

Министерство образования и науки Российской Федерации

Федеральное государственное автономное образовательное учреждение  
высшего образования

Санкт-Петербургский государственный политехнический университет

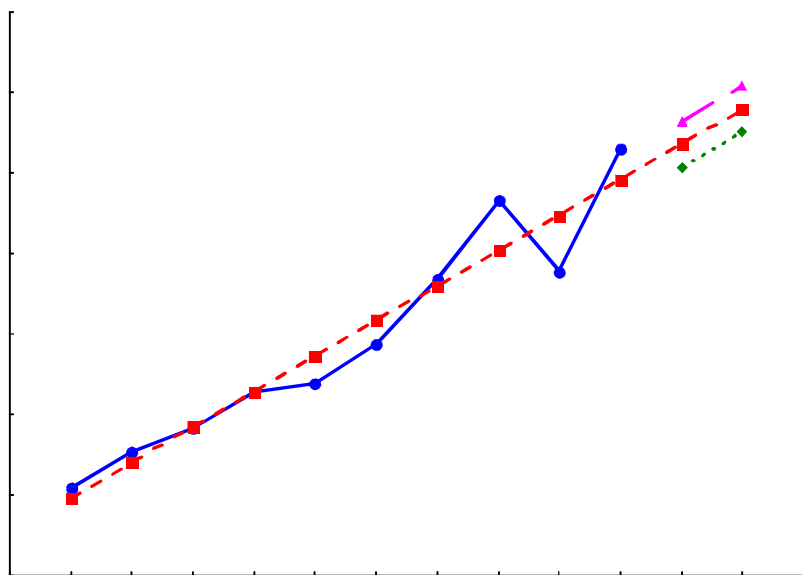
Н.В.Куприенко О.А.Пономарева Д.В. Тихонов

## СТАТИСТИКА.

### Временные ряды.

### Анализ тенденций и прогнозирование

### Учебное пособие



Санкт-Петербург

Издательство Политехнического университета

2015

Н. В. Куприенко **Статистика. Временные ряды. Анализ тенденций и прогнозирование:** учеб. пособие. / Н. В. Куприенко, О. А. Пономарева, Д. В. Тихонов. – СПб.: Изд-во Политехн. ун-та, 2015. – 123 с.

В учебном пособии рассматриваются возможности использования пакета прикладных программ (ППП) STATISTICA для реализации статистических методов анализа временных рядов в объеме, достаточном для решения широкого круга практических задач. Рекомендуется студентам инженерно-экономического института, изучающим дисциплину «Статистика».

Пособие может быть использовано студентами дневной, вечерней и заочной форм обучения, а также бакалаврами и магистрами при написании выпускных работ; аспирантами, научными и практическими работниками, столкнувшимися с необходимостью использования статистических методов обработки исходных данных. Пособие содержит сведения по ППП STATISTICA, не публиковавшиеся на русском языке.

Табл.11 Рис. 101 Библиогр.: 9 назв.

Печатается по решению редакционно-издательского совета Санкт-Петербургского государственного политехнического университета.

© Куприенко Н. В.,  
Пономарева О. А.,  
Тихонов Д. В., 2015

© Санкт-Петербургский  
государственный политехнический  
университет, 2015

# СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
1. ОСНОВНЫЕ ПОНЯТИЯ И ТЕРМИНЫ В АНАЛИЗЕ ВРЕМЕННЫХ РЯДОВ	5
1.1. Определение временных рядов	5
1.2. Классификация временных рядов	5
1.3. Правила построения временных рядов	7
2. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ В СРЕДЕ STATISTICA	100
2.1. Ввод исходных данных	100
2.2. Графическое представление временных рядов	144
2.3. Показатели изменения уровней временного ряда	233
2.4. Средние показатели динамики	400
2.5. Периодизация временных рядов	467
3. ВЫЯВЛЕНИЕ И АНАЛИЗ ОСНОВНОЙ ТЕНДЕНЦИИ ВРЕМЕННОГО РЯДА	533
3.1. Компоненты временного ряда	533
3.2. Методы анализа основной тенденции временных рядов	544
3.2.1. Проверка временного ряда на наличие тренда	544
3.2.2. Механическое выравнивание временного ряда. Скользящие средние	577
3.3. Аналитическое сглаживание временного ряда. Уравнение тренда	622
3.4. Выбор тренда	800
3.5. Оценка автокорреляции в остатках уравнения тренда	877
4. ЭКСТРАПОЛЯЦИЯ И ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ПРОГНОЗА	933
4.1. Экстраполяция на основе тренда	933
4.2. Доверительные интервалы прогноза	955
4.3. Графическое представление результатов прогнозирования	986
5. АВТОКОРРЕЛЯЦИЯ В ДИНАМИЧЕСКИХ РЯДАХ. АВТОРЕГРЕССИОННЫЕ МОДЕЛИ.	1011
6. КОРРЕЛЯЦИЯ РЯДОВ ДИНАМИКИ	1056
ЗАКЛЮЧЕНИЕ	1133
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	1144
<i>Приложение 1. Содержание курсового проекта</i>	<i>115</i>
<i>Приложение 2. Образец титульного листа</i>	<i>117</i>
<i>Приложение 3. Таблица критических значений t-критерия Student'a</i>	<i>118</i>
<i>Приложение 4. Таблица значений F-распределения</i>	<i>119</i>
<i>Приложение 5. Таблица критических значений коэффициента корреляции Пирсона</i>	<i>121</i>
<i>Приложение 6. Таблица значений интегральной функции нормального распределения</i>	<i>122</i>

# ВВЕДЕНИЕ

Пособие продолжает ряд учебных материалов, посвященных использованию пакета прикладных программ STATISTICA для решения различных аналитических задач. В отличие от предыдущих [2,3], в данном пособии уже не приводятся общесистемные приемы работы в среде ППП STATISTICA. Основное внимание уделено реализации методов статистического анализа временных рядов. На сквозном числовом примере достаточно подробно рассмотрен комплекс процедур, обеспечивающий решение такой популярной и важной в повседневной практике задачи, как детальный анализ и прогнозирование тенденций развития.

При изучении курса «Статистика» студенты 3 курса всех направлений выполняют курсовую или лабораторную работу, посвященную всестороннему анализу временных рядов. Целью выполнения работы является освоение методов анализа временных рядов в ППП STATISTICA. Выполнение работы предусматривает: оценку скорости и интенсивности изменения уровней изучаемого временного ряда, т.е. расчет абсолютных, относительных и средних показателей динамики; выявление основной тенденции ряда методами эмпирического и аналитического сглаживания; построение и оценку уравнения тренда; изучение автокорреляции и построение авторегрессионной модели; экстраполяцию на основе трендовой и авторегрессионной моделей; рассмотрение корреляционной зависимости временных рядов.



# 1. ОСНОВНЫЕ ПОНЯТИЯ И ТЕРМИНЫ В АНАЛИЗЕ ВРЕМЕННЫХ РЯДОВ

## 1.1. Определение временных рядов

Важнейшей задачей практической статистики, а также менеджеров разных уровней является построение и анализ временных рядов.

*Ряд динамики (временной ряд)<sup>1</sup> – это расположенные в хронологическом порядке значения того или иного показателя, изменение которого отражает ход развития изучаемого явления.*

Временной ряд состоит из двух элементов: моменты или периоды времени (годы, кварталы, месяцы), к которым относятся статистические данные, называемые уровнями ряда. Общепринятое формальное представление динамического ряда:

$$y_1, y_2, \dots, y_t, \dots, y_n$$

где  $y_t$  – уровень ряда, численное значение показателя в момент (период) времени  $t$ ;  $n$  – число уровней ряда.

Процесс развития социально-экономических явлений во времени заключается главным образом в том, что происходит изменение воздействия на них многих факторов социального, экономического, технического и другого порядка. Время, таким образом, становится собирательным фактором, вмещающим в себя многие факторы развития. Экономические явления, как и все другие явления общественной жизни, с течением времени изменяются под влиянием целого ряда причин, но с внешней стороны это проявляется как результат воздействия времени.

Построение и анализ рядов динамики позволяют выявить закономерности развития явлений общественной жизни и их особенности.

## 1.2. Классификация временных рядов

Для выбора адекватной процедуры анализа конкретного динамического ряда необходимо знать их общую классификацию.

По времени, отражаемому в динамических рядах, разделяют

---

<sup>1</sup> Далее понятия «временной ряд» и «ряд динамики» используются как синонимы.

*моментные и интервальные.*

В *моментных* рядах динамики уровни ряда представлены показателями, значения которых фиксируются на какой-то момент времени. Примером моментного ряда могут служить данные табл. 1.1.

Таблица 1.1

Численность постоянного населения России  
на конец года, тыс. чел. [1]

Год	Численность постоянного населения	В том числе	
		Мужчин	Женщин
1991	148,3	69,5	78,8
1996	148,3	69,5	78,8
2004	144,2	67,0	77,2
2005	143,5	66,6	76,9
2006	142,8	66,2	76,6
2007	142,2	65,8	76,4
2008	142,0	65,7	76,3
2009	141,9	65,6	76,3
2010	141,9	65,6	76,3

Следует отметить, что *уровни моментных рядов динамики суммировать не имеет смысла*, т.к. результат суммирования содержательно не интерпретируется в силу наличия многократного повторного счета. Разность уровней моментного ряда динамики имеет определенный смысл, характеризуя изменение показателя за период между двумя смежными уровнями.

В *интервальных* рядах уровни ряда выражают размеры явления за определенный промежуток времени (сутки, неделю, месяц и т.д.). Отличительной особенностью интервальных рядов динамики абсолютных величин является возможность суммировать уровни следующих друг за другом периодов, поскольку их можно рассматривать как итог за более длительный период времени. Примером интервального ряда служат данные табл. 1.2. Каждый уровень этого ряда динамики отражает число человек, поступивших и окончивших высшие учебные заведения (ВУЗ) за конкретный период.

По *полноте времени*, отражаемого в рядах динамики, их можно разделить на ряды *полные и неполные*. В *полных* динамических рядах даты или периоды следуют друг за другом с равным интервалом (так называемые *равноотстоящие динамические ряды*). В *неполных* дина-

мических рядах в последовательности времени равный интервал не соблюдается (*не равноотстоящие динамические ряды*).

Таблица 1.2

Соотношение приема и выпуска в высших учебных заведениях  
России, тыс. чел. [1]

Годы	Принято в ВУЗы	Выпущено ВУЗаи
2008	1642	1358
2009	1544	1442
2010	1399	1468
2011	1207	1443

В табл. 1.1 представлен пример не равноотстоящего ряда, а в табл. 1.2 – равноотстоящего ряда динамики.

По способу выражения уровней временные ряды могут быть рядами *абсолютных, средних и относительных*, величин. Все рассмотренные ранее ряды показывают изменение во времени абсолютных показателей. Ряд динамики средних величин представлен в табл. 1.3.

Таблица 1.3

Общая площадь жилых помещений,  
приходящаяся в среднем на одного жителя России  
на конец года, кв. м. [1]

Год	2003	2004	2005	2006	2007	2008	2009
Площадь	20,2	20,5	20,9	21,3	21,5	22,0	22,4

В табл. 1.4 приведен пример динамического ряда относительных показателей.

Таблица 1.4

Ввод в действие жилых домов  
в г. Санкт-Петербург (базисный темп роста, проценты) [1]

Год	2003	2004	2005	2006	2007	2008	2009
% к 2003 г.	100,0	115,6	129,3	135,2	150,0	182,7	148,1

### 1.3. Правила построения временных рядов

Весьма важной проблемой построения динамических рядов является проблема сопоставимости уровней, их образующих. Очевидно, что динамический ряд, пригодный для содержательного анализа, должны составлять уровни ( $y_t$ ), во всех отношениях сопоставимые. Причин несопоставимости уровней ряда довольно много. Рассмотрим

наиболее типичные из них.

Уровни динамических рядов должны быть сопоставимы *по кругу охватываемых объектов (по полноте охвата)* и *по методике расчета* показателей. Требование неизменной полноты охвата различных частей исследуемого объекта означает, что уровни динамического ряда за отдельные периоды (моменты) времени должны выражать размер его по одному и тому же кругу структурных составляющих. Не следует объединять в один динамический ряд, например, данные в расчете на одного рабочего и одного работающего, данные по студентам дневной формы обучения с данными, относящимися ко всему контингенту студентов.

Сопоставимость по кругу охватываемых объектов в ряде случаев может быть достигнута через использование приема *смыкания динамических рядов*, при котором абсолютные уровни заменяются относительными, обычно выражаемыми в процентах. Приведем пример смыкания динамического ряда. Имеем статистику по продукции ООО «БалтТорг» (название условное), в которую входило 10 предприятий, а в 2006 г. влилось еще 2 предприятия. Реализация продукции этой фирмы выражалась следующим рядом динамики:

Таблица 1.6

Реализованная продукция фирмы ООО «БалтТорг», млн. руб.

Год	2003	2004	2005	2006	2007	2008	2009	2010
Продукция 10 предприятий	120	125	130	150	-	-	-	-
Продукция 12 предприятий	-	-	-	170	175	180	192	200

Для получения сомкнутого ряда, отражающего динамику реализованной продукции этой фирмы, принимаем за 100% данные 2006 г. как для последующих, так и предыдущих лет. Следовательно, за 100% для последующих лет будет взята величина 170 млн. руб., а для предыдущих – 150 млн. руб. Произведя вычисления, получим сомкнутый ряд динамики реализации продукции фирмы (в процентах к 2006 г.):

Таблица 1.7

Динамика реализации продукции ООО «БалтТорг»

Год	2003	2004	2005	2006	2007	2008	2009	2010
Относительный уровень в % к 2006 г.	80	82,2	86,7	100	102,9	105,9	112,9	117,6

В моментных динамических рядах может возникнуть несопос-

тавимость по моменту регистрации для явлений с сезонным характером. Так, например, численность скота летом обычно больше, чем зимой. Следовательно, недопустимо строить динамический ряд, в котором уровни зафиксированы на моменты времени, например на 01.01 и на 01.07 одного года.

Несопоставимость из-за различия единиц измерения сама по себе очевидна. Многие явления могут учитываться в разных единицах, что естественно может привести к проявлению различной динамики. Например, производство паровых котлов в отечественной статистике традиционно измеряется в штуках и тысячах  $m^2$  поверхности нагрева; производство турбин – в штуках и тысячах киловатт мощности и т. д. Естественно, что большая полнота анализа в подобных случаях может быть достигнута совместным анализом параллельных динамических рядов в различных единицах измерения.

Несопоставимость единиц измерения может возникнуть при использовании *ценовых единиц измерения*. В этом случае следует иметь в виду, что, во-первых, цены со временем изменяются, а, во-вторых, существует несколько видов цен (например, цены производителей и цены потребителей).

Несопоставимость статистических данных может возникнуть также *из-за различного толкования понятия единицы совокупности*, характеризваемой рядом динамики, поскольку к определению единиц можно подойти по-разному и с течением времени подход может измениться.

Несопоставимость статистических показателей динамики может быть обусловлена и *различной структурой совокупности* за разные годы. Например, показатели естественного движения населения (рождаемость, смертность) в значительной мере связаны с возрастной структурой населения, которая различна в разные годы. Для приведения данных к сопоставимому виду в этом случае используют так называемую *стандартизацию структуры*. В качестве стандартной структуры может приниматься, например, структура одного из периодов времени, а все показатели других периодов рассчитываются путем приведения их к стандартной структуре. Показатели стандартной структуры становятся вполне сравнимыми.

## 2. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ В СРЕДЕ STATISTICA

### 2.1. Ввод исходных данных

Решение любой задачи статистического анализа с использованием пакетов прикладных программ начинается с процедуры ввода данных. В предыдущих изданиях серии учебных пособий [2,3] подробно описаны процедуры создания рабочих книг и листов, их открытия, сохранения, переименования, а также редактирования данных, строк, столбцов и ячеек в ППП STATISTICA. Поэтому в данном пособии стандартные приемы, связанные с этими процедурами, будут опущены.

В качестве исходных данных в учебном пособии будут рассмотрены динамические ряды, представленные в табл. 2.1.

Таблица 2.1

Динамика среднедушевых денежных доходов населения России и выездов за границу россиян с целью туризма в 1993-2010 г.г.

Год	Среднедушевые денежные доходы населения РФ, руб.	Число граждан, выезжающих за границу с целью туризма, тыс.чел.	Год	Среднедушевые денежные доходы населения РФ, руб.	Число граждан, выезжающих за границу с целью туризма, тыс.чел.
1993	45,27	1600	2002	3947,2	5051,31
1994	206,6	2100	2003	5170,4	5678,5
1995	515,9	2607	2004	6410,3	6557,1
1996	769,5	3508	2005	8111,9	6784,7
1997	940,6	4143	2006	10196	7752,8
1998	1010,2	3330	2007	12602,7	9369
1999	1658,9	2809	2008	14940,6	11313,7
2000	2281,1	4485	2009	16838,3	9555,2
2001	2062	4191	2010	18552,6	12605

Для ввода этих данных в файл, после запуска STATISTICA вы-

бираем процедуру *File/New*<sup>2</sup>. Далее на закладке *Spreadsheet* в поле *Placement* ставим метку *In a new Workbook*, в поле *Number of variables* заносим число переменных (динамических рядов, в нашем случае – 2), в поле *Number of cases* заносим число моментов (интервалов) времени в рассматриваемых динамических рядах (в нашем случае число лет – 18) (рис. 2.1).

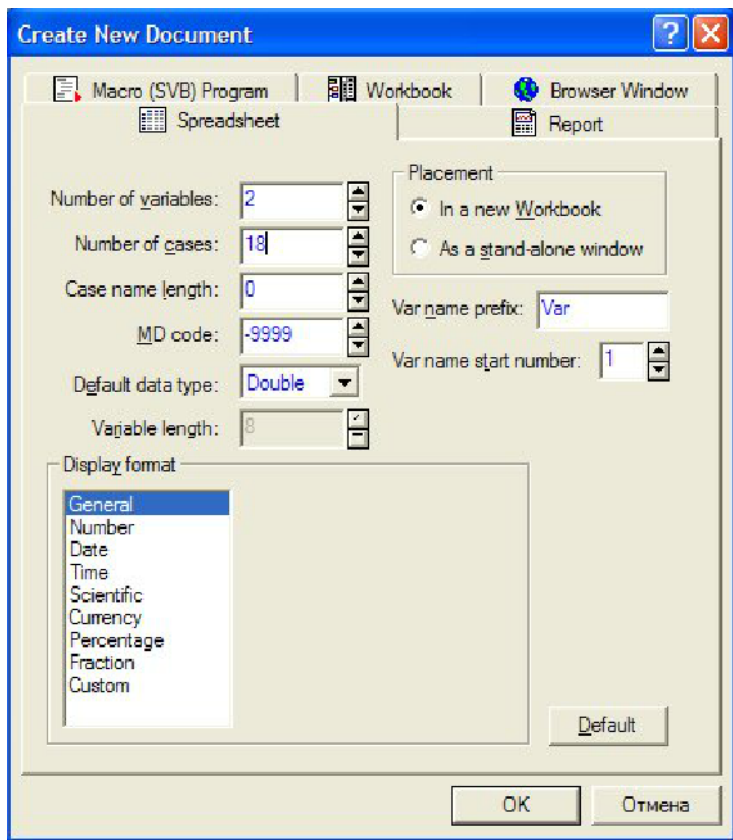


Рис. 2.1. Окно создания рабочей книги

<sup>2</sup> Здесь и далее подчеркнутый курсив – это имена кнопок и опций в интерфейсе программы.

После создания рабочей книги, сохраняем ее под уникальным названием в соответствующую папку.

**ВНИМАНИЕ!!!** Не желательно давать файлам стандартные имена типа *Workbook1*, поскольку это затруднит поиск файлов при их потере. Рекомендуется назвать файл своей фамилией латинскими буквами и сохранить в указанную преподавателем папку, а также создать резервную копию на съемных носителях.

Далее рекомендуется переименовать столбцы, соответствующие определенным явлениям и скопировать (ввести вручную) в них исходные данные (рис. 2.2).

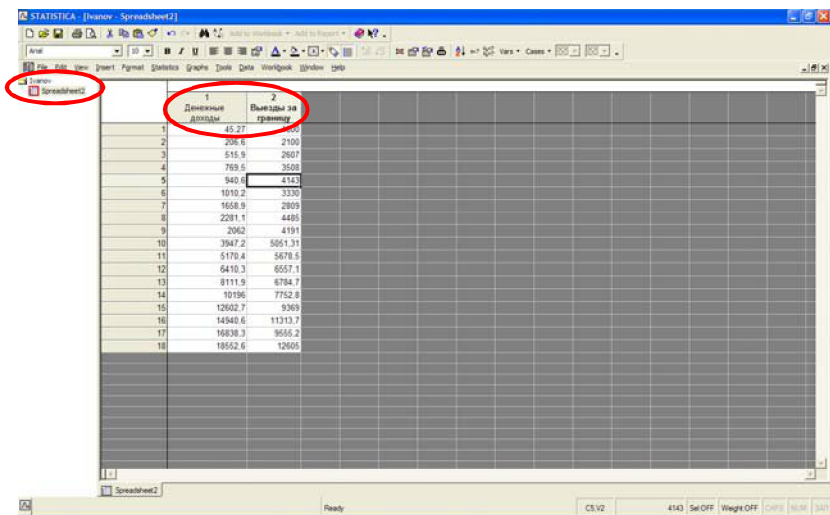


Рис. 2.2. Вид рабочего листа с исходными данными

Основным отличием динамических рядов от других типов статистических данных является наличие временной оси, которую необходимо также добавить на созданный рабочий лист вместо номеров ячеек. Сделать это можно двумя способами:

1) два раза щелкнуть левой кнопкой мыши по номеру ячейки и вручную ввести данные, далее, нажав кнопку *enter*, перейти в следующей ячейке и т.д.;

2) скопировать данные из созданного столбца в *Microsoft Excel*



[illegible]

1

2

Дополнение данных

Получение из таблицы

Год	1	2
1993	41,37	1600
1994	294,5	2120
1995	515,9	2500
1996	189,5	3020
1997	348,5	4140
1998	1010,2	3130
1999	1008,3	2920
2000	2281,1	4480
2001	2062	4181
2002	3947,2	6061,31
2003	5175,4	5478,9
2004	6410,3	6657,1
2005	8111,5	8784,7
2006	10186	11124,8
2007	12602,7	9369
2008	14268,6	11151,7
2009	16639,3	9564,7
2010	18012,6	9265

13

ни можно осуществлять между рабочими листами STATISTICA, выбирая только нужное число строк.

**ВНИМАНИЕ!!!** Ось времени должна присутствовать во всех таблицах и графиках, касающихся анализа временных рядов.

## 2.2. Графическое представление временных рядов

Перед началом анализа необходимо проверить настройки системы, позволяющее сделать работу более удобной. Выбираем меню Tools/Options (рис. 2.5). В появившемся окне выбираем закладку Output Manager. Необходимо проверить, чтобы закладки стояли именно таким образом, как на рис. 2.6, особенно важно наличие метки в поле Workbook containing the datafile, иначе результаты расчетов будут сохраняться как отдельные окна и рабочие листы могут быть утеряны.

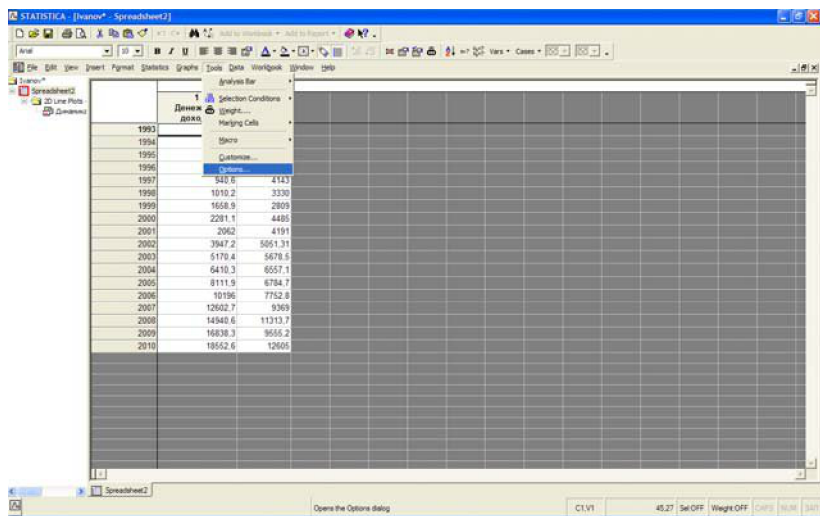


Рис. 2.5. Выбор процедуры Tools/Options

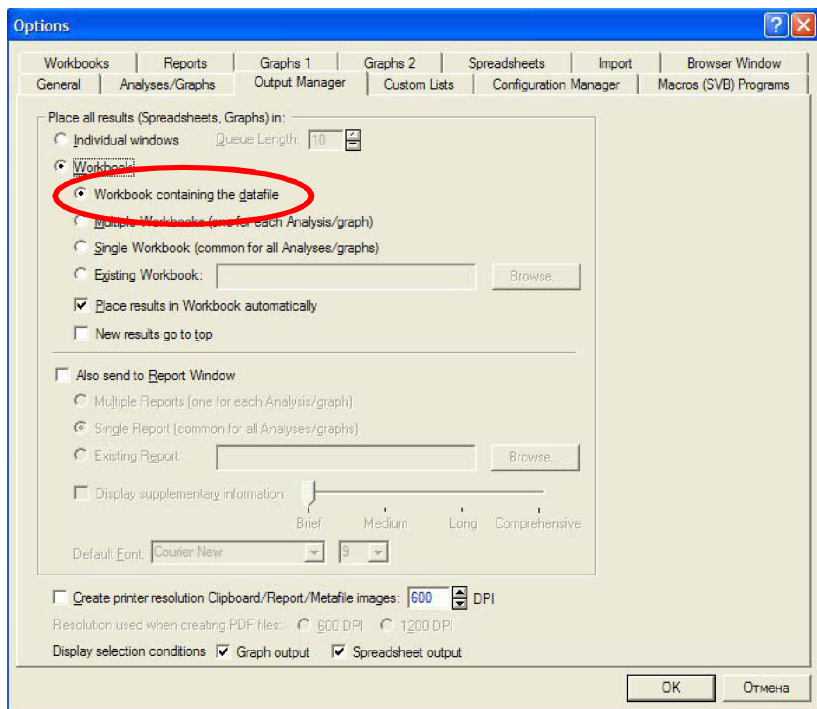


Рис. 2.6. Вид закладки *Output Manager* процедуры *Tools/Options* с настроенными полями

Графическое изображение временных рядов позволяет наглядно представить основные закономерности развития изучаемого явления или процесса.

В системе STATISTICA для графического представления рядов динамики используется меню *Graphs/2D Graphs/Line plots (Variables)* (рис. 2.7).

В появившемся окне кнопка *Variables* (рис. 2.8) позволяет выбрать необходимые переменные для построения графического изображения (рис. 2.9). Во всех разделах пособия, кроме раздела 6, для примера все операции будут осуществляться с переменной «Выезды за границу», операции с другой переменной выполняются аналогично.

В поле *Graph type* необходимо выбрать пункт *Regular*, так как график будет строиться для одной переменной. Пункт *Multiple* позво-

ляет на одном графике отобразить две и более переменные. Если при исследовании нескольких переменных выбрать пункт Regular, система построит по одному отдельному графику для каждой переменной.

Метка в поле Ignore missing data ставится в том случае, если в исходном динамическом ряду существуют пустые ячейки, и мы не хотим их учитывать.

Метка в поле Ignore out of frame points означает, что на графике не будут отображаться точки, которые не воспроизвести из-за выбранного масштаба осей. Ее необходимо убрать (рис. 2.10)

На закладке Options 1 в поле Custom title можно задать название графика (рис. 2.11).

Далее щелкаем мышкой по кнопке OK (рис. 2.12).

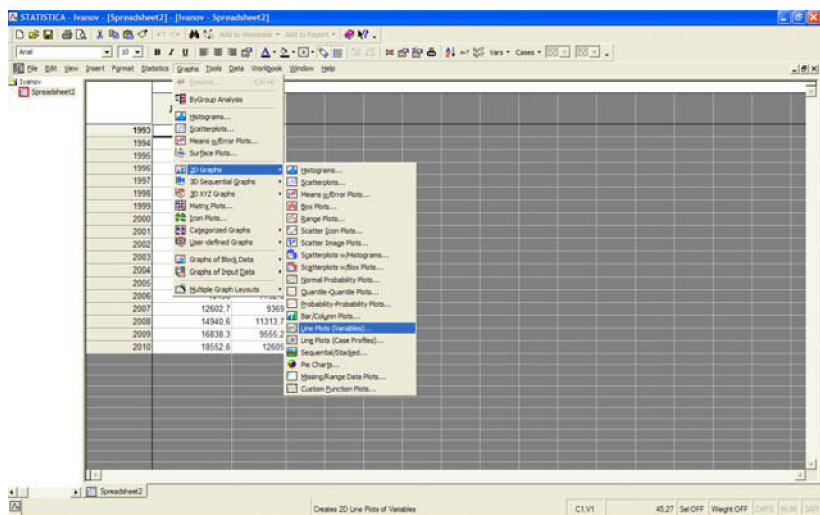


Рис. 2.7. Выбор процедуры Graphs/2D Graphs/Line plots (Variables)

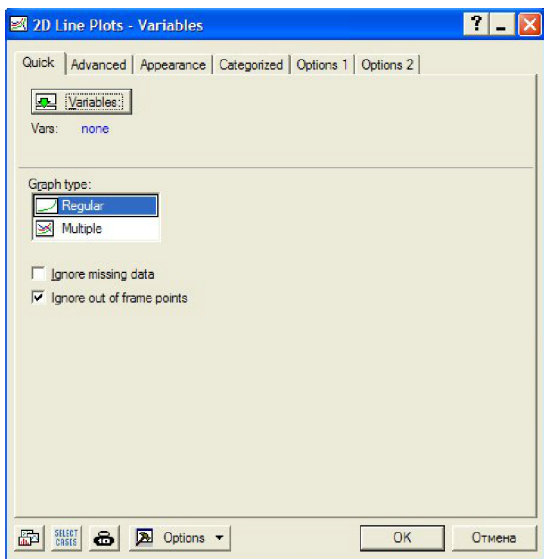


Рис. 2.8. Вид закладки *Quick* процедуры *Graphs/2D Graphs/Line plots (Variables)*

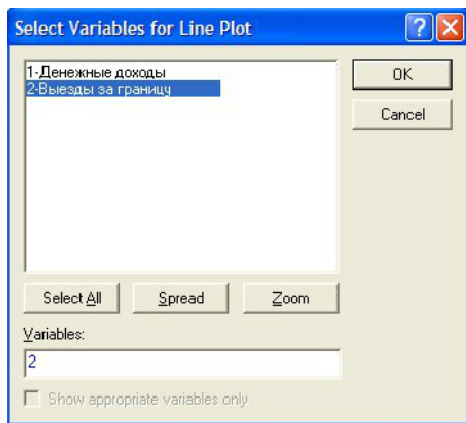


Рис. 2.9. Окно выбора переменной

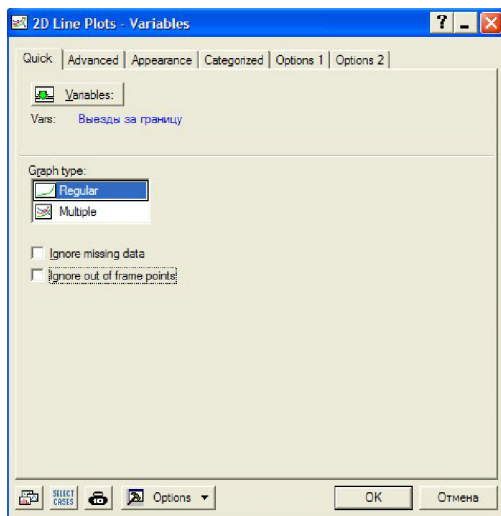


Рис. 2.10. Вид закладки Quick процедуры *Graphs/2D Graphs/Line plots (Variables)* с выбранными настройками

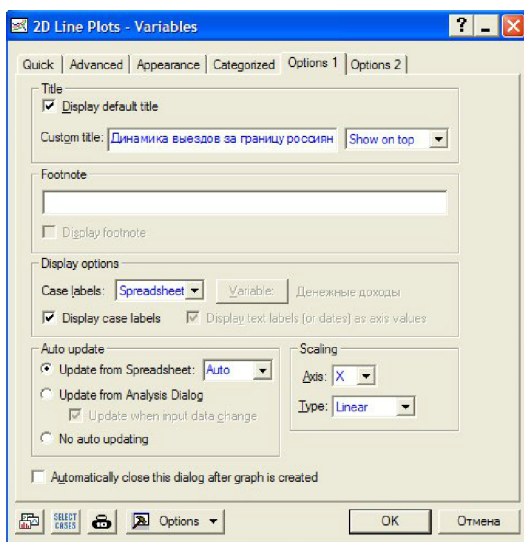


Рис. 2.11. Вид закладки Options 1 процедуры *Graphs/2D Graphs/Line plots (Variables)*



Рис. 2.12. Графическое представление переменной «Выезды за границу»

STATISTICA предоставляет широкие возможности по настройке и редактированию различных типов графиков. Если щелкнуть правой кнопкой мыши в свободном поле графика и выбрать в появившемся подменю свойства графика *Graph Properties (All Options)*, то на экране появится диалоговое окно, состоящее из 17 закладок для настройки определенных функций графика (рис. 2.13., 2.14.).

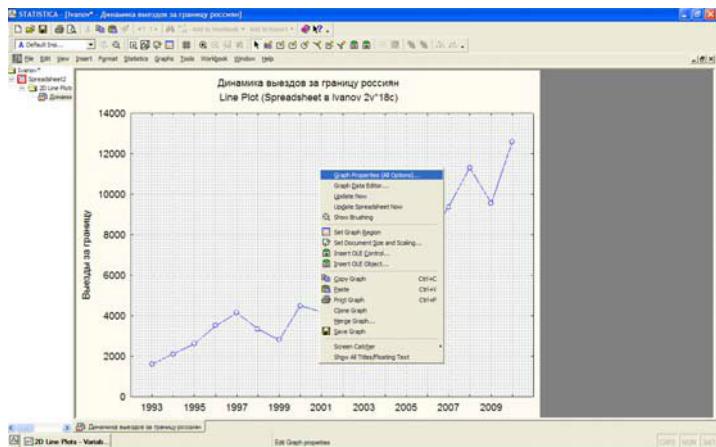


Рис. 2.13. Выбор свойств графического изображения

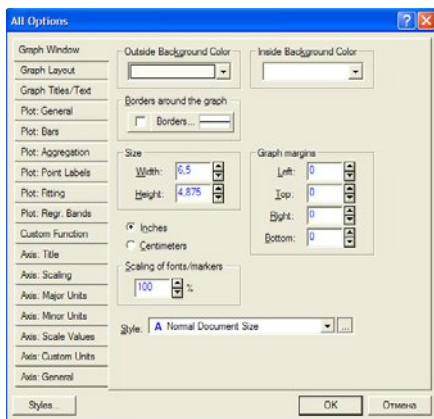


Рис. 2.14. Вид окна с основными функциями настройки графического изображения

Например, с помощью закладки *Graph Titles/Text* можно настроить характеристики титульной подписи графика. Для этого во всплывающем меню *Title* можно выбрать соответствующий заголовок и отредактировать или удалить его. В нашем случае выберем и удалим заголовок, содержащий характеристики рабочей книги (рис. 2.15).

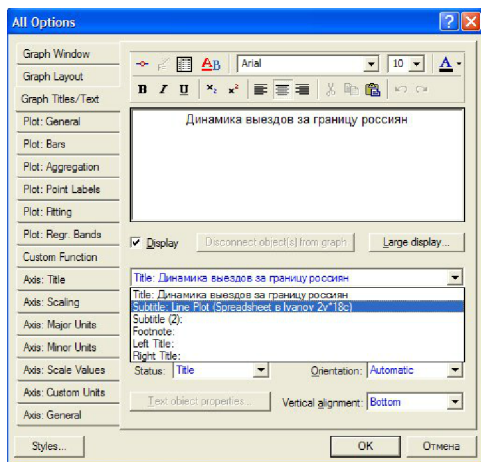


Рис. 2.15. Выбор редактируемого заголовка графика



С помощью закладки *Axis: Title* можно настроить подписи осей. Выбрав в поле *Axis* соответствующую ось (*Y left*) и редактируем название, вписывая единицы измерения (рис. 2.16).

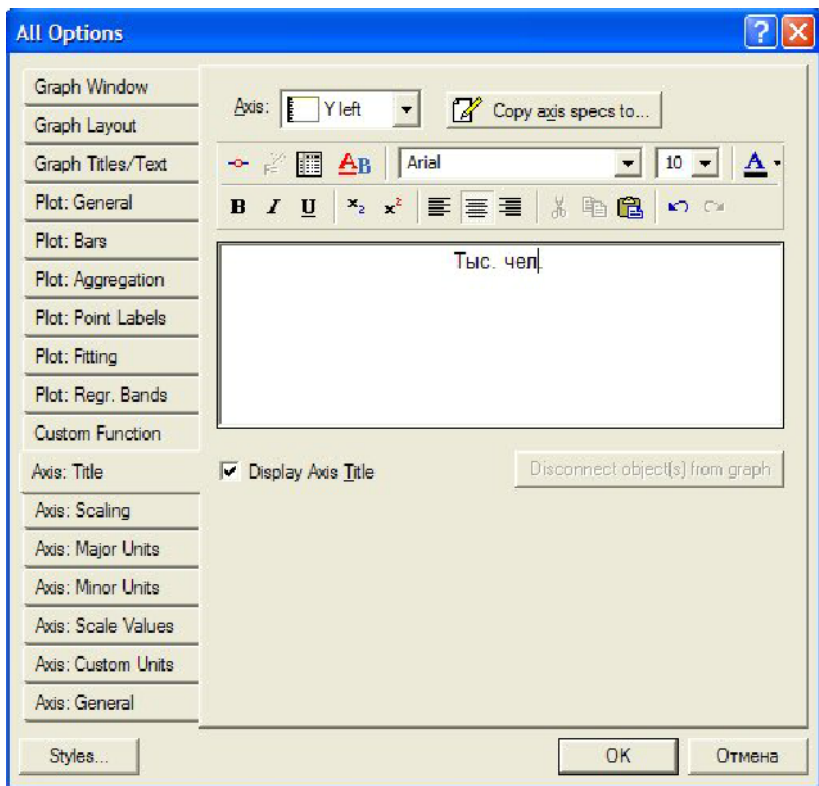


Рис. 2.16. Настройка подписей осей графика

Закладка *Axis: Scale Values* позволяет настроить шкалы осей. Выбрав в поле *Axis* ось *X* можно задать параметры отображения уровней ряда: в поле *Layout* выбираем *Perpendicular*, в поле *Skip values* выбираем *Off* (рис. 2.17). Отметим, что настройку параметров оси можно вызвать двойным щелчком левой кнопки мыши по ней.

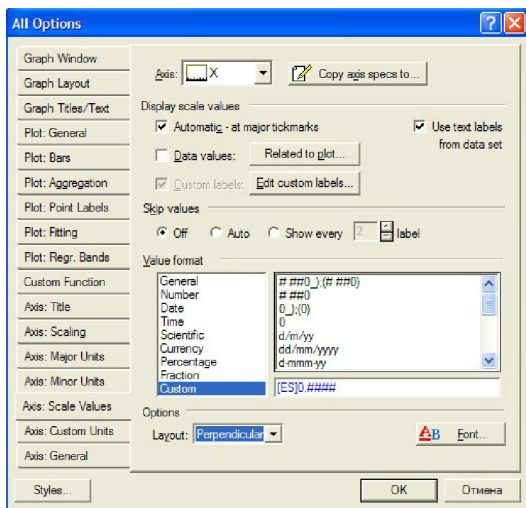


Рис. 2.17. Настройка шкал осей графика

В результате еще нескольких настроек график обретает следующий вид (рис. 2.18).



Рис. 2.18. Графическое представление переменной «Выезды за границу» после редактирования

Остальные настройки пользователь может освоить самостоятельно

тельно, они соответствуют настройкам основных *Windows*-приложений.

## 2.3. Показатели изменения уровней временного ряда

Анализ динамических рядов социально-экономических явлений обычно начинают с рассмотрения статистик, расчет которых не требует какой-либо предварительной обработки данных. Речь идет о так называемых показателях динамического ряда, позволяющих пояснить характер, скорость, интенсивность и направление развития изучаемого явления за определенный временной период.

В результате того или иного сопоставления уровней динамического ряда формируется система абсолютных и относительных показателей динамики, к числу которых относятся абсолютные приросты (и их среднее значение), темпы роста (и их среднее значение), темпы прироста, абсолютное значение одного процента прироста. Сравнимый уровень динамического ряда называется *текущим*, а уровень, с которым производится сравнение, *базисным*. В зависимости от того, что принимается за базу сравнения, будут получены различные показатели динамики. Приняв за базу сравнения некоторый постоянный уровень, например  $y_1$  получим серию *базисных показателей*, которые характеризуют окончательный результат всех изменений в уровнях ряда от первого периода (или момента времени) до текущего периода. Следует иметь в виду, что в реальных задачах за базу сравнения может быть принят уровень ряда, относящийся к периоду (моменту), выходящему за пределы анализируемого динамического ряда (например, начальный момент периода, с которого начинается некоторый новый этап развития). Если производится сравнение текущих уровней  $y_t$  с непосредственно предшествующими  $y_{t-1}$ , то получают *цепные показатели динамики*.

*Абсолютным приростом* называется разность между значением уровня данного периода и предшествующего (либо базисного):

$$\Delta_t = y_t - y_{t-1}, \quad (2.1)$$

где  $y_t$  – уровень ряда динамики в момент времени  $t$ ;  $y_{t-1}$  – уровень ряда динамики в момент времени  $t-1$ ;  $\Delta_t$  – цепной абсолютный прирост.

За весь период, описываемый временным рядом, абсолютный

прирост  $\Delta$  выразится как алгебраическая сумма частных цепных приростов (2.1) или, что очевидно, как разность между последним и первым уровнями ряда:

$$\Delta = \sum_{t=2}^n \Delta_t = \sum_{t=2}^n (y_t - y_{t-1}) = y_n - y_1, \quad (2.2)$$

где  $y_n$  – последний уровень ряда;  $y_1$  – первый уровень.

Абсолютный прирост может быть как положительным, так и отрицательным. Он показывает, насколько уровень текущего периода выше или ниже предшествующего и выражает абсолютную скорость роста или снижения уровней ряда.

Коэффициент роста – это отношение последующего уровня к предыдущему или какому-либо другому, принятому за базу сравнения. Коэффициент роста оценивает, во сколько раз уровень текущего периода выше или ниже уровня предыдущего (базисного) периода. Коэффициент роста, выраженный в процентах, называется темпом роста: показывает сколько процентов каждый уровень составляет по отношению к уровню предыдущего (базисного) периода.

Темп роста вычисляется по формулам:

$$Tr_t = \frac{y_t}{y_{t-1}} \cdot 100, \quad (2.3)$$

где  $Tr_t$  – цепной темп роста.

$$Tr'_t = \frac{y_t}{y_{const}} \cdot 100, \quad (2.4)$$

где  $Tr'_t$  – базисный темп роста;  $y_{const}$  – база сравнения.

$$Tr = \frac{y_n}{y_1} \cdot 100, \quad (2.5)$$

где  $Tr$  – темп роста за весь период.

Величина темпа роста больше 100 показывает увеличение уровня текущего периода по сравнению с предыдущим (базисным). Величина темпа роста, равная 100, показывает, что уровень текущего пе-

риода по сравнению с предыдущим (базисным) не изменился. Темп роста всегда имеет положительный знак. Цепные темпы роста характеризуют интенсивность изменения уровней ряда.

Темп прироста – это отношение абсолютного прироста к базе сравнения. Этот показатель характеризует относительную скорость изменения уровня ряда в единицу времени:

$$Tp_t = \frac{\Delta_t}{y_{t-1}} \cdot 100, \quad (2.6)$$

где  $Tp_t$  – темп прироста.

Темп прироста за весь период:

$$Tp = \frac{y_n - y_1}{y_1} \cdot 100 = \frac{y_n}{y_1} \cdot 100 - 100, \quad (2.7)$$

Темп прироста, выраженный в процентах, показывает, на сколько процентов увеличился или уменьшился текущий уровень по сравнению с предыдущим (базисным), принятым за 100%, или, иначе, сколько процентов составляет абсолютный прирост данного уровня по отношению к предыдущему (базисному) уровню.

Темп прироста и темп роста можно связать следующим соотношением:  $Tp_t = Tr_t - 100$ . При темпах роста, меньше 100% или единицы (уменьшение уровней ряда), получаем отрицательные темпы прироста, т.е. темпы снижения.

Выбор показателей изменения уровней динамических рядов и их формы зависит от характеристик используемых исходных данных: разрядности, величины изменений, наличия отрицательных значений и пр. Рассмотрим алгоритм расчета вышеописанных показателей в системе STATISTICA. Для этого потребуется произвести некоторые дополнительные преобразования имеющихся переменных.

В разделе главного меню Statistics выбираем подменю Advanced Linear/Nonlinear Models и в нем команду Time Series/Forecasting (рис. 2.19).

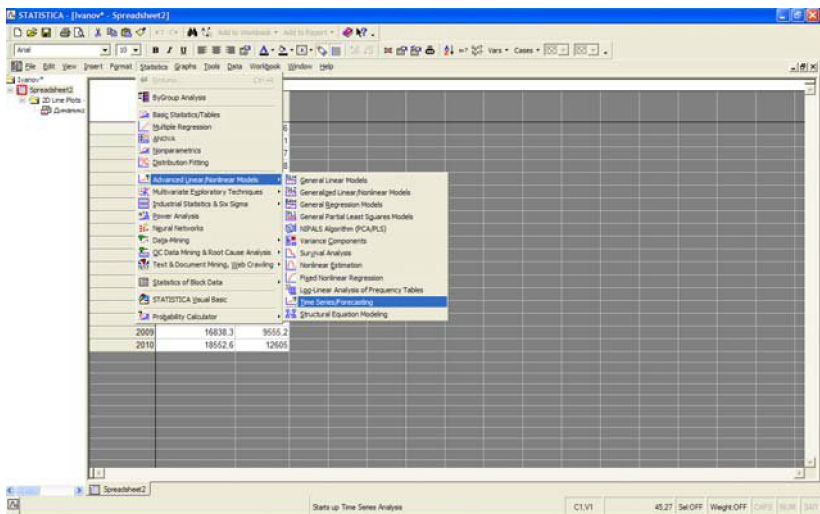


Рис. 2.19. Выбор процедуры *Time Series/Forecasting*

Открывается модуль, в котором реализованы практически все современные методы анализа и прогнозирования временных рядов, и который будет постоянно использоваться при выполнении курсового проекта (рис. 2.20). С помощью кнопки *Variables* выбираем необходимую переменную и получаем следующий вид окна (рис. 2.21).

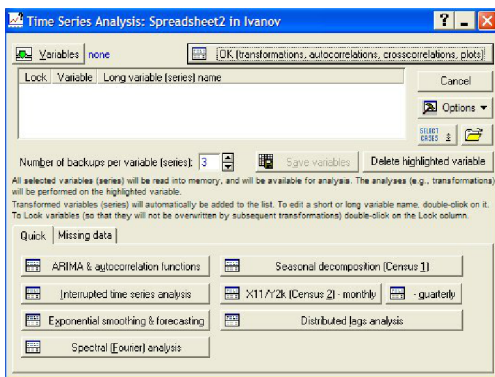


Рис. 2.20. Окно процедуры *Time Series/Forecasting*

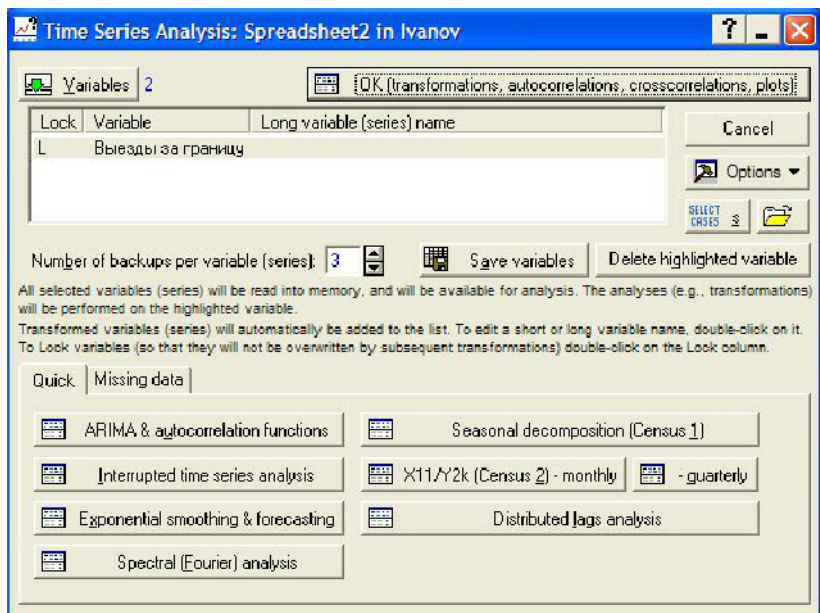


Рис. 2.21. Окно процедуры *Time Series/Forecasting* с выбранной переменной «Выезды за границу»

Рассмотрим рабочую область данного модуля:

1. В верхней части панели, в информационном поле записываются имена анализируемых и преобразованных переменных. Рядом с информационным полем расположены стандартные кнопки системы (*Cancel*, *Options*). Сверху от информационного поля уже известная нам кнопка *Variables* для выбора переменных и кнопка *OK (transformations, autocorrelations, crosscorrelations, plots)*.

2. Нижняя часть панели (под пояснительным текстом) – функциональная. На закладке *Quick* расположены кнопки, открывающие специальные диалоговые окна анализа, а закладка *Missing Data* содержит опции обработки пустых ячеек.

После выбора переменной для анализа в информационном поле появляются ее имя, а слева значок «*L*» в графе *Lock*, который означает, что выбранные переменные являются закрытыми на «ключ» и не могут быть удалены без прерывания анализа. Дальнейшая работа происходит именно с этими переменными, которые можно преобразовывать,

анализировать, трансформировать, но нельзя удалять из текущего анализа.

В процессе работы динамические ряды подвергаются многократному преобразованию. Однако, не все результаты этих преобразований необходимы в работе. Кнопка Delete highlighted variable позволяет удалять из диалогового окна лишнюю информацию (выделенную переменную). Напротив, некоторые переменные используются для дальнейшего анализа и их необходимо сохранять (кнопка Save variables).

Для расчета цепных показателей изменения уровней динамического ряда необходимо иметь две переменные: первая переменная – исходный динамический ряд, состоящий из уровней сравниваемых периодов  $y_t$ , а вторая переменная – динамический ряд, состоящий из уровней непосредственно предшествующих периодов  $y_{t-1}$ , т.е. база сравнения. Очевидно, что одна переменная у нас имеется, для получения второй переменной нажмем на кнопку OK (transformations, autocorrelations, crosscorrelations, plots), открывающую диалоговое окно преобразования переменных Transformations of Variables. Выбираем закладку Shift и ставим метку в поле Shift (Lag) Series Forward со значением лага, равному 1, так как мы хотим получить временной ряд, смещенный относительно исходного на один временной период (лаг) (рис. 2.22).

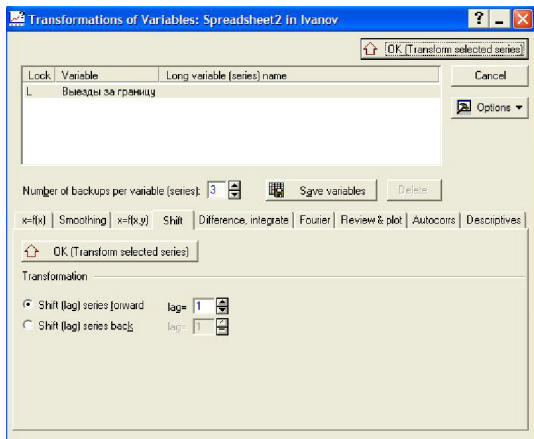


Рис. 2.22. Вид закладки Shift процедуры Transformations of Variables. Далее нажимаем кнопку OK (Transform Selected Series).



Система автоматически строит график полученного динамического ряда, который целесообразно удалить (рис. 2.23), как неиспользуемый в дальнейшем анализе.

Следует напомнить, что окно процедуры после ее запуска сворачивается и находится в виде кнопки в левом нижнем углу экрана. Для возвращения необходимо щелкнуть по ней левой кнопкой мыши.

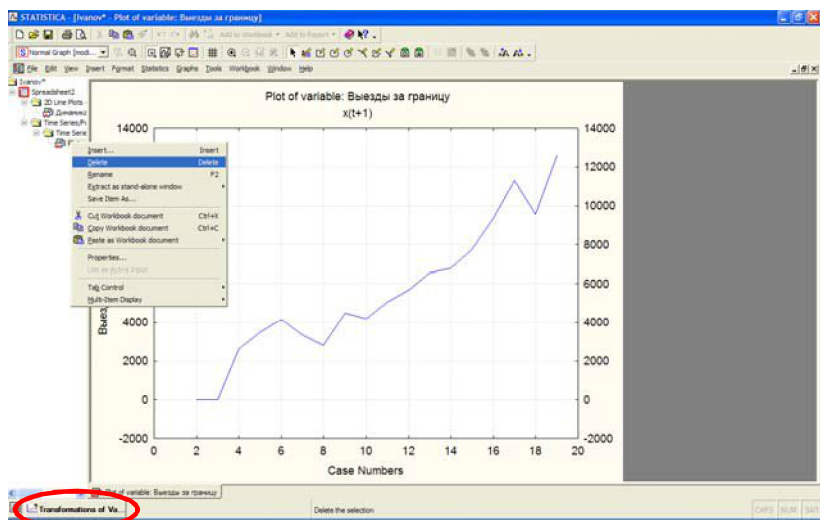


Рис. 2.23. Удаление элемента рабочей книги

В информационном поле диалогового окна в графе *Variables* появляется преобразованная переменная «Выезды за границу ( $t+1$ )» (рис. 2.24).

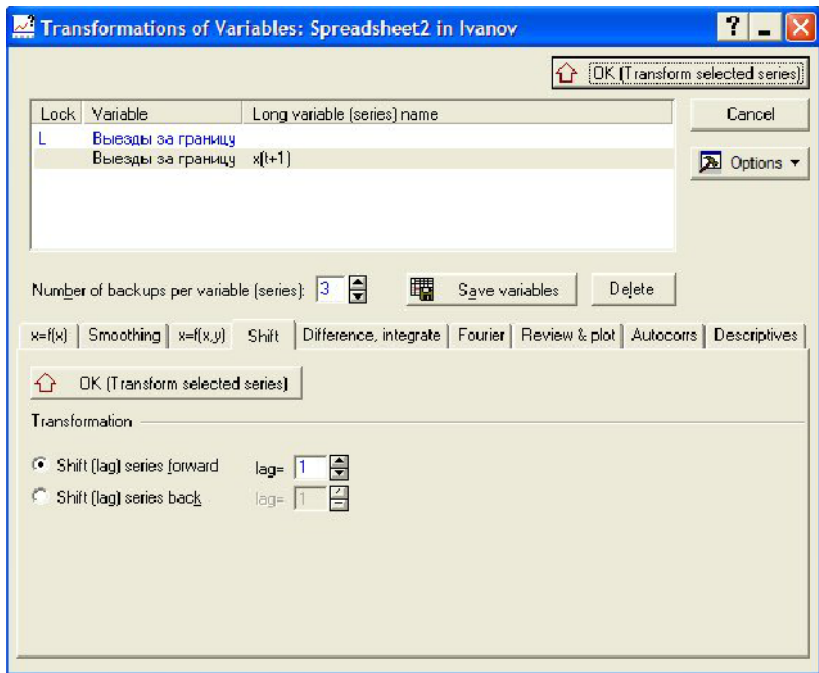


Рис. 2.24. Окно процедуры *Transformations of Variables* с исходной и смещенной переменными

После этого нажимаем на кнопку *Save variables* для сохранения результатов преобразования. Результаты сохраняются в отдельной электронной таблице в виде двух переменных. В нашем примере это «Выезды за границу» – исходный динамический ряд и «Выезды за границу\_1» – ряд, отстающий от исходного на один лаг (рис. 2.25.)

Во избежание потери данных предлагается поместить данную таблицу в исходную рабочую книгу путем нажатия на кнопку *Add to Workbook* и выбором соответствующей рабочей книги (Ivanov) (рис. 2.26.).

Отметим, что новый рабочий лист автоматически становится активным (рис. 2.27.).

Spreadsheet2 in Ivanov		
	1 Выезды за границу	2 Выезды за границу_1
1	1600,000	
2	2100,000	1600,000
3	2607,000	2100,000
4	3508,000	2607,000
5	4143,000	3508,000
6	3330,000	4143,000
7	2809,000	3330,000
8	4485,000	2809,000
9	4191,000	4485,000
10	5051,310	4191,000
11	5678,500	5051,310
12	6557,100	5678,500
13	6784,700	6557,100
14	7752,800	6784,700
15	9369,000	7752,800
16	11313,700	9369,000
17	9555,200	11313,700
18	12605,000	9555,200
19		12605,000

Рис. 2.25. Рабочая область с исходной переменной и смещенной переменной

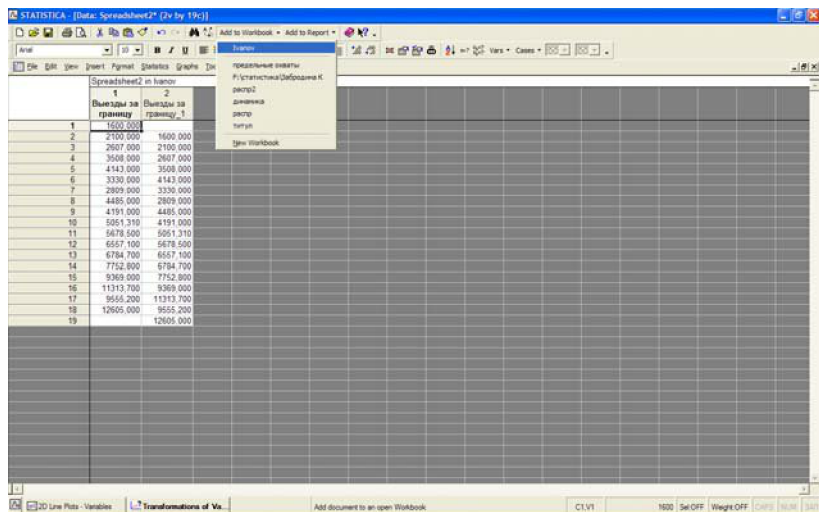


Рис. 2.26. Добавление рабочего листа в рабочую книгу

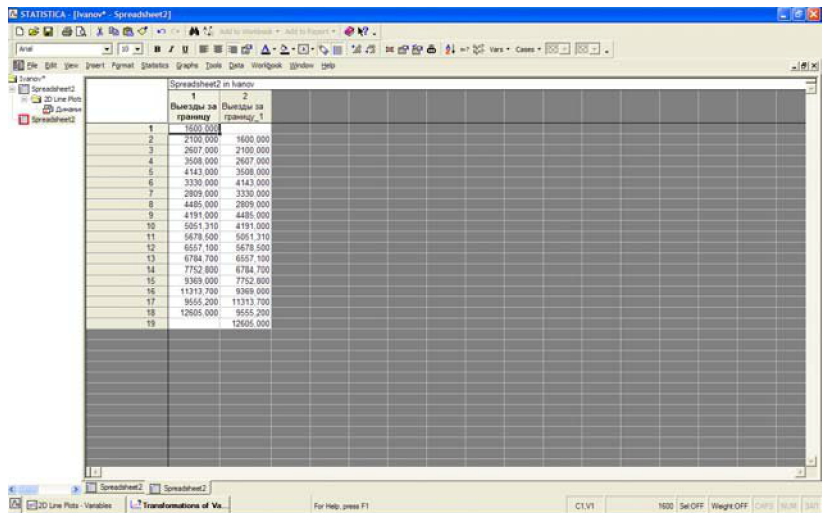


Рис. 2.27. Рабочая область с новым активным рабочим листом

Напомним, что на каждом рабочем листе должна присутствовать ось времени, поэтому копируем ее с исходного рабочего листа (рис. 2.28) и вставляем на активный лист (рис. 2.29). Отметим также, что последнюю строку можно удалить, так как она не содержит данных для расчета показателей. Для этого щелкаем правой кнопкой мыши по любому номеру ячейки и выбираем процедуру Case Management/Delete Cases (рис. 2.30), затем указываем номера строк, которые необходимо удалить: от (*From Case*) и до (*To Case*) и нажать кнопку OK (рис. 2.31).

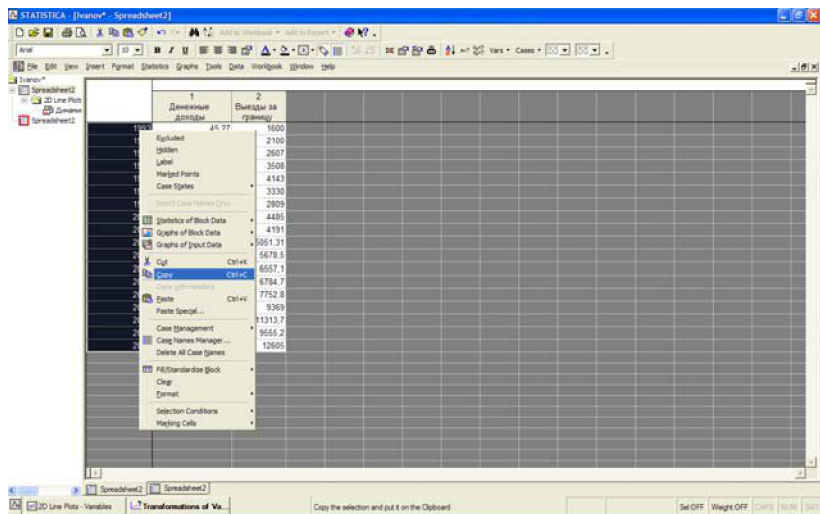


Рис. 2.28. Копирование оси времени

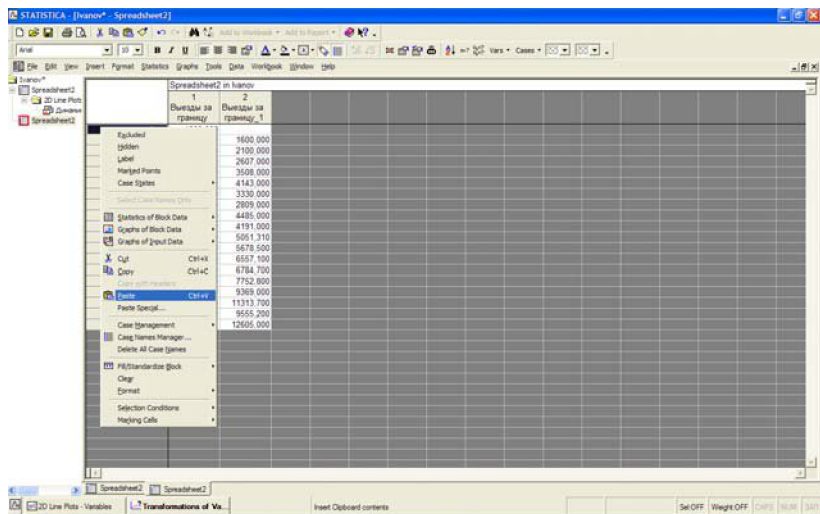


Рис. 2.29. Вставка оси времени

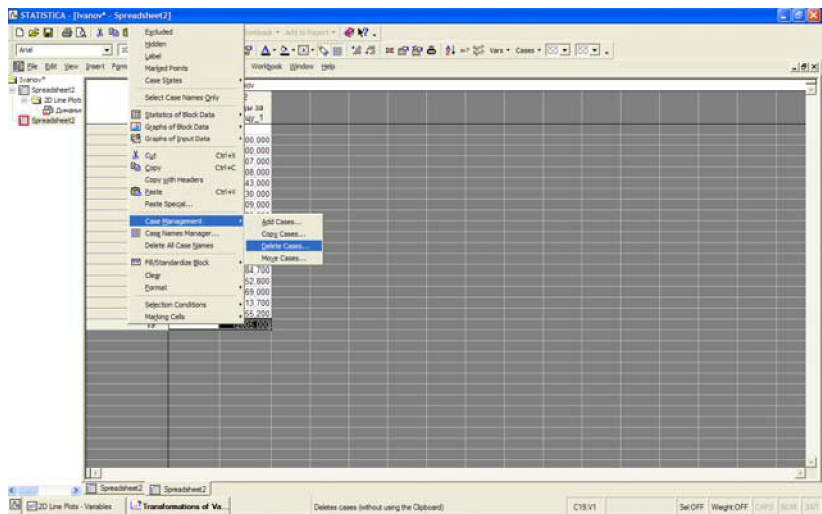


Рис. 2.30. Выбор функции удаления строки



Рис. 2.31. Выбор строк для удаления

После того как получены необходимые для дальнейших вычислений переменные, можно приступить непосредственно к расчету показателей динамики. В курсовом проекте предлагается рассчитать следующие показатели изменения уровней динамического ряда:

- абсолютные приросты цепные (APRc) и базисные (APRb);
- темпы роста цепные (TRc) и базисные (TRb);
- темпы прироста цепные (TPRc) и базисные (TPRb).

Отметим, что существуют и другие показатели для оценки динамических рядов, однако представленные наиболее часто используются в практическом анализе. При этом графическое представление

цепных (базисных) показателей, как абсолютных приростов, так и темпов роста, одинаково демонстрирует тенденцию в поведении уровней ряда.

Для расчета показателей добавим на рабочий лист шесть новых переменных, соответствующих числу цепных и базисных показателей.

Дважды щелкаем левой кнопкой мыши по серому полю справа от заголовка переменной. Появляется меню (рис. 2.32), в котором предлагается выбрать число добавляемых переменных (*Add variables*) и наблюдений (*Add cases*). В нашем случае это 6 переменных и ноль наблюдений. Нажимаем OK и попадаем в меню настройки параметров переменной (рис. 2.33).

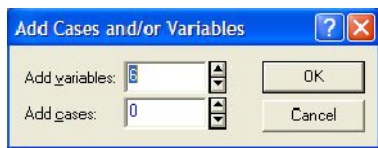


Рис. 2.32. Процедура добавления переменных

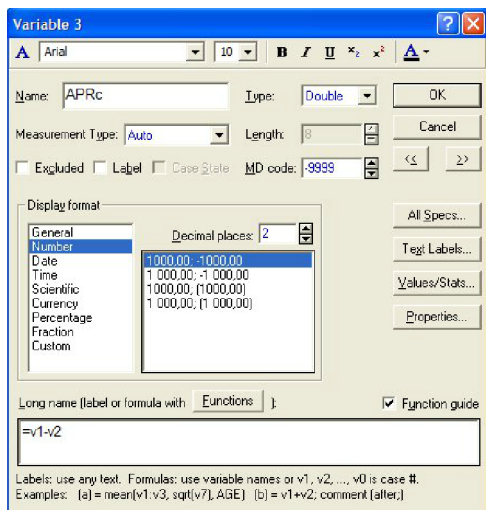


Рис. 2.33. Настройка параметров переменной

В качестве примера рассмотрим процедуру расчета цепных аб-

солютных приростов. В поле Name вводим имя переменной в любом формате (APRc), далее выбираем формат переменной – нас интересует числовой (значит, в поле *Display Format* выбираем *Number*). Важно обратить внимание, чтобы в поле *Decimal places* не значился ноль (что будет означать округление до целого); число, фиксируемое в данном поле, означает число знаков после запятой (нам достаточно 2). Если выбрать формат данных *Percentage*, то показатели будут рассчитаны не в коэффициентах, а в процентах

Система позволяет вводить в электронные таблицы новые данные, на основе имеющихся, путем вычислений с использованием формул, которые задаются в поле *Long Name (Label Or Formula With Functions)*.

Все математические операторы воспринимаются таким же образом, как и в других Windows-приложениях. Пробелов между элементами уравнения ставить не надо. В формулу можно вводить и численные значения и функции, полный набор которых вызывается кнопкой *Function* или автоматически раскрывается при вводе первой буквы соответствующей функции.

Условием получения новых данных является постановка знака равенства перед любой формулой.

Переменные в системе вводятся в следующем формате: любая переменная обозначается буквой «v» с соответствующим номером, обозначающим порядковый номер столбца, в котором эта переменная содержится. При этом регистр буквы не важен.

Записываем формулу для расчета цепных абсолютных приростов «=v1-v2», где «v1» – исходный динамический ряд, а «v2» – динамический ряд, отстающий от исходного на один год.

Далее нажимаем кнопку  $\geq$ , находящуюся справа вверху под кнопкой *Cancel* и позволяющую переместиться к настройке следующей переменной. При появлении следующего окна выбираем «Да». Это означает подтверждение расчета новой переменной и согласие с тем, что в случае невозможности расчета каких-либо ячеек, они будут оставлены пустыми (рис. 2.34). После того, как все переменные настроены, нажимаем *OK*. Чтобы попасть в меню редактирования любой переменной достаточно дважды щелкнуть левой кнопкой мыши по заголовку столбца.



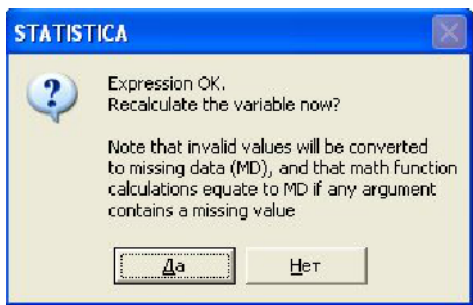


Рис. 2.34. Окно подтверждения пересчета новой переменной

Остальные показатели рассчитываются аналогичным образом, по соответствующим формулам. При расчете базисных показателей в поле формул вносится значение уровня, выбранного за базу сравнения (в нашем случае первого) (рис. 2.35).

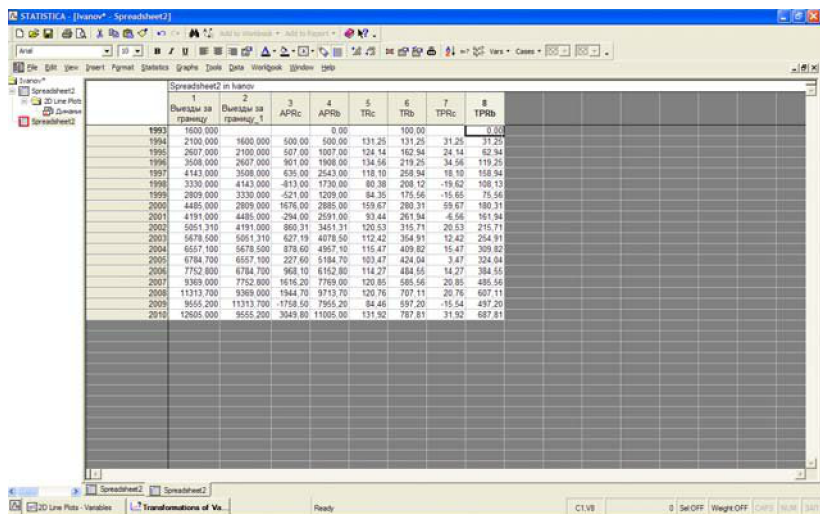


Рис. 2.35. Таблица с рассчитанными показателями изменения уровней динамического ряда

Все рассчитанные показатели могут быть представлены графически. Для этого воспользуемся процедурой построения графиков,

рассмотренной ранее (раздел 2.2). Целесообразно строить отдельное графическое изображение для каждого показателя (тип *Regular*), в качестве переменных выбирая сразу все показатели (рис. 2.36, 2.37).

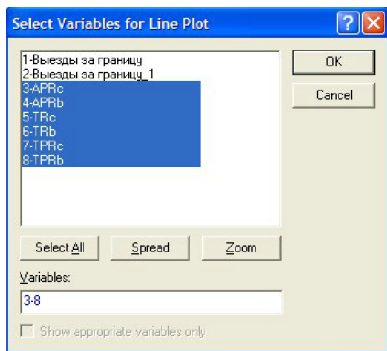


Рис. 2.36. Выбор нескольких переменных для построения графиков

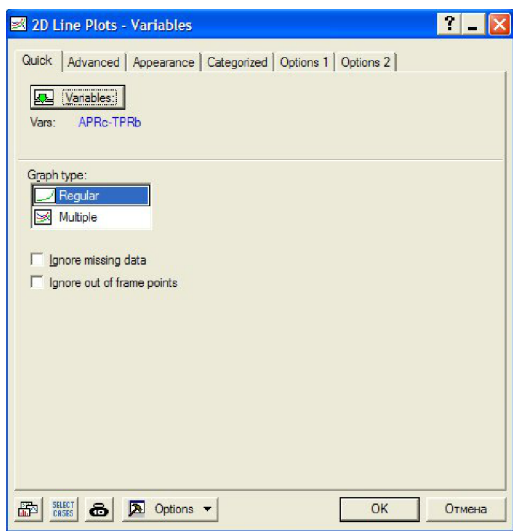


Рис. 2.37. Выбор типа графика

Приведем графические изображения цепных и базисных темпов роста (рис. 2.38, 2.39).

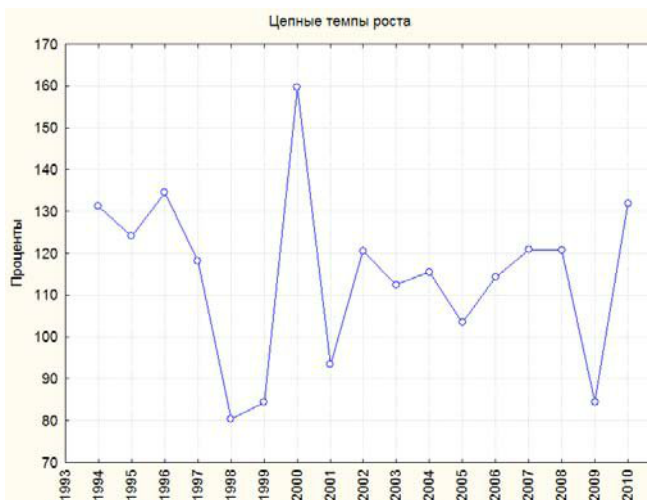


Рис. 2.38. Цепные темпы роста переменной «Выезды за границу»

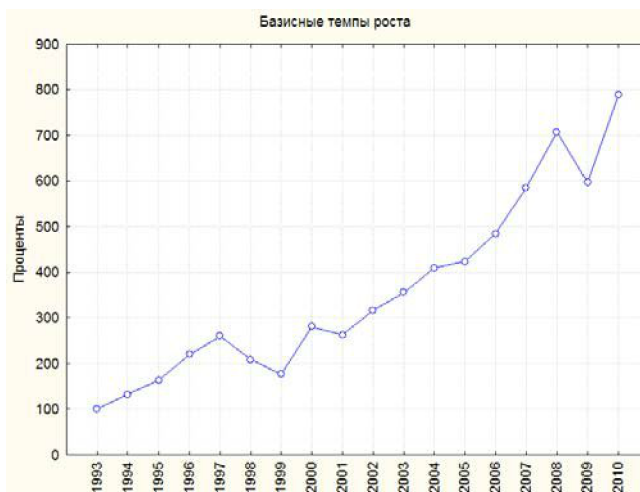


Рис. 2.39. Базисные темпы роста переменной «Выезды за границу»

Рассчитанные абсолютные и относительные показатели динамики варьируют, изменяются во времени. Это обстоятельство вызыва-

ет необходимость расчета обобщающих характеристик, которыми являются средние показатели.

## 2.4. Средние показатели динамики

Как уже отмечалось, средние показатели необходимы для получения обобщающих оценок изменения уровней временного ряда. Часто использование средних показателей становится просто необходимым. Например, сельскохозяйственное производство в огромной степени зависит от погодных условий конкретного года, и сравнение годовых показателей становится нецелесообразным. Правильнее сравнивать среднегодовые уровни, среднегодовые абсолютные приросты и темпы роста, рассчитанные за несколько лет. При сравнительном анализе изменения тех или иных показателей по разным странам, регионам или, например, при сопоставлении темпов роста заработной платы и производительности труда также целесообразно использовать средние показатели рядов динамики.

Анализируя временные ряды и рассчитанные показатели, можно определить средний уровень ряда, средний абсолютный прирост и средний темп роста (средний темп прироста определяется на основании темпа роста).

Средний уровень ряда рассчитывается по-разному для моментных и интервальных рядов динамики. Средний уровень интервального ряда вычисляется по формуле средней арифметической простой:

$$\bar{y} = \frac{y_1 + y_2 + y_3 + \dots + y_n}{n} = \frac{\sum_{i=1}^n y_i}{n}, \quad (2.8)$$

где  $\bar{y}$  – средний уровень интервального ряда;  $n$  – общее число уровней ряда.

Если отдельные периоды интервального ряда динамики имеют неодинаковую длину, то для определения среднего уровня следует воспользоваться средней арифметической взвешенной. Для неполных интервальных рядов иногда определяют полусумму уровней на начало и конец периода и принимают ее за характеристику среднего уровня всего периода. Такой средний уровень является грубой оценкой и

применяется редко.

Средний уровень моментного ряда определяется по формуле, получившей название средней хронологической:

$$\bar{y} = \frac{\frac{1}{2} y_1 + y_2 + y_3 + \dots + \frac{1}{2} y_n}{n-1}, \quad (2.9)$$

где  $\bar{y}$  – средний уровень моментного ряда;  $n$  – число уровней ряда;  $y_n$  – уровень последнего года (периода);  $y_1$  – уровень первого года (периода).

В знаменателе формулы – число уровней без единицы, поскольку в числителе первый и последний уровни берутся в половинном размере.

Обобщающим показателем скорости изменения явления во времени служит средний абсолютный прирост – среднее значение цепных абсолютных приростов за равные промежутки времени.

Если абсолютные приросты обозначить через  $\Delta_1, \Delta_2, \Delta_3, \dots$ , то средний абсолютный прирост, обозначаемый через  $\bar{\Delta}_y$ , может быть найден по формуле:

$$\bar{\Delta}_y = \frac{\Delta_1 + \Delta_2 + \dots + \Delta_n}{n-1} = \frac{\sum_{t=1}^{n-1} \Delta_t}{n-1}. \quad (2.10)$$

или

$$\bar{\Delta}_y = \frac{y_n - y_1}{n-1}. \quad (2.11)$$

При исчислении среднего темпа роста нужно учитывать, что интенсивность развития явлений идет по правилам сложных процентов, где накладывается прирост на прирост. Поэтому средний темп роста принято вычислять по формуле средней геометрической на основании цепных темпов роста.

Если через  $Tr_1, Tr_2, Tr_3, \dots, Tr_n$  обозначить цепные темпы роста

за равные промежутки, то средний темп роста выразится формулой:

$$\overline{Tr} = \sqrt[n]{Tr_1 \cdot Tr_2 \cdot Tr_3 \cdot \dots \cdot Tr_{n-1}} = \sqrt[n]{\prod_{t=1}^{n-1} Tr_t}, \quad (2.12)$$

где  $\overline{Tr}$  – средний темп роста.

Или

$$\overline{Tr} = \sqrt[n]{\frac{y_n}{y_1}}. \quad (2.13)$$

Для расчета средних темпов прироста пользуются уже известным соотношением:  $\overline{Tr}_{pr} = \overline{Tr} - 100$ .

Интерпретация всех выше описанных показателей обязательно должна сопровождаться указанием временного отрезка, за который рассчитана характеристика, а также единицы времени, к которой относятся уровни ряда, например: среднегодовой абсолютный прирост числа выезжающих за границу с целью туризма с 1993 по 2010 годы; среднемесячный темп роста объема продаж за последние 10 лет и т. п.

Рассмотрим алгоритм расчета описанных выше показателей в STATISTICA. В нашем примере рассчитаем показатели для всего динамического ряда.

В главном меню выбираем процедуру Statistics/Basic Statistics/Tables (рис. 2.40).

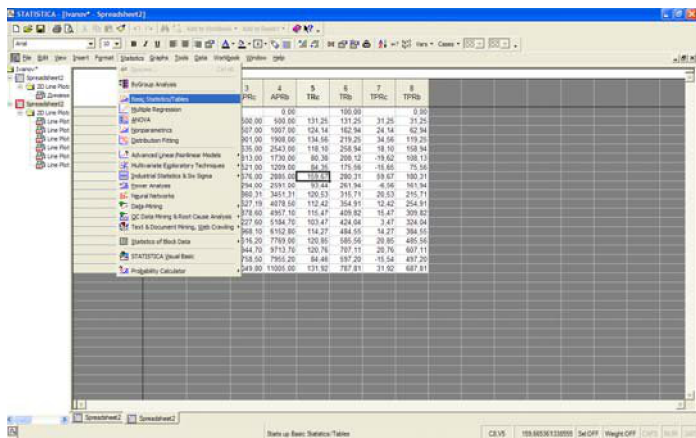


Рис. 2.40. Окно STATISTICA с выбранной функцией Basic Statistics/Tables

Затем, в появившемся контекстном окне, выбираем первый пункт Descriptive statistics – описательные статистики (рис. 2.41.).

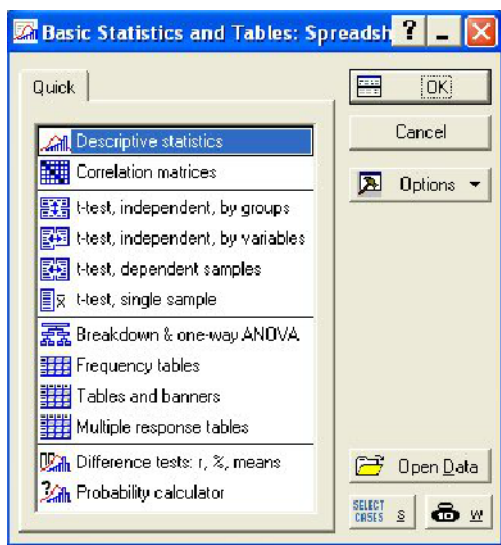


Рис. 2.41. Окно выбора процедуры в меню Basic Statistics/Tables

Процедура *Descriptive statistics* предлагает пользователю широкий набор функций, а также опций к ним. Работа начинается с выбора переменной или переменных, по которым будет проведен анализ. Для этого из базы данных, вызываемой кнопкой *Variables*, выбираются нужные переменные, в рассматриваемом примере это «Выезды за границу», APRc, TRc, (исходный динамический ряд, цепные абсолютные приросты, цепные темпы роста) (рис. 2.42).

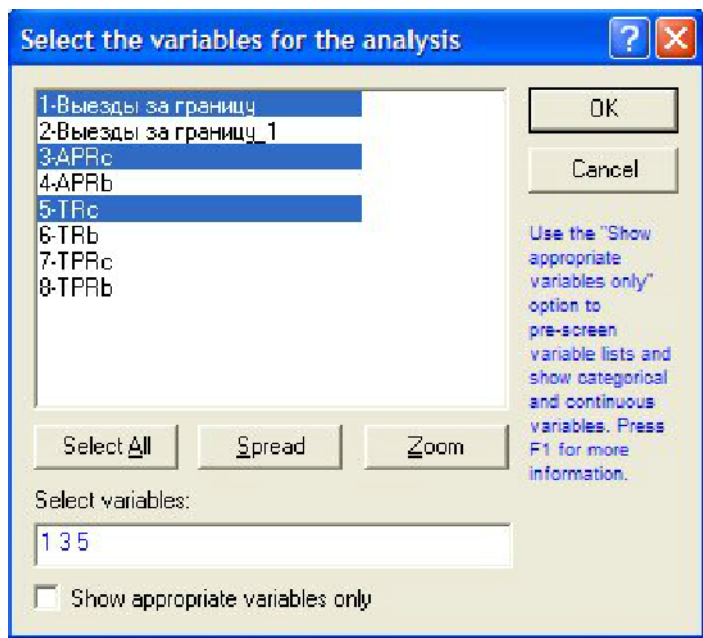


Рис. 2.42. Окно выбора переменных для расчета средних показателей динамического ряда

Для расчета основных статистических характеристик ряда распределения необходимо воспользоваться закладкой *Advanced* (рис. 2.43).

Закладка *Advanced* предлагает пользователю сформировать набор вычисляемых статистик, отвечающих целям анализа. При этом есть возможность выбрать все характеристики (кнопка *Select all stats*).



В нашем случае необходимо выбрать только две статистики.

1. Mean – средняя арифметическая, этот показатель используется для нахождения среднего уровня ряда и определения среднего абсолютного прироста.

2. Для определения среднего темпа роста используется показатель геометрической средней – Geom. mean. Для нахождения среднего темпа прироста из последнего показателя вычитаем единицу.

Для запуска процедуры необходимо воспользоваться кнопкой Summary: Descriptive statistics или Summary.

Результаты расчета основных статистических характеристик представлены на рис. 2.44.

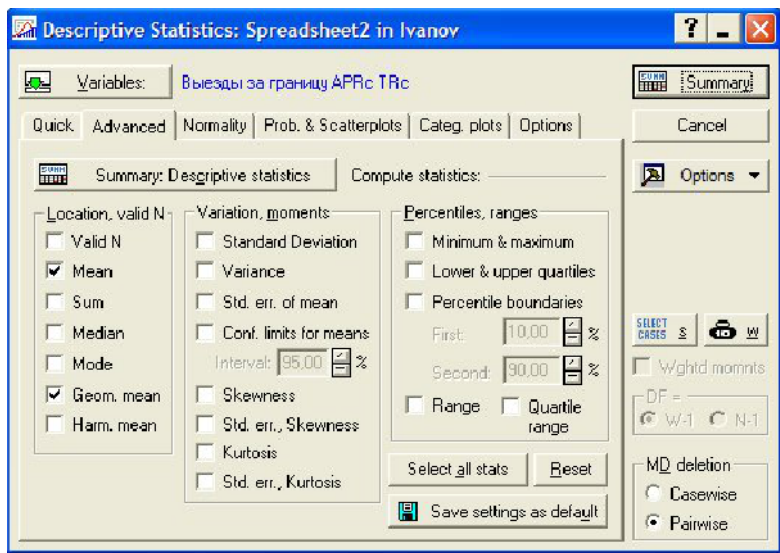


Рис. 2.43. Закладка Advanced процедуры Descriptive Statistics

Variable	Descriptive Statistics (S	
	Mean	Geometric Mean
<b>Выезды за границу</b>	5746,684	4917,469
APRc	647,353	
TRc	114,708	112,910

Рис. 2.44. Результаты расчета средних показателей динамики

Отметим, что программа рассчитывает все отмеченные виды средних для всех переменных (за исключением средней геометрической для переменных с отрицательными элементами). Однако каждый показатель рассчитывается либо по средней арифметической, либо по геометрической. Поэтому необходимо удалить неверно рассчитанные средние (рис. 2.45).

Variable	Descriptive Statistics (€)	
	Mean	Geometric Mean
Выезды за границу	5746,684	
APRc	647,353	
TRc		112,910

Рис. 2.44. Результаты расчета средних показателей динамики после корректировки

На основе рассчитанных показателей можно сделать некоторые выводы.

Например, в среднем ежегодно число россиян, выезжающих за границу с целью туризма в период с 1993 по 2010 годы увеличивалось на 12,9% (средний темп прироста) или на 647,35 тыс. чел. (средний абсолютный прирост).

Средний уровень ряда говорит о том, что в период с 1993 по 2010 годы в среднем почти 6 миллионов россиян ежегодно выезжали за границу. Безусловно, этот показатель интересно рассматривать в сравнении с другими периодами времени. Или на фоне показателей других стран.

## 2.5. Периодизация временных рядов

Периодизация ряда динамики – это разделение его на временные этапы, в течение которых сохранялась неизменной основная тенденция развития явления. Это, своего рода, типологическая группировка во времени.

Динамические ряды отражают развитие какого-либо явления или процесса за длительные периоды времени. На анализируемом временном отрезке могут происходить существенные качественные

изменения условий развития изучаемого объекта, что, в свою очередь, приводит к изменению основной тенденции изменения уровней ряда.

Большие системы (биологические или макроэкономические), обладающие значительной инерцией, более устойчивы, чем, например, отдельные компании или организации. Однако при анализе любого временного ряда имеет смысл оценить необходимость проведения периодизации.

Выделение однородных временных отрезков необходимо: при расчете средних показателей динамики, поскольку средняя величина отражает типический уровень только тогда, когда она рассчитана по качественно однородной совокупности; при построении моделей ряда; при осуществлении экстраполяции, предполагающей продление в будущее тенденции, сформировавшейся в прошлом.

Проведение периодизации должно основываться, прежде всего, на всестороннем анализе внутренних причин и внешних условий существования и развития объекта изучения. На смену тенденции развития социально-экономических явлений могут повлиять определенные исторические события, смена системы управления и регулирования экономики, изменение хозяйственного механизма, кардинальные изменения в технике и технологии производства и т. п. Таким образом, теоретический анализ – отправная точка в периодизации рядов динамики.

Провести периодизацию рядов динамики часто помогает анализ их графических изображений (большую наглядность обеспечивают графики, построенные на основе базисных темпов роста) и показателей динамики, описанных в предыдущем параграфе.

О необходимости периодизации говорит смена знака при показателях скорости изменения уровней ряда или многократное увеличение (уменьшение) значений характеристик интенсивности. Моменты времени, в которые наблюдаются эти изменения, можно считать границами выделяемых периодов.

Анализируя все перечисленные аспекты для рассмотренного динамического ряда, можно выделить некоторые тенденции:

- с 1993 по 1997 год наблюдался устойчивый и стабильный рост числа выездов россиян за границу с целью туризма;

- экономический кризис 1998 года вернул данный показатель на уровень 1995 года и рост продолжился только в 2000 году, однако нестабильная ситуация в мире не позволила стремительно развиваться

туризму, поэтому устойчивый рост продолжился только с 2001 года;

- с 2001 года наблюдался поступательный рост числа выездов россиян за границу, а с 2005 года этот рост стал более интенсивным;

- экономический кризис 2008 года вызвал падение показателя более чем на 15%, но в отличие от кризиса 1998 года, индустрия восстановилась уже к 2010 году.

Таким образом, можно выделить три периода:

- 1993-1997 – период начального роста;

- 1998-2000 – период стагнации;

- 2001-2010 – период активного роста.

Здесь необходимо отметить, что в данном динамическом ряду основная тенденция за 18 лет не менялась, то есть число выездов за границу россиян с целью туризма увеличивалось, однако имели место некоторые колебания тенденции, которые и отмечены выше. По сути, периодизация в таких случаях не нужна, однако колебания стоит использовать при разделении на периоды в тех случаях, когда исходных данных периода, отражающего текущие особенности основной тенденции достаточно для прогноза. Дело в том, что период упреждения, то есть период, на который делается прогноз, не должен превышать треть длины анализируемого ряда. Поэтому, если мы хотим осуществить прогнозирование на 2 года, необходимы статистические данные минимум за 6 лет. Также стоит отметить, что ошибка прогнозирования тем больше, чем меньше уровней ряда содержит статистика.

Далее в проекте используются лишь данные последнего периода, с помощью которых будет осуществляться прогнозирование.

Для дальнейшей работы создадим отдельную переменную, содержащую данные последнего периода.

Для этого создадим новый рабочий лист (меню Workbook/Insert (рис. 2.45), в появившемся окне нажимаем OK – это означает, что мы хотим создать документ STATISTICA и поместить его следующим за текущим рабочим листом (рис. 2.46); затем еще раз нажимаем OK – это означает, что мы создаем именно рабочий лист (рис. 2.47); после этого появляется электронная таблица размером 10x10 в дереве рабочей книги (рис. 2.48); оставим в ней только одну переменную с названием «ВГ2001-2010» (Выезды за границу в 2001-2010 г.г.), скопируем туда данные с исходного рабочего листа и ось времени (рис. 2.49); сделаем лист активным (для этого щелкаем по нему в дереве рабочей книги правой кнопкой мыши и выбираем Use as Active Input) (рис.

2.50); построим графическое изображение новой переменной (рис. 2.51).

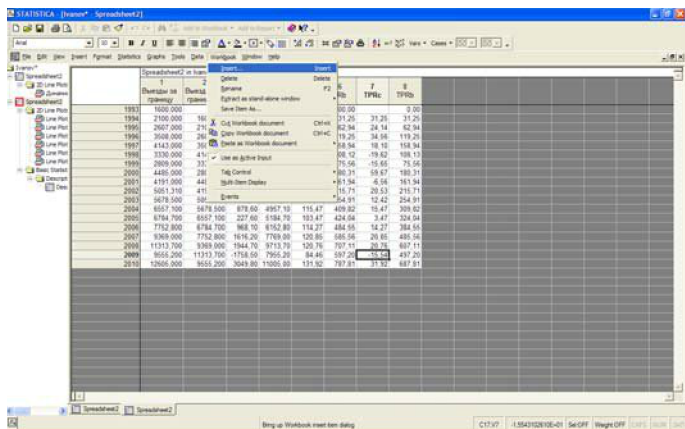


Рис. 2.45. Выбор функции добавления объекта STATISTICA



Рис. 2.46. Выбор типа добавляемого объекта и места его положения

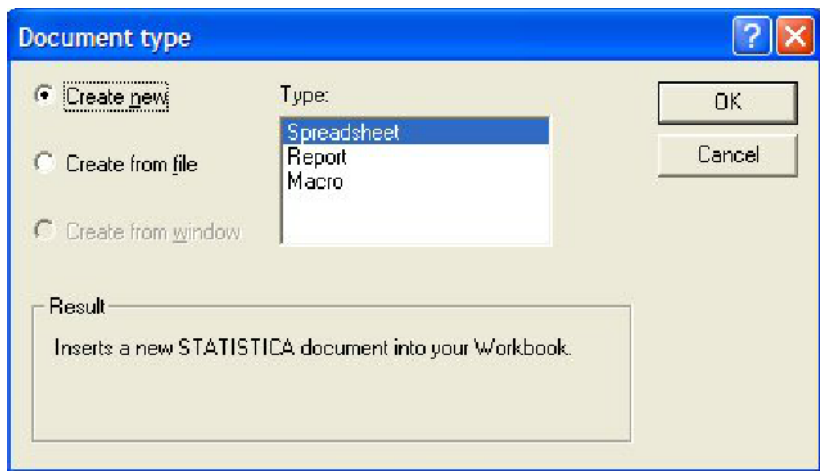


Рис. 2.47. Добавление рабочего листа

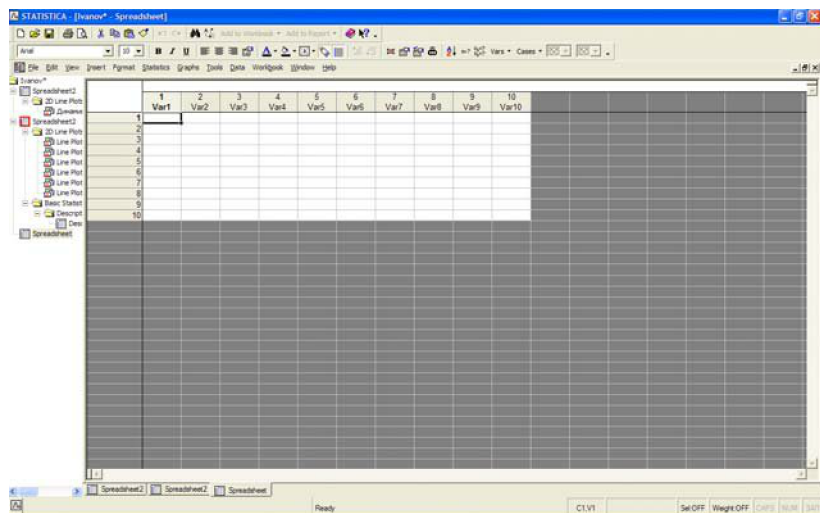


Рис. 2.48. Окно с электронной таблицей 10x10

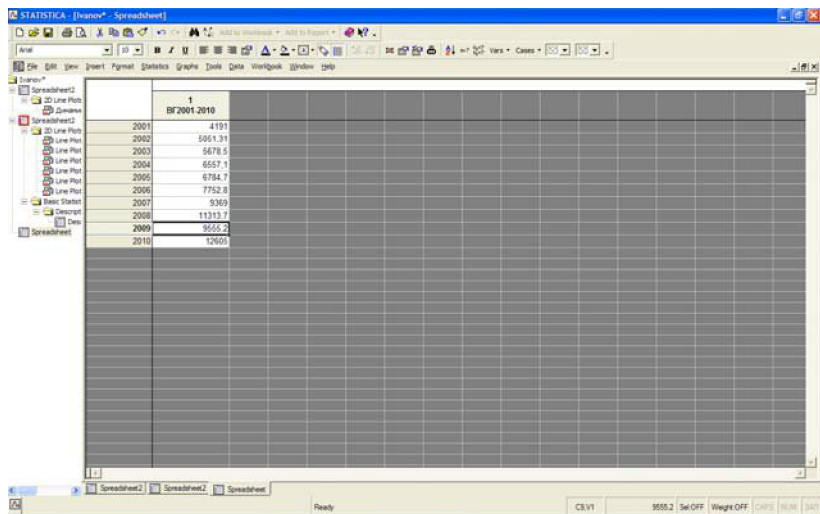


Рис. 2.49. Рабочая область с новым активным рабочим листом и переменной «ВГ2001-2010»

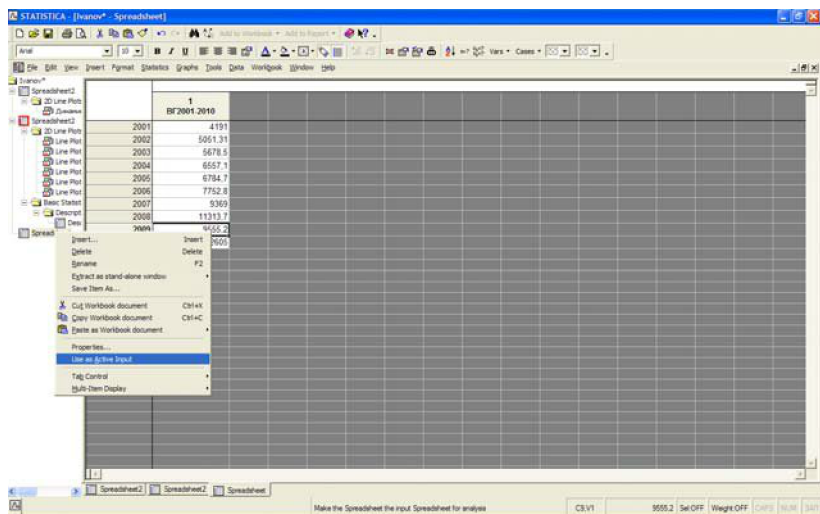


Рис. 2.50. Выбор активного рабочего листа

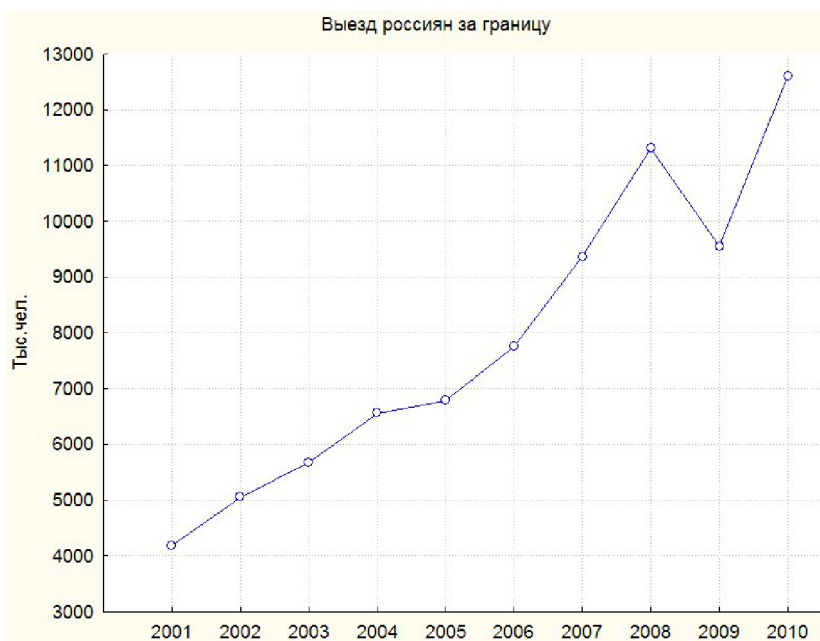


Рис. 2.51. Динамика выезда россиян за границу в 2001-2010 г.г.



### 3. ВЫЯВЛЕНИЕ И АНАЛИЗ ОСНОВНОЙ ТЕНДЕНЦИИ ВРЕМЕННОГО РЯДА

#### 3.1. Компоненты временного ряда

Одной из важнейших задач статистического анализа рядов динамики является выявление и описание основной тенденции развития изучаемого явления, закономерности изменения уровней ряда. Иногда характер тенденции достаточно отчетливо проявляется на графике и в системе статистических показателей, описанных в предыдущих параграфах. Однако часто встречаются динамические ряды, в которых основная тенденция не является очевидной, поскольку на уровни ряда влияет большое число разнообразных факторов.

Основная тенденция развития того или иного явления складывается под воздействием долговременно действующих внутренних и внешних причин и условий, благодаря которым, в основном, формируется величина уровня ряда. Одновременно с этим, уровни динамического ряда колеблются (отклоняются от основной тенденции) под воздействием краткосрочных случайных или систематически (циклически) действующих факторов. Чем сильнее их влияние, тем сложнее вскрыть основную закономерность развития объекта.

При анализе рядов динамики могут быть выделены четыре компоненты уровней ряда:

$$y = f(T, C, S, \varepsilon), \quad (3.1)$$

где  $T$  – трендовая компонента, отражающая основную тенденцию развития, так называемый тренд;  $C$  – циклическая (конъюнктурная) компонента;  $S$  – сезонная компонента;  $\varepsilon$  – случайная компонента.

Выделение и изучение отдельных компонент временного ряда называется декомпозицией ряда динамики. Изучение каждой компоненты предполагает использование специальных приемов и методов.

Все компоненты присутствуют далеко не в каждом динамическом ряду. Чаще всего в практических исследованиях встречаются ряды, содержащие трендовую и случайную компоненты. Чтобы видеть влияние сезонной составляющей, нужно иметь ряд, уровни которого относятся к месяцам или кварталам. Проявление циклической компо-

ненты, как правило, характерно для больших динамических рядов, что связано с экономическими (бизнес) циклами. Чем меньше влияние на уровни ряда нетрендовых компонент, тем проще выделить тренд – основную тенденцию изучаемого ряда, описание и прогнозирование которой является центральной задачей изучения временных рядов.

## 3.2. Методы анализа основной тенденции временных рядов

*Тенденция* – это объективно существующее свойство того или иного процесса, которое лишь приближенно описывается трендом определенного вида. Трендом называют и саму основную тенденцию развития, и конкретное ее описание с помощью уравнения регрессии.

Для выявления и анализа общей тенденции развития изучаемого явления необходимо абстрагироваться от влияния нетрендовых факторов. Достичь этого, в определенной степени, позволяют приемы сглаживания или выравнивания временного ряда.

Различают механическое и аналитическое выравнивания. Последнее позволяет формализовать тенденцию, представить ее в виде конкретной математической функции.

Суть различных приемов, с помощью которых осуществляется сглаживание, сводится к замене фактических уровней динамического ряда расчетными, имеющими значительно меньшую колеблемость, чем исходные данные. Уменьшение колеблемости уровней позволяет тенденции развития проявиться более отчетливо. В ряде случаев сглаживание ряда может рассматриваться как важное вспомогательное средство, облегчающее применение других методов и, в частности, более строгих методов выделения тенденции.

### 3.2.1. Проверка динамического ряда на наличие тренда

Прежде чем приступить к решению задачи аналитического сглаживания динамических рядов (аналитического описания общей тенденции развития регрессионными моделями), необходимо проверить существенность трендовой составляющей динамического ряда. В качестве примера проведем проверку динамического ряда на наличие тренда с помощью фазочастотного критерия Валлиса-Мура.

Этот критерий позволяет отличить закономерные отклонения последовательности уровней ряда от чисто случайной последователь-

ности. Если тренд отсутствует, то знаки разностей значений уровней  $y_{t+1} - y_t$  образуют случайную последовательность.

С помощью критерия Валлиса-Мура проверяется гипотеза: последовательность знаков разностей имеет случайный характер. Альтернативной является гипотеза: последовательность знаков разностей значительно отличается от случайной. Последовательность одинаковых знаков называется «фазой». При расчетах первая и последняя фаза опускаются. В предположении случайности ряда статистика  $z$  распределена нормально.

$$z = \frac{\left| h - \frac{2n-7}{3} \right| - 0.5}{\sqrt{\frac{16n-29}{90}}}, \quad (3.2)$$

где  $h$  – число фаз (за вычетом первой и последней);  $n$  – число уровней ряда. При  $n > 30$  так называемая поправка Йейтса (0,5 в числителе) на непрерывность может быть опущена.

Проверим на наличие тренда переменную «ВГ2001-2010». Отметим, что абсолютный прирост в 2001 году не указывается, поскольку для этой переменной это начальный уровень ряда.

Таким образом, рассматривается 10 уровней ряда ( $n = 10$ ),  $h = 1$  (три фазы, но первая и последняя фазы отбрасываются).

Подставляем значения в формулу и получаем расчетное значение критерия  $z$  равно 1,95.

$$z = \frac{\left| 1 - \frac{2 \cdot 10 - 7}{3} \right| - 0.5}{\sqrt{\frac{16 \cdot 10 - 29}{90}}} = 1,95$$

Таблица 3.1

Данные об изменениях абсолютных приростов

Год	Абсолютный прирост (цепной)	Знак изменения	№ фазы
2001	-		-
2002	860,31	+	1
2003	627,19	+	
2004	878,60	+	
2005	227,60	+	
2006	968,10	+	
2007	1616,20	+	
2008	1944,70	+	
2009	-1758,50	-	2
2010	3049,80	+	3

Теоретическое значение критерия при доверительной вероятности 95% равно 1,96 (нормальное распределение). Расчетное значение критерия не превышает табличное, тем самым на 5%-м уровне значимости принимается нулевая гипотеза об отсутствии тренда и отвергается альтернативная. Следовательно, можно утверждать, что в динамическом ряду отсутствует тренд.

*Однако критерий Валлиса-Мура имеет ограничение в виде  $n > 12$ . Простейшее предположение об увеличении числа уровней ряда хотя бы до 12 даже при еще одной смене фазы приведет к отклонению нулевой гипотезы и выводе о присутствии тренда в динамическом ряду. Пренебрежение к условиям применения критерия привело к абсурдному выводу, так как графическое изображение четко демонстрирует наличие тренда.*

Данный критерий основан на анализе его авторами большого числа статистических данных и выведен эмпирически, а, значит, не может быть окончательной истиной, но может использоваться как один из методов оценки наличия тренда.

Для проверки рядов на наличие тренда может быть использован критерий Кокса-Стюарта.

Присутствие тренда можно оценить, анализируя автокорреляционную функцию (см. раздел 5). Если коэффициенты автокорреляции уровней значимы, и функция медленно убывает – это говорит о наличии тренда.

### 3.2.2. Механическое выравнивание временного ряда. Скольльзящие средние

Один из наиболее простых приемов сглаживания заключается в расчете скользящих, или, как иногда их называют, подвижных средних. Применение последних, позволяет сгладить периодические и случайные колебания и тем самым выявить присутствующую в развитии тенденцию.

Пусть динамический ряд состоит из уровней  $y_t$ ,  $t = 1, \dots, n$ . Для каждых  $m$  последовательных уровней этого ряда ( $m < n$ ) можно подсчитать среднюю величину. Вычислив значение средней для первых  $m$  уровней, переходят к расчету средней для уровней  $y_2, \dots, y_{m+i}$ , затем  $y_3, \dots, y_{m+2}$  и т. д. Таким образом, интервал сглаживания, т. е. интервал, для которого подсчитывается средняя, как бы скользит по динамическому ряду с шагом, равным единице. Если  $m$  нечетное число, а предпочтительнее брать именно нечетное число уровней, поскольку в этом случае расчетное значение уровня окажется в центре интервала сглаживания и им легко заменить конкретное фактическое значение, то для определения скользящей средней можно записать следующую формулу:

$$\hat{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{\sum_{i=t-p}^{t+p} y_i}{m}, \quad (3.3)$$

где  $\hat{y}_t$  — значение скользящей средней для момента  $t$ ,  $y_i$  — фактическое значение уровня в момент  $i$ ;  $i$  — порядковый номер уровня в интервале сглаживания;  $m$  — интервал сглаживания (период скольжения).

Величина  $p$  легко определяется из продолжительности интервала сглаживания. Поскольку  $m = 2p + 1$  при нечетном  $m$ , то

$$p = \frac{m-1}{2}. \quad (3.4)$$

Выбор периода скольжения имеет большое значение, особенно, если в изучаемом ряду имеются циклические или сезонные колебания. В этом случае период скольжения должен быть равным, либо кратным

периоду колеблемости. Если циклических колебаний не наблюдается, то рекомендуется выполнить несколько вариантов выравнивания: начать с расчета скользящей средней с минимальным периодом скользящего и постепенно увеличивать период сглаживания, пока основная тенденция не проявится достаточно отчетливо. Средние, рассчитанные по большому периоду, лучше сглаживают случайные колебания. Но использование многочленных скользящих средних может быть ограничено незначительной продолжительностью исходного ряда. Необходимо учитывать, что использование метода скользящих средних приводит к получению укороченного временного ряда.

В рамках курсового проекта требуется провести сглаживание динамического ряда 3-х и 5-ти членными скользящими средними (если это позволяет длина динамического ряда).

Для начала работы воспользуемся уже известным нам меню *Statistics/Advanced Linear/Nonlinear Models/Time Series/Forecasting*. Напомним, что работать необходимо с переменной, по которой будет вестись прогнозирование.

В появившемся окне указываем переменную, с которой будем работать («ВГ2001-2010»), нажимаем на кнопку *OK (transformations, autocorrelations, crosscorrelations, plots)* и затем выбираем закладку *Smoothing* (рис. 3.1.).

На данной закладке мы выбираем опцию *N*-точечной скользящей средней *N-pts Moving Average*. Для получения 3-х членной скользящей средней в поле интервала сглаживания *N* ставим цифру 3 (*N = 3*). Далее нажимаем на кнопку *OK (Transform selected series)*, система автоматически строит графическое изображение, выраженное скользящей средней с заданными параметрами, которое мы удаляем за ненадобностью. Далее возвращаемся к анализу (кнопка в левом нижнем углу экрана). В информационном поле находятся не только исходная переменная, но и вновь созданная 3-х членный скользящий средний (*3pt.mov.aver*). Выбираем в информационном поле исходную переменную, меняем период скользящего (*N=5*) (рис. 3.2), опять нажимаем на кнопку *OK (Transform selected series)*, удаляем построенное изображение и возвращаемся к анализу.

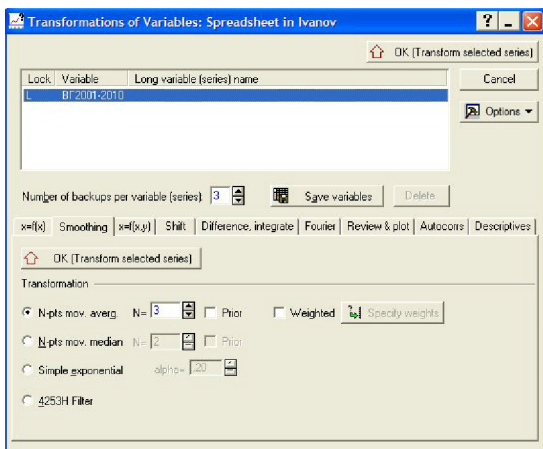


Рис. 3.1. Вид закладки *Smoothing* процедуры *Transformation of Variables*

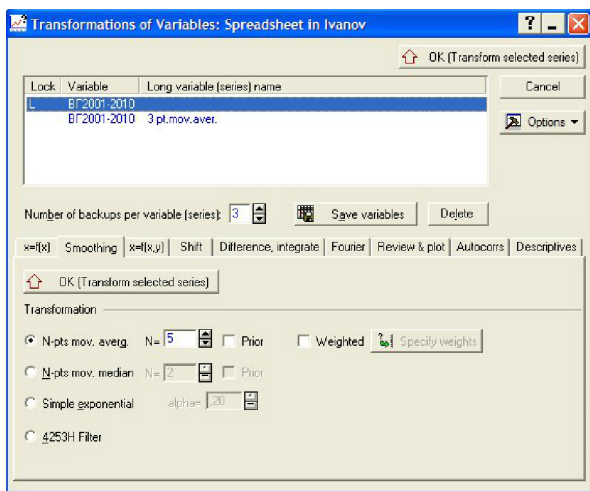


Рис. 3.2. Вид закладки *Smoothing* процедуры *Transformation of Variables* с построенной 3-х членной скользящей средней

В информационном поле содержится уже 3 переменных (рис. 3.3). Далее нажимаем кнопку *Save variables*, и система переносит расчетные значения на отдельный рабочий лист, который необходимо

добавить в исходную рабочую книгу (см. рис. 2.26), а затем скопировать туда ось времени (рис. 3.4).

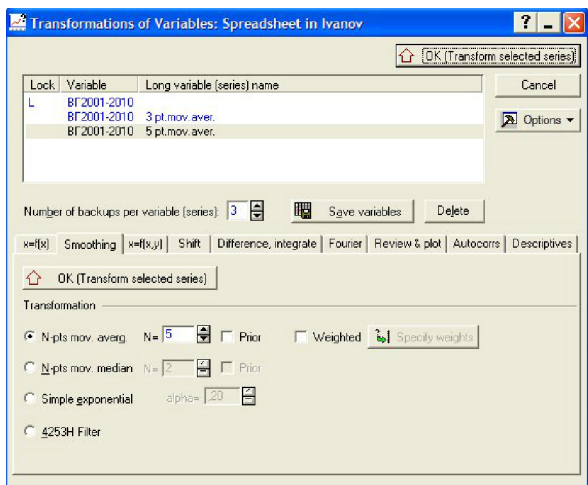


Рис. 3.3. Информационное поле с 3-х и 5-ти членными скользящим средними

	Spreadsheet in Ivanov		
	1 BG2001-2010	2 BG2001-2010_3-х член. скол. сред.	3 BG2001-2010_5-ти член. скол. сред.
2001	4191,000		
2002	5051,310	4973,603	
2003	5678,500	5762,303	5652,522
2004	6557,100	6340,100	6364,882
2005	6784,700	7031,533	7228,420
2006	7752,800	7968,833	8355,460
2007	9369,000	9478,500	8955,080
2008	11313,700	10079,300	10119,140
2009	9555,200	11157,967	
2010	12605,000		

Рис. 3.4. Исходный и сглаженные динамические ряды

Для наглядного отображения механического выравнивания представим сглаживание скользящими средними графически. Для этого выделяем столбцы с исходными данными и сглаженными рядами,



щелкаем по ним правой кнопкой мыши и выбираем процедуру Graphs of Block Data/Line Plot: Entire Columns (рис. 3.5). В результате получаем график, на котором исходный динамический ряд сглажен 3-х и 5-ти членными скользящими средними (рис. 3.6).

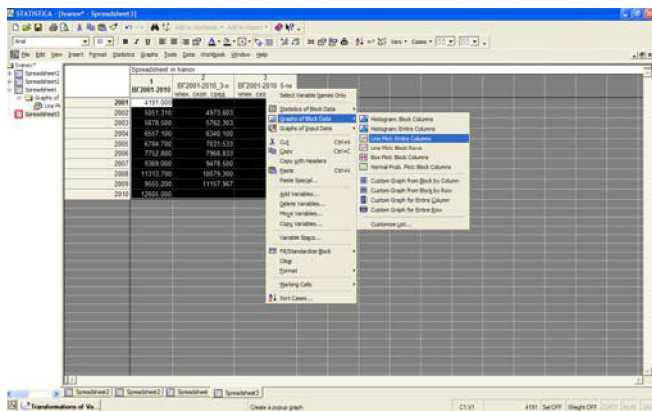


Рис. 3.5. Построение графиков на основе данных столбцов

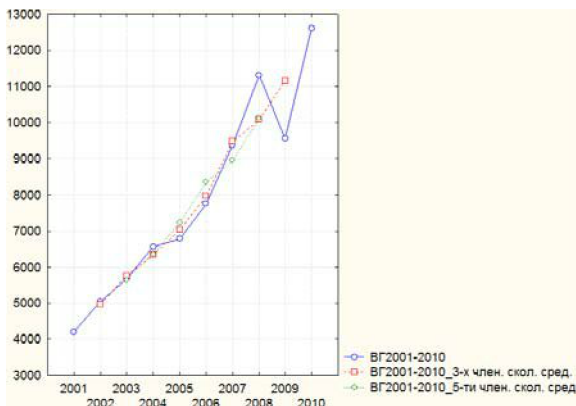


Рис. 3.5. Динамический ряд, сглаженный 3-х и 5-ти членными скользящими средними

На графике отчетливо видно, что выравнивание позволяет сгладить небольшие колебания 2005 и 2006 года, а также аппроксимировать значительные отклонения 2008 и 2009 года.

Простые скользящие средние – относительно грубый статисти-

ческий прием выявления тенденции. В ряде случаев сглаживание с помощью простой скользящей средней оказывается настолько сильным, что тенденция развития проявляется лишь в самом общем виде, а отдельные важные для экономического анализа детали теряются.

Более тонкий прием, базирующийся на той же самой идее, что и простые скользящие средние, заключается в применении взвешенных скользящих средних, экспоненциального сглаживания. Поскольку процедура экспоненциального сглаживания чаще используется при решении задач прогнозирования, в данном учебном пособии мы не будем ее рассматривать<sup>3</sup>.

### 3.3. Аналитическое сглаживание временного ряда. Уравнение тренда.

Кривые роста, описывающие закономерности развития явлений во времени – это результат аналитического выравнивания динамических рядов. Выравнивание ряда с помощью тех или иных функций в большинстве случаев оказывается удобным средством описания эмпирических данных. Это средство при соблюдении ряда условий можно применить и для прогнозирования. Процесс выравнивания состоит из следующих основных этапов:

- выбора типа кривой, форма которой соответствует характеру поведения динамического ряда;
- определения численных значений (оценка) параметров кривой;
- апостериорного контроля качества выбранного тренда.

В современных ППП все перечисленные этапы реализуются одновременно, как правило, в рамках одной-двух процедур.

Аналитическое сглаживание с использованием той или иной функции позволяет получить выровненные, или, как их иногда не вполне правомерно называют, теоретические значения уровней динамического ряда, т. е. уровни, которые наблюдались бы, если бы динамика явления полностью совпадала с кривой. Эта же функция с некоторой корректировкой или без нее, применяется в качестве модели для экстраполяции (прогноза).

---

<sup>3</sup> Статистика для менеджеров с использованием Microsoft Excel, 4-е изд. : Пер. с англ. – М. : Издательский дом «Вильямс», 2004., 1312 с.

Вопрос о выборе типа кривой является основным при выравнивании ряда. При всех прочих равных условиях ошибка в решении этого вопроса оказывается более значимой по своим последствиям (особенно для прогнозирования), чем ошибка, связанная со статистическим оцениванием параметров.

Поскольку форма тренда объективно существует, то при выявлении ее следует исходить из материальной природы изучаемого явления, исследуя внутренние причины его развития, а также внешние условия и факторы на него влияющие. Только после глубокого содержательного анализа можно переходить к использованию специальных приемов, разработанных статистикой.

Весьма распространенным приемом выявления формы тренда является графическое изображение временного ряда. Но при этом велико влияние субъективного фактора, даже при отображении выровненных уровней.

Из сказанного выше можно сделать вывод о том, что выбор формы кривой для выравнивания представляет собой задачу, которая не решается однозначно, а сводится к получению ряда альтернатив. Окончательный выбор не может лежать в области формального анализа, тем более, если предполагается с помощью выравнивания не только статистически описать закономерность поведения уровня в прошлом, но и экстраполировать найденную закономерность в будущее. Вместе с тем различные статистические приемы обработки данных наблюдения могут принести существенную пользу, по крайней мере, с их помощью можно отвергнуть заведомо непригодные варианты и тем самым существенно ограничить поле выбора.

Рассмотрим наиболее используемые типы уравнений тренда:

1. Линейная форма тренда:

$$\hat{y}_t = a_0 + a_1 t, \quad (3.5)$$

где  $\hat{y}_t$  – уровень ряда, полученный в результате выравнивания по прямой;  $a_0$  – начальный уровень тренда;  $a_1$  – средний абсолютный прирост, константа тренда.

Для линейной формы тренда характерно равенство так называемых первых разностей (абсолютных приростов) и нулевые вторые разности, т. е. ускорение.

2. Параболическая (полином 2-ой степени) форма тренда:

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2 \quad (3.6)$$

Для данного типа кривой постоянными являются вторые разности (ускорение), а нулевыми – третьи разности.

Параболическая форма тренда соответствует ускоренному или замедленному изменению уровней ряда с постоянным ускорением. Если  $a_2 < 0$  и  $a_1 > 0$ , то квадратическая парабола имеет максимум, если  $a_2 > 0$  и  $a_1 < 0$  – минимум. Для отыскания экстремума первую производную параболы по  $t$  приравнивают 0 и решают уравнение относительно  $t$ .

3. Логарифмическая форма тренда:

$$\hat{y}_t = a_0 + a_1 \ln t, \quad (3.7)$$

где  $a_1$  – константа тренда.

Логарифмическим трендом может быть описана тенденция, проявляющаяся в замедлении роста уровней ряда динамики при отсутствии предельно возможного значения. При достаточно большом  $t$  логарифмическая кривая становится мало отличимой от прямой линии.

4. Мультипликативная (степенная) форма тренда:

$$\hat{y}_t = a_0 t^{a_1} \quad (3.8)$$

5. Полином 3-ей степени:

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \quad (3.9)$$

Естественно, кривых, описывающих основные тенденции, гораздо больше. Однако формат учебного пособия не позволяет описать все их многообразие. Показанные далее приемы построения моделей позволят пользователю самостоятельно использовать другие функции,

в частности обратные.

В практике исследования социально-экономических явлений исключительно редко встречаются динамические ряды, характеристики которых полностью соответствуют признакам эталонных математических функций. Это обусловлено значительным числом факторов разного характера, влияющих на уровни ряда и тенденцию их изменения.

Для решения поставленной задачи по аналитическому сглаживанию динамических рядов в системе STATISTICA нам потребуется создать дополнительную переменную на листе с исходными данными переменной «ВГ2001-2010», который следует сделать активным.

Нам предстоит построить уравнение тренда, которое по существу является уравнением регрессии, в котором в качестве фактора выступает «время». Создаем переменную «Т», содержащую интервалы времени с 2001 по 2010. Переменная «Т» будет состоять из натуральных чисел от 1 до 10, соответствующих указанным годам.

В итоге получается следующий рабочий лист (рис. 3.6)

Далее рассмотрим процедуру, позволяющую строить регрессионные модели как линейного, так и нелинейного типа. Для этого выбираем: Statistics/Advanced Linear/Nonlinear Models/Nonlinear Estimation (рис. 3.7). В появившемся окне (рис. 3.8) выбираем функцию User-specified Regression, Least Squares (построение моделей регрессии пользователем вручную, параметры уравнения находятся по методу наименьших квадратов (МНК)).

В следующем диалоговом окне (рис. 3.9) нажимаем на кнопку Function to be estimated, чтобы попасть на экран для задания модели вручную (рис. 3.10).

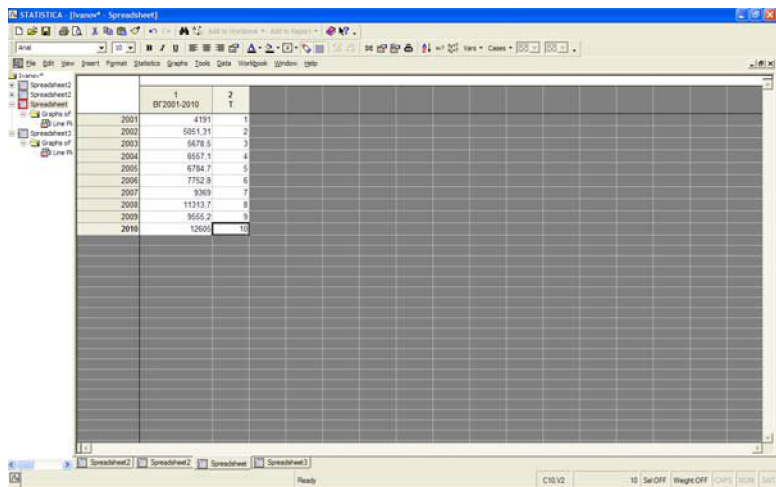


Рис. 3.6. Рабочий лист с созданной переменной времени

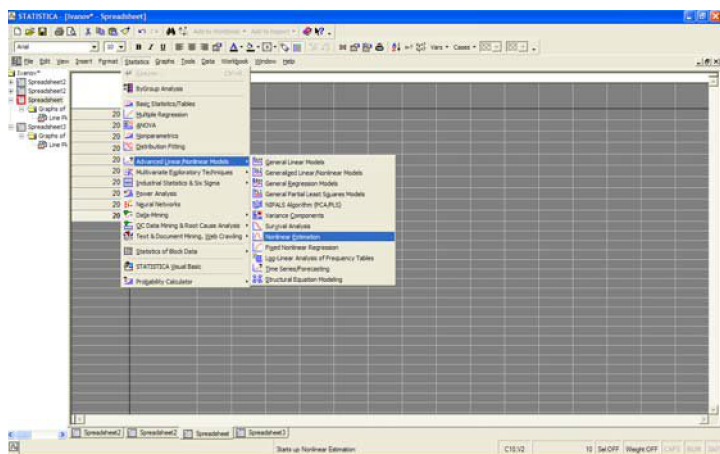


Рис. 3.7. Запуск процедуры *Statistics/Advanced Linear/Nonlinear Models/Nonlinear Estimation*

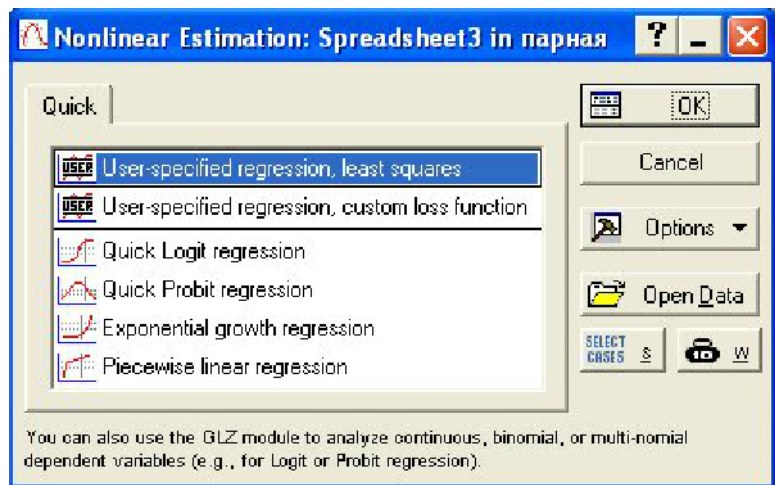


Рис. 3.8. Окно процедуры *Nonlinear Estimation*

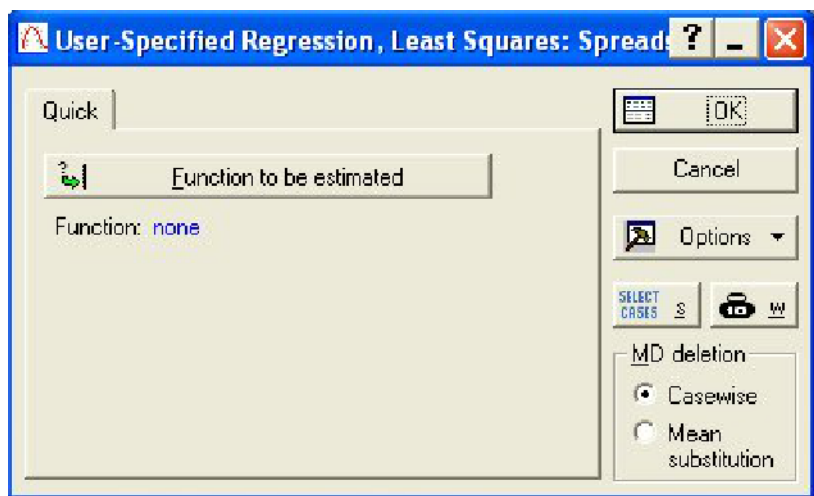


Рис. 3.9. Окно процедуры *User-Specified Regression, Least Squares*

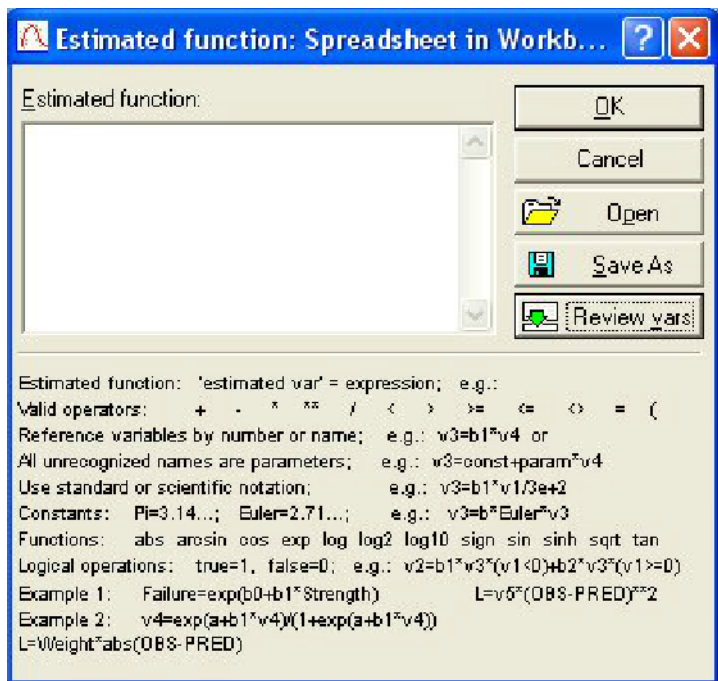


Рис. 3.10. Окно для реализации процедуры задания уравнения тренда вручную

В верхней части экрана находится поле для ввода функции, в нижней части располагаются примеры ввода функций для различных ситуаций.

Прежде чем сформировать интересные нас модели, необходимо пояснить некоторые условные обозначения. Переменные уравнений задаются в формате «v№», где «v» обозначает переменную (*от* англ. «*variable*»), а «№» – номер столбца, в котором она расположена в таблице на рабочем листе с исходными данными. Если переменных очень много, то справа находится кнопка *Review vars*, позволяющая выбирать их из списка по названиям и просматривать их параметры с помощью кнопки *Zoom* (рис. 3.11).



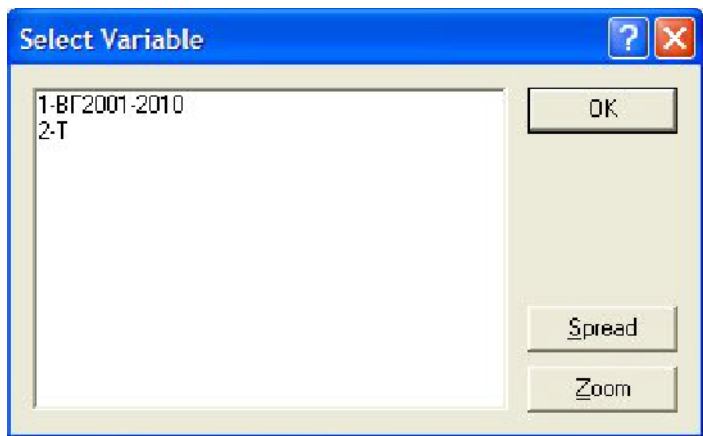


Рис. 3.11. Окно выбора переменной с помощью кнопки *Review vars*

Параметры уравнений обозначаются любыми латинскими буквами, не обозначающими какое-либо математическое действие. Для упрощения работы предлагается обозначать параметры уравнения так, как в описании уравнений тренда – латинской буквой «a», последовательно присваивая им порядковые номера. Знаки математических действий (вычитания, сложения, умножения и пр.) задаются в обычном для *Windows*-приложений формате. Пробелы между элементами уравнения не требуются.

Итак, рассмотрим первую модель тренда – линейную,  
 $\hat{y}_t = a_0 + a_1 t$ .

Следовательно, после набора она будет выглядеть следующим образом:

$$v1 = a0 + a1 * v2 ,$$

где  $v1$  – это столбец на листе с исходными данными, в котором находятся значения исходного динамического ряда;  $a0$  и  $a1$  – параметры уравнения;  $v2$  – столбец на листе с исходными данными, в котором находятся значения интервалов времени (переменная  $T$ ) (рис. 3.12).

После этого дважды нажимаем кнопку OK.

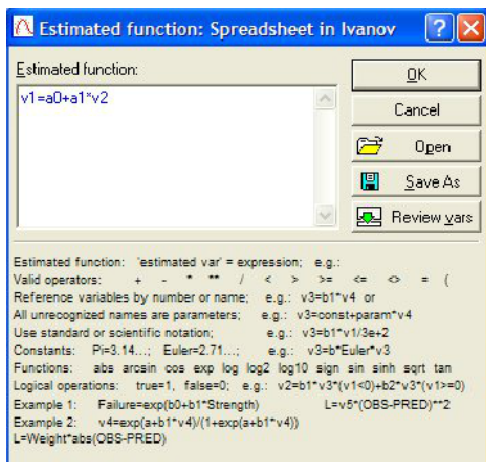


Рис. 3.12. Окно процедуры задания уравнения линейного тренда

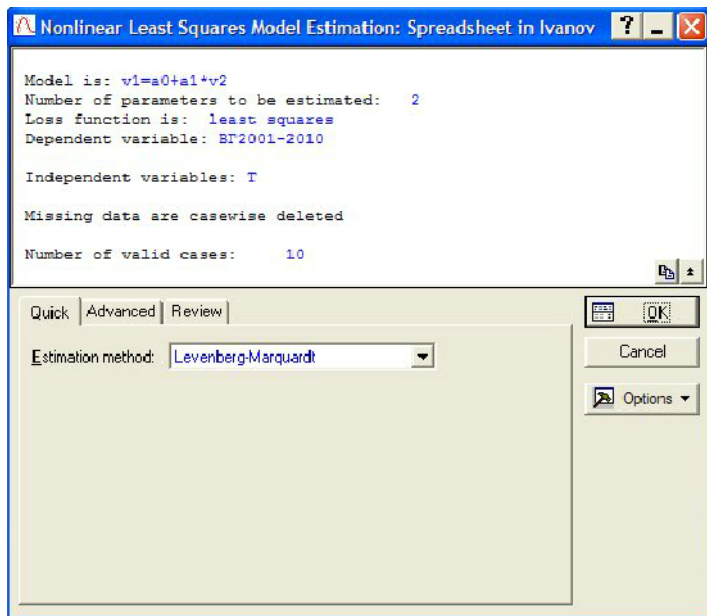


Рис. 3.13. Закладка Quick процедуры оценки уравнения тренда.

В появившемся окне (рис. 3.13) можно выбрать метод оценки параметров уравнения регрессии (*Estimation method*), если это необходимо. В нашем случае нужно перейти к закладке *Advanced* и нажать на кнопку *Start values* (рис. 3.14). В этом диалоге задаются стартовые значения параметров уравнения для их нахождения по МНК, т.е. их минимальные значения. Изначально они заданы как 0,1 для всех параметров. В нашем случае можно оставить эти значения в том же виде, но если значения в наших исходных данных меньше единицы, то необходимо задать их в виде 0,001 для всех параметров уравнения тренда (рис. 3.15). Далее нажимаем кнопку *OK*.

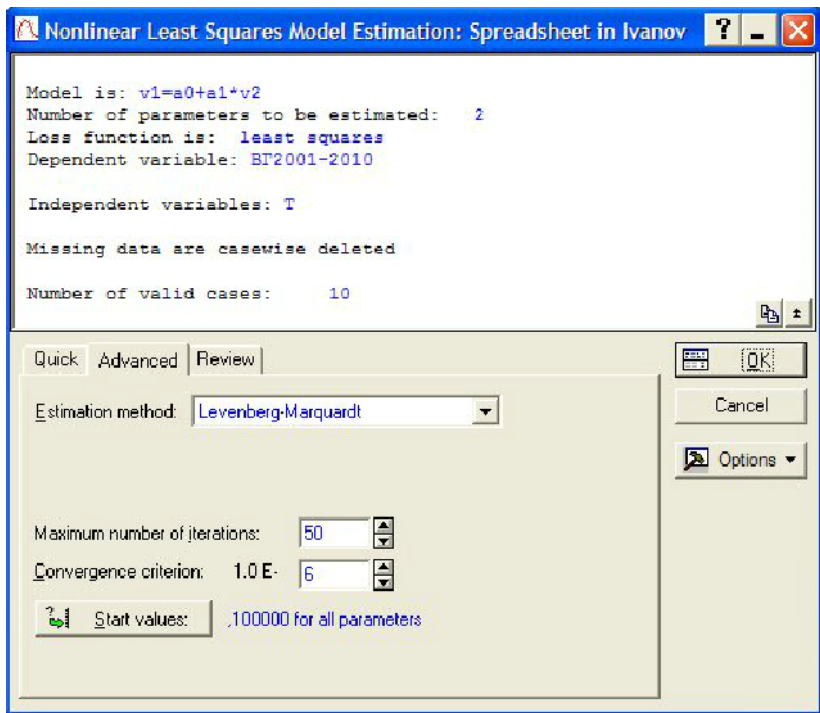


Рис. 3.14. Закладка *Advanced* процедуры оценки уравнения тренда

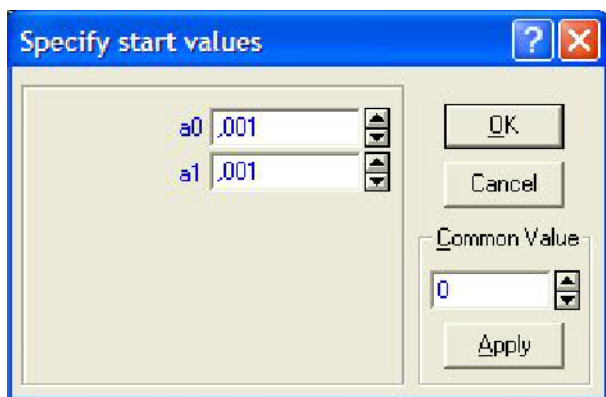


Рис. 3.15. Окно задания стартовых значений параметров уравнения тренда

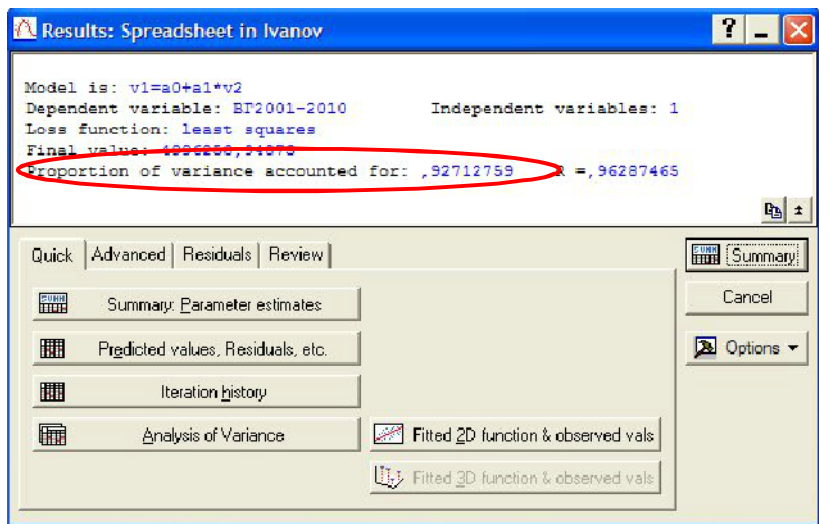


Рис. 3.16. Закладка Quick окна результатов регрессионного анализа

На закладке Quick (рис.3.16) очень важным является значение строчки Proportion of variance accounted for, которое соответствует коэффициенту детерминации; это значение лучше записать отдельно, так как в дальнейшем оно выводиться не будет, и пользователю при-

дется рассчитывать коэффициент вручную, при этом достаточно трех знаков после запятой<sup>4</sup>. Далее нажимаем кнопку *Summary: Parameter estimates* для получения данных о параметрах линейного уравнения тренда (рис. 3.17).

Model is: $v1=a0+a1*v2$ (Spreadsheet in Ivanov)						
Dep. Var. : БГ2001-2010						
Level of confidence: 95.0% ( alpha=0.050)						
	Estimate	Standard error	t-value df = 8	p-level	Lo. Conf Limit	Up. Conf Limit
a0	3058,063	539,8596	5,66455	0,000473	1813,145	4302,982
a1	877,776	87,0063	10,08865	0,000008	677,139	1078,413

Рис. 3.17. Результаты расчета параметров линейной модели тренда

Столбец *Estimate* – числовые значения параметров уравнения; *Standard error* – стандартная ошибка параметра; *t-value* – расчетное значение *t*-критерия; *df* – число степеней свободы ( $n-2$ ); *p-level* – расчетный уровень значимости; *Lo. Conf. Limit* и *Up. Conf. Limit* – соответственно нижняя и верхняя граница доверительных интервалов для параметров уравнения с установленной вероятностью (указана как *Level of Confidence* в верхнем поле таблицы).

Уравнение линейно модели тренда имеет вид  $\hat{y}_t = 3058,063 + 877,776 \cdot t$ .

Параметры уравнения тренда в STATISTICA, как и в большинстве других программ, рассчитываются по методу наименьших квадратов (МНК).

Метод позволяет получить значения параметров, при которых обеспечивается минимизация суммы квадратов отклонений фактических уровней от сглаженных, т. е. полученных в результате аналитического выравнивания:

$$S = \sum_{t=1}^n [y_t - f(t)]^2 \rightarrow \min \quad (3.10)$$

<sup>4</sup> Во избежание разночтений в тексте все характеристики моделей будут приводиться с той точностью, которую обеспечивает вычислительный алгоритм. В реальных задачах обычно проводится округление.

Математический аппарат метода наименьших квадратов описан в большинстве работ по математической статистике, поэтому нет необходимости подробно на нем останавливаться.

Далее возвращаемся к анализу и нажимаем на кнопку *Analysis of Variance* (дисперсионный анализ) на той же закладке *Quick* (см. рис. 3.16).

Опишем содержание появившейся таблицы (рис. 3.18).

Model is: $v_1 = a_0 + a_1 \cdot v_2$ (Spreadsheet in Ivanov)					
Dep. Var. : ВГ2001-2010					
Effect	1 Sum of Squares	2 DF	3 Mean Squares	4 F-value	5 p-value
Regression	685428780	2,00000	342714390	548,7536	0,000000
Residual	4996259	8,00000	624532		
Total	690425039	10,00000			
Corrected Total	68561733	9,00000			
Regression vs. Corrected Total	685428780	2,00000	342714390	44,9876	0,000021

Рис. 3.18. Результаты дисперсионного анализа линейной модели тренда

В верхней, заголовочной строке таблицы выдаются пять оценок: *Sum of Squares* – сумма квадратов; *df* – число степеней свободы; *Mean Squares* – средний квадрат; *F-value* – расчетное значение *F*-критерия; *p-value* – расчетный уровень значимости *F*-критерия (данный критерий не является критерием Фишера).

В левом столбце указывается источник вариации:

*Regression* – вариация, объясненная уравнением тренда; *Residual* – вариация остатков – отклонений фактических значений от выровненных (необъясненная вариация); *Total* – общая вариация переменной; *Corrected Total* – скорректированная общая вариация переменной; *Regression vs. Corrected Total* – соотношение вариации по тренду и общей вариации.

На пересечении столбцов и строк получаем однозначно определенные показатели, расчетные формулы для которых представлены в табл. 3.2,

Расчет показателей вариации трендовых моделей

Source	df	Sum of Squares	Mean squares	F-value
Regression	$m+1$	$SSR = \sum_{t=1}^n \hat{y}_t^2$	$MSR = \frac{SSR}{m+1}$	$F = \frac{MSR}{MSE}$
Residual	$n-m-1$	$SSE = \sum_{t=1}^n (y_t - \hat{y}_t)^2$	$MSE = \frac{SSE}{n-m-1}$	
Total	$n$	$SST = \sum_{t=1}^n y_t^2$		
Corrected Total	$n-1$	$SSCT = \sum_{t=1}^n (y_t - \bar{y})^2$		
Regression vs. Corrected Total	$m$	$SSR$	$MSR$	$F_{adj} = \frac{MSR}{MSCT}$

где  $\hat{y}_t$  – выровненные (ожидаемые) значения уровней динамического ряда;  $y_t$  – фактические значения уровней динамического ряда;  $\bar{y}$  – среднее значение уровней динамического ряда; ;  $n$  – число уровней ряда;  $m$  – число параметров уравнения тренда при факторах;  $(n-m-1)$  – число степеней свободы остаточной дисперсии (*df residual*).

SSR (Regression Sum of Squares) – сумма квадратов теоретических (прогнозных) значений уровней ряда; SSE (Residual Sum of Squares) – сумма квадратов отклонений теоретических от фактических значений  $y$  (для расчета остаточной, необъясненной дисперсии); SST (Total Sum of Squares) – сумма первой и второй строк (сумма квадратов фактических значений уровней ряда); SSCT (Corrected Total Sum of Squares) – сумма квадратов отклонений фактических значений  $y$  от средней величины (для расчета общей дисперсии); Regression vs. Corrected Total Sum of Squares – повторение первой строчки; MSE (Residual Mean Squares) – остаточная, необъясненная дисперсия; MSCT (Mean Squares Corrected Total) – скорректированная общая дисперсия (в таб-

$$\text{лице отсутствует)} \quad MSCT = \frac{\sum_{t=1}^n (y_t - \bar{y})^2}{n-1}.$$

Далее опять же на закладке Quick (см. рис. 3.16) нажимаем кнопку Predicted values, Residuals, etc. После ее нажатия система строит таблицу, состоящую из трех столбцов (рис. 3.19).

Observed – наблюдаемые значения (то есть уровни исходного динамического ряда);

Predicted – ожидаемые значения (полученные по уравнению тренда для данных моментов времени);

Residuals – остатки (разность между фактическими и ожидаемыми значениями).

Model is: v1=a0+a1*v2 (Spreadsheet)			
Dep. Var. : ВГ2001-2010			
	Observed	Predicted	Residuals
2001	4191,00	3935,84	255,16
2002	5051,31	4813,62	237,69
2003	5678,50	5691,39	-12,89
2004	6557,10	6569,17	-12,07
2005	6784,70	7446,94	-662,24
2006	7752,80	8324,72	-571,92
2007	9369,00	9202,49	166,51
2008	11313,70	10080,27	1233,43
2009	9555,20	10958,05	-1402,85
2010	12605,00	11835,82	769,18

Рис. 3.19. Таблица наблюдаемых, прогнозных значений и остатков для линейной модели тренда

Далее целесообразно построить графическое изображение, на котором линия линейного тренда будет наложена на исходный динамический ряд – это позволит визуально оценить степень соответствия.

Для этого на рабочем листе (см. рис. 3.19) выделяем столбцы Observed и Predicted, щелкаем по ним правой кнопкой мыши и выбираем функции Graphs of Block Data/ Line Plot: Entire Columns. Постро-



енному графику лучше всего присвоить название и подписать легенду, с тем, чтобы далее легко его идентифицировать (рис. 3.20).

**ВНИМАНИЕ!!!** Полученные графики и таблицы целесообразно переименовывать в дереве рабочей книги из-за их большого количества и во избежание путаницы.

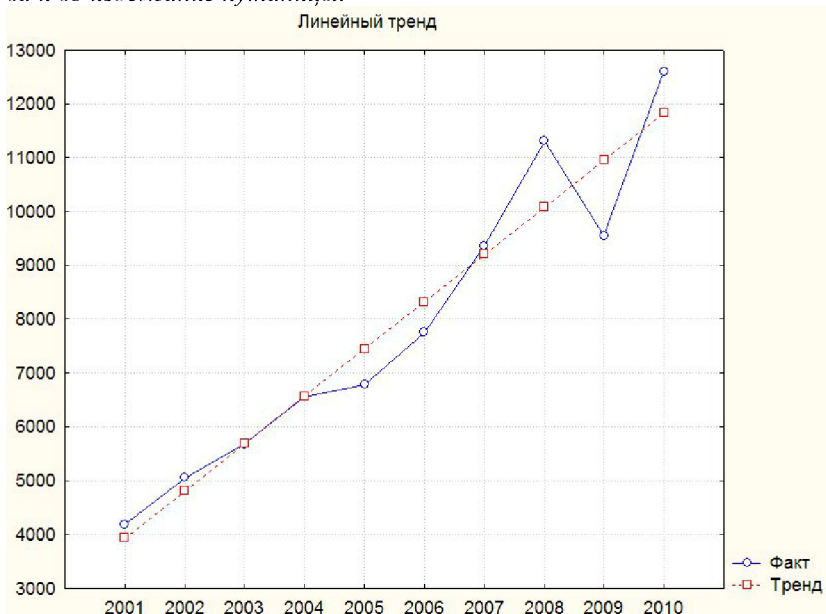


Рис. 3.20. Исходный динамический ряд и линейный тренд

Итак, мы получили все необходимые результаты расчета параметров тренда, выраженного линейной моделью, а также построили график данного ряда, совмещенный с линией тренда.

На практике чаще всего строят целый ряд функций, описывающих тренд, а затем выбирают лучшую на основе того или иного формального критерия. В курсовом проекте предлагается построить несколько уравнений тренда, исходя из характера анализируемого ряда. В качестве примера приведем варианты построения параболической и логарифмической моделей на этапах ввода расчетной формулы в систему, получения уравнения тренда и итогового графика (рис. 3.21 – 3.26). Отметим, что для попадания в меню ввода формулы достаточно

нажать на кнопку Cancel текущего анализа (см. рис. 3.16).

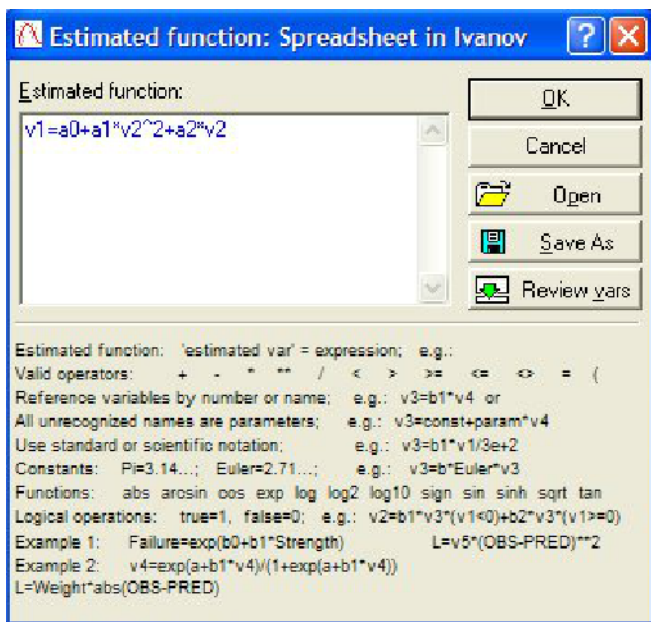


Рис. 3.21. Окно процедуры задания уравнения параболической модели тренда

Model is: v1=a0+a1*v2^2+a2*v2 (Spreadsheet in Ivanov)						
Dep. Var. : БГ2001-2010						
Level of confidence: 95.0% ( alpha=0.050)						
	Estimate	Standard error	t-value df = 7	p-level	Lo. Conf Limit	Up. Conf Limit
a0	3647,107	955,2770	3,817852	0,006561	1388,235	5905,978
a1	26,775	35,3467	0,757488	0,473480	-56,807	110,356
a2	583,254	398,9640	1,461922	0,187154	-360,146	1526,654

Рис. 3.22. Результаты расчета параметров параболической модели тренда

$$\hat{y}_t = 3647,107 + 26,775 \cdot t^2 + 583,254 \cdot t$$



Рис. 3.23. Исходный динамический ряд и параболическая модель тренда

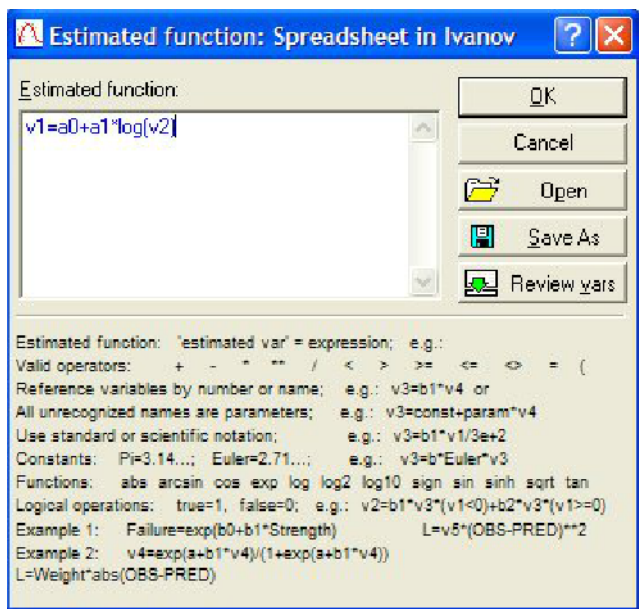


Рис. 3.24. Окно процедуры задания уравнения логарифмической модели тренда

Model is: $v1=a0+a1*\log(v2)$ (Spreadsheet in Ivanov)						
Dep. Var. : ВГ2001-2010						
Level of confidence: 95.0% ( alpha=0.050)						
	Estimate	Standard error	t-value df = 8	p-level	Lo. Conf Limit	Up. Conf Limit
a0	2792,807	985,1602	2,834876	0,021985	521,024	5064,591
a1	3371,878	592,4577	5,691339	0,000459	2005,668	4738,088

Рис. 3.25. Результаты расчета параметров логарифмической модели тренда

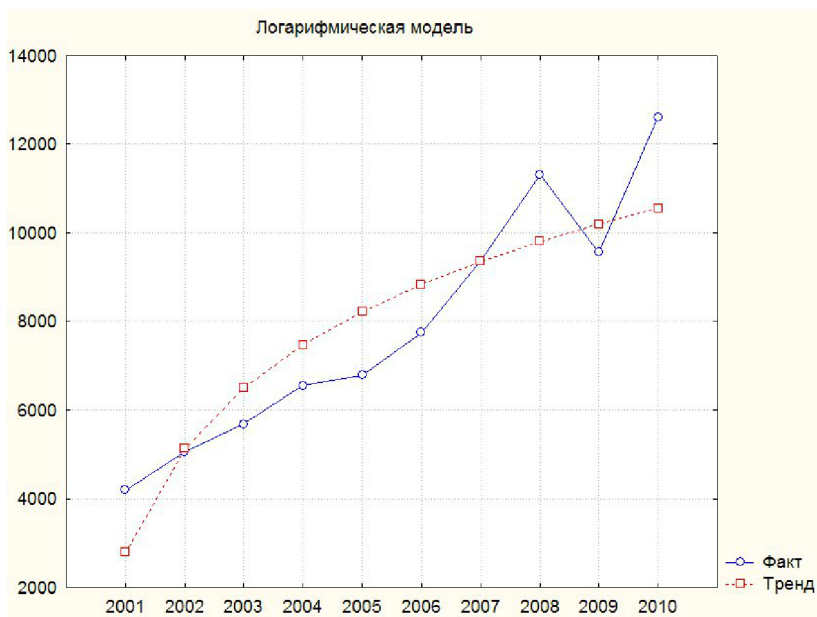


Рис. 3.26. Исходный динамический ряд и логарифмическая модель тренда

### 3.4. Выбор тренда

Как уже отмечалось, проблема выбора формы кривой – одна из основных проблем, с которой сталкиваются при построении модели тренда. Решение этой проблемы во многом определяет результаты

экстраполяции тренда. В большинстве специализированных программ для выбора лучшего уравнения тренда предоставляется возможность воспользоваться несколькими критериями, приведем некоторые из них:

1. Минимальное значение среднеквадратической ошибки тренда:

$$S = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n - m - 1}}, \quad (3.11)$$

где  $y_t$  – фактические значения уровней ряда динамики;  $\hat{y}_t$  – значения уровней ряда, определенные по уравнению тренда;  $n$  – число уровней ряда;  $m$  – число параметров при факторах в уравнении тренда.

2. Минимальное значение остаточной дисперсии:

$$\sigma_{ост}^2 = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n - m - 1}. \quad (3.12)$$

3. Минимальное значение средней ошибки аппроксимации (MAPE – Mean Absolute Percentage Error):

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{n}. \quad (3.13)$$

4. Минимальное значение среднего абсолютного отклонения (MASD – Mean Absolute Deviation):

$$MASD = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n}. \quad (3.14)$$

5. Максимальное значение коэффициента детерминации

$$R^2 = \frac{\sigma_{обц}^2 - \sigma_{ост}^2}{\sigma_{обц}^2}, \quad (3.15)$$

где  $\sigma_{\text{общ}}^2$  – общая дисперсия;  $\sigma_{\text{ост}}^2$  – остаточная дисперсия.

6. Максимальное значение F- критерия Фишера:

$$F = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{m} : \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m - 1} . \quad (3.16)$$

В данном пособии для идентификации тренда используется формальный метод, который основывается на использовании численного критерия. В качестве такого критерия рассматривается максимальный коэффициент детерминации, который показывает, какая доля общей дисперсии результативного признака обусловлена вариацией признака – фактора или, по-другому, долю объясненной дисперсии признака результата.

Напоминаем, что значение коэффициента детерминации необходимо было выписать (см. раздел 3.3.). Если этого не было сделано, то нужно рассчитать его значение по формуле (3.15).

Далее необходимо проанализировать построенные трендовые модели с точки зрения их адекватности реальным тенденциям исследуемого временного ряда через оценку надежности полученных уравнений по F-критерию Фишера и параметров уравнений по t-критерию Стьюдента.

Поскольку F-критерий основан на соотношении факторной и остаточной дисперсий, то вполне логично его использование для оценки качества модели. Если факторная дисперсия значительно больше остаточной, это означает, что уравнение тренда объясняет существенную часть вариации уровней ряда. Статистическая значимость уравнения одновременно означает статистическую значимость коэффициента детерминации.

Если  $F_{\text{расч.}} \geq F_{\text{табл.}}$ , то делается вывод о статистической значимости уравнения в целом.

Рассмотрим оценку значимости уравнения на примере линейного тренда. Согласно таблице дисперсионного анализа (см. рис. 3.18)  $F_{\text{расч.}} = 44,9876$ . Для определения теоретического значения F-

критерия можно воспользоваться встроенным вероятностным калькулятором STATISTICA. Для этого запускаем процедуру Statistics/Probability Calculator/Distributions (рис. 3.27). В появившемся окне в левом столбце выбираем распределение Фишера  $F(Fisher)$ , ставим метку в поле (1-Cumulative p), далее в поле  $p$  (теоретический уровень значимости) заносим 0,05 (поскольку установленная вероятность равна 95%), в поле  $df1$  заносим число степеней свободы факторной дисперсии (равно числу параметров трендового уравнения, для линейного – 2), в поле  $df2$  заносим число степеней свободы остаточной дисперсии (число уровней ряда минус число параметров уравнения, в нашем случае – 8) и нажимаем кнопку Compute. В поле  $F$  появляется теоретического значения  $F$ -критерия (в нашем случае – 4,458970) (рис. 3.28). Отметим, что число степеней свободы для каждого уравнения можно взять из таблицы дисперсионного анализа (см. рис. 3.18). Таким образом, линейную модель тренда следует считать статистически значимой. То же самое необходимо проделать и для других моделей.

Также определить теоретическое значение  $F$ -критерия можно с помощью таблиц с  $F$ -распределением, представленных в Приложении 4.

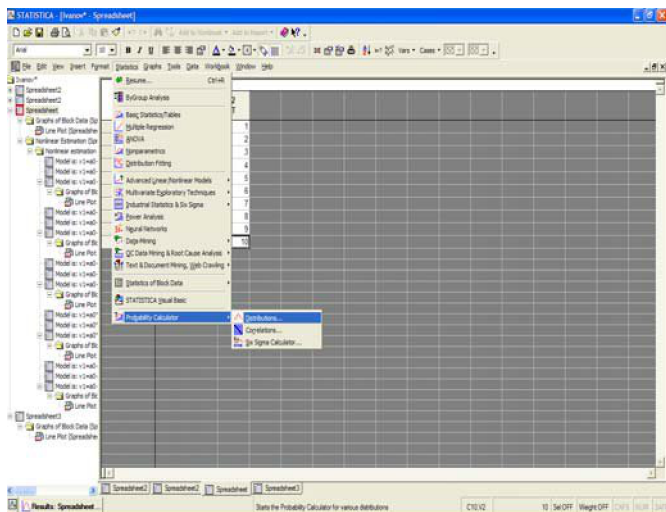


Рис. 3.27. Запуск процедуры Statistics/Probability Calculator/Distributions

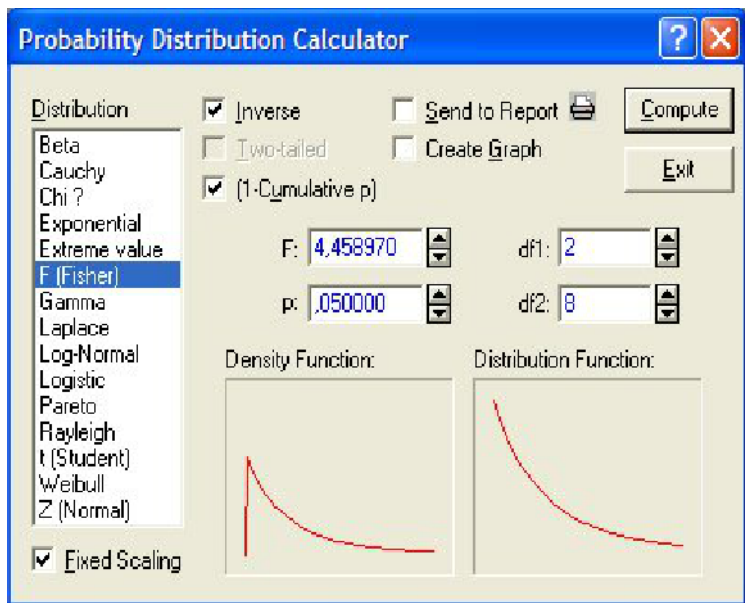


Рис. 3.28. Вероятностный калькулятор с расчетом критерия Фишера

Оценка статистической значимости параметров модели означает проверку нулевых гипотез о равенстве параметров генеральной совокупности нулю, т.е.:

$$H_0: A_0 = 0, \quad H_0: A_1 = 0.$$

Проверка производится с использованием  $t$ -статистики, которая в этом случае представляет собой отношение значения параметра к его стандартной (среднеквадратической) ошибке  $S$ :

$$t_{a_0} = \frac{a_0 - A_0}{S_{a_0}} \quad \text{и} \quad t_{a_1} = \frac{a_1 - A_1}{S_{a_1}}, \quad (3.17)$$

поскольку  $A_0 = 0$  и  $A_1 = 0$ , то



$$t_{a_0} = \frac{a_0}{S_{a_0}}, \quad t_{a_1} = \frac{a_1}{S_{a_1}}, \quad (3.18)$$

где  $S_{a_0}$  – стандартная ошибка параметра  $a_0$ :  $S_{a_0} = \frac{\sigma_y \sqrt{1-r^2}}{\sqrt{n-2}}$ ;  $S_{a_1}$  – стандартная ошибка параметра  $a_1$ :  $S_{a_1} = \frac{\sigma_y \sqrt{1-r^2}}{\sigma_x \sqrt{n-2}}$ .

Фактические значения  $t$ -критерия сравниваются с табличными (с учетом уровня значимости  $\alpha$  и числа степеней свободы). Параметры признаются статистически значимыми, т.е. сформированными под воздействием неслучайных факторов, если  $t_{\text{факт}} \geq t_{\text{табл}}$ .

Фактические значения  $t$ -критерия можно взять из таблицы расчета параметров уравнения тренда (см. рис. 3.17) в соответствующем столбце, там же находятся и значения числа степеней свободы (df). Для получения теоретических значений  $t$ -критерия опять воспользуемся встроенным вероятностным калькулятором, однако теперь в столбце слева выберем распределение Стьюдента  $t$  (Student). Ставим метку в полях (1-Cumulative p) и Two-tailed, далее в поле p (теоретический уровень значимости) заносим 0,05, в поле df заносим число степеней свободы и нажимаем кнопку Compute. В поле t появляется теоретического значения  $t$ -критерия (в нашем случае – 2,306004) (рис. 3.29). Поскольку  $t_{\text{факт}}$  для параметров линейного уравнения соответственно равны  $t_{a_0} = 5,66455$  и  $t_{a_1} = 10,08865$ , то их следует признать статистически значимыми.

То же самое следует определить и для других трендовых моделей.

Отметим, что для нахождения теоретических значений  $t$ -статистики можно воспользоваться соответствующими таблицами (Приложение 3), которые, как правило, присутствуют во всех учебниках и пособиях по статистике.

Результаты оценки уравнения могут быть разными. Возможен вариант, когда уравнение в целом статистически значимо, а некоторые параметры уравнения незначимы. Это означает, что описанная зависимость может служить основой для принятия некоторых управленче-

ских решений, но полученное уравнение тренда нельзя использовать для прогнозирования. Уравнение признается моделью и может быть использовано в целях прогнозирования, если статистически значимы все параметры и уравнение в целом.

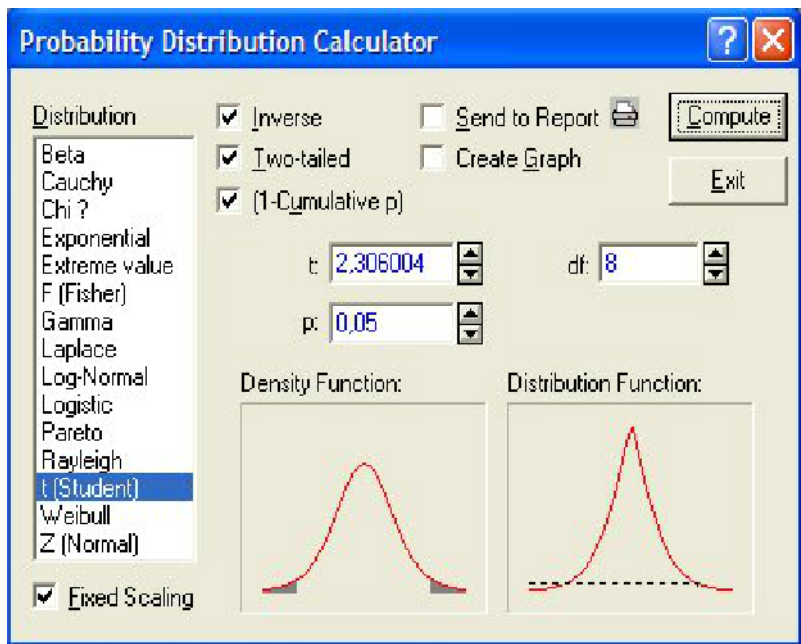


Рис. 3.29. Вероятностный калькулятор с расчетом критерия Стьюдента

Система дает подсказки пользователю:

- 1) красным цветом выделяются строки, соответствующие статистически значимым параметрам уравнения тренда;
- 2) таким же образом выделяется значимость  $F$ -критерия в строке Regression дисперсионного анализа.

После анализа значимости уравнений тренда и параметров уравнений, а также выбора критерия сравнения (коэффициент детерминации), рекомендуется составить следующую таблицу (табл. 3.3).

Таблица 3.3

## Итоговые характеристики построенных уравнений тренда

№	Модель	Уравнение	$R^2$	Значимость уравнения	Значимость параметров уравнения
1	Линейная	$\hat{y}_t = 3058,063 + 877,776 \cdot t$	0,927	+	+
2	Полином 2-ой степени	$\hat{y}_t = 3647,107 + 26,775 \cdot t^2 - 283,254 \cdot t$	0,932	+	-
3	Логарифмическая	$\hat{y}_t = 2792,87 + 3371,878 \cdot \ln t$	0,802	+	+
4	Степенная	$\hat{y}_t = 3157,897 \cdot t^{0,559}$	0,893	+	+
5	Полином 3-ей степени	$\hat{y}_t = 3741,849 - 1,104 \cdot t^3 + 44,994 \cdot t^2 + 499,222 \cdot t$	0,933	+	-

Сопоставив значения коэффициентов детерминации для различных типов кривых можно сделать вывод о том, что максимальное значение  $R^2$  получено по полиномам 3-й и 2-й степени, однако анализ значимости параметров уравнения говорит о невозможности использования полиномов 2-й и 3-й степени для прогнозирования. Исходя из этого, рассматривать стоит только три модели, которые имеют значимые оценки уравнения и параметров. Наибольший коэффициент детерминации из оставшихся моделей имеет линейная.

### 3.5. Оценка автокорреляции в остатках уравнения тренда

Важнейшим элементом оценки качества выбранной модели является анализ автокорреляции в остатках, т.е. в отклонениях исходных значений динамического ряда от рассчитанных по уравнению тренда. Если аппроксимация удовлетворительная, то случайные составляющие – отклонения от тренда  $\varepsilon = y_t - \hat{y}_t$  – в своей последовательности должны быть лишены автокорреляции, то есть в остатках должна отсутствовать тенденция.

Рассчитаем коэффициенты автокорреляции в остатках выбранного (лучшего) уравнения тренда. Значения остатков можно получить

из таблицы *Predicted & Residuals Values*, столбца *Residual Value*, созданной в главе, посвященной аналитическому сглаживанию динамических рядов (см. рис. 3.19). Для этого делаем данный лист по линейной модели (в нашем случае лучшей) активным и запускаем знакомую процедуру *Statistics/Advanced Linear/Nonlinear Models/Times Series/Forecasting*, а в качестве переменной выбираем *Residuals* (рис. 3.30). Далее нажимаем кнопку *OK (transformations, autocorrelations, crosscorrelations, plots)* и переходим на закладку *Autocorr* (рис. 3.31).

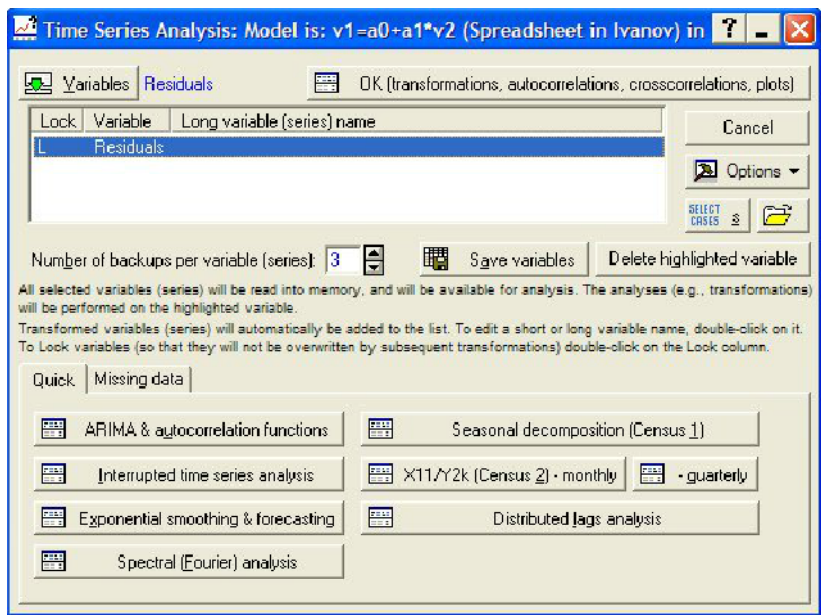


Рис. 3.30. Выбор остатков для анализа автокорреляции

В тематической области *Autocorrelations & Crosscorrelations* в левой стороне диалогового окна расположены три кнопки: *Autocorrelations*, *Partial autocorrelations*, *Crosscorrelations*, при нажатии на которые STATISTICA выводит на экран табличные и графические результаты расчета соответственно автокорреляций, частных автокорреляций и кросс-корреляций с величиной лага от 1 до числа, указанных в поле *Number of lags*.

Установим число лагов равное 3. Значение в поле *p-level for highlighting*, определяющее уровень значимости оставим без изменений ( $\alpha = 0,05$ ). Далее убираем метку с поля *White noise standard errors*. Если метка в этом поле поставлена, то STATISTICA изменяет стандартный алгоритм расчета стандартных ошибок на модифицированный вариант, основанный на предположении о том, что все автокорреляции равны нулю. Далее нажимаем кнопку *Autocorrelations*.

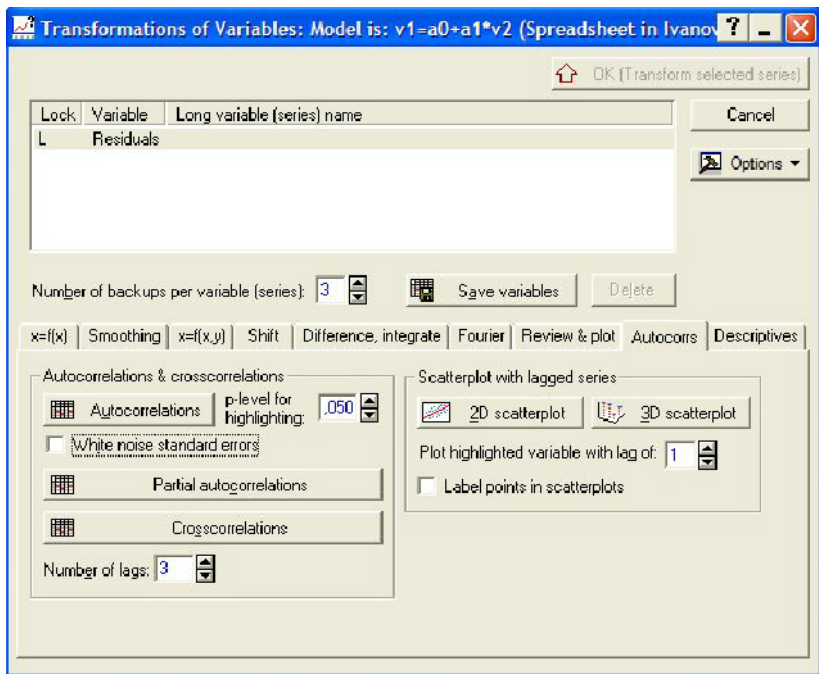


Рис. 3.31. Закладка *Autocorr* процедуры *Transformations of Variables*

В дереве рабочей книги появятся график и таблица с одинаковым названием *Autocorrelation Functions*. Таблица (рис. 3.32) содержит расчетные значения коэффициентов автокорреляции (столбец *Autocorrelations*), стандартных ошибок коэффициентов автокорреляции (*Standard Errors*), значения критерия *Box & Ljung* и его расчетного уровня значимости *P*. При этом система также дает подсказки пользователю: если строка выделена красным цветом, то коэффициент кор-

реляции с данным лагом является статистически значимым.

На графике (рис. 3.33) представлена автокорреляционная функция – последовательность значений коэффициентов автокорреляции с учетом разной величины лага. Программа представляет и значения стандартных ошибок коэффициентов автокорреляции (*S.E.*), что позволяет оценить статистическую значимость коэффициентов с помо-

щью *t*-статистики:  $t_{расч} = \frac{Corr}{S.E.}$ . Горизонтальные прямоугольники

обозначают коэффициенты автокорреляции. Графическое представление рассчитанных коэффициентов автокорреляции наглядно демонстрирует, что они статистически незначимы, поскольку значения ни одного из них не выходят на границы доверительных интервалов, обозначенных на графике красной пунктирной линией.

Autocorrelation Function (Model is: v1=a0+ Residuals)				
Lag	Auto-Corr.	Std.Err.	Box & Ljung Q	p
1	-0.451282	0.316228	2.715408	0.099393
2	-0.018266	0.375141	2.720413	0.256623
3	-0.008320	0.375230	2.721600	0.436581

Рис. 3.32. Таблица коэффициентов автокорреляции

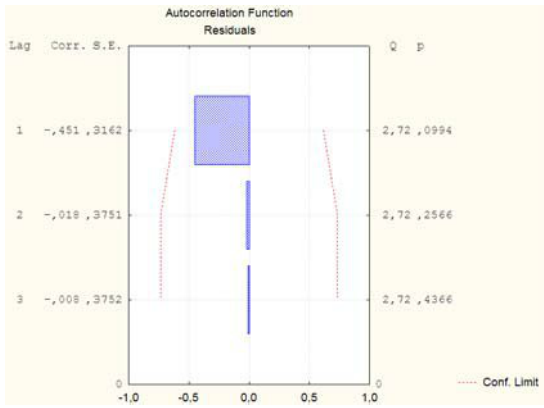


Рис. 3.33. Автокорреляционная функция остатков

Оценка значимости также может быть произведена с использованием теоретических значений самого коэффициента корреляции  $|r_{факт}| \geq r_{теор}$ . Для того, чтобы получить теоретические значения коэффициентов корреляции воспользуемся вероятностным калькулятором: Statistics/Probability Calculator/Correlations... (рис. 3.34). В появившемся окне (рис. 3.35) ставим метку в поле Two-tailed (двусторонний критерий), ставим метку в поле Compute r from p (посчитать коэффициент корреляции исходя из уровня значимости), в поле N заносим число уровней ряда (в нашем случае 10), в поле p заносим уровень значимости ( $\alpha = 0,05$ ), далее нажимаем на кнопку Compute, и в поле r появляется теоретическое значение коэффициента корреляции. Сопоставив его с расчетными значениями на трех лагах, делаем вывод о незначимости коэффициентов корреляции и отсутствии автокорреляции в остатках.

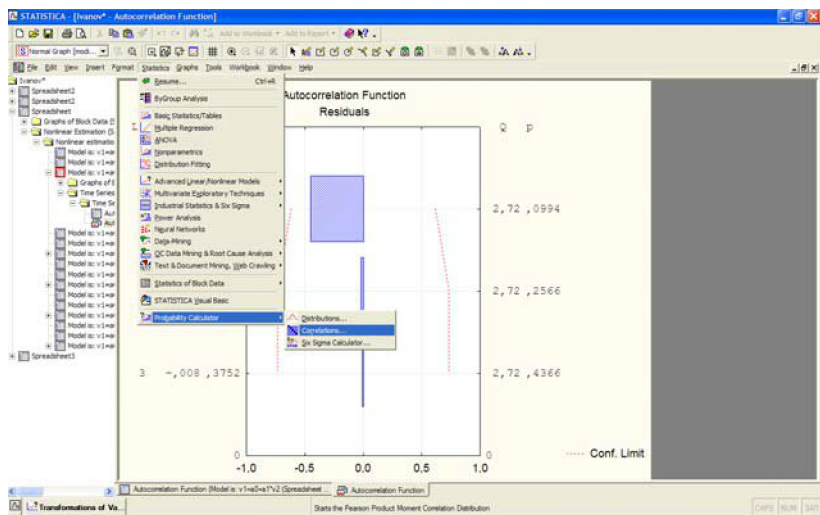


Рис. 3.34. Запуск вероятностного калькулятора для оценки значимости автокорреляции

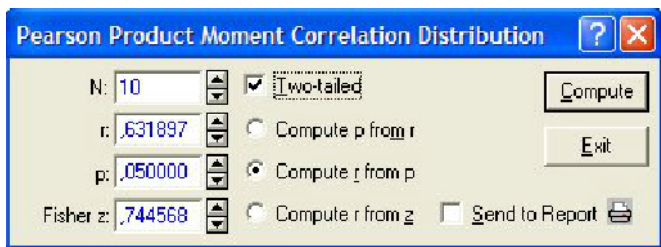


Рис. 3.35. Оценка теоретических значений коэффициентов корреляции Пирсона

Статистическая значимость коэффициентов автокорреляции проверяется также на основе  $t$ -статистики, которая рассчитывается как модуль отношения величины коэффициента автокорреляции (*Auto-Corr.*) к его стандартной ошибке (*Std.Err.*). Если  $t_{\text{факт}} \geq t_{\text{теор}}$  (распределение Стьюдента), величина коэффициента статистически значима, что говорит о наличии автокорреляции.

Следовательно, автокорреляция в остатках анализируемого уравнения тренда отсутствует, что свидетельствует о возможности использования его для прогнозирования.

Отметим, что если бы в линейной модели была обнаружена автокорреляция в остатках, следовало бы выбрать другую модель (со значимыми параметрами и следующим по ранжиру коэффициентом детерминации) и проанализировать автокорреляцию в остатках.



## 4. ЭКСТРАПОЛЯЦИЯ И ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ПРОГНОЗА

### 4.1. Экстраполяция на основе тренда

Один из наиболее распространенных методов прогнозирования заключается в экстраполяции, т.е. в продлении в будущее тенденции, наблюдавшейся в прошлом. Экстраполяция тенденций динамических рядов сравнительно широко применяется в практических исследованиях в силу ее простоты, возможности осуществления на основе относительно небольшого объема информации, наконец, ясности принимаемых допущений. Отсутствие необходимости в иной информации, помимо отдельно рассматриваемого динамического ряда, часто оказывается решающим аргументом при выборе этого метода прогнозирования.

При таком подходе к прогнозированию предполагается, что размер признака, характеризующего явление, формируется под воздействием множества факторов, причем не представляется возможным выделить порознь их влияние. В связи с этим ход развития связывается не с какими-либо конкретными факторами, а с течением времени.

Экстраполяция базируется на следующих допущениях:

- 1) развитие явления может быть с достаточным основанием охарактеризовано плавной (эволюторной) траекторией – трендом;
- 2) общие условия, определявшие тенденцию развития в прошлом, не претерпят существенных изменений в будущем.

Таким образом, экстраполяция дает описание некоторого общего будущего развития объекта прогнозирования. Причем, если развитие в прошлом носило перманентно скачкообразный характер, то при достаточно продолжительном периоде наблюдений скачки оказываются «зафиксированными» в самом тренде, и последний опять-таки можно применить в прогнозировании.

Выше были сформулированы основные условия, наличие которых дает возможность осуществлять экстраполяцию тренда. В практике прогнозирования может возникнуть вопрос, а как поступить, если условия формирования тренда заметно изменяются и этого следует ожидать и в будущем? В этом случае возможны различные подходы к решению проблем. В частности, в ряде случаев тренд можно “исправить”, сокращая период наблюдения, отсекая члены ряда, сформиро-

вавшиеся при явно других условиях и искажающие новую тенденцию. Однако далеко не всегда можно провести четкую границу во времени, разделяющую новые и старые условия развития исследуемого явления.

По-видимому, самым правильным было бы рассматривать экстраполяцию не как конечный результат прогнозирования, а как некоторый отправной момент, на основе которого с привлечением дополнительной информации, не содержащейся в самом динамическом ряду, разрабатывают прогноз. Вместе с тем часто ее результат с соответствующей корректировкой или без нее рассматривается и как окончательный прогноз.

Если при анализе развития объекта прогноза есть основания принять два базовых допущения экстраполяции, о которых говорилось выше, то процесс прогнозирования заключается в подстановке соответствующей величины периода упреждения в формулу, описывающую тренд.

Проведем прогнозирование числа выездов за рубеж с целью туризма на основе экстраполяции выбранной формы тренда (линейной):

$$\hat{y}_t = 3058,063 + 877,776 \cdot t.$$

Напомним, что у анализируемой переменной 10 уровней ряда, обозначенных натуральными числами. Соответственно прогноз выездов за границу россиян в 2011 ( $t=11$ ) и в 2012 году ( $t=12$ ) составит:

$$\hat{y}_{11} = 3058,063 + 877,776 \cdot 11 = 12713,6 \text{ (тыс.чел.)},$$

$$\hat{y}_{12} = 3058,063 + 877,776 \cdot 12 = 13591,38 \text{ (тыс.чел.)}.$$

Экстраполяция дает возможность получить точечное значение прогноза, что может быть признано удовлетворительным только при наличии функциональной зависимости. Однако для экономических явлений характерна корреляционная зависимость и переменные, как правило, являются непрерывными. Следовательно, указание точечных значений прогноза, строго говоря, лишено содержания. Отсюда следует, что прогноз должен быть дан в виде интервала значений, т.е. необходимо определение доверительного интервала прогноза.

## 4.2. Доверительные интервалы прогноза

При определении прогнозных значений того или иного явления с помощью экстраполяции наибольший интерес представляет, по-видимому, не сама экстраполяция – это более или менее механический прием, а определение доверительных интервалов прогноза.

Доверительные интервалы могут быть определены двояко: формально и неформально. Что касается последнего, то это дело экспертного суждения, которое выносится при качественном осмыслении результатов прогноза, сопоставлении их с другими имеющимися у эксперта данными. При этом, естественно, эксперт должен учитывать не только степень колеблемости фактических уровней вокруг тренда в прошлом, но и возможность деформации тренда в будущем (соответственно могут быть получены различные варианты прогноза).

Формальный доверительный интервал учитывает лишь ту неопределенность, которая связана с ограниченностью числа наблюдений и соответствующей неточностью найденных оценок параметров кривой. Основной вопрос, – в какой мере в будущем сохранится найденная тенденция, – естественно, не может быть решен с помощью таких доверительных интервалов. Это дело содержательного экономического анализа и экспертной оценки.

Основное внимание в данном учебном пособии уделим оценке формальных доверительных интервалов, базирующихся на статистическом анализе.

Соответствующая погрешность имеет следующие источники:

- 1) выбор формы кривой, характеризующей тренд, содержит элемент субъективизма. Во всяком случае, часто нет твердой основы для того, чтобы утверждать, что выбранная форма кривой является единственно возможной, а тем более лучшей для экстраполяции в данных конкретных условиях;

- 2) оценивание параметров кривых (иначе говоря, оценивание тренда) производится на основе ограниченной совокупности наблюдений, каждое из которых содержит случайную компоненту. В силу этого параметрам кривой, а, следовательно, и ее положению в пространстве свойственна некоторая неопределенность;

- 3) тренд характеризует средний уровень ряда на каждый момент времени. Отдельные наблюдения, как правило, отклонялись от него в прошлом. Естественно ожидать, что подобного рода отклонения будут

происходить и в будущем.

Вполне возможны случаи, когда форма кривой, описывающей тенденцию, выбрана неправильно или когда тенденция развития в будущем может существенно измениться и не следовать тому типу кривой, который был принят при выравнивании. В последнем случае основное допущение экстраполяции не соответствует фактическому положению вещей. Найденная кривая лишь выравнивает динамический ряд и характеризует тенденцию только в пределах периода, охваченного наблюдением. Экстраполяция такого тренда неизбежно приведет к ошибочному результату, причем ошибку такого рода нельзя оценить заранее. В связи с этим можно лишь отметить то, что, по-видимому, следует ожидать рост такой погрешности (или вероятности ее возникновения) при увеличении периода упреждения.

Погрешность, связанная со вторым и третьим источниками, может быть отражена в виде доверительного интервала прогноза при принятии некоторых допущений о свойстве ряда. С помощью такого интервала точечный прогноз преобразуется в интервальный.

Интуитивно понятно, что в основу расчета доверительного интервала прогноза должен быть положен измеритель отклонений наблюдаемых значений признака от значений, полученных по уравнению тренда. Чем выше эта колеблемость, тем менее определенно положение тренда в пространстве “уровень — время” и тем шире должен быть интервал для вариантов прогноза при одной и той же степени доверия. Традиционно в качестве такого измерителя колеблемости используется среднее квадратическое (стандартное) отклонение (формула 3.11).

Полученные в ходе статистического оценивания параметры не свободны от погрешности, связанной с тем, что объем информации, на основе которой производилось оценивание, ограничен, и в некотором смысле эту информацию можно рассматривать как выборку. Во всяком случае, смещение периода наблюдения только на один шаг или добавление, или устранение членов ряда в силу того, что каждый член ряда содержит случайную компоненту, приводит к изменению численных оценок параметров. Отсюда расчетные значения  $\bar{y}_t$  несут на себе груз неопределенности, связанной с ошибками в значении параметров.

В общем виде доверительный интервал для тренда определяется как:

$$\bar{y}_i \pm t_\alpha \cdot \frac{S_y}{\sqrt{n}}, \quad (4.1)$$

где  $S_y$  – стандартная ошибка тренда;  $\hat{y}_i$  – расчетное значение уровня ряда;  $t_\alpha$  – значение коэффициента доверия, которое берется из таблицы распределения Стьюдента или таблицы нормального распределения (вид распределения зависит от размера выборки, т.е. числа уровней ряда) с учетом установленной вероятности.

В STATISTICA при расчете доверительных интервалов прогноза величину среднего квадратического отклонения  $S_y$  можно определить, воспользовавшись таблицей дисперсионного анализа (см. рис. 3.17). Рассчитанное в ячейке *Residual Mean Squares* значение соответствует подкоренному выражению в формуле (3.11) для  $S_y$ , то есть остаточной дисперсии. Остается только извлечь из него квадратный корень ( $S_y = 790,27$  тыс.чел.).

Значение коэффициента доверия  $t=2,306$  было определено при оценке статистической значимости параметров линейной модели тренда.

Таким образом, доверительный интервал прогноза на 2011 год определяется как:

$$12713,6 - 2,306 \cdot \frac{790,27}{\sqrt{10}} \leq y_{11} \leq 12713,6 + 2,306 \cdot \frac{790,27}{\sqrt{10}}$$

$$12137,31 \leq y_{11} \leq 13289,88 .$$

На 2012 год:

$$13591,38 - 2,306 \cdot \frac{790,27}{\sqrt{10}} \leq y_{12} \leq 13591,38 + 2,306 \cdot \frac{790,27}{\sqrt{10}}$$

$$13015,09 \leq y_{12} \leq 14167,66 .$$

Этот прогноз можно интерпретировать следующим образом: число выездов россиян за границу с целью туризма в 2011 году с веро-

ятностью 95% можно ожидать от 12137,31 тыс. чел. до 13289,88 тыс. чел.

### 4.3. Графическое представление результатов прогнозирования

Завершающим этапом прогнозирования является построение графических изображений, дающих представление о точности прогноза и наглядно демонстрирующих размах доверительных интервалов.

Для этого создаем новый рабочий лист уже известным способом (см. рис. 2.45-2.47) и создаем на нем четыре переменные (столбца). В первые два столбца копируем данные о фактических и ожидаемых значениях числа выездов за границу за 2001-2010 годы (их можно взять из соответствующей таблицы, см. рис. 3.19). Далее расширяем ось времени до горизонта прогнозирования (2011 и 2012) года. В столбец с прогнозными данными в пустые строки вносим точечные прогнозы по уравнению тренда, а в третий и четвертый столбцы вписываем значения с границами доверительных интервалов (рис. 4.1). Далее строим графическое изображение всех переменных (рис. 4.2).

	1 ВГ2001-2010	2 Тренд	3 Нижняя граница	4 Верхняя граница
2001	4191	3935,84		
2002	5051,31	4813,61		
2003	5678,5	5691,39		
2004	6557,1	6569,17		
2005	6784,7	7446,94		
2006	7752,8	8324,72		
2007	9369	9202,49		
2008	11313,7	10080,27		
2009	9555,2	10958,05		
2010	12605	11835,82		
2011		12713,60	12137,310	13289,880
2012		13591,38	13015,090	14167,660

Рис. 4.1. Таблица с данными прогнозирования

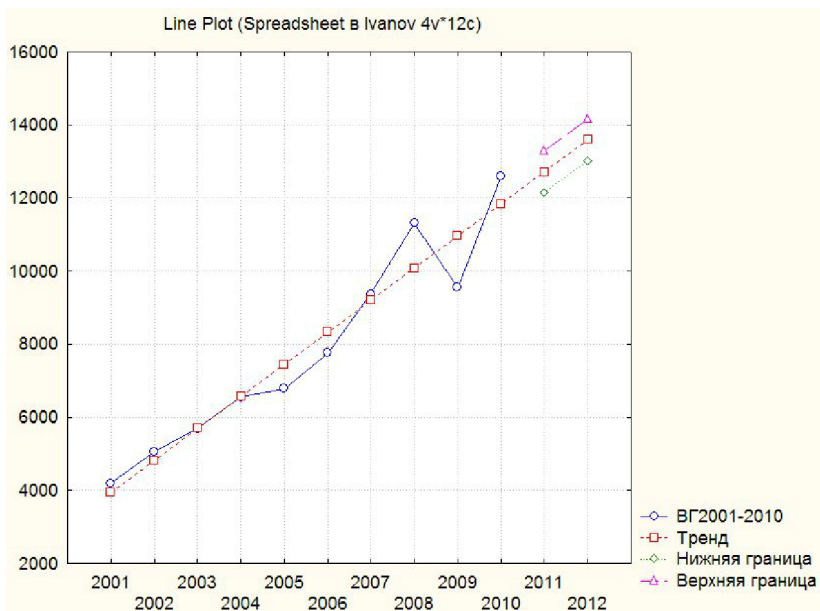


Рис. 4.2. Доверительные интервалы прогноза

Заметим, что при выполнении курсовой работы (проекта) студенты получают и фактические данные на горизонт прогнозирования, которые также можно нанести на график, вписав их в первый столбец соответствующей таблицы, что позволит сопоставить реальное развитие анализируемого явления с прогнозом и сделать выводы о точности прогноза.

Например, число выездов россиян с целью туризма в 2011 и 2012 годах составило соответственно 14052 тыс. чел. и 14816 тыс. чел., что не попадает в рассчитанные интервалы прогноза. Отчасти это обусловлено высокой точностью выбранной модели, а значит небольшими доверительными интервалами, отчасти самим видом теоретической кривой: в реальных условиях немногие социально-экономические процессы имеют линейный характер. Однако заметно, что развитие явления между 2010 и 2012 годом и прогноз визуально похожи (рис. 4.3), это говорит о том, что прогноз можно было уточнить за счет изучения смещения ряда или внешних поправочных коэффициентов.

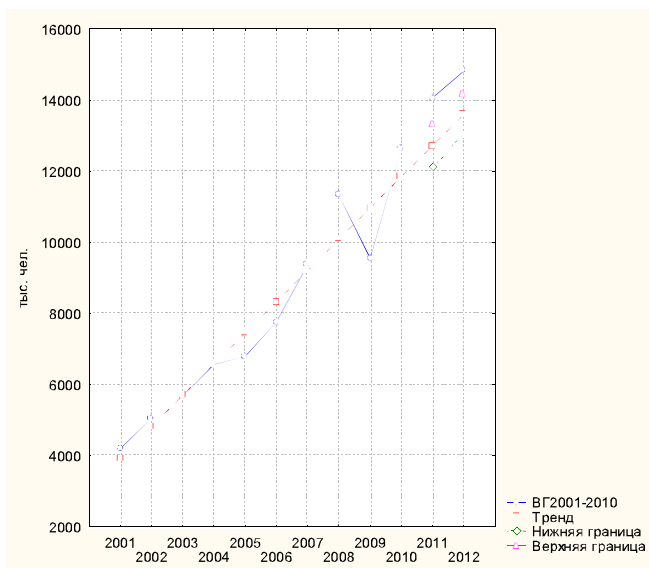


Рис. 4.3. Наблюдаемые значения признака, прогнозные значения и доверительные интервалы прогноза



## 5. АВТОКОРРЕЛЯЦИЯ В ДИНАМИЧЕСКИХ РЯДАХ. АВТОРЕГРЕССИОННЫЕ МОДЕЛИ.

Еще одним подходом к описанию основной тенденции временного ряда и прогнозированию является авторегрессионная модель. Ее построению предшествует оценка наличия автокорреляции в изучаемом ряду.

Автокорреляция – это зависимость между последовательными значениями (уровнями) временного ряда. Автокорреляция первого порядка (*first-order autocorrelation*) оценивает степень зависимости между соседними значениями временного ряда. Автокорреляция второго порядка (*second-order autocorrelation*) оценивает тесноту связи между значениями, разделенными двумя временными интервалами, и т.д. Интервал времени, разделяющий зависимые уровни динамического ряда, называется лагом (*lag*). Автокорреляционная зависимость может быть представлена как зависимость между уровнями исходного ряда:

$$y_1, y_2, y_3, \dots, y_n$$

и того же ряда, но смещенного на  $i$  периодов (моментов) времени:

$$y_{1-i}, y_{2-i}, y_{3-i}, \dots, y_{n-i}.$$

Интервал смещения  $i$  – временной лаг ( $i = 1, i = 2, i = 3$  и т. д.).

Наличие автокорреляции проверяется на основе коэффициентов автокорреляции. При этом в качестве результативного признака принимается переменная, содержащая фактические значения уровней исходного ряда динамики, а в качестве факторного признака – переменная, содержащая фактические уровни смещенного ряда. Величина временного лага определяет порядок коэффициента автокорреляции.

Для целей выполнения курсовой работы (проекта) в качестве исходных данных рекомендуется использовать исходный динамический ряд (до периодизации), если качество исходных данных позволяет сделать это. Необходимо сделать активным рабочий лист с исходным динамическим рядом и запустить процедуру, рассмотренную при анализе автокорреляции в остатках (см. рис. 3.30-3.33). Разница будет

заключаться в том, что в данном случае коэффициенты автокорреляции будут рассчитаны для уровней ряда. В итоге получаем таблицу (рис. 5.1) и график (рис. 5.2).

Autocorrelation Function (Spreadsheet2 in Ivanov)					
Выезды за границу					
Lag	Auto-Corr.	Std.Err.	Box & Ljung Q	p	
1	0,743538	0,235702	11,70739	0,000623	
2	0,632789	0,342028	20,71689	0,000032	
3	0,427806	0,401839	25,10932	0,000015	

Рис. 5.1. Таблица со значениями коэффициентов автокорреляции для динамического ряда

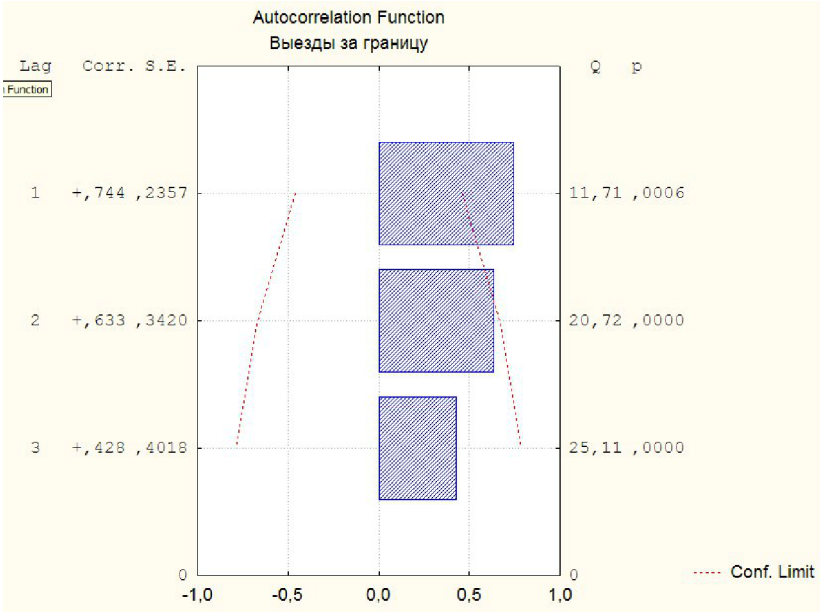


Рис. 5.2. Автокорреляционная функция уровней ряда

Анализируя полученные данные, можно сделать вывод о присутствии автокорреляции первого и второго порядка, при этом автокорреляция первого порядка выше  $r_{y_t y_{t-1}} = 0,743538$ .

После подтверждения наличия автокорреляции в динамическом ряду, может идти речь о построении авторегрессионной модели.

Авторегрессионная модель первого порядка (lag=1):

$$\hat{y}_t = a_0 + a_1 y_{t-1}.$$

Авторегрессионная модель второго порядка:

$$\hat{y}_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} \quad \text{и т.д.}$$

Поскольку в нашем примере максимальное значение имеет коэффициент автокорреляции первого порядка, построим авторегрессионную модель, сместив исходный ряд с лагом, равным 1, т.е. модель первого порядка.

Для вычисления параметров уравнения авторегрессии потребуется создать дополнительную переменную  $y_{t-1}$ . Воспользуемся тем же приемом, что и при расчете показателей изменения уровней динамического ряда. Выберем меню Statistics/Advanced Linear/Nonlinear Models/Times Series/Forecasting, в качестве переменной используем исходные данные. Нажимаем кнопку OK (transformations, autocorrelations, crosscorrelations), переходим на закладку Shift и в поле Shift (Lag) Series Forward ставим 1. Далее нажимаем OK (Transform selected series) и сохраняем полученные данные с помощью кнопки Save variables. Добавляем данные в нашу рабочую книгу с помощью кнопки Add to Workbook. В результате получаем таблицу с двумя переменными, первая – исходный ряд, а вторая – ряд, смещенный на 1 период. Длина переменных должна быть одинаковой, для чего нужно удалить первую и последнюю строки, содержащие пустые элементы (рис. 5.3).

Процедура расчета численных значений коэффициентов уравнения авторегрессии идентична определению параметров уравнений для различных трендовых моделей и осуществляется с помощью меню Statistics/Advanced Linear/Nonlinear Models/Nonlinear Estimation. При этом в качестве зависимой переменной выбирается исходный ряд, в качестве независимой – ряд, сдвинутый на 1 лаг назад, так как нас ин-

тересует зависимость текущего уровня от предыдущего. Результаты на рис. 5.4., 5.5.

Spreadsheet2 in Ivanov		
	Выезды за границу	Выезды за границу_1
1994	2100,000	1600,000
1995	2607,000	2100,000
1996	3508,000	2607,000
1997	4143,000	3508,000
1998	3330,000	4143,000
1999	2809,000	3330,000
2000	4485,000	2809,000
2001	4191,000	4485,000
2002	5051,310	4191,000
2003	5678,500	5051,310
2004	6557,100	5678,500
2005	6784,700	6557,100
2006	7752,800	6784,700
2007	9369,000	7752,800
2008	11313,700	9369,000
2009	9555,200	11313,700
2010	12605,000	9555,200

Рис. 5.3. Исходный и смещенный динамические ряды с удаленными данными для построения авторегрессионной модели

Model is: $v1=a0+a1*v2$ (Spreadsheet7 in Ivanov)						
Dep. Var. : Выезды за границу						
Level of confidence: 95.0% ( alpha=0.050)						
	Estimate	Standard error	t-value df = 15	p-level	Lo. Conf Limit	Up. Conf Limit
a0	408,5116	609,7947	0,66992	0,513093	-891,235	1708,258
a1	1,0447	0,1015	10,29489	0,000000	0,828	1,261

Рис. 5.4. Результаты расчета параметров авторегрессионной модели

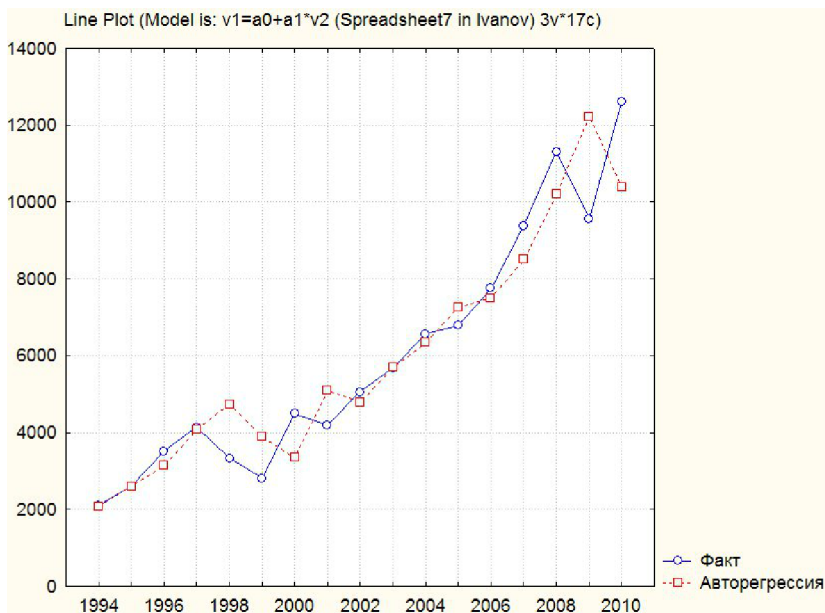


Рис. 5.5. Графическое представление динамического ряда и авторегрессионной функции

Соответственно, уравнение авторегрессии имеет вид:

$$\hat{y}_t = 408,5116 + 1,0447y_{t-1}.$$

Поскольку в нашем примере один из параметров уравнения авторегрессии статистически незначим, оно не может быть использовано для прогнозирования.

При наличии возможности прогнозирования по авторегрессионной модели все действия, включая расчет прогноза, доверительных интервалов и построение графического изображения осуществляются аналогично экстраполяции трендовых моделей, но не стоит забывать, что в качестве фактора в уравнении авторегрессии используется не период времени, а значение предшествующего уровня.

## 6. КОРРЕЛЯЦИЯ РЯДОВ ДИНАМИКИ

При изучении тенденции развития явления во времени часто возникает необходимость определить степень зависимости между динамическими рядами.

Корреляционная связь между уровнями двух динамических рядов называется кросс-корреляцией. Оценка тесноты связи в задачах исследования кросс-корреляции производится с использованием стандартного коэффициента корреляции Пирсона. Однако применение традиционных методов корреляции и регрессии к анализу зависимости временных рядов имеет определенные особенности.

Особое значение приобретает теоретический, содержательный анализ изучаемых явлений и их возможных взаимосвязей во избежание оценки «ложной корреляции». Однонаправленность трендов и высокое значение коэффициента корреляции вовсе не означает наличие причинно-следственной зависимости между рядами. Поэтому, прежде чем приступать к количественной оценке корреляционной зависимости, необходимо теоретически обосновать ее наличие.

Вторая особенность обусловлена тем, что одним из условий применения корреляционно-регрессионного анализа является независимость наблюдений. В контексте изучения временных рядов – это отсутствие связи между уровнями ряда, т.е. автокорреляции. Наличие тренда (автокорреляции) в анализируемых динамических рядах может существенно исказить оценку. Поэтому для получения адекватного результата, необходимо предварительно исключить тенденцию из анализируемых рядов.

Существует несколько способов исключения автокорреляции (тенденции). Один из них основан на переходе от корреляции уровней ряда к корреляции остатков, отклонений фактических уровней от тренда. При этом:

- определяют форму тренда и производят аналитическое выравнивание каждого из связанных рядов;
- рассчитывают отклонения фактических уровней от соответствующих выровненных уровней по каждому ряду;
- определяют численное значение коэффициента корреляции по полученным отклонениям.

Практика показывает, что часто в отклонениях от тренда сохраняется автокорреляция. Прежде чем приступить к расчету коэффициента корреляции по остаткам, необходимо проверить наличие в них автокорреляции.

Наряду с коррелированием остатков, способом обойти автокорреляцию уровней может быть метод коррелирования последовательных разностей или тех цепных показателей динамических рядов, которые являются константами их трендов. Данный подход к исключению автокорреляции вполне оправдан. На длительном временном отрезке искажение корреляции при наличии тренда может быть весьма существенным, благодаря кумулятивному эффекту. В разностях между соседними уровнями воздействие тренда незначительно, т. к. в большей мере они отображают влияние колеблемости.

Еще один прием устранения автокорреляции основан на включении переменной времени в уравнение регрессии в качестве аргумента. Математически доказано, что непосредственное введение в уравнение регрессии фактора времени устраняет автокорреляцию, аналогично использованию отклонений фактических уровней от тренда. Простота реализации этого подхода явилась причиной его широкого применения в практических исследованиях.

В рамках курсового проекта студентам будут предложены для анализа два динамических ряда. Вспомним, что в данном учебном пособии вначале были предложены два динамических ряда, содержащие данные о числе выездов россиян за границу и среднедушевых доходах по годам. Очевидно, что динамика выезда за границу зависит от уровня материального положения граждан, а, следовательно, можно рассмотреть корреляционную зависимость между динамическими рядами.

Для демонстрации методов оценки связи между временными рядами необходимо сделать активным рабочий лист, содержащий данные двух исходных динамических рядов.

При изучении связи между временными рядами следует помнить, что между изменением уровней одного ряда, как отклика на изменение уровней другого, может существовать определенный временной лаг.

Коэффициенты кросс-корреляции в STATISTICA рассчитываются точно так же, как и коэффициенты автокорреляции. Только на закладке Autocorr. выбирается кнопка Crosscorrelations. Напомним, что это осуществляется с помощью меню Statistics/Advanced Linear/Nonlinear Models/Time Series Forecasting. В качестве переменных выбирается сразу оба динамических ряда (рис. 6.1). Поле Number of lags=3 (рис. 6.2). Необходимо теоретически обосновать, какой из динамических рядов является признаком-фактором, а какой признаком-

результатом, т.е. обосновать причинно-следственную связь. В нашем случае результат – это переменная «Выезды за границу», а фактор – «Денежные доходы». На рис. 6.2. выделена результивная переменная. После нажатия кнопки *Crosscorrelations* система предложит выбрать переменную, которая является фактором в уравнении.

После нажатия кнопки *OK*, система строит таблицу с рассчитанными коэффициентами кросс-корреляции и их стандартными ошибками (рис. 6.4) и графическое изображение, интерпретация которого схожа с графическим изображением коэффициентов автокорреляции (рис. 6.5).

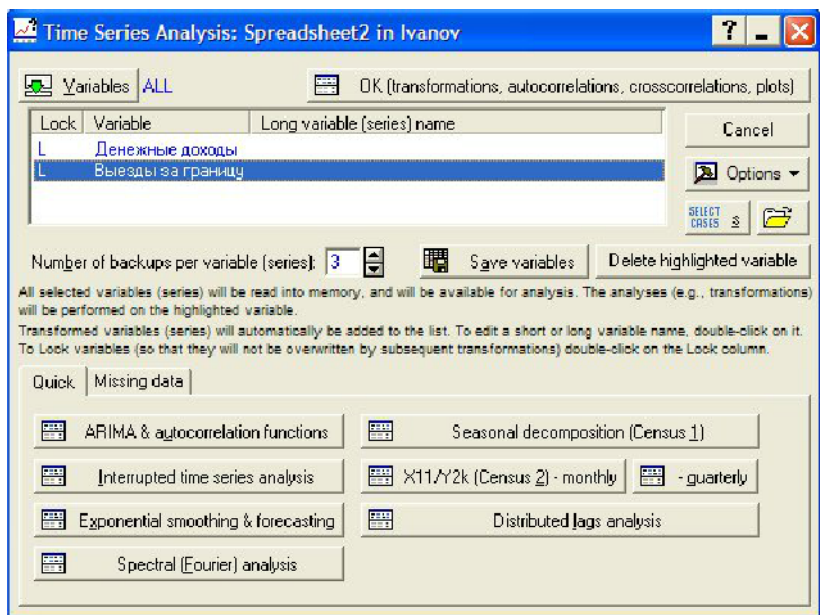


Рис. 6.1. Выбор переменных для построения коэффициентов кросс-корреляции



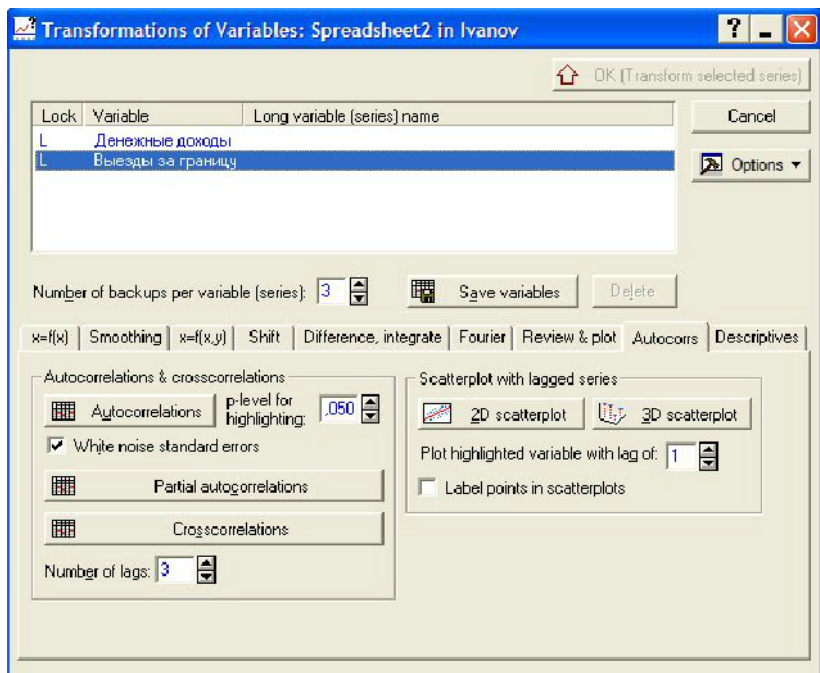


Рис. 6.2. Окно расчета коэффициентов кросс-корреляции

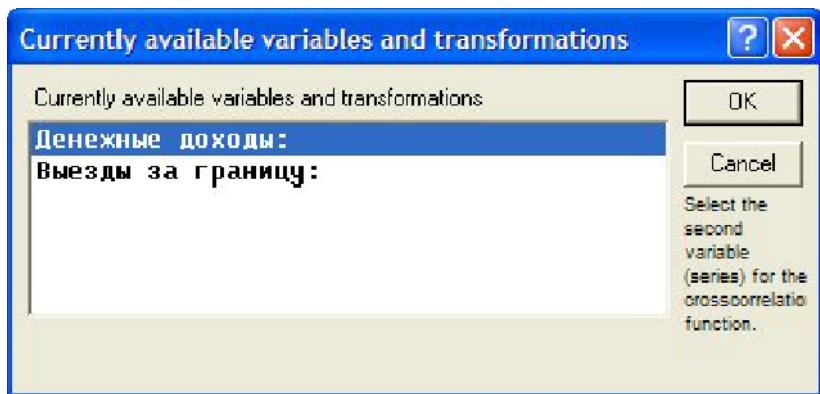


Рис. 6.3. Выбор факторной переменной

CrossCorrelation Function		
First : Выезды за гр		
Lagged: Денежные д		
Lag	Cross Corr.	Std.Err.
-3	0,487231	0,258199
-2	0,693249	0,250000
-1	0,789954	0,242536
0	0,974599	0,235702
1	0,781874	0,242536
2	0,588305	0,250000
3	0,402894	0,258199

Рис. 6.4. Таблица коэффициентов кросс-корреляции

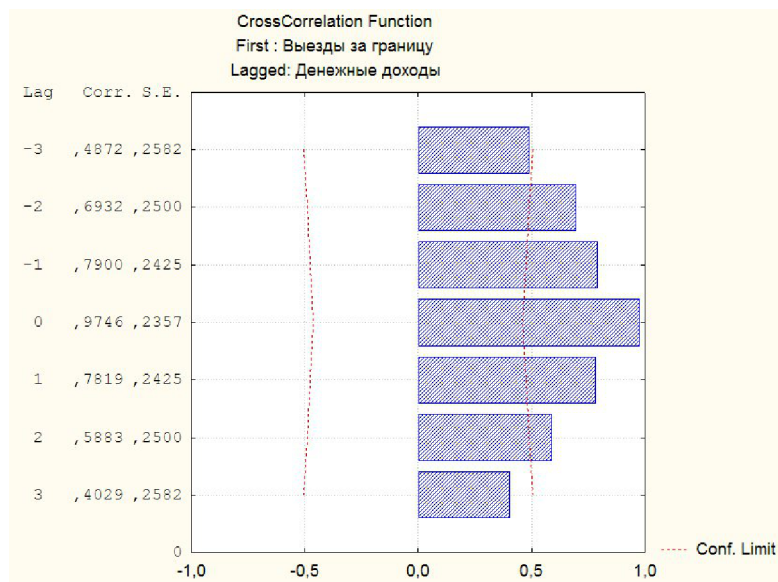


Рис. 6.5. Графическое изображение коэффициентов кросс-корреляции

На основании рассчитанных коэффициентов кросс-корреляции определяется лаг наиболее существенной взаимосвязи между динамическими рядами, то есть тот лаг, которому соответствует максимальный значимый коэффициент кросс-корреляции. В нашем случае максимальное значение достигается при  $i = 0$  и составляет  $r = 0,974599$ . Это говорит о том, что динамические ряды не нужно смещать относительно друг друга, то есть россияне тратят денежные доходы на поездки за границу сразу, а не делают накопления. При этом менее тесная, но также значимая связь наблюдается на лагах от -2 до 2.

Описанный выше прием непосредственного включения в уравнение связи фактора времени, позволяет не только оценить зависимость между рядами, но и получить модель для прогнозирования:

$$\hat{y}_t = a_0 + a_1 x_{t-i} + a_2 t,$$

где  $i$  – лаг наибольшей взаимосвязи между рядами;  $y_{t-i}$  – признак-результат (переменная «Выезды за границу») на временном интервале  $t-i$ ;  $x$  – признак-фактор (переменная «Денежные доходы») на временном интервале  $t-i$ .

На листе следует создать переменную с фактором времени (рис. 6.6), а затем уже известным способом построить факторно-временную модель. Для примера построим модель зависимости переменной «Выезды за границу» от переменной «Денежные доходы» (рис. 6.7).

При наличии сдвига необходимо создать новый рабочий лист со смещенными относительно друг друга переменными, аналогично авторегрессионной модели и так же добавить переменную времени.

	1 Денежные доходы	2 Выезды за границу	3 Т
1993	45,27	1600	1
1994	206,6	2100	2
1995	516,9	2607	3
1996	769,5	3608	4
1997	940,6	4143	5
1998	1010,2	3330	6
1999	1658,9	2809	7
2000	2281,1	4486	8
2001	2062	4191	9
2002	3947,2	6061,31	10
2003	5170,4	5678,6	11
2004	6410,3	6557,1	12
2005	8111,9	6784,7	13
2006	10196	7752,8	14
2007	12602,7	9369	15
2008	14940,6	11313,7	16
2009	16838,3	9555,2	17
2010	18552,6	12605	18

Рис. 6.6. Окно с новым рабочим листом для построения факторно-временной модели

Model is: $v_2 = a_0 + a_1 \cdot v_1 + a_2 \cdot v_3$ (Spreadsheet2 in Ivanov)						
Dep. Var. : Выезды за границу						
Level of confidence: 95.0% ( alpha=0.050)						
	Estimate	Standard error	t-value df = 15	p-level	Lo. Conf Limit	Up. Conf Limit
a0	1920,541	474,4740	4,047727	0,001052	909,2237	2931,859
a1	0,372	0,0758	4,906273	0,000190	0,2102	0,533
a2	171,749	87,8062	1,956000	0,069352	-15,4055	358,903

Рис. 6.7. Расчет параметров факторно-временной функции

В данном случае модель выглядит следующим образом:  
 $\hat{y}_t = 1920,541 + 0,372 \cdot x_t + 171,749 \cdot t$ .

При условии статистической значимости уравнения и параметров модель может быть использована для прогнозирования. В нашем случае этого не наблюдается.

Отметим, что при возможности прогнозирования все действия аналогичны прогнозированию на основе трендовых моделей.

## **ЗАКЛЮЧЕНИЕ**

В учебном пособии кратко освещаются теоретические вопросы статистического анализа и основ прогнозирования временных рядов, подробно продемонстрирована реализация задач с использованием ППП STATISTICA.

В рамках решения данных задач рассматриваются вопросы расчета и интерпретации показателей изменения уровней временных рядов, определения тенденции в рядах динамики, периодизации данных, аналитического выравнивания временных рядов, автокорреляции и корреляции рядов, а также прогнозирования на основе разных моделей временных рядов.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Федеральная служба государственной статистики. Режим доступа [<http://www.gks.ru>].

2. Куприенко Н. В. Статистические методы анализа связей. Корреляционно-регрессионный анализ: учеб. пособие / Н. В. Куприенко, О. А. Пономарева, Д. В. Тихонов. – СПб. : Изд-во Политехн. ун-та, 2009. – 116 с.

3. Куприенко Н. В. Статистика. Распределение и выборочное наблюдение в среде STATISTICA / Н. В. Куприенко, О. А. Пономарева, Д. В. Тихонов. – СПб. : Изд-во Политехн. ун-та, 2011. – 137 с.

4. Сигел Э. Практическая бизнес-статистика : Пер. с англ / Э. Сигел. – М. : Изд. дом «Вильямс», 2008. – 1056 с.

5. Статистика для менеджеров. 4-е изд.: Пер. с англ / Д. Левин и др. – М. : Изд. дом «Вильямс», 2005. – 1312 с.

6. Теория статистики. : учеб. / Под ред. Р. А. Шмойловой. – М. : Финансы и статистика, 2009. – 656 с.

7. Теория статистики: учеб. / Под ред. проф. Г. Л. Громыко. – М. : ИНФРА-М, 2013. – 480 с.

8. [www.statsoft.ru](http://www.statsoft.ru) (сайт компании StatSoft Russia – документация по ППП STATISTICA).

9. [www.exponenta.ru](http://www.exponenta.ru) (примеры решения практических задач в ППП STATISTICA).

Курсовая работа  
«Анализ временных рядов»

*Общие сведения.*

Целью работы является освоение методики и приобретение практических навыков анализа рядов динамики, выявления и описания тенденций развития тех или иных объектов или процессов, построения моделей временных рядов и прогнозирования. Курсовая выполняется на основе материалов, публикуемых в официальных статистических изданиях или иных источниках. Исходные данные могут быть предложены преподавателем или выбраны студентом самостоятельно, исходя из области его интересов.

*Требования, предъявляемые к работе.*

В каждом разделе курсовой работы должны быть кратко изложены основные теоретические положения по соответствующим вопросам. В разделах, посвященных расчету и анализу конкретных статистических характеристик и моделей, необходимо привести формулы. Основной акцент в работе следует сделать на содержательной интерпретации полученных результатов. Работа оформляется в соответствии с требованиями ГОСТа.

*Структура работы.*

*Введение.*

Во введении целесообразно раскрыть понятие временного ряда, рассмотреть цели и практическое значение ведения и исследования рядов динамики, дать характеристику исходных данных, с указанием источника информации, дать графическое представление рядов (для выполнения работы необходимы данные по двум рядам динамики, построенным по взаимосвязанным показателям).

1. Расчет и анализ показателей изменения уровней временных рядов.

В данном разделе следует рассчитать и представить в табличной форме значения абсолютных и относительных показателей изменения уровней анализируемых временных рядов. Выбрав какой-либо год, необходимо дать содержательный комментарий значений каждого показателя для этого периода. В отдельной таблице представляются и затем анализируются средние характеристики временных рядов.

2. Построение трендовых моделей и прогнозирование

В данном разделе реализуется механическое выравнивание (ме-

тодом скользящего среднего) и аналитическое выравнивание рядов, позволяющее получить модели тренда. Необходимо представить построение нескольких уравнений тренда для каждого ряда и оценку качества каждого из них, оценив статистическую значимость параметров и уравнения в целом, а также проверив наличие автокорреляции в остатках моделей (сравнительный анализ моделей рекомендуется представить в табличной форме).

По результатам оценки выбирается уравнение, отвечающее всем требованиям модели, пригодной для прогнозирования (параметры и уравнение в целом статистически значимы, автокорреляция в остатках отсутствует). Если удастся получить требуемую модель, то далее проводится прогноз на 2-3 периода с расчетом доверительных интервалов прогноза и графическим представлением результатов.

### 3. Построение авторегрессионных моделей и прогнозирование

Построение авторегрессионных моделей предваряется оценкой наличия автокорреляции в анализируемых временных рядах, для чего строятся и анализируются автокорреляционные функции. Результаты анализа позволят определить, уравнение какого порядка целесообразно построить для каждого ряда.

Далее проводится построение и оценка качества авторегрессионных моделей. Если удастся получить модели пригодные для прогнозирования, реализуется прогноз с расчетом доверительных интервалов. Результаты прогноза сопоставляются с прогнозом по трендовым моделям и фактическими данными.

### 4. Корреляция рядов динамики

В данном разделе отрабатывается методика анализа и построения моделей связи между временными рядами. Следует помнить об особенностях применения классического корреляционно-регрессионного анализа при коррелировании рядов динамики, что следует отразить в теоретических заметках по данному разделу. Далее строится и оценивается модель связи между временными рядами, рассматриваются варианты возможного ее использования.

### Заключение

В заключении следует представить основные результаты исследования и выводы.



**Образец титульного листа**

**Министерство образования и науки Российской Федерации**  
**Федеральное государственное автономное образовательное учреждение**  
**высшего образования**  
**Санкт-Петербургский государственный политехнический университет**

---

**Инженерно-экономический институт**

**Кафедра предпринимательства и коммерции**

**Курсовая работа**

по дисциплине «СТАТИСТИКА»

на тему «*Анализ временных рядов*»

Выполнил: студент \_\_\_\_\_ группы

\_\_\_\_\_  
(подпись)

\_\_\_\_\_  
(Фамилия И.О.)

Принял: \_\_\_\_\_  
(должность, ученая степень)





\_\_\_\_\_  
(подпись)

\_\_\_\_\_  
(Фамилия И.О.)

\_\_\_\_\_  
(Дата)

Санкт-Петербург  
20\_\_

Таблица критических значений  $t$  - критерия  $Student'a$ <sup>5</sup>  
 1 - односторонний критерий , 2 - двусторонний критерий

	 0 						 -c 0 		
1	.50	.20	.10	.05	.02	.01	.005	.002	.001
2	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	1.000	3.078	6.314	12.70	31.82	63.63	127.3	318.3	636.6
2	.816	1.866	2.920	4.303	6.965	9.925	14.08	22.32	31.59
3	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
6	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.711	1.415	1.895	2.365	2.998	3.499	4.020	4.785	5.408
8	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.537
11	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.689	1.333	1.740	2.110	2.576	2.898	3.222	3.646	3.965
18	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.257	3.189
22	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
$\infty$	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

<sup>5</sup> Источник: Abridged from Table 12 // Biometrika Tables for Statisticians 1962. Vol. 1 / Ed. E.S. Pearson, H.O. Hartley. London: Cambridge University Press, 1962.

Таблица значений F-распределения<sup>6</sup>,  $\alpha = 0.05$ 

	Число степеней свободы								
	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
$\infty$	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Приложение 4 (продолжение)

	Число степеней свободы									
	10	12	15	20	24	30	40	60	120	$\infty$
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
$\infty$	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

<sup>6</sup> Источник: Merrington M. , Thompson C. M. – Tables of Percentage Points of the Inverted Beta (D-Distribution// Biometrika. 1943. 33. P 73-88. Reproduced by permission of the Biometrika Trustees. Приложения 4 - 8 приведены с сохранением знаков источников.

Таблица критических значений коэффициента корреляции Пирсона

Достоверные (критические) значения $r$	
<b>df=(N-2)</b>	<b><math>p=0,05</math></b>
3	0,88
4	0,81
5	0,75
6	0,71
7	0,67
8	0,63
9	0,60
10	0,58
11	0,55
12	0,53
13	0,51
14	0,50
15	0,48
16	0,47
17	0,46
18	0,44
19	0,43
20	0,42

Значения интеграла вероятностей  $F(t) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt$

$t$	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
0,0	0000	0080	0160	0239	0319	0399	0478	0558	0638	0717
0,1	0797	0876	0955	1034	1114	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2961	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	6827	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7995	8030
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9089
1,7	9109	9127	9146	9164	9182	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9425	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9606	9615	9625	9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9755	9762	9768	9774	9780
2,3	9785	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9840	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9909	9912	9915	9917	9920	9924	9926	9927	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9950	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	99730	99739	99747	99755	99763	99771	99779	99786	99793	99800
3,1	99807	99813	99819	99825	99831	99837	99842	99847	99853	99858
3,2	99863	99867	99872	99876	99880	99884	99889	99892	99896	99900
3,3	99903	—	—	—	—	—	—	—	—	—
3,4	99933	—	—	—	—	—	—	—	—	—
3,5	99953	—	—	—	—	—	—	—	—	—
4,0	99994	—	—	—	—	—	—	—	—	—
5,0	99999	—	—	—	—	—	—	—	—	—

Куприенко Николай Владимирович  
Пономарева Ольга Алексеевна  
Тихонов Дмитрий Владимирович

СТАТИСТИКА.  
Временные ряды.  
Анализ тенденций и прогнозирование  
*Учебное пособие*

Лицензия ЛР №020593 от 07.08.97  
Налоговая льгота – Общероссийский классификатор  
продукции ОК 005 – 93, т.2; 95 3005 – учебная литература

---

Подписано в печать .03.2015. Формат 60×84/16.  
Печать офсетная. Усл. печ. л. . Уч.-изд. л. Тираж 300.  
Заказ 300 экз .

---

Отпечатано с готового оригинал-макета, предоставленного авторами,  
в типографии Издательство Политехнического университета.  
195251, Санкт-Петербург, Политехническая ул., 29