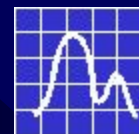


**Пример:
Анализ и прогнозирование
объема продаж сетей
автозаправочных
станций в США**

Семенов Дмитрий



StatSoft® Russia

**Рассматривались данные
по месячным объемам продаж бензина
на автозаправочных
станциях в США**



**Временной ряд взят на сайте
www.economagic.com**

**Временной ряд по продажам
рассматривался на интервале
январь 1967 года-
январь 2001 года**

**Данные в
STATISTICA**



Данные: RETAI...	
ТЕКСТОВЫЕ ЗНАЧЕНИЯ	
	1 ПРОДАЖИ
Янв-1967	1697
Фев-1967	1599
Мар-1967	1765
Апр-1967	1803
Май-1967	1891
Июн-1967	1986
Июл-1967	2009
Авг-1967	1969
Сен-1967	1893
Окт-1967	1900
Ноя-1967	1914
Дек-1967	1936
Янв-1968	1858

Визуализация данных

The image shows a software interface with a menu and a dialog box. The menu is titled 'Графика' and contains several options. The dialog box is titled '2M линейные графики' and contains various settings for a 2D line chart.

Графика Сервис Окно ?

- Быстрые статистические графики...
- Статистические 2M графики**
 - Гистограммы...
- Статистические 3M последовательные графики
 - Диаграммы рассеяния...
- Статистические XYZ графики
 - Диаграммы рассеяния с гистограммами...
- Статистические матричные графики...
- Статистические пиктографики...
- Статистические категоризованные г...

Размещение нескольких графиков
 Пустые графические окна

Пользовательские графики
Блочные статистические графики
Статистические графики пользовате...
Связи данных и графика...

122	Фев-1977
-----	----------

2M линейные графики ? X

Переменные:
Перем.: ПРОДАЖИ

Тип графика:

- Простой
- Составной
- С двойн. осью Y
- Трассировочный XY
- Агрегированный

ПОДГОНКА

Нет

- Линейная
- Логарифмическая
- Экспоненциальная
- Слайны

Другая: нет

Показать точки: Авто

Не учитывать ПД

НАБЛЮДЕНИЯ

Все наблюдения

От: 1

До: 409

АГРЕГИРОВАННЫЕ ДАННЫЕ

Индекс 5

Столбцы знач. диапазонов

OK Отмена Параметры...

График временного ряда



**Целесообразно рассматривать
динамику показателя на двух
различных временных
интервалах:**

- январь 1967 - декабрь 1978**
- январь 1979 - январь 2001**

Интерактивно выбираем рабочую область для анализа - временной интервал, на котором будет проходить дальнейшее исследование

Условия выбора наблюдений

Изменить/добавить условия выбора:

Включ., если:

Искл., если:

Заголовок (для файла условий выбора):

Состояние

ВКЛ

ВЫКЛ

Операторы: = <> < > <= >= NOT AND OR
Переменные: имена или v1, v2...
Номер наблюдения: v0
Примеры: v1 = 0 OR v2 >= 0
(v7<1 OR v9='YES') and v4<>0

OK

Отмена

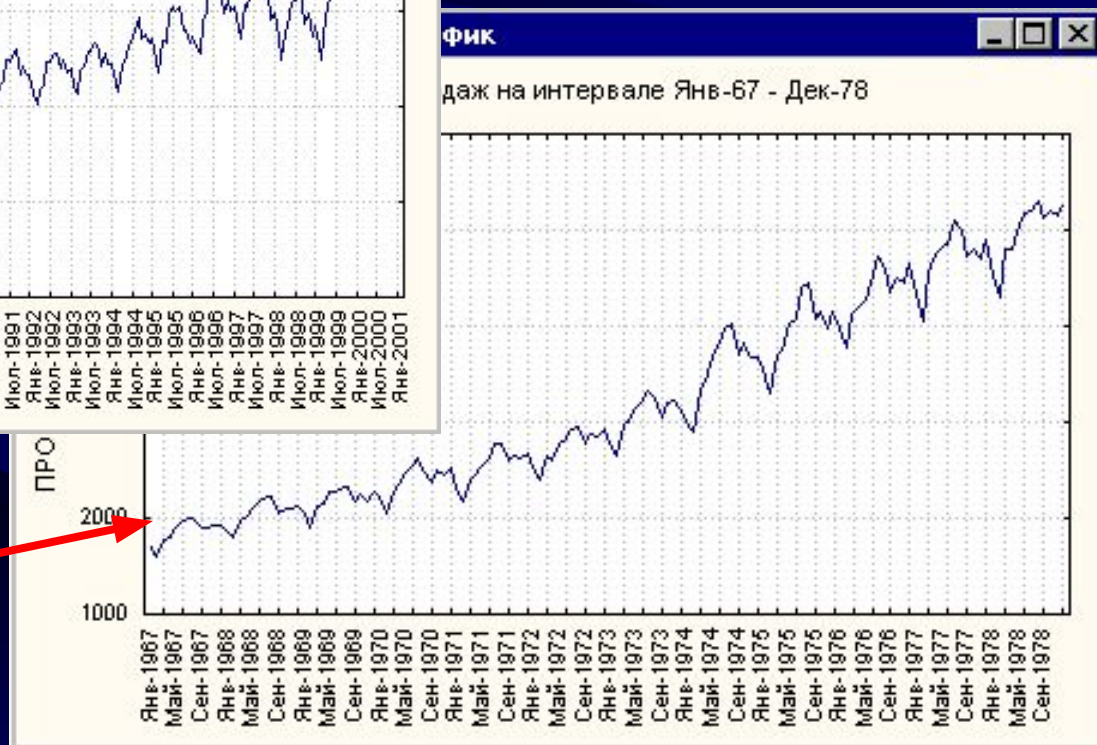
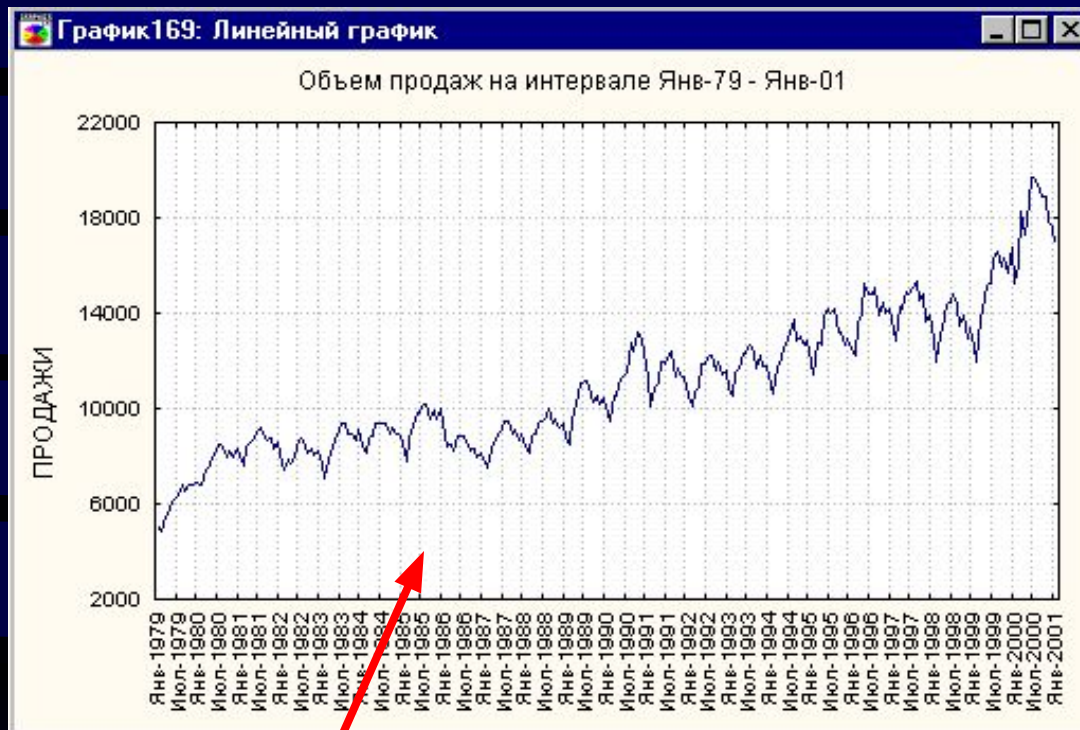
Переменные

Открыть

Сохранить...

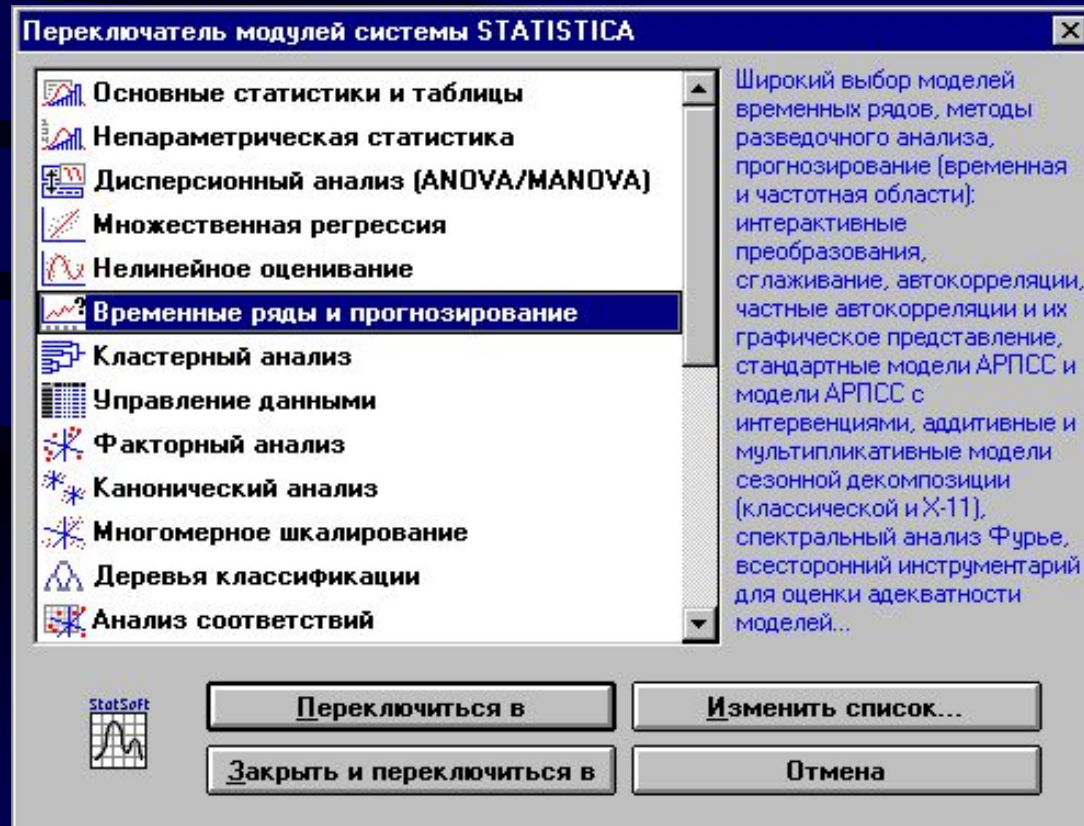
Задаем условие
на номер
наблюдения или
на значение
переменной

Динамика показателя на разных временных интервалах



Динамика различается

Применяем методы модуля “Временные ряды и прогнозирование”



Обозначим этапы исследования:

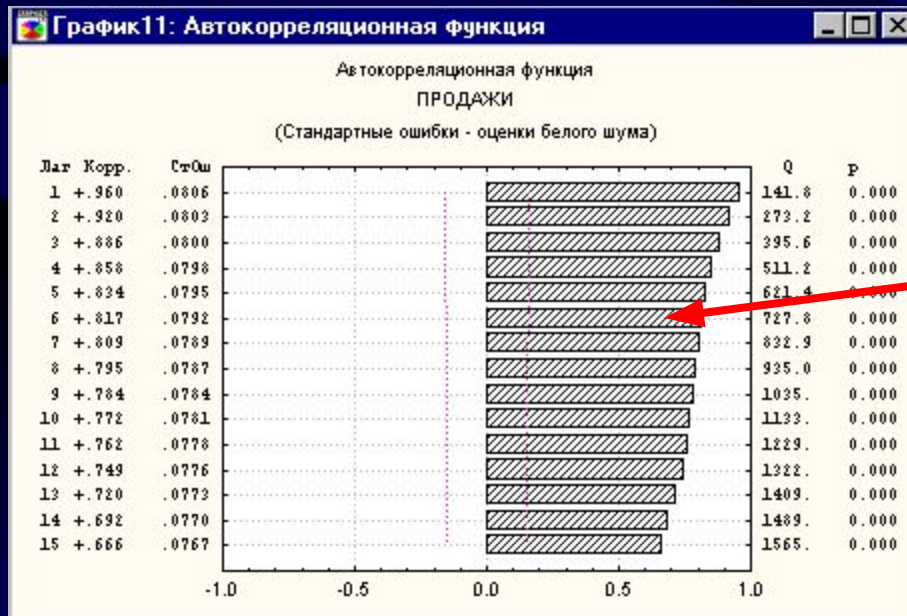
- **провести анализ динамики объемов продаж на каждом временном интервале**
- **построить модель, адекватно отражающую изменение показателя во времени**
- **сделать прогноз на основе полученной модели на 12 месяцев**

**Исключим из рассмотрения последние
12 месяцев в каждом интервале.
На этих данных будем проверять
точность прогноза,
построенного с помощью
модели**

Исследование временного ряда на первом интервале: январь 1967 - декабрь 1977

Шаг 1:

Проверка ряда на стационарность - анализ автокорреляционной функции:



Наблюдается сильная
корреляция между
соседними членами ряда,
причем значения
коэффициентов
автокорреляции
убывают очень медленно.
Отсюда следует, что
ряд **нестационарный**.

Преобразуем исходный ряд к следующему виду: $Dy(t)=y(t)-y(t-1)$

Преобразования временных рядов

Преобразовать переменную: **ПРОДАЖИ**

Преобразования: $x=f(x)$

Прибавить константу ($x=x+C$) C= -1599.

Степень ($x=x^{**}C$) C= 2.00

Обратная степень ($x=x^{**}1/C$) C= 2.00

Натуральный логарифм ($x=\ln(x)$)

Экспонента ($x=\exp(x)$)

Вычесть среднее ($x=x-M$) M= 3047.6

Стандартиз. ($x=(x-M)/SD$) M= 3047.6 S= 955.86

Оценить среднее и станд. отклонение из данных

Вычесть тренд ($x=x-(a+b*t)$) a= 1445.8 b= 24.086

Автокорр. ($x=x-(a+b*x(\text{лаг}))$) a= 0. b= 1. лаг= -1


Оценить a/b из данных

Сглаживание

N-точ. скольз. средним N= 2 По пред Взвеш.

N-точ. скольз. медианой N= 2 По пред

Простое экспоненциальное альфа= .20

 **Преобразования для спектрального анализа**

Преобразования двух рядов

Разность ($x=x-y(\text{лаг})$) лаг= 0

Остатки ($x=x-(a+b*y(\text{лаг}))$) a= 0. b= 1. лаг= 0

Оценить a и b из данных

Сдвинуть начальную точку ряда

Сдвиг вперед лаг= 1

Сдвиг назад лаг= 1

Фильтрация и другие методы

4253H фильтр

Разность ($x=x-x(\text{лаг})$) лаг= 1

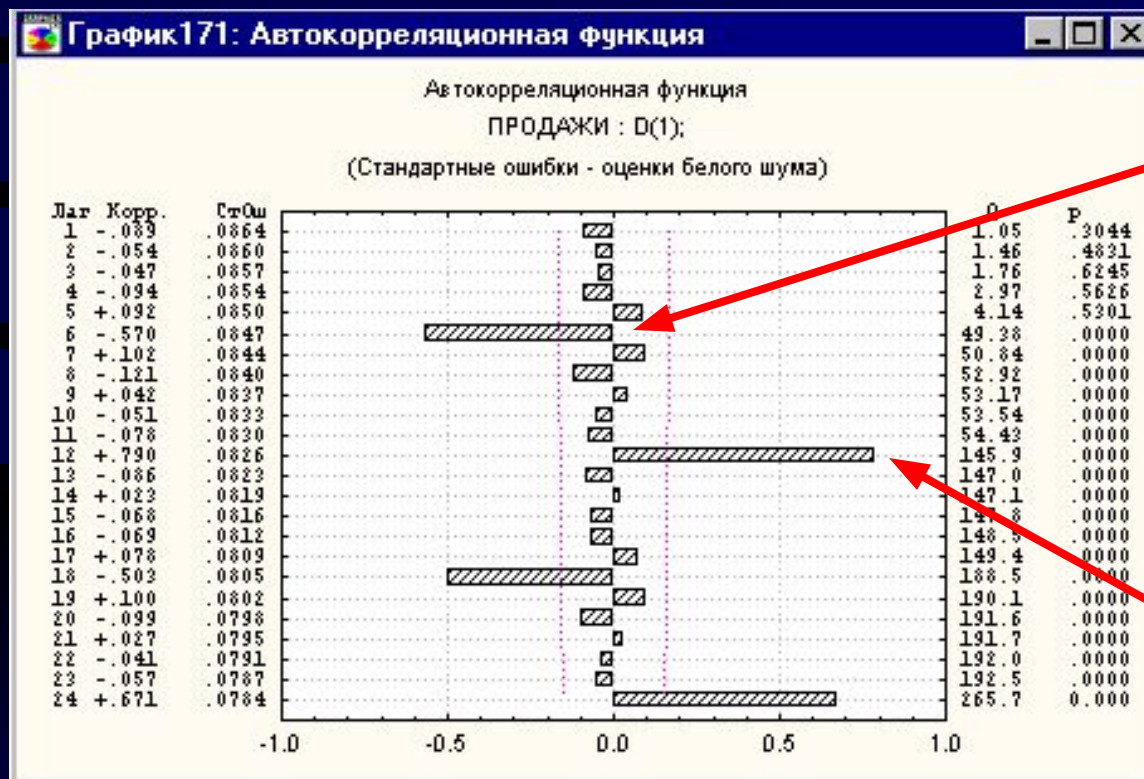
Суммир. ($x=x+x(\text{лаг})$) лаг= 1

OK (Преобразовать)

Отмена

Выберите преобразование для выделенной переменной (ряда); в открытом диалоговом окне доступна также процедура простого экспоненциального сглаживания.

Строим график автокорреляционной функции преобразованного ряда:

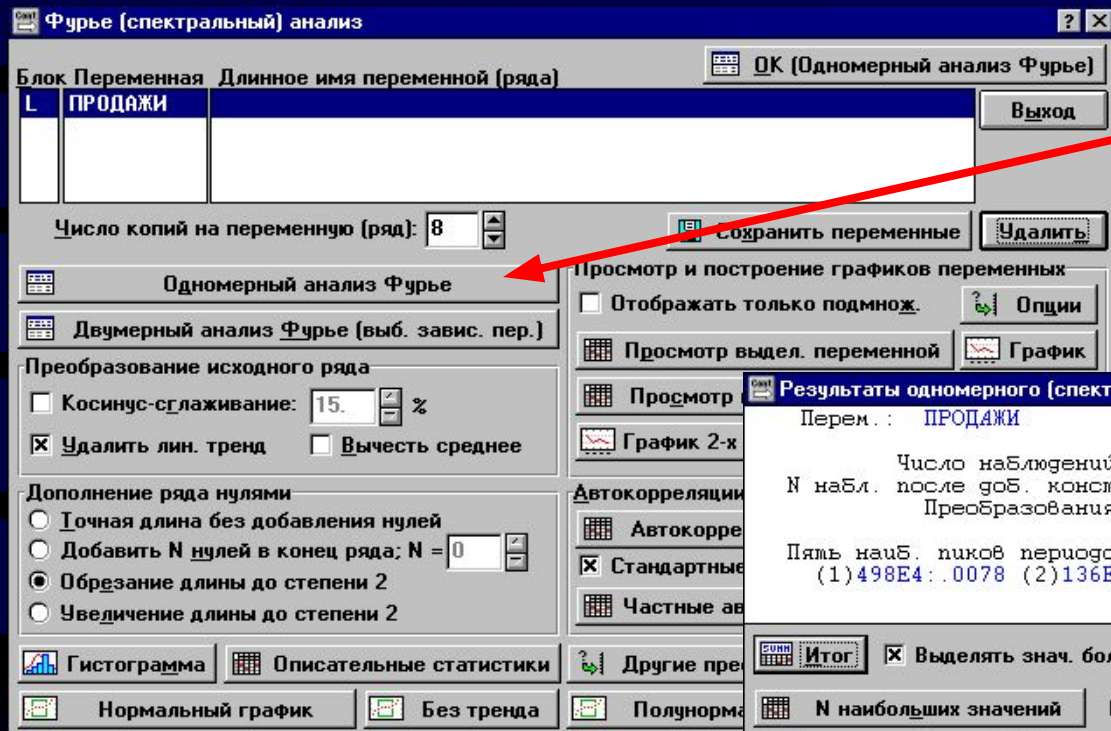


Значимые коэффициенты корреляции наблюдаются между значениями показателя, отстоящими на 6 месяцев друг от друга.

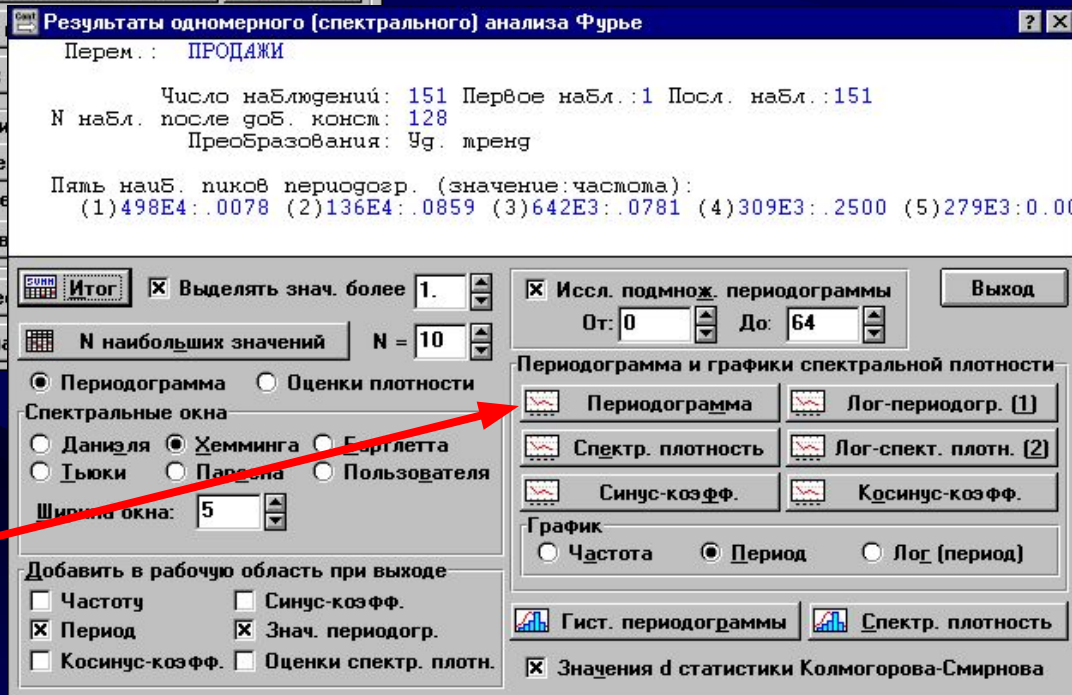
**Исследуемый временной ряд
обнаруживает свойства
периодичности.**

**Нужно выделить сезонную
составляющую и
скорректированный ряд,
который несет информацию
об общем характере динамики.**

Шаг 2: Определение периода сезонной составляющей.

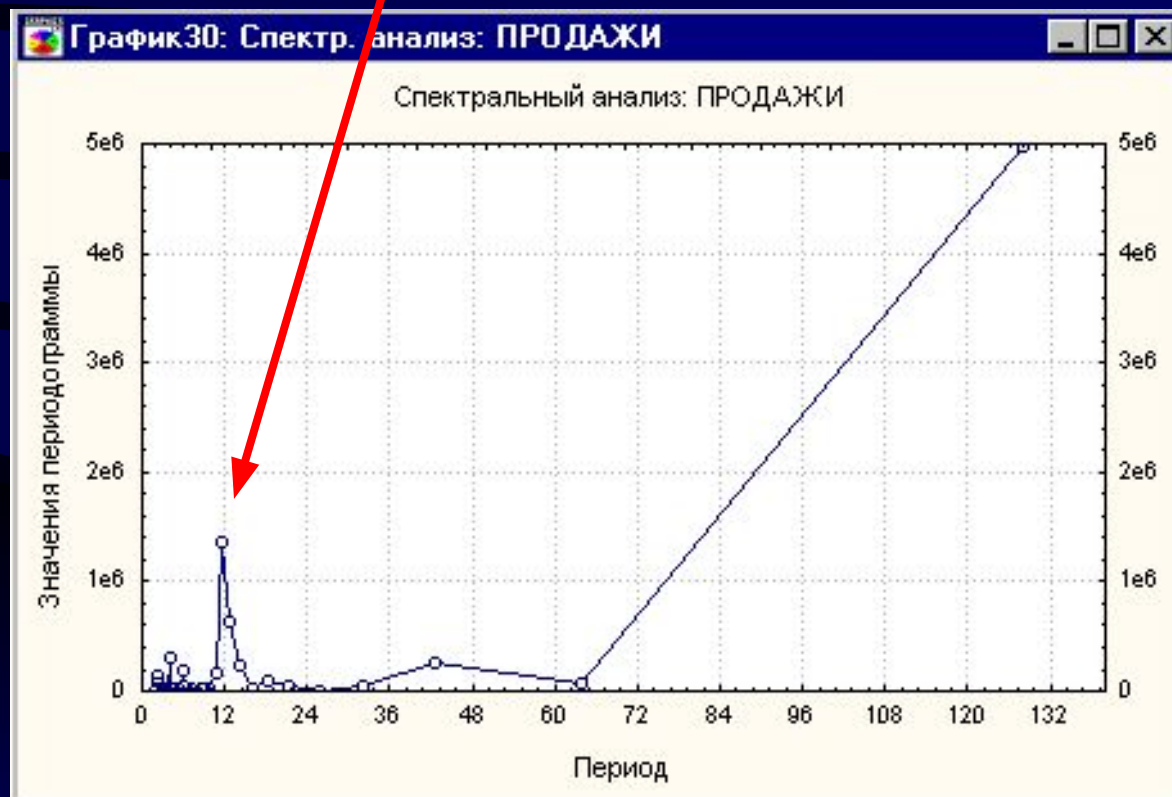


Применяем метод
спектрального анализа
Фурье



На графике периодограммы
изображены вклады
каждого значения
периода

**Периодограмма содержит ярко выраженный пик на значении периода, равном 12.
В исходном временном ряде имеется годовой цикл.**



Применяем метод сезонной декомпозиции ряда

Устанавливаем
сезонный лаг
равным 12

Предполагаем,
что модель
мультипликативна,
поскольку
размахи через период
возрастают

Классическая сезонная декомпозиция (метод Census I)

Блок Переменная Длинное имя переменной (ряда)

L	ПРОДАЖИ	
	ПРОДАЖИ	Сезон. составл. (сезон = 12);
	ПРОДАЖИ	Скоррект. (сезон=12);

Число копий на переменную (ряд): 8

Сезонная модель

Аддитивная Мультипликат. Сезонный лаг: 12 Цент. скользящее среднее (для четного лага)

Добавить в раб. область при нажатии ОК

Скользящие средние
 Отношения/Разности
 Сезонные составляющие
 Сезонную корректировку ряда
 Сглаженный тренд-цикл
 Нерегулярную компоненту

Просмотр и построение графиков переменных

Отображать только подмнож.

Автокорреляции

Альфа (выдел.): .050

Стандартные ошибки белого шума

Число лагов: 15

Результаты сезонной декомпозиции:



Сезонная составляющая.
Наблюдается годовая периодичность.

Скорректированный ряд.
Ярко выраженная
возрастающая
тенденция в динамике
показателя.



Шаг 3: На каждом временном интервале будем описывать динамику объемов продаж моделями Авторегрессии и скользящего среднего

Анализ временных рядов

Переменные: ПРОДАЖИ

ОК (преобразования, авто- и кросскорреляции, графики)

Блок	Переменная	Длинное имя переменной (ряда)
L	ПРОДАЖИ	

Число копий на переменную (ряд): 8

Сохранить переменные | Удалить переменную

ARСС (Бокс и Дженкинс) и автокорреляции

- ARСС и автокорреляционные функции
- Анализ прерванных временных рядов
- Экспоненциальное сглаживание и прогноз
- Сезонная декомпозиция (Census 1)
- X11 (Census 2) - месячный | кварталный
- Анализ распределенных лагов
- Фурье (спектральный) анализ

Замена пропущенных данных

- Общим средним
- Интерполяцией по ближайшим точкам
- Средним N ближайших значений; N: 1
- Медианой N ближайших значений; N: 2
- Значениями, предсказ. лин. регрессией

Построение модели АРСС:

Значение сезонного лага равно 12.

Перед оценкой приводим ряд к стационарному виду с помощью взятия первой разности.

Одномерная АРСС

Блок Переменная Длинное имя переменной (ряда)

L	ПРОДАЖИ	Сезон. составл. (сезон = 12);
	ПРОДАЖИ	Скоррект. (сезон=12);
	ПРОДАЖИ	

Число копий на переменную (ряд): 8

Параметры модели АРСС

Оценить константу

Сезонный лаг: 12

p - авторегрессии: 0 P - Сезонных: 1

q - скользящее среднее: 0 Q - Сезонных: 1

Преобразовать переменную (ряд) перед анализом

Нат. логарифм Возвести в степень: 2.0

Разность 1. Лаг: 1 Порядок: 1

2. Лаг: 0 Порядок: 0

Другие преобразования и графики

Оценивание методом максимал. правдоподобия

Приближенный (МакЛеода и Сейза)

Точный (Меларда) Обратные итерации: 0

Нач. значения, число итераций, сходимость

Просмотр и построение графиков переменных

Отображать только подмнож. Опции

Просмотр выдел. переменной График

Просмотр нескольких перемен. График

График 2-х списков перемен. в разных масштабах.

Автокорреляции

Автокорреляции Альфа (выдел.): 0.050

Стандартные ошибки белого шума

Частные автокорр. Число лагов: 15

Гистограмма Описательные статистики

Нормальный Без тренда Полунорм.

Сохранить переменные Удалить

Выход

ОК (Начать оценивание параметров)

Добавляем в модель скользящее среднее.

Оценивание коэффициентов модели:

Итеративная
процедура
оценивания.

Итерация	Потери	Параметры	
* 4	732980.	.932698	.341983
* 5	676950.	.995202	.400982
* 6	674185.	.998978	.404799
* 7	673908.	.999365	.405208
* 8	673633.	.999757	.405639
* 9	673605.	.999797	.405685
* 10	673578.	.999836	.405730

Процесс оценивания параметров сошелся

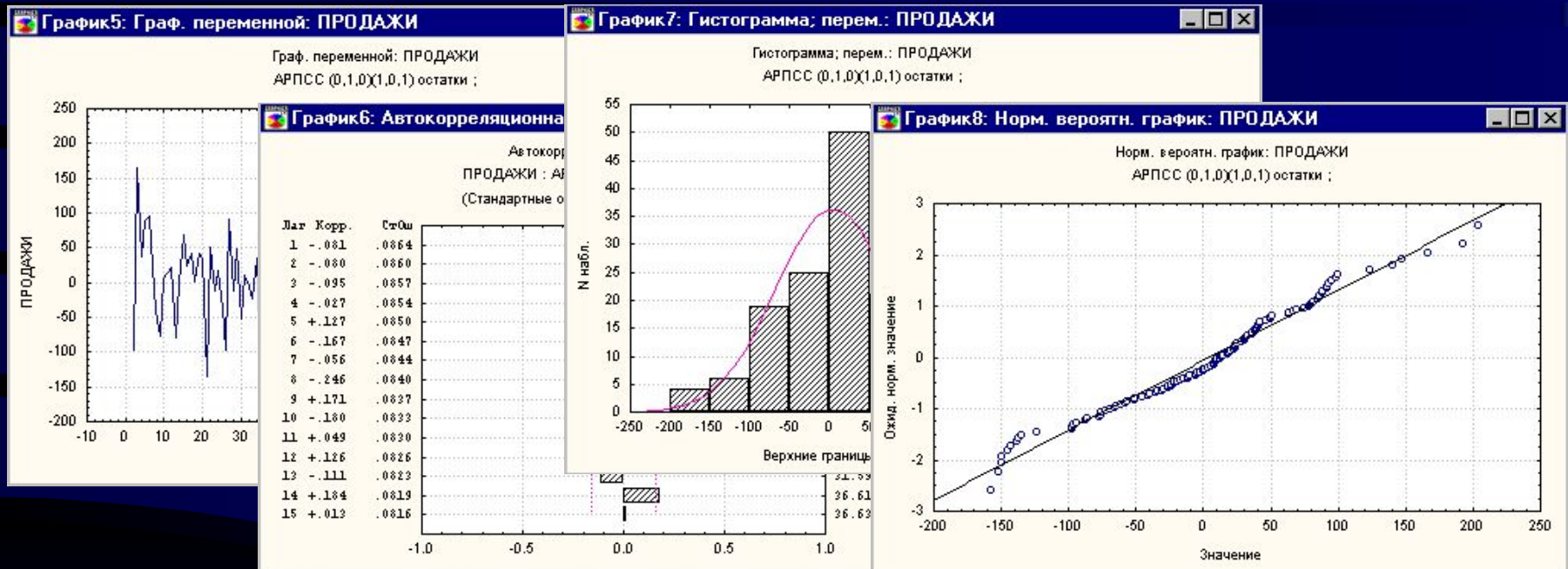
Отмена OK

ВРЕМЕН. РЯДЫ	Преобразование: D(1) Модель(0,1,0)(1,0,1) Сезонный лаг: 12 MS Остаток=5221.5					
Параметр	Парам.	Асимпт. Ст. ошиб.	Асимпт. t(129)	p	Нижняя 95% дов.	Верхняя 95% дов.
Ps(1)	.999836	.000000	97476E12	0.000000	.999836	.999836
Qs(1)	.405730	.077473	5.	.000001	.252448	.559012

Коэффициенты модели статистически значимы
на доверительном уровне 5%.

Шаг 4: Исследование адекватности модели.

Анализируем остатки:



Распределение остатков достаточно хорошо описывается нормальным распределением, следовательно, модель адекватно отражает динамику исследуемого процесса.

Шаг 5: Построение прогноза на основе полученной модели.

Численные
оценки
параметров
модели.

Строим
прогноз
на
12 месяцев.

Результаты одномерной АРПСС

Переменная: ПРОДАЖИ
Преобразования: D(1)
Модель: (0,1,0)(1,0,1) Сезонный лаг: 12
Число набл.: 131 Начальн SS= 3324E3 Оконч SS= 6736E2(20.26%) MS= 5221.5
Параметры (p/Ps-авторегрессии, q/Qs-скольз. средн.); выделение: p<.05
Ps(1) Qs(1)
Оценка: .99984 .40573
Ст. ошиб.: .00000 .07747

Оценки параметров Печать результатов Выход

Ковариации и корреляции параметров

Прогноз

Прогноз График ряда и прогнозов

Число наблюдений: 12
Уровень доверия: .9
Начать с наблюдения: 133

Добавить прогнозы к исход. ряду при Выходе

Графики остатков

Гистограмма Нормальный
Без тренда Полнормальный

При нажатии кнопки Выход преобразованный ряд и остатки добавятся к переменным в памяти.

Просмотр и построение графиков переменных

Отображ. только подмнож. Опции

Просмотр исходного ряда График (1)
Просмотр преобраз. ряда График (2)
Просмотр остатков График (3)

Автокорреляция остатков

Автокорреляции Альфа (выдел.): .050
 Стандартные ошибки белого шума
Частные автокор. Число лагов: 15

График исходного ряда и прогноза



Прогноз на
12 месяцев

Границы
доверительных
интервалов

Шаг 6: Проверка точности прогноза



Рассмотрим переход между двумя временными интервалами



Разбиение выборки на два интервала было корректным, поскольку характер динамики в точке перехода меняется.

Результаты для временного интервала январь 1979 - январь 2001

График27: Норм. вероятн. график: ПРОДАЖИ

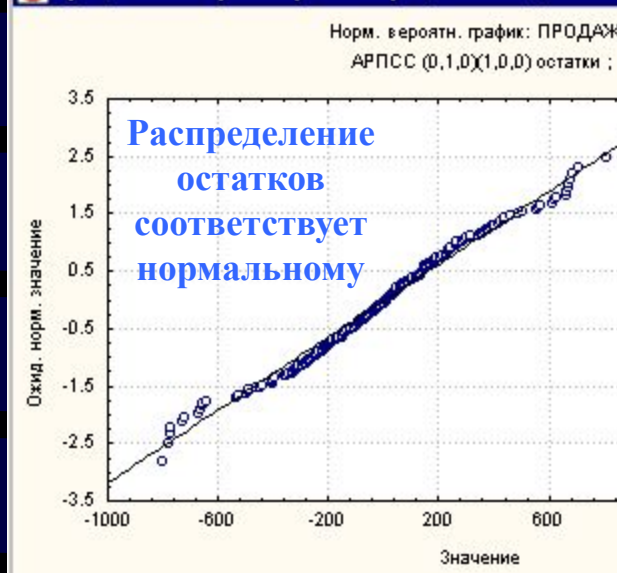


График26: Линейный график



Исход.: ПРОДАЖИ (retail.sta)

ВРЕМЕН. РЯДЫ	Преобразования: D(1) Модель(0,1,0)(1,0,0) Сезонный лаг: 12 MS			
Параметр	Парам.	Асимпт. Ст. ошиб.	Асимпт. t(261)	p
Конст.	62.90756	76.96186	.81739	.414453
Ps(1)	.86752	.03788	22.90091	0.000000

Спецификация модели
изменилась

Шаг 7: Строим прогноз согласно полученной модели:



**Можно ли сделать это
проще с помощью
других методов
В
STATISTICA ?**

Применим методы анализа прерванных временных рядов

Прерванная АРПСС (анализ интервенций)

Блок Переменная Длинное имя переменной (ряда)

L	ПРОДАЖИ	
	ПРОДАЖИ	D(1);
	ПРОДАЖИ	АРПСС (0,1,0)(1,0,0) остатки ;
	ПРОДАЖИ	D(1);

Число копий на переменную (ряд): 8

Сохранить переменные

Параметры модели АРПСС

Оценить константу Сезонный лаг: 12

p - авторегрессии: 0 P - Сезонных: 1

q - скольз.средн.: 0 Q - Сезонных: 0

Преобразовать переменную (ряд) перед анализом

Нат. логарифм Возвести в степень: 2

Разность 1. Лаг: 1 Порядок: 1

2. Лаг: 0 Порядок: 0

Другие преобразования и графики

Оценивание методом максимал. правдоподобия

Приближенный (МакЛеода и Сейлза)

Точный (Меларда) Обратные итерации: 0

Нач. значения, число итераций, сходимость

Используйте опцию АРПСС и автокорреляционные функции на стартовой панели для определения и подгонки АРПСС модели до интервенции.

Задание времени и типа интервенций

Интервенция	N набл.	Тип интервенции
<input checked="" type="checkbox"/>	1: 149	Скачкообр., устойчивое
<input type="checkbox"/>	2: 0	Скачкообр., устойчивое
<input type="checkbox"/>	3: 0	Скачкообр., устойчивое
<input type="checkbox"/>	4: 0	Скачкообр., устойчивое
<input type="checkbox"/>	5: 0	Скачкообр., устойчивое
<input type="checkbox"/>	6: 0	Скачкообр., устойчивое

Просмотреть типы и

Указываем
тип
интервенции и
номер наблюдения,
с которого
характер
зависимости
меняется.

Задание типа интервенции

Тип интервенции

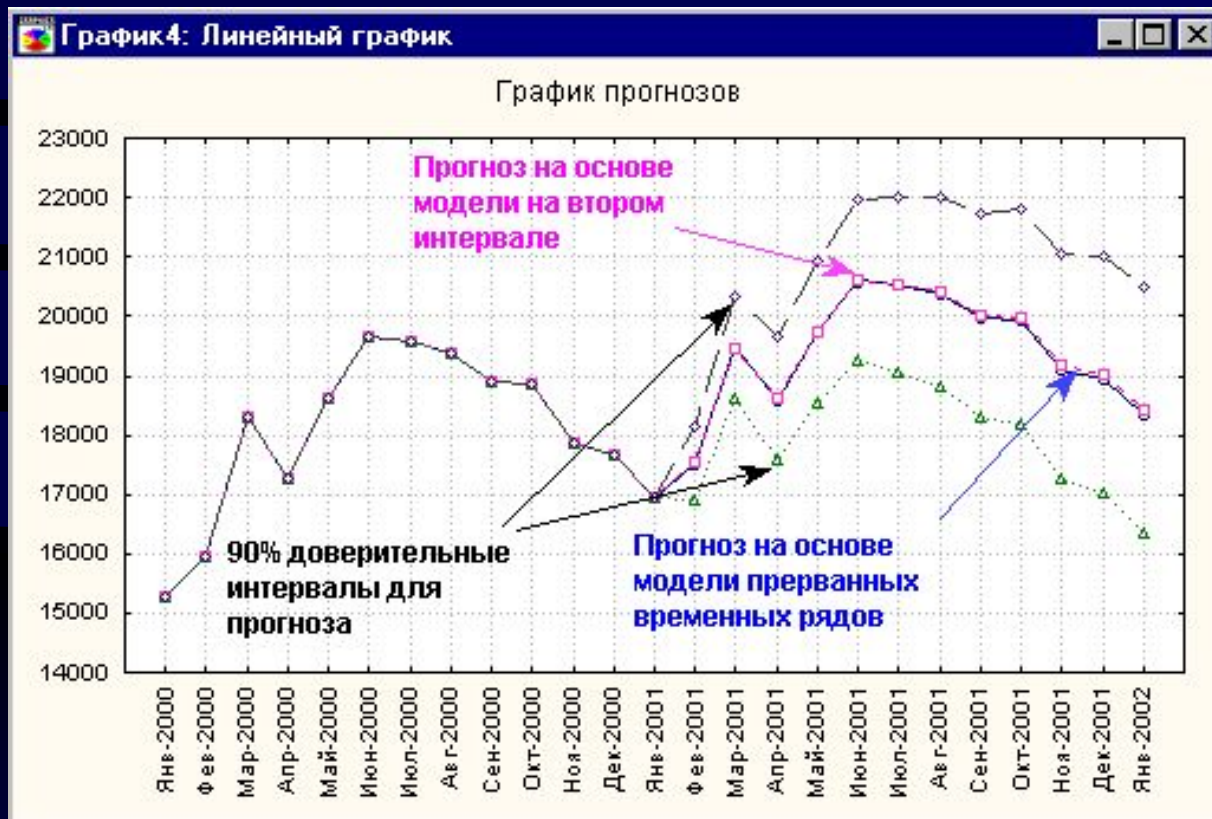
Скачкообр., устойчивое

Скачкообр., устойчивое

Постепенное, устойчивое

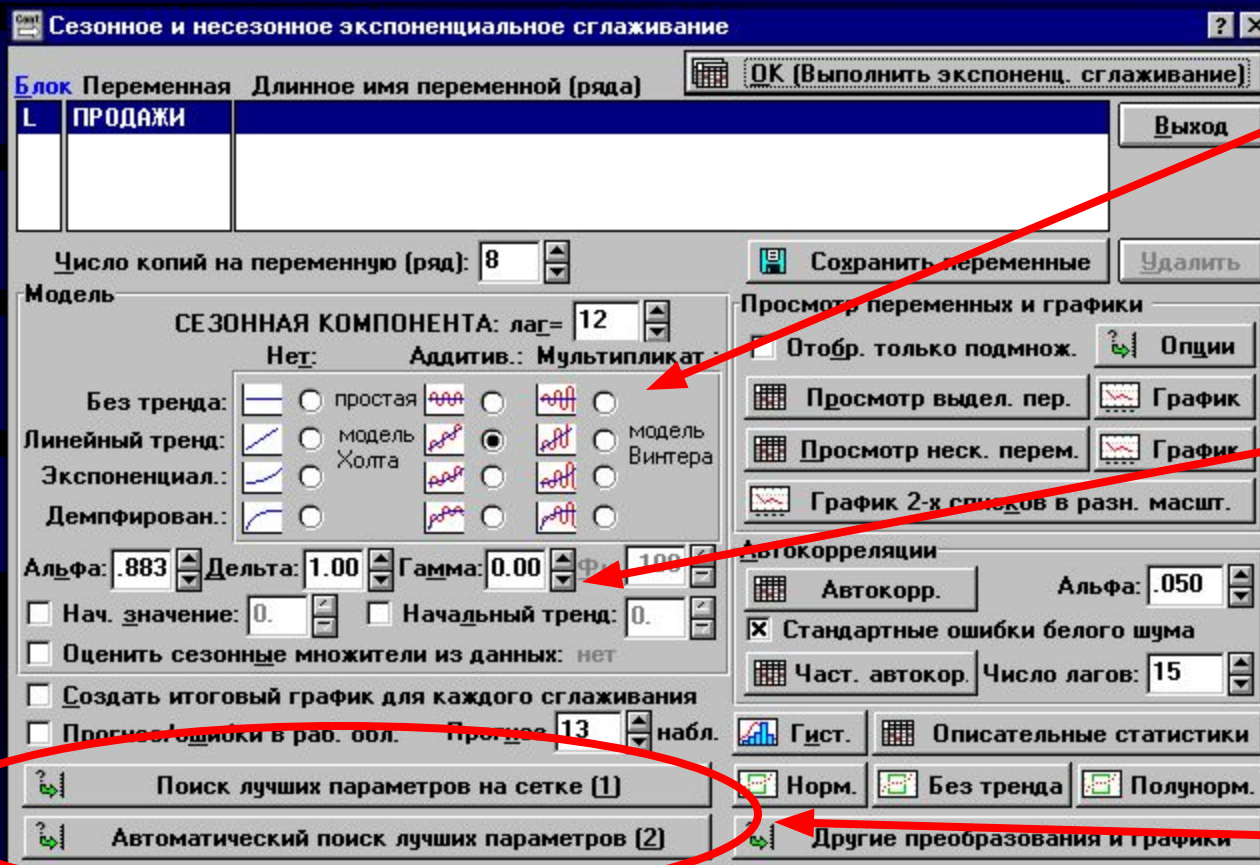
Скачкообр., временное

Построение прогноза двумя способами и сравнение результатов:



**В пределах
заданных
доверительных
интервалов
точность
прогнозов
одинакова.**

Пользуемся методом экспоненциального сглаживания



Выбираем тип
модели
сглаживания.

Параметры
сглаживания
можно задать
вручную.

Можно
воспользоваться
методом
автоматического
поиска параметра.

Сравниваем результаты прогнозов на последний год выборки с наблюдаемыми:



**Сглаживание
отражает
общую тенденцию,
но менее
чувствительно
к отдельным
колебаниям.**

Строим прогноз с помощью метода экспоненциального сглаживания:



На графике
показаны ряды
наблюдаемых
величин
и прогноз
на 12 месяцев.

Система *STATISTICA*
предоставляет пользователю
все необходимые методы
визуализации данных, анализа
и прогнозирования:

- **построение и оценивание ARIMA моделей**
- **анализ прерванных временных рядов**
- **сезонная корректировка**
- **экспоненциальное сглаживание**
- **методы спектрального анализа Фурье**