

ЛАБОРАТОРНАЯ РАБОТА № 13

ВРЕМЕННЫЕ РЯДЫ

Рассмотрим динамику повышения урожайности зерновых культур в СССР с 1945 по 1989 год. В файле *Zerno.sf* введены данные в соответствии с таблицей:

Год	Урожайность	Год	Урожайность	Год	Урожайность
1945	5,6	1960	10,9	1975	10,9
1946	4,6	1961	10,7	1976	17,5
1947	7,3	1962	10,9	1977	15,0
1948	6,7	1963	8,3	1978	18,5
1949	6,9	1964	11,4	1979	14,2
1950	7,9	1965	9,5	1980	14,9
1951	7,4	1966	13,7	1981	12,6
1952	8,6	1967	12,1	1982	15,2
1953	7,8	1968	14,0	1983	15,9
1954	7,7	1969	13,2	1984	14,4
1955	8,4	1970	15,6	1985	16,2
1956	9,9	1971	15,4	1986	18,0
1957	8,4	1972	14,0	1987	18,3
1958	11,1	1973	117,6	1988	17,0
1959	10,4	1974	15,4	1989	18,8

В строке меню выберите *Describe, Time Series, Descriptive Methods*. Раскроется окно.

Descriptive Methods

Col_1
Col_2

Data: Col_1

Sampling Interval

Once Every: 1

☒ Year(s) (4-digit)
 ☐ Hour(s)
 ☐ Quarter(s)
 ☐ Minute(s)
 ☐ Month(s)
 ☐ Second(s)
 ☐ Day(s)
 ☐ Other

Starting At: 1945,0

(Seasonality:)

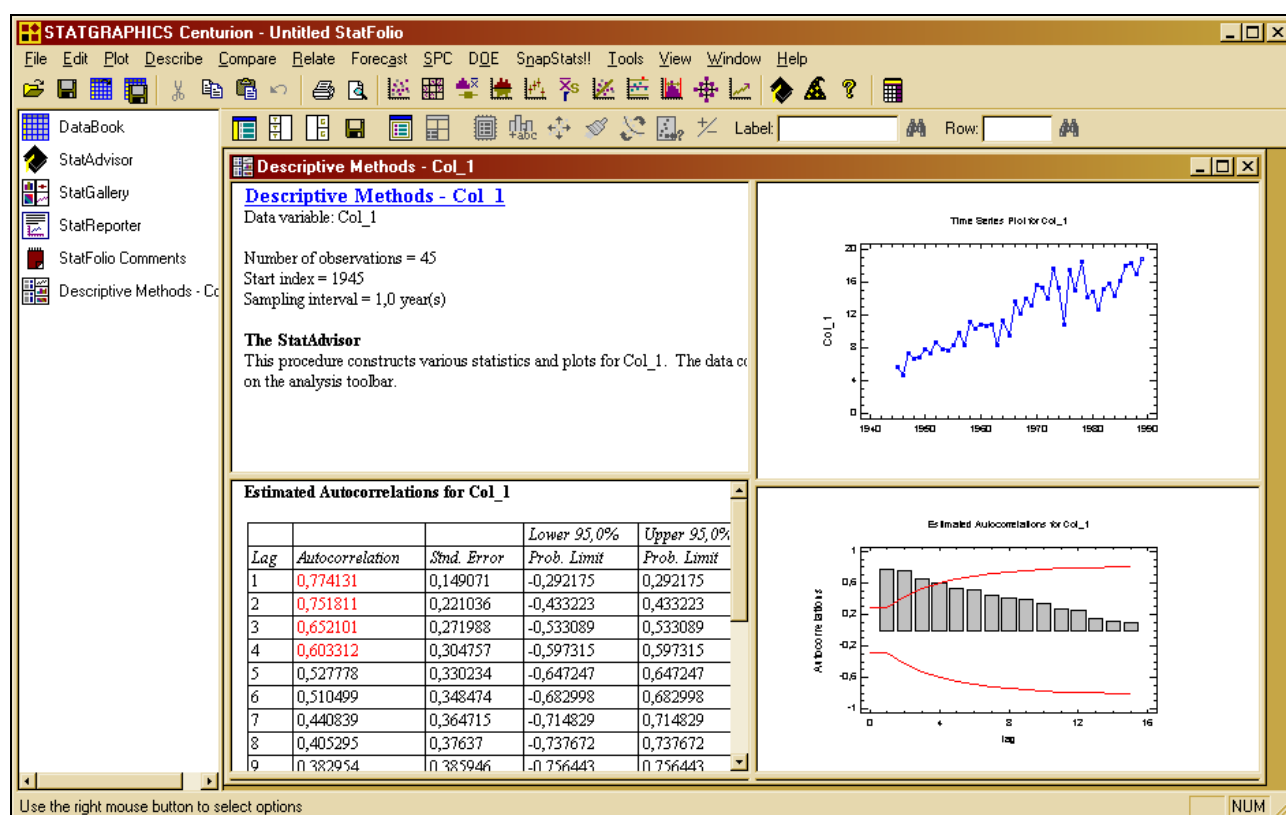
(Trading Days Adjustment:)

(Select:)

☒ Sort column names

OK Cancel Delete Transform... Help

В этом окне надо выбрать данные, а затем способ ввода данных (в нашем случае – по годам) и начальную точку (1945). Нажав кнопку ОК, получим:



Из графика *Time series Plot* хорошо видно, что урожайность за эти годы возрастала. Видно, что временной ряд не является стационарным. На графике авто-

корреляции видно, что часть столбиков выходит за границы 95 % интервала. Это говорит о том, что данные коррелированы. В окне *Estimated Autocorrelation* также видно наличие автокорреляции – красным цветом выделены автокорреляции для соответствующих лагов. По графику автокорреляции (медленное спадание автокорреляционной функции лесенкой) видно, что стоит пытаться выделить тренд.

Попытаемся выявить наличие тренда, используя формальные статистические критерии. Для этого используются тесты на нерегулярность. Нажмите кнопку *Tables*, в появившемся окне выберите *Test for Randomness* (тесты на нерегулярность), нажмите кнопку ОК. Раскроется окно с результатами. Здесь рассмотрены три теста на нерегулярность временного ряда.

Tests for Randomness of adjusted RESIDUALS	
(1) Runs above and below median	
Median = 0,39108	
Number of runs above and below median = 34	
Expected number of runs = 22,0	
Large sample test statistic $z = 3,59307$	
P-value = 0,000326885	
(2) Runs up and down	
Number of runs up and down = 35	
Expected number of runs = 29,0	
Large sample test statistic $z = 2,00832$	
P-value = 0,0446094	
(3) Box-Pierce Test	
Test based on first 14 autocorrelations	
Large sample test statistic = 35,6461	
P-value = 0,00117857	

В первом тесте *Runs above and below median* рассчитывается, сколько раз значение анализируемой переменной лежит выше или ниже медианы распределения. Медиана равна 0,39108, рассчитанное число равно 34, а ожидаемое число для случайной последовательности должно быть равно 22. При этом значение *p-value* равно 0,000326885, это меньше 0,01, поэтому гипотеза о нерегулярности временного ряда отвергается с 99 % доверительной вероятностью.

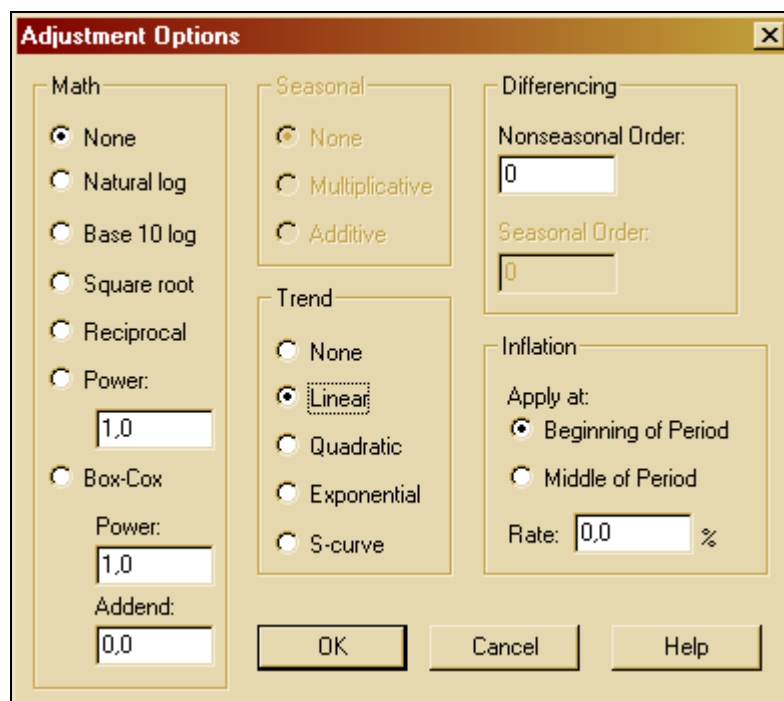
Во втором тесте *Runs up and down* вычисляется, сколько раз временной ряд повышался или понижался. Для нашей последовательности данная величина равняется 35, а ожидаемое для случайной последовательности значение 29. При

этом значение p -value равно 0,044, это меньше 0,05, поэтому гипотеза о нерегулярности временного ряда отвергается с 95 % доверительной вероятностью.

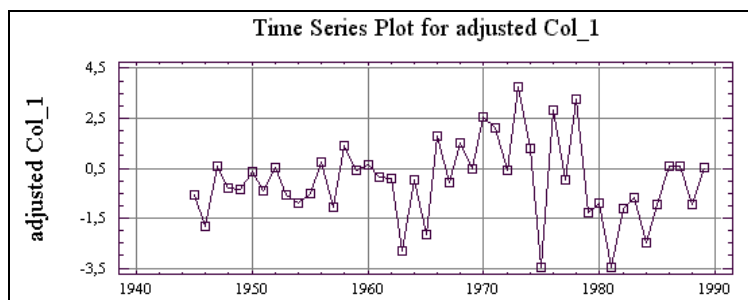
Третий тест *Box-Pierce Test* основан на анализе суммы квадратов первых 14 автокорреляционных коэффициентов. Для него p -value также меньше 0,01, поэтому и в этом случае гипотеза о нерегулярности временного ряда отвергается с 99 % доверительной вероятностью.

Вывод. На основании результатов трех тестов, каждый из которых чувствителен к различным аспектам отклонения от случайной последовательности, можно с большой уверенностью утверждать, что проанализированный временной ряд имеет выраженную регулярную компоненту (тренд).

Попробуем выделить этот тренд. Щелкните **правой** кнопкой мыши, в контекстном меню выберите ***Analysis Options***, раскроется окно:



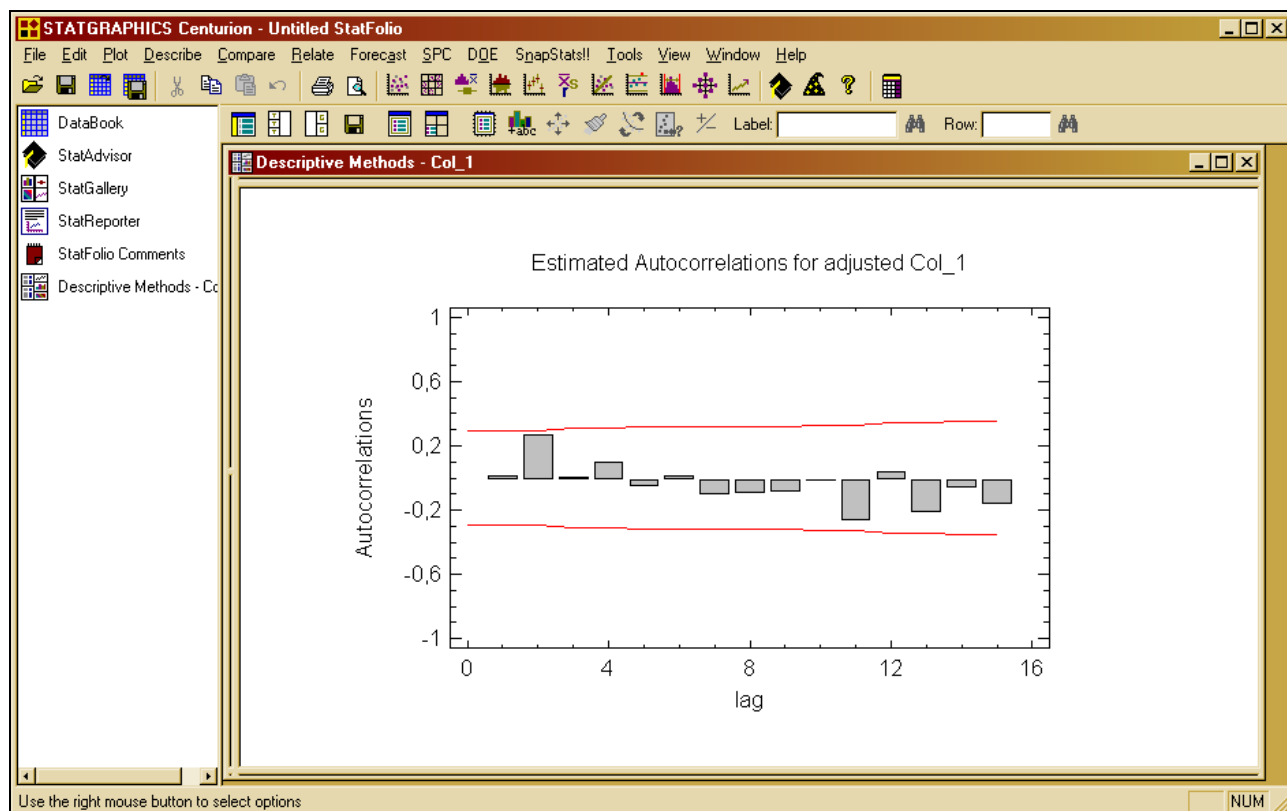
В этом окне можно выбирать различные модели тренда, выберите ***Linear***, посмотрите получившийся график остатков. Должно получиться следующее.



Сохраните теперь остатки временного ряда после удаления из него подобранной модели тренда. Для этого нажмите на кнопку ***Save Results***, раскроется окно:

В этом окне поставьте галочку в строке ***Adjusted data*** (данные после преобразования), нажмите ОК. Теперь нужно убедиться в том, что остаток можно рассматривать как последовательность нормально распределенных независимых случайных величин с постоянными математическими ожиданиями, т. е., как гауссовский белый шум.

Во-первых, необходимо убедиться в том, что остатки некоррелированы. Постройте функцию автокорреляции, убедитесь в этом. Для этого рассмотрите график.



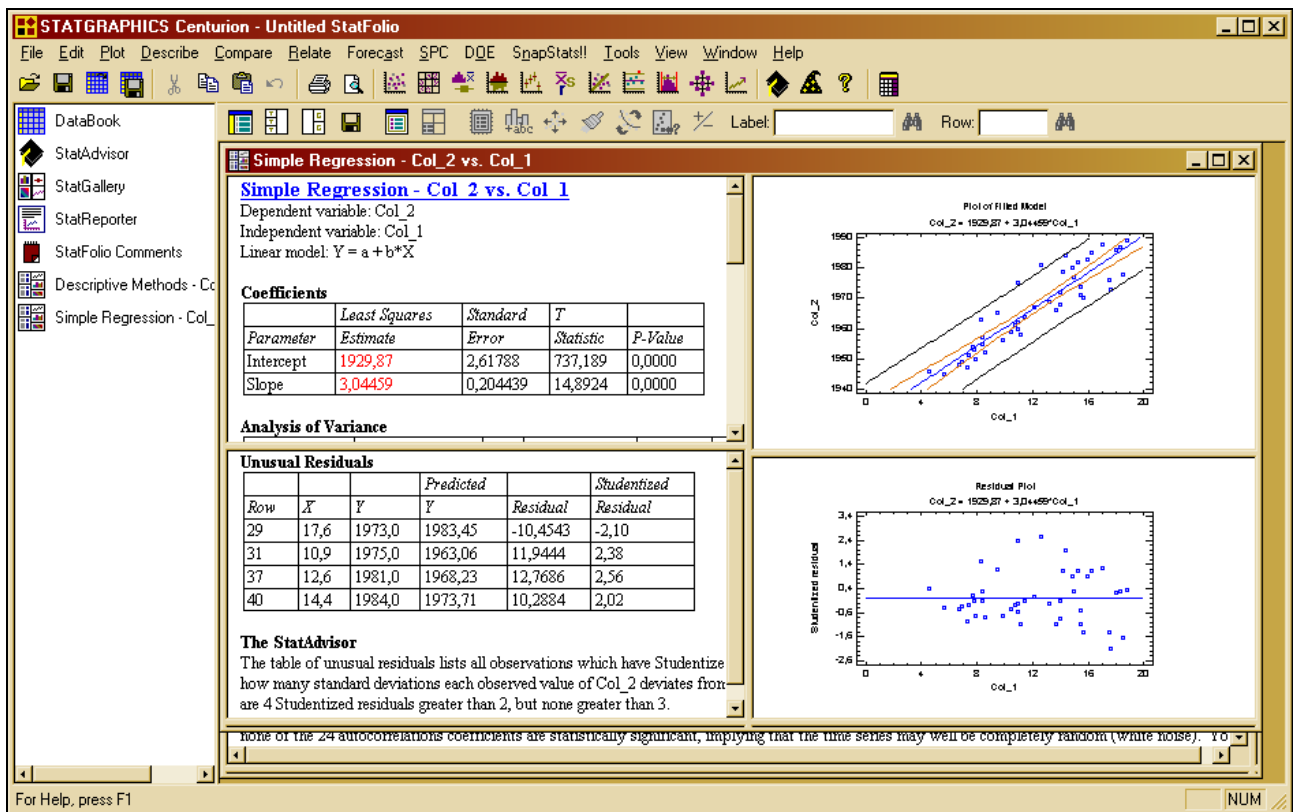
Так как все столбики укладываются в трубку, можно считать, что автокорреляции нет. Для надежности разверните окно:

Lag	Autocorrelation	Std. Error	Lower 95,0% Prob. Limit	Upper 95,0% Prob. Limit
1	0,0152124	0,149071	-0,292175	0,292175
2	0,271674	0,149106	-0,292242	0,292242
3	0,00403054	0,159727	-0,31306	0,31306
4	0,0962038	0,15973	-0,313065	0,313065
5	-0,0593073	0,161012	-0,315578	0,315578
6	0,011199	0,161497	-0,316528	0,316528
7	-0,107084	0,161514	-0,316562	0,316562
8	-0,0974813	0,163084	-0,31964	0,31964
9	-0,0871001	0,164374	-0,322167	0,322167
10	-0,0173881	0,165396	-0,324171	0,324171
11	-0,265881	0,165437	-0,324251	0,324251
12	0,0408506	0,174675	-0,342357	0,342357
13	-0,219755	0,174887	-0,342773	0,342773
14	-0,0662649	0,180919	-0,354596	0,354596
15	-0,164889	0,181458	-0,355651	0,355651

Убедитесь в том, что автокорреляции нет, а также в нормальности распределения остатков. *Результаты покажите преподавателю.*

Для того чтобы записать теперь **аналитически** функцию линейного тренда, вспомните, как выделяется тренд (линейная регрессия), выделите самостоя-

тельно линейный тренд. Подумайте, что будет в качестве аргумента (X). Должно получиться следующее:



Сохраните теперь остатки временного ряда после удаления из него подобранной модели тренда. Для этого нажмите на кнопку **Save Results**, раскроется окно:

Save Results Options

Save

- ☒ Predicted Values
- ☐ Lower Limits for Predictions
- ☐ Upper Limits for Predictions
- ☐ Lower Limits for Forecast Means
- ☐ Upper Limits for Forecast Means
- ☐ Residuals
- ☐ Studentized Residuals
- ☐ Leverages

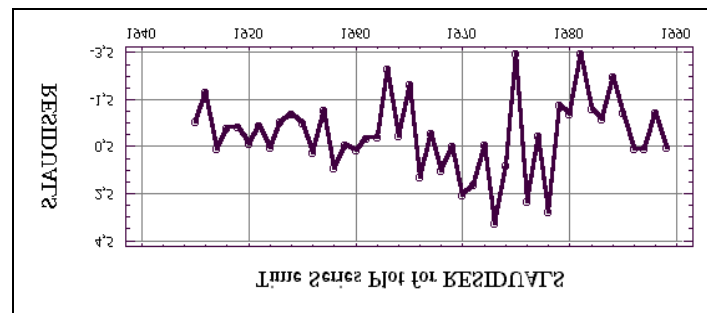
Target Variables

PREDICTED
LOWERPLIMS
UPPERPLIMS
LOWERCLIMS
UPPERCLIMS
RESIDUALS
SRESIDUALS
LEVERAGES

OK
Cancel
Help

в этом окне выберите **Residuals** (остатки) и **Predicted Values** (если вы хотите, сохраните значения, которые дает модель тренда. Если хотите, можно изменить в правой части названия соответствующих столбцов, в которых будут храниться данные. Нажмите кнопку OK.

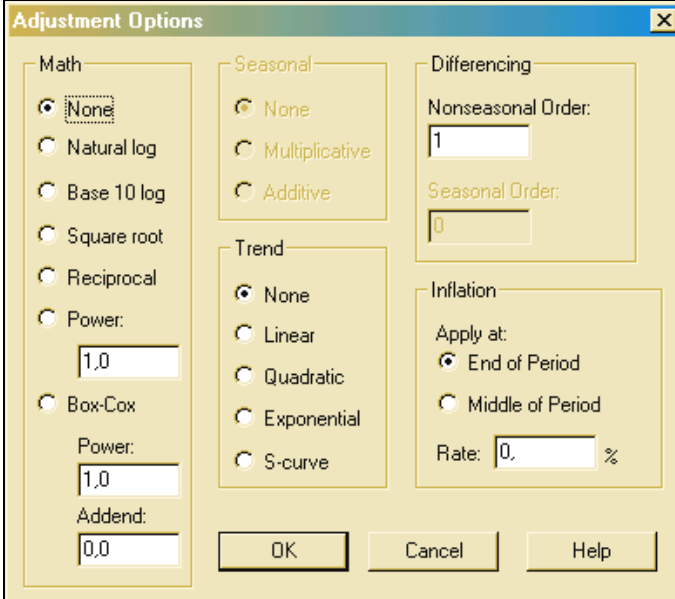
Постройте теперь самостоятельно график остатков для временного ряда. Должно получиться следующее:



Сравните, совпадают ли остатки ряда, полученные здесь, с остатками, полученными в окне временных рядов. Подумайте, как это сделать. *Результаты покажите преподавателю.*

Наряду с методом наименьших квадратов, для удаления тренда можно использовать и ряд других методов. Одним из них является метод перехода от исходного ряда к ряду разностей соседних значений ряда. В более общем виде эта идея описывается с помощью применения к ряду разностных операторов различных порядков (применение операторов дифференцирования). Особенно они пригодны для выделения трендов, если не видна подходящая аналитическая модель тренда.

Попробуем применить в этом примере этот метод. Снова в строке меню выберите ***Describe, Time Series, Descriptive Methods***. В раскрывшемся окне первичного анализа щелкните правой кнопкой, выберите ***Analysis Options***. Раскроется окно:

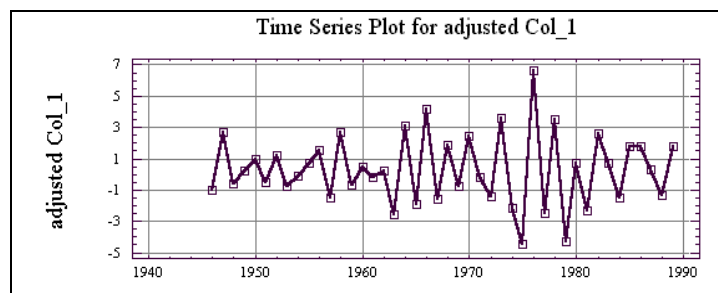


The 'Adjustment Options' dialog box contains the following settings:

- Math:** ☒ None, ☐ Natural log, ☐ Base 10 log, ☐ Square root, ☐ Reciprocal, ☐ Power: 1.0, ☐ Box-Cox, Power: 1.0, Addend: 0.0
- Seasonal:** ☒ None, ☐ Multiplicative, ☐ Additive
- Trend:** ☒ None, ☐ Linear, ☐ Quadratic, ☐ Exponential, ☐ S-curve
- Differencing:** Nonseasonal Order: 1, Seasonal Order: 0
- Inflation:** Apply at: ☒ End of Period, ☐ Middle of Period, Rate: 0.0 %

Buttons: OK, Cancel, Help

В этом окне в поле **Differencing, Nonseasonal Order** поставьте 1 (порядок дифференцирования), нажмите ОК. Вызовите теперь график временного ряда после дифференцирования.

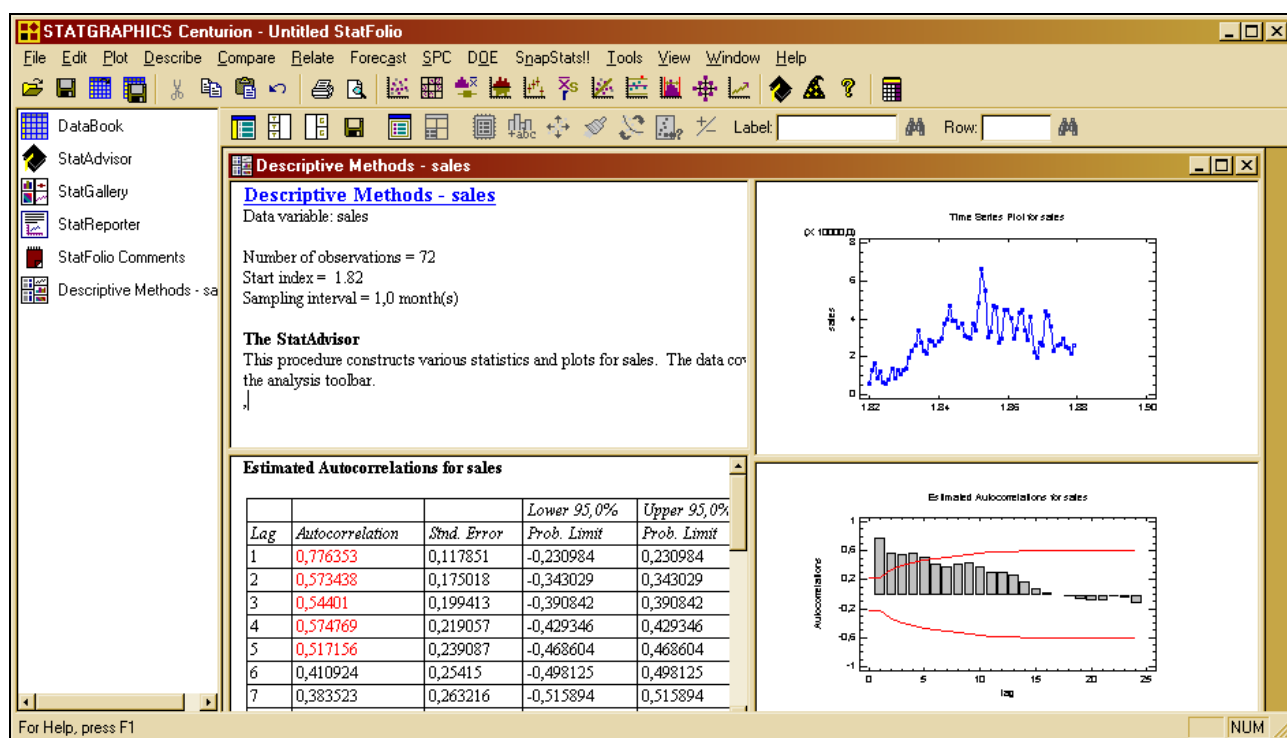


Проанализируйте самостоятельно, являются ли эти остатки независимыми случайными нормально распределенными величинами. Результаты покажите преподавателю. Объясните, какой метод (тренд или дифференцирование) кажется вам предпочтительнее в данном случае.

В общем случае нельзя сказать, какой из методов удаления тренда предпочтительнее. Все зависит от заранее неизвестной структуры случайной компоненты временного ряда ε_t .

Сейчас будем работать с файлом *Carsales.sf*, откройте его. Это — данные объемов месячных продаж автомобилей ведущих американских фирм за 6 лет с января 1982 по декабрь 1987 года. Для анализа вам нужна переменная **sales**.

Самостоятельно, воспользовавшись описательными методами для временных рядов, постройте график временного ряда. Должно получиться следующее:



Постройте самостоятельно коррелограмму. Попробуйте выделить тренд (линейный и нелинейный). Следите за остатками, будут ли они похожи на белый шум.

В нашем случае для решения задачи можно применить другой метод. Из графика автокорреляционной функции видно, что дисперсия для нашего временного ряда велика. Для того чтобы устранить этот недостаток, часто данные логарифмируют. Прологарифмируем ряд. Щелкните правой кнопкой, выберите *Analyses Options*, там выберите *Natural Log*. Эта операция уменьшает дисперсию. График автокорреляционной функции говорит о том, что здесь стоит попробовать или выделить тренд или продифференцировать ряд. Какой из методов дает приемлемые остатки? Покажите результат преподавателю.

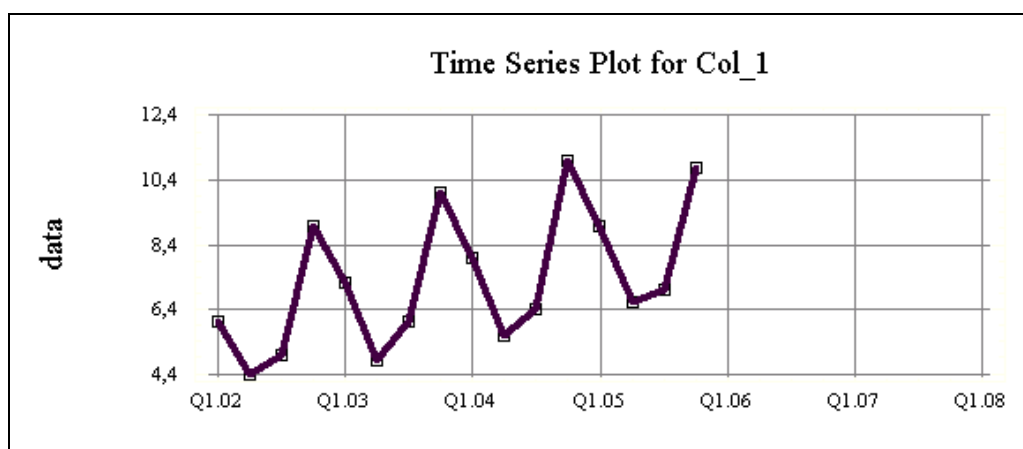
ВЫДЕЛЕНИЕ СЕЗОННОЙ КОМПОНЕНТЫ

Многие временные ряды содержат сезонные компоненты. Сезонные компоненты ряда могут представлять интерес как сами по себе, так и выступать в роли мешающего фактора. Во втором случае ставится задача — выделить и устранить сезонную компоненту из ряда.

Пример. Пусть имеются данные об объемах потребления электроэнергии жителями региона за 16 кварталов. Этот пример носит учебный характер, т. к. выборка достаточно мала.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
y_t	6,0	4,4	5,0	9,0	7,2	4,8	6,0	10	8,0	5,6	6,4	11	9,0	6,6	7,0	10,8

Проведите самостоятельно первоначальный анализ, постройте график, коррелограмму и периодограмму.



Понятно, что данный ряд не является случайным. Видна периодичность (период — 4 квартала), причем амплитуда колебаний почти не изменяется. В этом случае воспользуемся аддитивной моделью сезонных колебаний. Если амплитуда менялась, следовало бы попробовать использовать мультипликативную модель. В строке меню выберите **Describe\Time Series Analyses\Sesonal Decomposition**.

В поле поставьте 4 (период). Нажмите кнопку OK. Нажмите кнопку **Tables**, выберите **Data table**. Щелкните правой кнопкой, выберите **Pane Options**, в раскрывшемся диалоговом окне выберите **Additive**.

Раскроется диалоговое окно со следующими данными.

Seasonal Decomposition - Col_1					
Data Table for Col_1					
Seasonal decomposition method: Additive					
Period	Data	Trend-Cycle	Seasonality	Irregular	Seasonally Adjusted
Q1.82	6,0				5,41875
Q2.82	4,4				6,37708
Q3.82	5,0	6,25	-1,25	0,04375	6,29375
Q4.82	9,0	6,45	2,55	-0,139583	6,31042
Q1.83	7,2	6,625	0,575	-0,00625	6,61875
Q2.83	4,8	6,875	-2,075	-0,0979167	6,77708
Q3.83	6,0	7,1	-1,1	0,19375	7,29375
Q4.83	10,0	7,3	2,7	0,0104167	7,31042
Q1.84	8,0	7,45	0,55	-0,03125	7,41875
Q2.84	5,6	7,625	-2,025	-0,0479167	7,57708
Q3.84	6,4	7,875	-1,475	-0,18125	7,69375
Q4.84	11,0	8,125	2,875	0,185417	8,31042
Q1.85	9,0	8,325	0,675	0,09375	8,41875
Q2.85	6,6	8,375	-1,775	0,202083	8,57708
Q3.85	7,0				8,29375
Q4.85	10,8				8,11042

Здесь *Trend-Cycle* — это центрированная скользящая средняя, *Seasonally Adjusted* — это данные без сезонной компоненты (т. е. тренд плюс случайная компонента). Сама сезонная компонента получается, если нажать *Tables*, а затем выбрать *Seasonal Indices*.

Seasonal Indices for Col_1	
Seasonal decomposition method: Additive	
Season	Index

1	0,58125
2	-1,97708
3	-1,29375
4	2,68958

Сохраните теперь остатки после выделения сезонной компоненты. Для этого нажмите на кнопку **Save Results**.

Save Results Options

Save

☐ Data

☐ Trend-cycle

☐ Seasonal indices

☐ Irregular

☒ Seasonally adjusted data

Target Variables

DATA

TREND

INDICES

IRREGULAR

SADJUSTED

OK

Cancel

Help

Datasheet

☐ A

☒ B

☐ C

☐ D

☐ E

☐ F

☐ G

☐ H

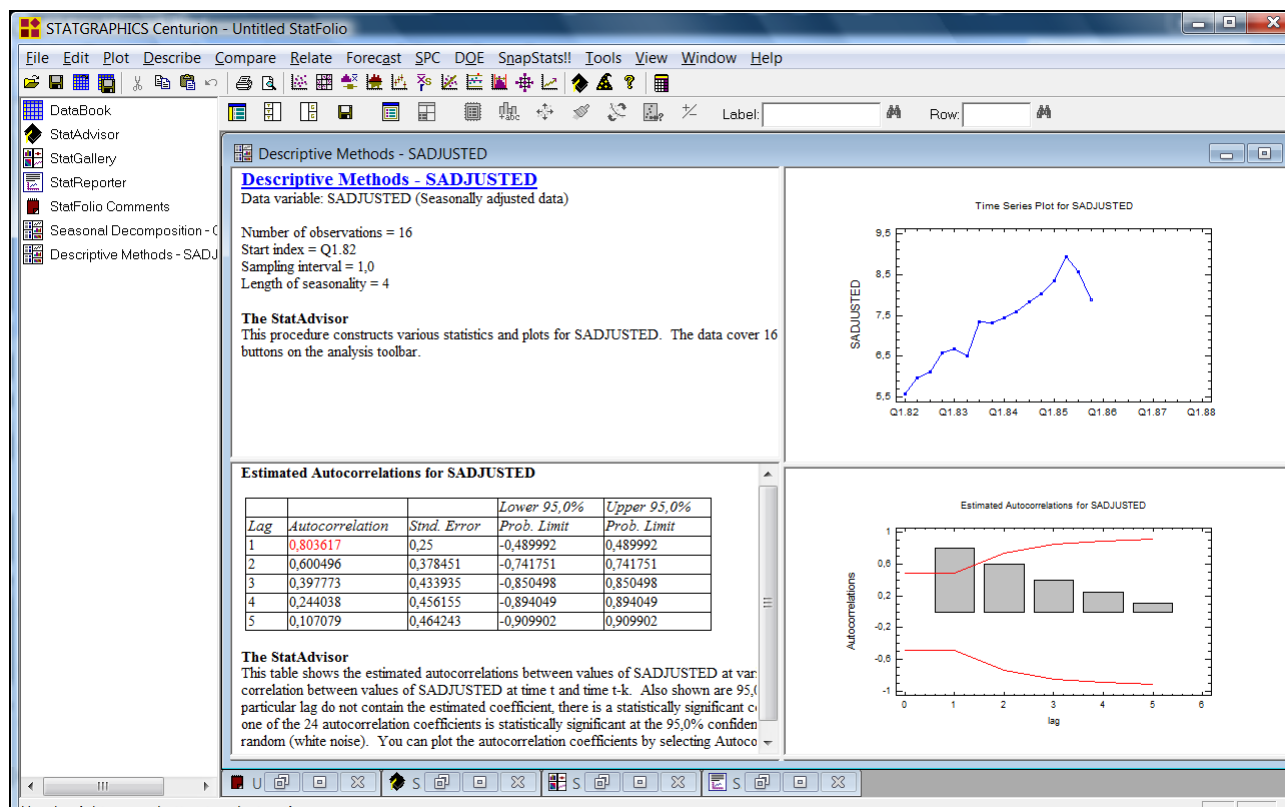
☐ I

☐ J

☐ Autosave

☒ Save comments

Выберите *Seasonally adjusted data*, нажмите OK. Проверим регулярность этих остатков. ***Describe, Time Series, Descriptive Methods.***



Теперь можно выделить тренд, сделайте это самостоятельно. *Результаты покажите преподавателю.*

В случае, если амплитуда колебаний временного ряда не постоянна, применяют **мультипликативную** модель.

Для анализа временного ряда и построения прогнозов в 70-е годы XX века был разработан мощный метод **ARIMA** (авторегрессия и проинтегрированный метод скользящего среднего). В основном варианте этого метода требуется подобрать три неизвестных параметра — p , q и r . Параметр r по существу, совпадает с параметром несезонного дифференцирования, о котором мы говорили выше, и он позволяет путем исключения полиномиального тренда сделать временной ряд стационарным. Пусть выбором r мы добились того, чтобы ряд из остатков стал стационарным. Далее мы будем подбирать параметры авторегрессии p и скользящего среднего q . Параметры должны быть выбраны так, чтобы кореллограмма остатков временного ряда быстро затухала (укладывалась в 95 %-й доверительный интервал). Есть вариант этого метода, в котором аналогично подбираются варианты для сезонной составляющей ps , qs и rs . Не вдаваясь в детали, посмотрим, что это нам даст для рассмотренного выше примера.

В строке меню выберите **Forecast, Automatic Model Selection**. В диалоговом окне выберите столбец с данными. В поле **Number of Forecasts** введите 4 (т. е. наш прогноз будет на четыре квартала).

Automatic Forecasting

Col_1
SADJUSTED

Data: Col_1

Sampling Interval

Once Every: 1

☐ Year(s) (4-digit)
 ☐ Hour(s)
 Starting At: Q1.82

☒ Quarter(s)
 ☐ Minute(s)

☐ Month(s)
 ☐ Second(s)
 (Seasonality:) 4

☐ Day(s)
 ☐ Other

(Trading Days Adjustment)

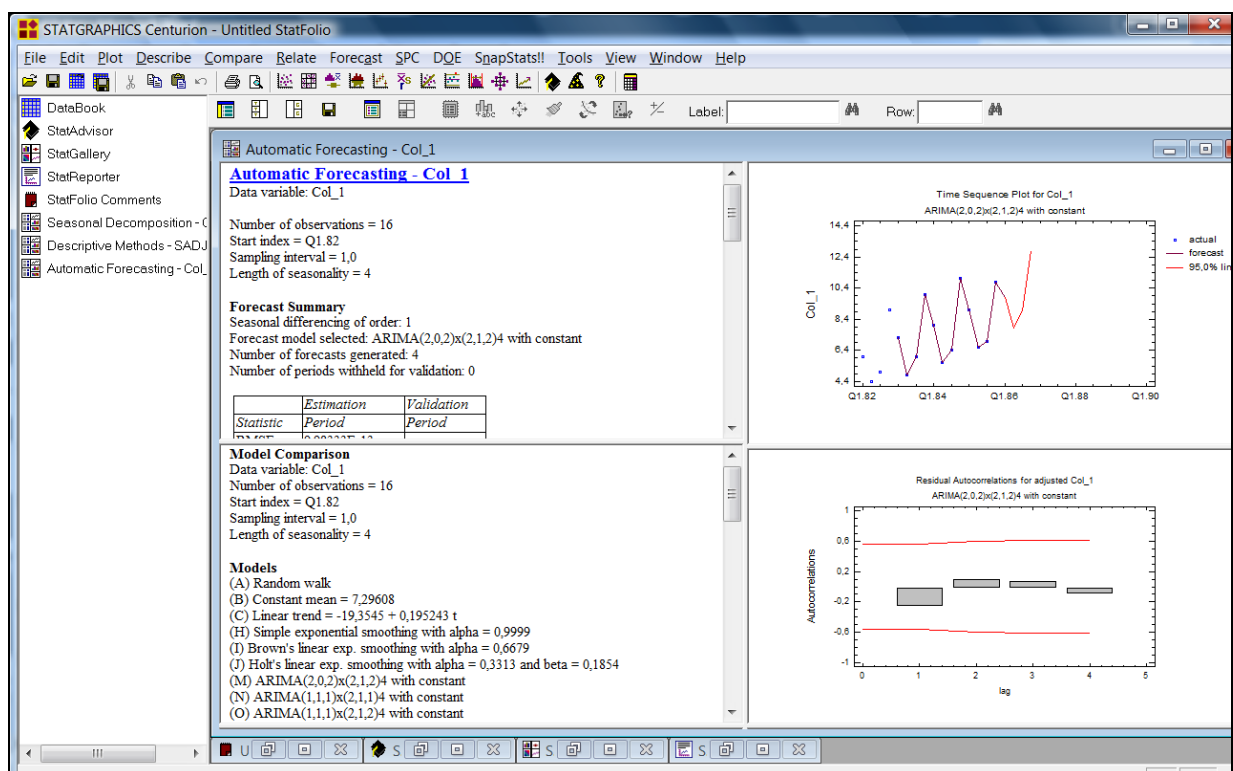
(Select)

☒ Sort column names

Number of Forecasts: 4 Withhold for Validation: 0

OK Cancel Delete Transform... Help

Нажмите **ОК**.



StatGraphics выбрал метод **ARIMA** с параметрами 2,0,2 и сезонными параметрами (2,1,2). Если раскрыть окно **Model Comparison**, можно увидеть, что

данная модель удовлетворяет всем требованиям (в окне против строки (M), в которой рассматриваются соответствующие параметры, можно видеть все ОК).

ЗАДАНИЕ 1. В файле **Среднее число яиц на несушку по США** находятся данные по месяцам, начиная с января 1938 года. Выделите сезонную компоненту, подумайте при этом, какой вид модели (мультипликативную или аддитивную) надо выбрать. Проанализируйте остатки, постарайтесь получить белый шум.

ЗАДАНИЕ 2. В файле **Series_g.sf** приведены данные о перевозках авиатранспорта по месяцам с 1 января 1950 г. Постройте модель, спрогнозируйте перевозки на 3 периода.