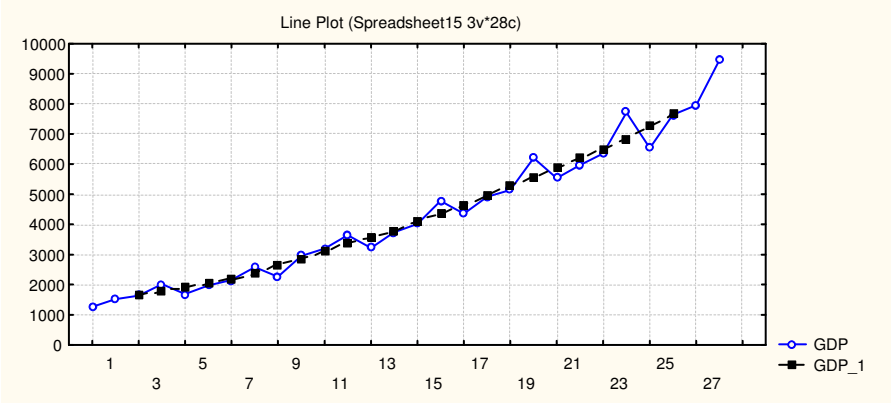

Рисунок 5.5 – Динамика и выровненные уровни ВВП России

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

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.*

5.1 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 5.6) 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.

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 5.7).

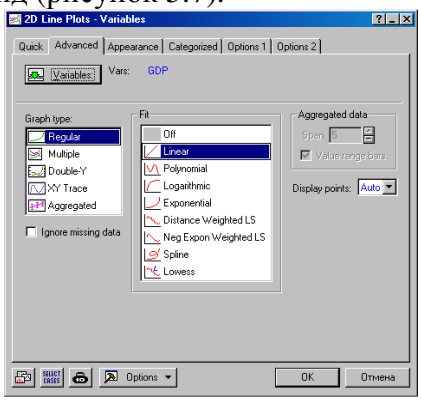

Рисунок 5.6 – Диалоговое окно *2D Line Plots - Variables*

Согласно данным, приведенным на рисунке 5.7, 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.

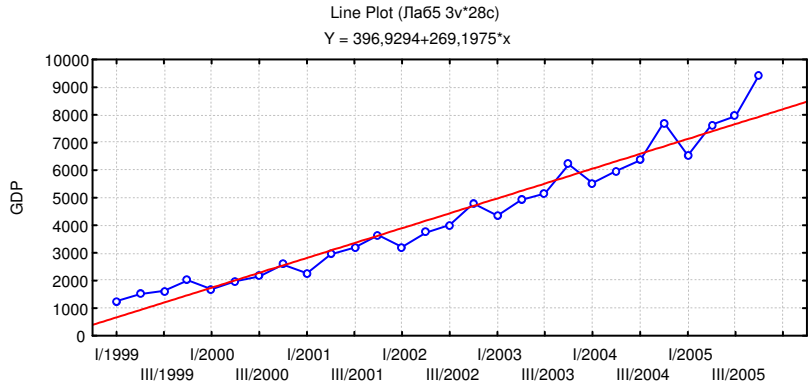

Рисунок 5.7 – Динамика ряда ВВП России, млрд. руб.

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.

Таблица 5.2 – Методы расчета параметров линейного тренда

Моменты (периоды) времени расстановлены от середины ряда	Моменты (периоды) времени расстановлены от начала ряда
$a_0 = \frac{\sum y_t}{n}$	$a_0 = \bar{y} - a_1\bar{t}$

$a_1 = \frac{\Sigma y_t t_t}{\Sigma t_t^2}$	$a_1 = \frac{\overline{yt} - \bar{y} \cdot \bar{t}}{\overline{t^2} - (\bar{t})^2}$

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.

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:

Data: Лаб 5.sta (4v by 28c)

	1 GDP	2 t1	3 t2
I/1999	1259,1	1	-13,5
II/1999	1514,7	2	-12,5
III/1999	1632,5	3	-11,5
IV/1999	2015,6	4	-10,5
I/2000	1688,1	5	-9,5
II/2000	1970	6	-8,5
III/2000	2152,2	7	-7,5
IV/2000	2582,9	8	-6,5
I/2001	2267,5	9	-5,5
II/2001	2950,4	10	-4,5
III/2001	3193,2	11	-3,5
IV/2001	3649,6	12	-2,5
III/2002	3219,7	13	-1,5
III/2002	3735,2	14	-0,5
III/2002	4012,3	15	0,5
III/2002	4785,1	16	1,5
I/2003	4354	17	2,5
II/2003	4916,8	18	3,5
III/2003	5147,8	19	4,5

Рисунок 5.8 – Динамика ВВП России в текущих ценах, трлн. руб. (приведена часть исходного окна)

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

Шаг 1: В главном меню выбреем *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).

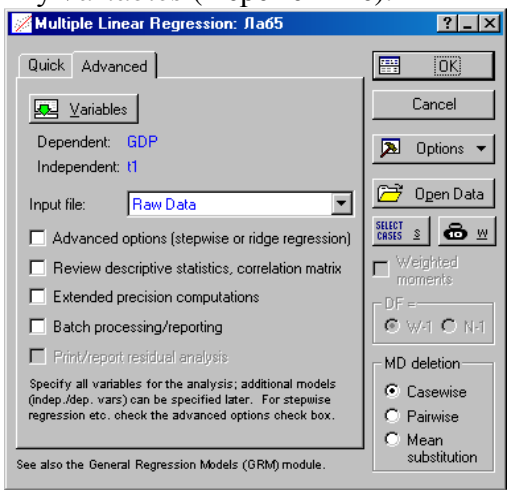

Рисунок 5.9 – Диалоговое окно *Multiple Linear Regression* (Множественная линейная регрессия)

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 *t1*. 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.

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

	Value
Multiple R	0,974
Multiple R?	0,949
Adjusted R?	0,947
F(1,26)	480,951
p	0,000
Std.Err. of Estimate	524,674

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 α=0,05 и степенях свободы v_1=1, v_2=26.

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

	Beta	*Std.Err. of Beta*	*B*	*Std.Err. of B*	*t(26)*	*p-level*
Intercept			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*

	Value
Multiple R	0,974
Multiple R?	0,949
Adjusted R?	0,947

F(1,26)	480,951
p	0,000
Std.Err. of Estimate	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 *t2*

	Beta	*Std.Err. of Beta*	*B*	*Std.Err. of B*	*t(26)*	*p-level*
Intercept			4300,293	99,154	43,370	0,000
t2	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.

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, чаше 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

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:

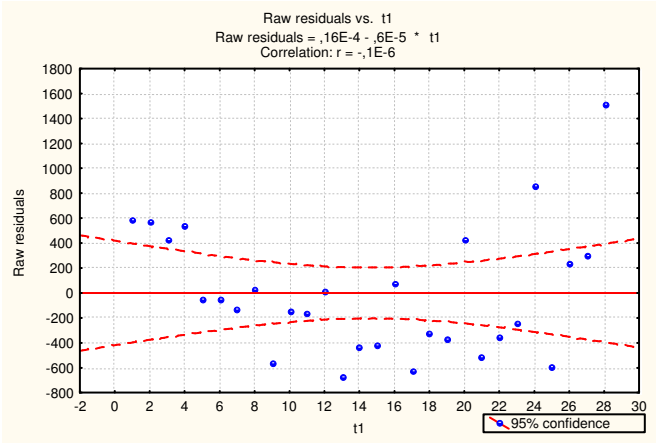

Рисунок 5.10 – Динамики случайного члена временного ряда

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

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).

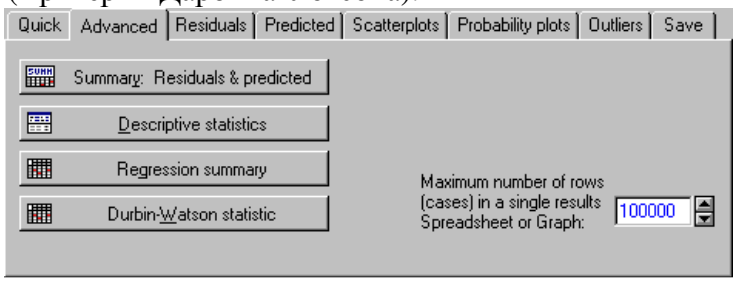

Рисунок 5.11 – Выбор критерия Дарбина-Уотсона (приведена часть исходного окна)

В результате оценки получаем следующую таблицу 5.7:

Таблица 5.7 – Значение критерия Дарбина-Уотсона

	Durbin-Watson	*Serial Corr.*
Estimate	1,317	0,230

Значение критерия отлично от двух (приложение 11В), при этом согласно таблице критических значений данной статистки при n=25 и k=2 верхняя граница $d_н$=1,21 и нижняя граница $d_в$=1,55. Отсюда получаем, что фактическое значение попадает в зону неопределенности $d_н<DW<d_в$, соответственно невозможно точно сделать вывод о наличии (отсутствии) автокорреляции (см. рисунок 5.12).

Есть положительная автокорреляция остатков. *H0* отклоняется с вероятностью P=(1-α) принимается *H1*	Зона неопределенности	Нет оснований отклонять H0 (автокорреляция остатков отсутствует)	Зона неопределенности	Есть отрицательная автокорреляция остатков. *H0* отклоняется с вероятностью P=(1-α) принимается *H1*
0	$d_н$	$d_в$ 2	4-$d_н$	4-$d_в$ 4

Рисунок 5.12 - Механизм проверки гипотезы о наличии автокорреляции остатков

6.3. Тест серий Бреуша-Годфри

Для реализации данного теста в пакете STATISTICA 6.0, необходимо выполнить следующие шаги:

Шаг 1. Воспользуемся результатами оценки тренда - $\tilde{y}_t = a_0 + a_1 t 2_t$, 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…* и сделаем установки как показано на рисунке 5.13.

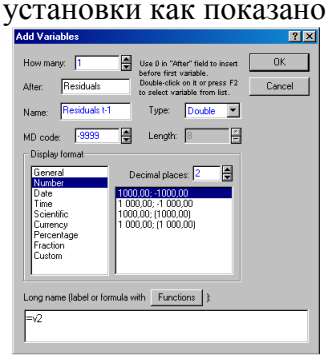

Рисунок 5.13 – Образование новой переменной *Residuals t-1*

В результате образования новой переменной получаем таблицу представленную на рисунке 5.14.

	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 5.14 – 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 5.15) в поле *Lag* укажем значение 1.

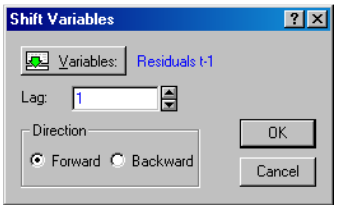

Рисунок 5.15 – Создание запаздывающей переменной

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 - $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$

	Beta	*Std.Err. of Beta*	*B*	*Std.Err. of B*	*t(26)*	*p-level*
Residuals t-1	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.

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 *GDP* в качестве не зависимых - *t2* и *GDP t-1*. Получаем следующие результаты:

Таблица 5.9 – Результаты построения модели вида $\tilde{y}_t = a_0 + a_1 t2_t + a_2 y_{t-1}$

	Beta	*Std.Err. of Beta*	*B*	*Std.Err. of B*	*t(24)*	*p-level*
Intercept			3221,338	977,606	3,295	0,003
GDP t-1	0,246	0,227	0,265	0,244	1,084	0,289
t2	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

	Durbin-Watson	*Serial Corr.*
Estimate	1,946	-0,186

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 *t2* соответствующее прогнозному периоду – в данном случае 14,5.

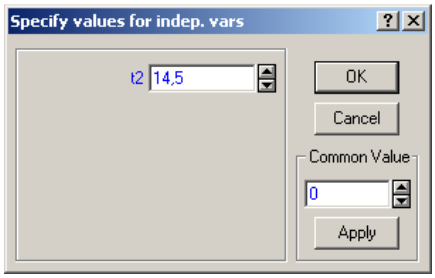

Рисунок 5.16 – Окно выбора прогнозного периода

После нажатия кнопки ОК, получаем следующие результаты.

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

	B-Weight	*Value*	*B-Weight * Value*
t2	269,197	14,500	3903,363
Intercept			4300,293
Predicted			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. руб.

Шаг 3. Аналогичным образом рассчитаем значения для 3, 4 и 5 кварталов 2006 года (*t2* = 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.

9. Задание для самостоятельного выполнения

Таблица 1 – Производство молока сельскохозяйственными организациями Оренбургской области, тонн

	2003	2004	2005	2006	2007
январь	16754,7	14866,6	13001,1	12915,1	13098,5
февраль	19609	17518,1	15893,1	14682,5	15050,1
март	22463,3	21565,5	19250,3	18113,9	18299,8
апрель	23061,4	23135	19467,4	19554,3	19475,1
май	35038,2	30788,1	27097,2	26996,3	26843,2
июнь	45883,2	41708,4	35390,2	33869	31989,5
июль	42740,6	39219,6	34090,4	30949,8	31646,6
август	35669,3	34178	30172	27582,4	26493,7
сентябрь	27270,5	26149,2	22846,9	21529,2	27291,1
октябрь	18208,7	17767,6	15241,1	15601,2	15671,1
ноябрь	14890,7	12572,6	13100	12029,7	12755,6
декабрь	13240,3	12304,1	11099	14915,2	14924,2

1) Введите данные в СПП STATISTICA 6.0 (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.

Практическое занятие 6 (ПЗ-6) Измерение взаимосвязей общественных явлений

1. Сущность и задачи корреляционно-регрессионного анализа.
2. Параметрические методы изучения связи.
3. Технология решения корреляционного и регрессионного анализа в MS Excel и в ППП "Statistica".

Задача 3. Экспоненциальное сглаживание временных рядов

4.1. Цели и задачи лабораторной работы

В данной лабораторной работе проведем построение и прогнозирование динамики с помощью метода экспоненциального сглаживания, при этом будут решаться следующие задачи:

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

4.2. 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.

4.3. 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).

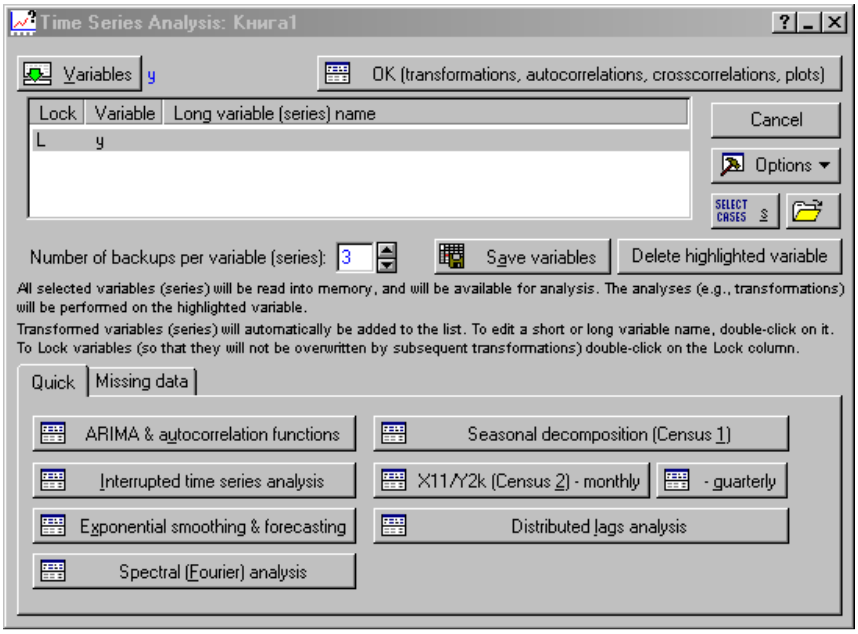

Рисунок 6.1 – Окно выбора методов анализа временных рядов

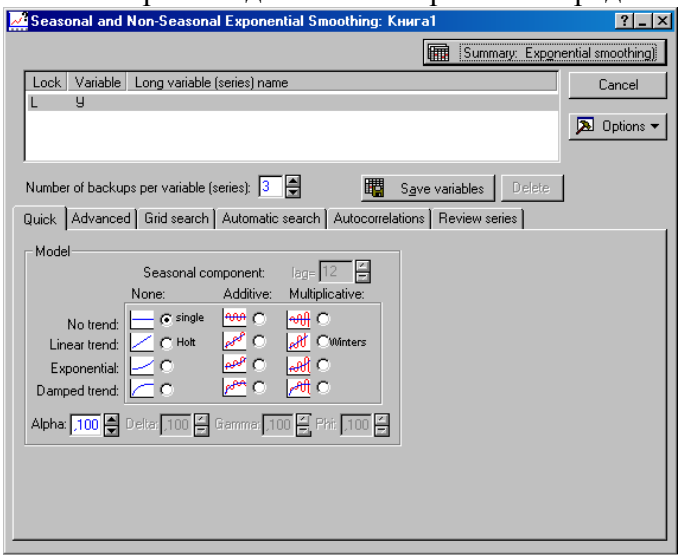

Рисунок 6.2 – Окно настроек моделей экспоненциального сглаживания

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 α=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*).

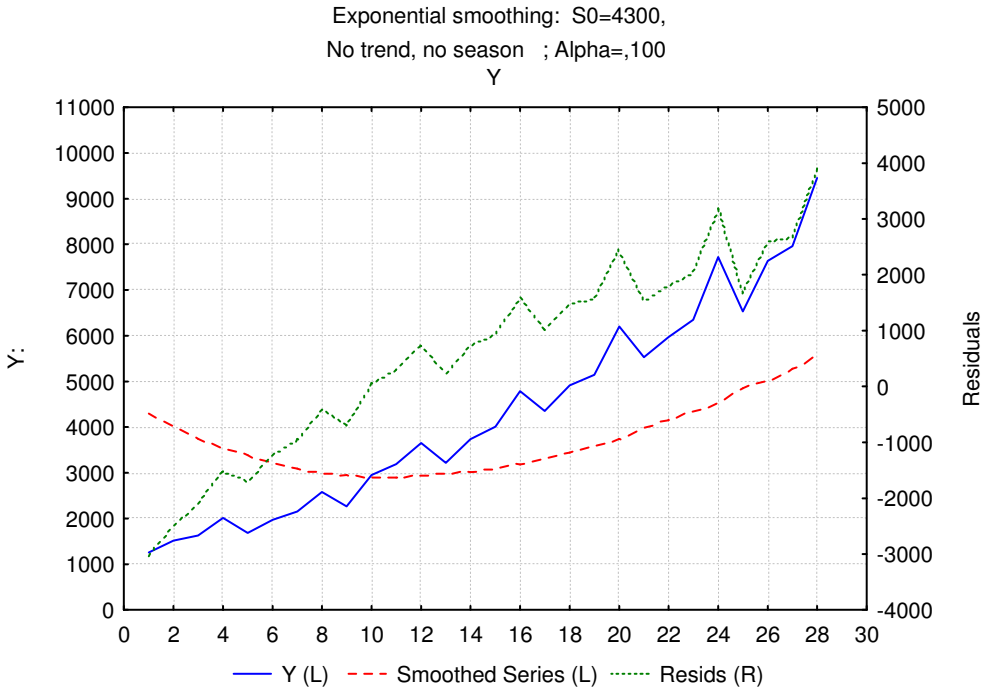

Рисунок 6.3 - Графики сглаженного временного ряда среднедушевых денежных доходов населения при α=0,1.

Аналогичные вычисления проведем для α= 0,5 и α = 0,9.

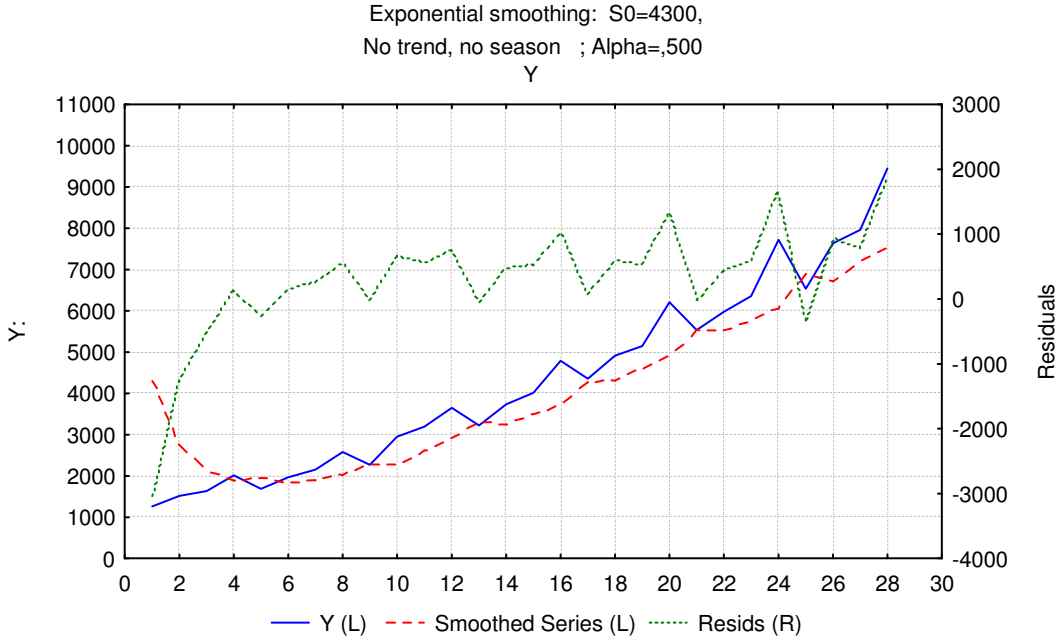

Рисунок 6.4 - Графики сглаженного временного ряда среднедушевых денежных доходов населения при α=0,5

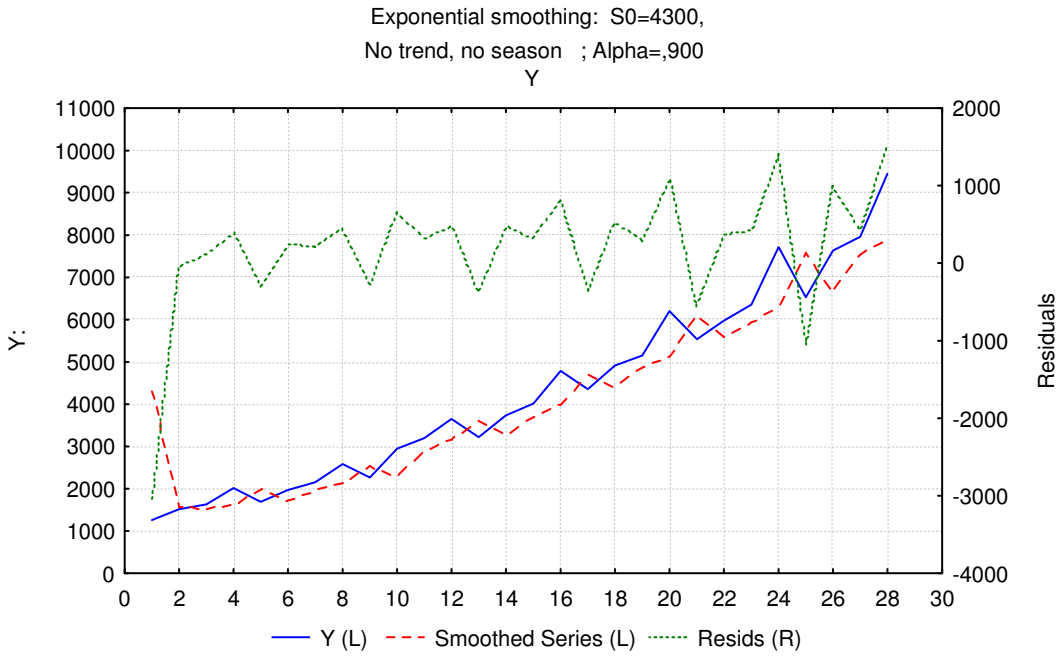

Рисунок 6.5 - Графики сглаженного временного ряда среднедушевых денежных доходов населения при α=0,9.

Анализируя приведенные графики нетрудно заметить, что чем меньше значение сглаживающей константы, тем меньше варьируют сглаженные значения. При малой величине α= 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 α= 0,9 дает гораздо меньший сглаживающий эффект, однако сглаженные значения в большей степени следуют фактическим значениям по всей длине исходного временного ряда.

Константа α= 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 6.0 необходимо в окне *Seasonal and Non-Seasonal Exponential smoothing* выбрать вкладку *Advanced* (Расширенные).

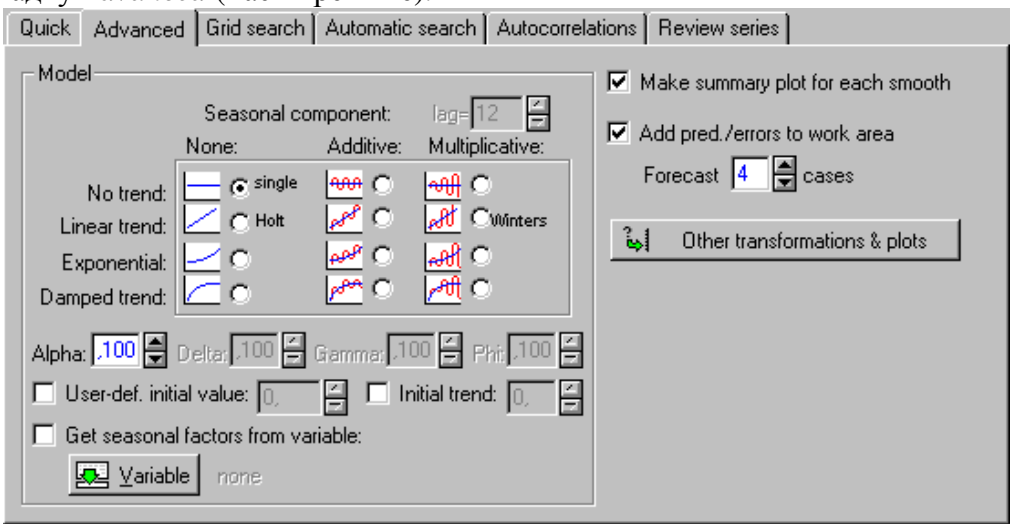

Рисунок 6.6 – Окно установок для прогнозирования (приведена часть исходного окна)

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.

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

Квартал/год	Модели			
	Фактические значения	α=0,1	α=0,5	α=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

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.

6.4. 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).

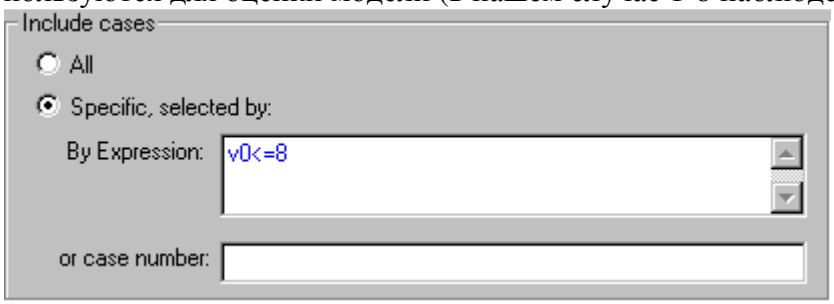

Рисунок 6.7 – Выбор наблюдений для оценки линейного тренда (приведена часть исходного окна)

После проведения процедуры получаем следующие результаты:

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

	Beta	*Std.Err. of Beta*	*B*	*Std.Err. of B*	*t(6)*	*p-level*
Intercept			1148,007	134,123	8,559	0,000
t	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* выбрем 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.

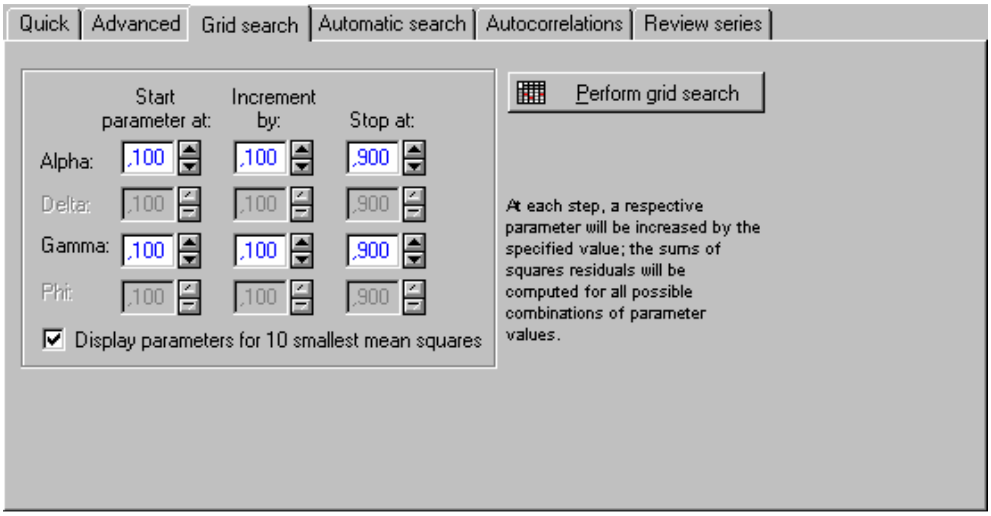

Рисунок 6.7 – Установки процедуры поиска оптимальных вариантов констант (приведена часть исходного окна)

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)

	Alpha	*Gamma*	*Mean Error*	*Mean Abs Error*	*Sums of Squares*	*Mean Squares*	*Mean % Error*	*Mean Abs Error*
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)

	Alpha	*Gamma*	*Mean Error*	*Mean Abs Error*	*Sums of Squares*	*Mean Squares*	*Mean % Error*	*Mean Abs Error*
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.

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

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.

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

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:

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

где: y_t - наблюдаемое значение в момент времени t,

y_t - прогноз (сглаженное значение).

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

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

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.

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

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:

$\bar{\varepsilon}$ **,%**	Интерпретация точности
< 10	Высокая
10 – 20	Хорошая
20 – 50	Удовлетворительная
> 50	Не удовлетворительная

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).

6.5. 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* ___ (Лаг 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)

	Alpha	*Delta*	*Gamma*	*Mean Error*	*Mean Abs Error*	*Sums of Squares*	*Mean Squares*	*Mean % Error*	*Mean Abs Error*
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 вкладку *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)

	Alpha	*Delta*	*Gamma*	*Mean Error*	*Mean Abs Error*	*Sums of Squares*	*Mean Squares*	*Mean % Error*	*Mean Abs Error*
239	0,3	0,9	0,5	64,752	182,265	**1381981**	**49356,45**	1,223	5,106410
238	0,3	0,9	0,4	74,993	**178,374**	1398731	49954,69	1,402	**4,926425**
242	0,3	0,9	0,8	**47,060**	189,521	1430175	51077,67	**0,908**	5,509370

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	-	-
Alpha=0,3 *Delta*=0,9 *Gamma*=0,5 (a_1 = 156,418 и y_1=1259,1)	8284,16	9344,05	9705,14	11145,25	5,11
Alpha=0,3 *Delta*=0,9 *Gamma*=0,4 (a_1 = 156,418 и y_1=1259,1)	8251,06	9290,60	9637,92	11063,27	4,93
Alpha=0,3 *Delta*=0,9 *Gamma*=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. Задания для самостоятельного выполнения

Таблица 1 – Производство молока сельскохозяйственными организациями Оренбургской области, тонн

	2003	2004	2005	2006	2007
январь	16754,7	14866,6	13001,1	12915,1	13098,5
февраль	19609	17518,1	15893,1	14682,5	15050,1
март	22463,3	21565,5	19250,3	18113,9	18299,8
апрель	23061,4	23135	19467,4	19554,3	19475,1
май	35038,2	30788,1	27097,2	26996,3	26843,2
июнь	45883,2	41708,4	35390,2	33869	31989,5
июль	42740,6	39219,6	34090,4	30949,8	31646,6
август	35669,3	34178	30172	27582,4	26493,7
сентябрь	27270,5	26149,2	22846,9	21529,2	27291,1
октябрь	18208,7	17767,6	15241,1	15601,2	15671,1
ноябрь	14890,7	12572,6	13100	12029,7	12755,6
декабрь	13240,3	12304,1	11099	14915,2	14924,2

Практическое занятие 7 (ПЗ-7) Российские статистические пакеты прикладных программ

1. Отечественные статистические пакеты прикладных программ.
2. ППП "STADIA"
3. История создания системы ЭВРИСТА
4. ППП "ОЛИМП"
5. ППП "МЕЗОЗАВР"

Практическое занятие 8 (ПЗ-8) Зарубежные статистические пакеты прикладных программ

1. SAS.
2. MINITAB.
3. SPSS.

4. Statistica

Задача 5. Оценка нелинейных 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.

10.3 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, равносторонняя 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 удается. В этом случае применяются методы нелинейной оптимизации на основе исходных переменных.

10.3 Рекомендуемая литература

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**

$$\tilde{y}_i = AK^{\alpha}L^{\beta} \qquad (10.1)$$

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 α% (β%).

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:

$$lny = lnA + \alpha lnK + \beta lnL$$

Далее заменяем $y' = lny$, $x_1 = lnK$, $x_2 = lnL$, получаем

$$\tilde{y}' = a_0 + a_1 x_1 + a_2 x_2 \qquad (10.2)$$

a_1 – оценка α; a_2 – оценка β; a_0 – оценка lnA

Для построения производственной функции Кобба-Дугласа в пакте STATISTICA воспользуемся данными по 14 субъектам Приволжского федерального округа (приложение 10А). 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 СПП STATISTICA 6.0 существует специальный модуль *Fixed Nonlinear Regression* (Фиксированная нелинейная регрессия) предназначенный для оценки нелинейных моделей посредством линеаризации 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).

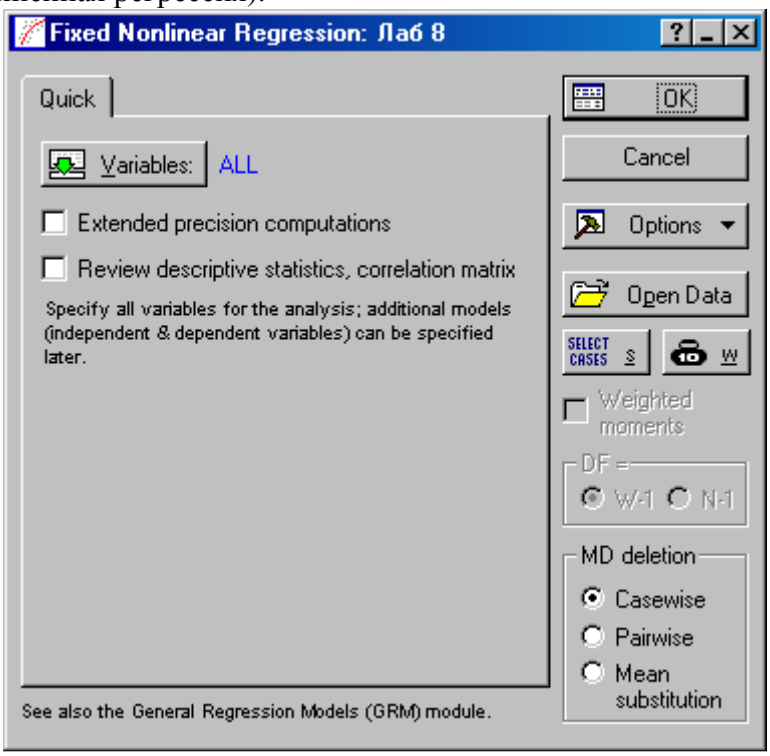

Рисунок 10.1 - Окно установок нелинейной регрессии

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.

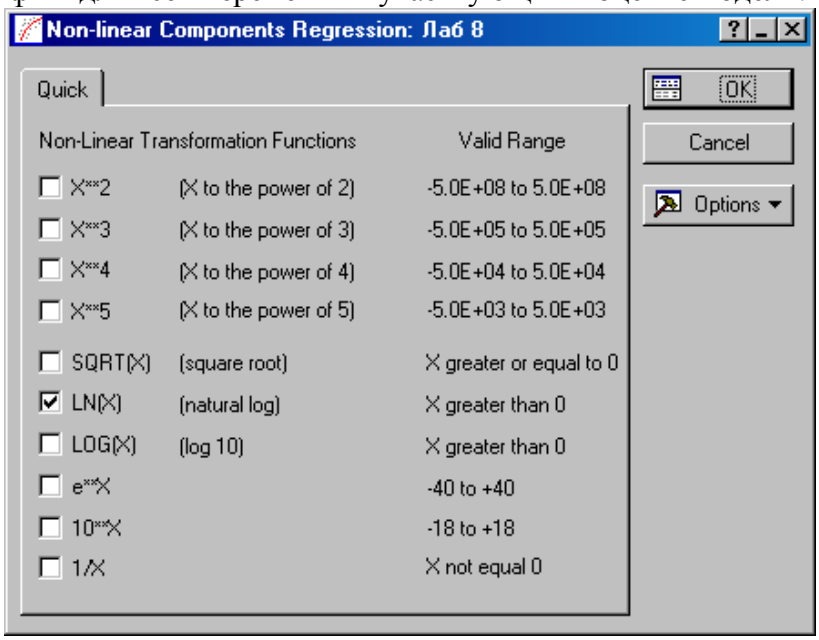

Рисунок 10.2- Установка процедур линеаризации исходных переменных

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*.

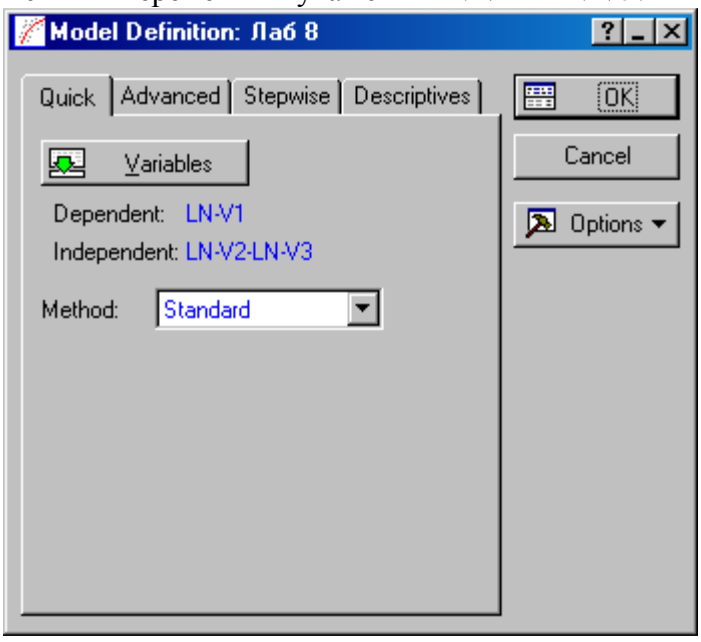

Рисунок 10.3 – Окно выбора зависимой и независимых переменных для построения производственной функции

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

	Value
Multiple R	0,991
Multiple R?	0,982
Adjusted R?	0,979
F(2,11)	308,655
p	0,000
Std.Err. of Estimate	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

	Beta	*Std.Err. of Betta*	*B*	*Std.Err. of B*	*t(11)*	*p-level*
Intercept			1,505	0,424	3,548	0,005
LN-V2	0,602	0,094	0,556	0,087	6,423	0,000
LN-V3	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. Образуем 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 убедится 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 1).

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).

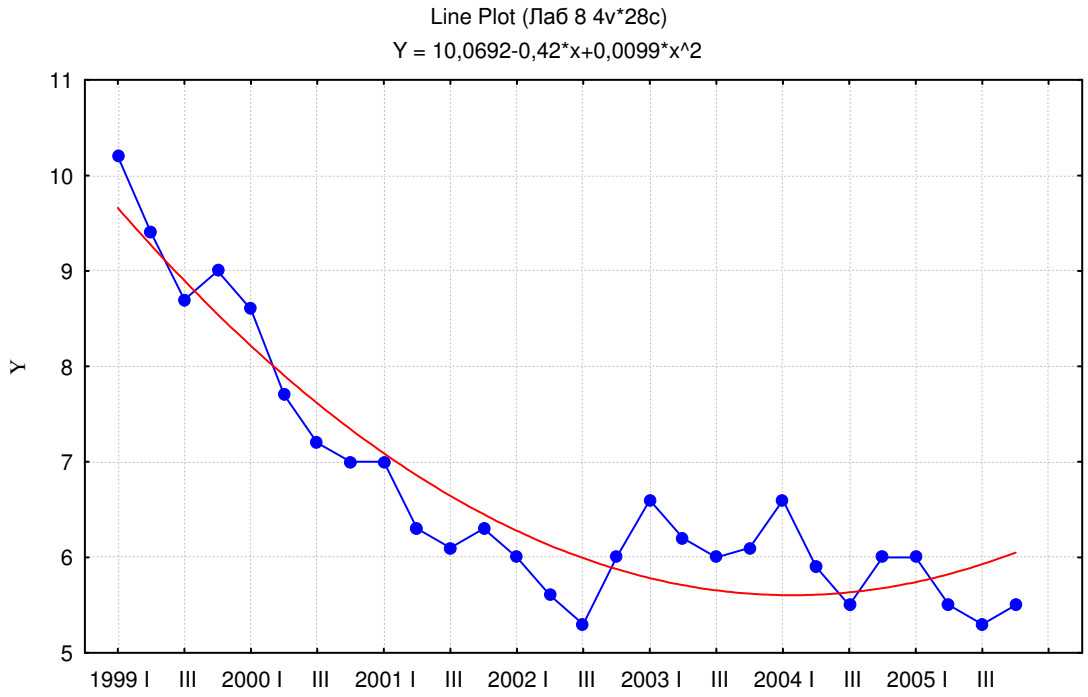

Рисунок 10.4 – Динамика численности безработных в РФ и выровненные на основе параболы второго порядка уровни

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:

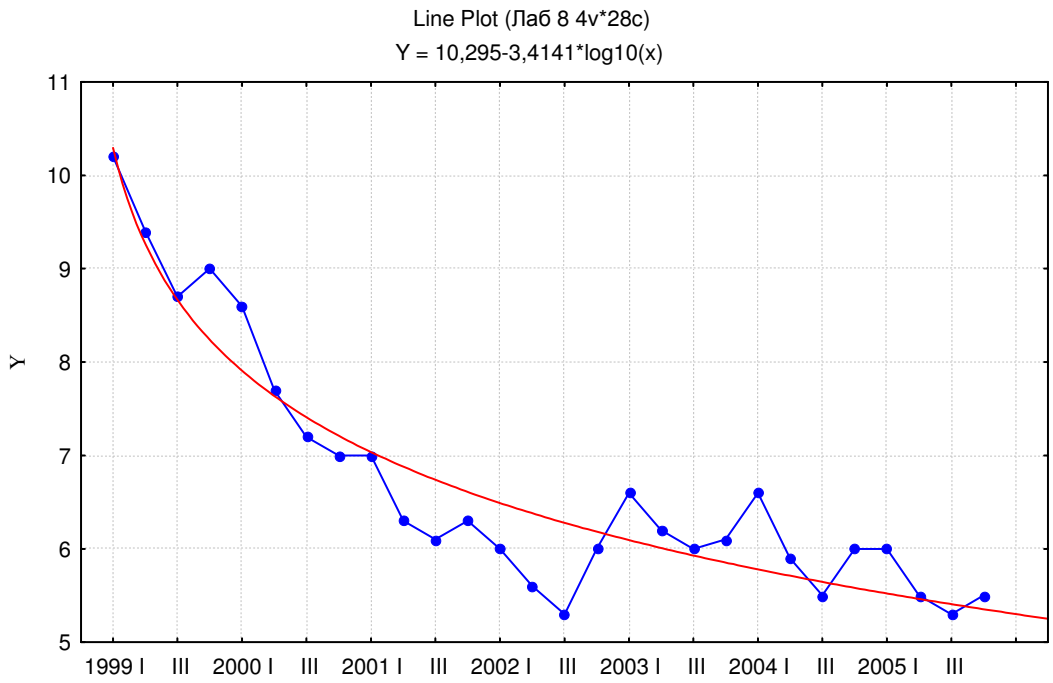

Рисунок 10.5 – Динамика численности безработных в РФ и выровненные на основе логарифмической прямой уровни

Согласно данным, приведенным на рисунке 5, выбранная функция (в отличие от параболы второго порядка) имеет тенденцию к снижению, что в прогнозном периоде может привести к пересечении с осью *OX*. С математический 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 прицидента 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:

3) 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;*
4) $\frac{1}{t}$ - после столбца *t^2* добавляем переменную *1/t* и указываем *=1/v2*;
5) *log10t* – после столбца *1/t* добавляем переменную *log10 t* и указываем *=log10(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.* - *log10t.*

Результаты оценки трех кривых приведем в таблице 10.3.

Таблица 10.3 - Сравнительные характеристики динамических моделей ряда численности безработных в РФ

Коэффициенты	Модель	Коэффициент детерминации R^2	Скорректированный коэффициент детерминации AR^2	F-критерии Фишера	Стандартное отклонение
Парабола второго порядка	$\tilde{y}_t = 10{,}06 - 0{,}42t_t + 0{,}01t^2_t$ (34,15) (-8,96) (6,30)	0,878	0,868	89,615	0,483
Гипербола	$\tilde{y}_t = 5{,}91 + 5{,}66\frac{1}{t_t}$ (34,76) (7,98)	0,710	0,699	63,748	0,729
Логарифмическая прямая	$\tilde{y}_t = 10{,}295 - 3{,}41 \log 10\, t_t$ (38,72) (-14,29)	0,887	0,883	204,159	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 гиперболе 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
0,031

Таблица 10.4 – Прогнозируемые значения числа безработных в РФ

Годы	Фактические значения	Прогнозные значения по параболе 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.

10.7 Задания для самостоятельного выполнения

На основе фактических данных оценить наилучшую нелинейную модель для соответствующего временного ряда и провести прогнозирование на 3 уровня вперед.

годы	1991	1992	1993	1994	1995	1996	1997	1998
Чисенность родженных, тыс. чел.	1860	1860	1736	1612	1705	1984	1860	1643
1999	2000	2001	2002	2003	2004	2005	2006	2007
1240	1457	1674	1736	1488	1178	1953	1798	1457
2008	2009	2010	2011	2012	2013	2014	2015	
2108	1829	1581	1612	1829	2201	1612	2046	