ЛАБОРАТОРНЫЙ ПРАКТИКУМ № 3
ПО КУРСУ «ЭКОНОМЕТРИКА»

ПАРНАЯ И МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Цель занятия – освоение навыков проведения парного и множественного регрессионного анализа в программной среде **Microsoft Excel** и **Statgraphics Plus for Windows.** В ходе занятия студенты должны научиться использовать вычислительные и графические возможности используемых пакетов, правильно и полно анализировать результаты моделирования, осуществлять спецификацию регрессионной модели, проводить ее проверку на мультиколлинеарность объясняющих переменных, гетероскедастичность и автокорреляцию остатков, знать, как следует поступать в случае их обнаружения.

Часть 1

МОДЕЛИ ПРОСТОЙ И МНОЖЕСТВЕННОЙ РЕГРЕССИИ

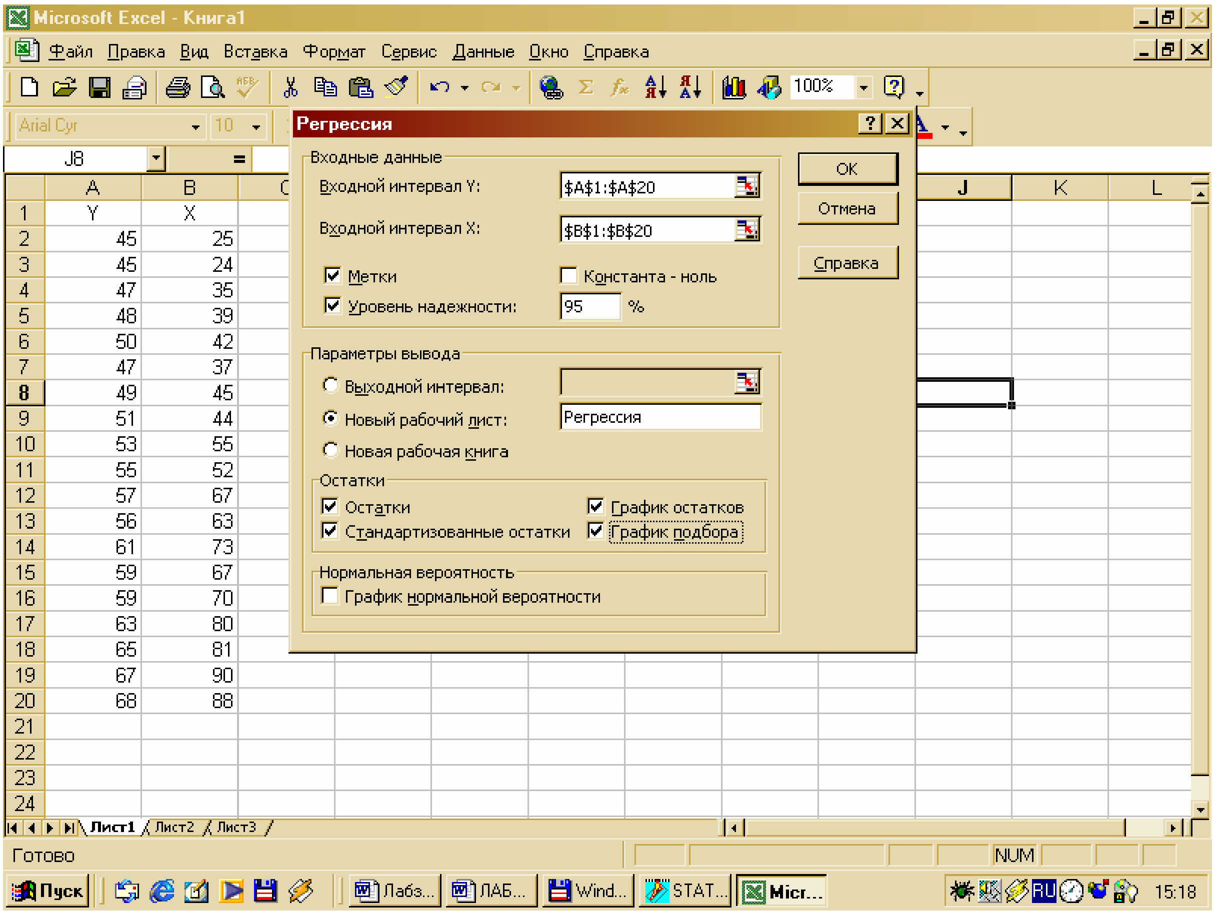
А. Пример модели простой регрессии.

Имеется выборка показателей хозяйственной деятельности одной из американских корпораций за19 кварталов. В таблице представлены следующие данные.

Номер квартала	Инвестиции, млн. долл. (Y)	Валовая прибыль, млн. долл. (X)
1	45	25
2	45	24
3	47	35
4	48	39
5	50	42
6	47	37
7	49	45
8	51	44
9	53	55
10	55	52
11	57	67
12	56	63
13	61	73
14	59	67
15	59	70
16	63	80
17	65	81
18	67	90
19	68	88

Порядок выполнения задания в пакете **Excel.**

1. Сформировать на рабочем листе исходные данные в виде столбцов массива.
2. Вызвать процедуру регрессионного анализа: **Сервис → Анализ данных → Регрессия**. В появившемся диалоговом окне необходимо задать параметры входных интервалов Y и X, выходного интервала, установить флажок в поле **Метки**, если в первой строке входного интервала записаны имена переменных. Выберите уровень надежности 95%. Обеспечьте вывод значений остатков и графиков, установив соответствующие флажки в области диалогового окна **Остатки**. Рекомендуемое состояние диалогового окна изображено ниже.
3. Нажмите кнопку **OK**.



На листе **Регрессия** будут приведены оценки параметров регрессионной модели, статистические характеристики точности и достоверности регрессионной связи, включая значения *F*-статистики и *t*-статистики, которые используются для проверки нулевой гипотезы в отношении отдельных параметров регрессии и регрессионного уравнения в целом, а также доверительные интервалы параметров регрессии для выбранного уровня надежности.

Для большей наглядности рекомендуется применить к выведенным результатам команду **Автоформат** меню **Формат**.

Проведите анализ поученного уравнения регрессии и сделайте соответствующие выводы. При проверке нулевой гипотезы в отношении параметров регрессии ориентируются на соответствующие *P*-значения (уровни значимости) в сопоставлении с критическими величинами последних, а при проверке нулевой гипотезы в отношении коэффициента детерминации – на значимость *F*.

Как показывают результаты, все проверяемые параметры данной модели статистически значимы.

Общепринятая текстовая запись полученной регрессионной модели при правильном выполнении процедуры будет иметь следующий вид (под оценками параметров регрессии приведены их стандартные ошибки):

$$\underset{(0{,}826)}{Y = 34{,}746} + \underset{(0{,}014)}{0{,}357X}, \qquad R^2 = 0{,}975, \qquad \overline{R}^2 = 0.974, \qquad F = 676{,}852.$$

Более полное исследование может быть выполнено на основе профессионального статистического пакета **Statgraphics Plus for Windows.**

Данные в рабочий лист **Statgraphics Plus for Windows** могут быть занесены заново либо скопированы из **Excel** (только числовые данные).

Для передачи данных в пакет **Statgraphics** из **Excel** скопируем содержимое в буфер, далее вызовем с рабочего стола пакет **Statgraphics**, актуализируем окно «untitled» (в версии 5.1 актуализация происходит автоматически) и перенесем в него информацию из буфера (кнопка «**Paste**» контекстного меню или сочетание клавиш **Shift + Insert**).

Диалоговое окно парного регрессионного анализа открывается при выборе пункта меню **Relate → Simple Regression**.

После нажатия кнопки **OK** на экран выводятся результаты моделирования, естественно, совпадающие с теми, которые были получены при использовании пакета **Excel.**

```
Regression Analysis - Linear model: Y = a + b*X
-----------------------------------------------------------------------------
Dependent variable: Y
Independent variable: X
-----------------------------------------------------------------------------
                               Standard          T
Parameter       Estimate         Error       Statistic        P-Value
-----------------------------------------------------------------------------
Intercept        34,7465       0,826142        42,0587         0,0000
Slope           0,357305      0,0137338        26,0164         0,0000
-----------------------------------------------------------------------------


                           Analysis of Variance
-----------------------------------------------------------------------------
Source             Sum of Squares     Df  Mean Square    F-Ratio      P-Value
-----------------------------------------------------------------------------
Model                     983,303      1      983,303     676,85       0,0000
Residual                  24,6969     17      1,45276
-----------------------------------------------------------------------------
Total (Corr.)              1008,0     18

Correlation Coefficient = 0,987674
R-squared = 97,5499 percent
R-squared (adjusted for d.f.) = 97,4058 percent
Standard Error of Est. = 1,2053
Mean absolute error = 0,963662
Durbin-Watson statistic = 1,93564 (P=0,3441)
Lag 1 residual autocorrelation = -0,0695226
```

Обратите внимание на то, что пакет дополнительно выдает статистику Дарбина-Уотсона с соответствующим уровнем значимости в скобках и коэффициент автокорреляции первого порядка. В нашем примере эти значения убедительно свидетельствуют о том, что гипотеза об отсутствии корреляции остатков не может быть отвергнута.

Для управления процедурой регрессионного анализа используется нижний ряд кнопок. Первая слева кнопка служит для возвращения в окно диалога. Вторая слева вызывает окно табличных опций.

Выбор в ней пункта **Forecasts** (Прогнозы) вызывает таблицу **Predicted Values** (Предсказанные Значения). В ней для наименьшего и наибольшего значений независимой переменной приводятся:

прогнозируемые значения зависимой переменной **(Predicted Y)**;

95%-е доверительные интервалы прогноза **(Prediction Limits)**;

95%-е доверительные интервалы так называемого среднего отклика **(Confidence Limits)**.

Ввести другие прогнозируемые значения независимой переменной и доверительной вероятности можно вызовом контекстного меню и выбором в раскрывшемся окне пункта **Pane Options.**

Имеется возможность сравнивать линейную регрессионную модель с альтернативными нелинейными моделями, которые могут быть вызваны через пункт контекстного меню **Analysis Options** (Аналитические опции);

Пункт **Unusual Residuals** (необычные остатки) позволяет выявлять выбросы, то есть наблюдения, имеющие аномальные (чрезмерно большие) остатки, а пункт **Influential Points** – наблюдения, влияющие сильнее определенного уровня на полученные оценки модели. Опция **Unusual Residuals** помогают обеспечить однородность выборки, используемой для построения регрессионной модели, путем исключения резко выделяющихся наблюдений.

Третья кнопка слева включает графические опции.

Наиболее иллюстративна опция **Plot of Fitted Model** (График эмпирической модели). Растягивание графика на весь экран включает дополнительные кнопки, одна из которых позволяет в ручном режиме исключать из выборки наблюдения с автоматическим пересчетом оценок параметров модели.

Следующая опция **Observed versus Predicted** (наблюдаемые против предсказанных) графически показывает соответствие эмпирических и предсказанных (расчетных) значений зависимой переменной остальные предназначены для анализа остатков модели и, в частности, выявления гетероскедастичности и автокорреляции.

Четвертая кнопка слева позволяет сохранять результаты моделирования в одном файле с данными.

Проверьте функционирование всех вышеупомянутых опций и их возможности, имея в виду возможность подключения через контекстное меню, вызываемое нажатием правой кнопки мыши, различных дополнительных опций (**Pane Options**).

B. *Пример модели множественной регрессии.*

Расширим решаемую задачу. Введем дополнительно две новые объясняющие переменные:

X1 – Прибыль от реализации услуг, млн. долларов;

X2 – Прибыль от реализации товаров, млн.долларов.

Расширенные выборочные данные по 19 предприятиям представлены ниже в таблице.

Инвестиции, млн. долл. (Y)	Валовая прибыль, млн. долл. (X)	Прибыль от реализации услуг, млн. долл. (X1)	Прибыль от реализации товаров, млн. долл. (X2)
45	25	15	11
45	24	14	9
47	35	22	12
48	39	21	16
50	42	17	25
47	37	20	18
49	45	12	31
51	44	21	22
53	55	17	35
55	52	18	33
57	67	19	40
56	63	20	41
61	73	21	50
59	67	26	47
59	70	22	48
63	80	26	52
65	81	30	53
67	90	29	59
68	88	32	58

A. Используйте пакет **Excel.** Введите данные в электронную таблицу.

Сначала постройте регрессионную модель с двумя объясняющими переменными: X и X1. Для этого в окне диалога **Регрессия** в поле **Входной интервал X** выделите столбцы, содержащие X1 и X2, предварительно записав их рядом. Другие действия аналогичны процедуре простой регрессии. После нажатия **OK** должна быть выведена следующая таблица.

Множественный R	0,990405707			
R-квадрат	0,980903464			
Нормированный R-квадрат	0,978516397			
Стандартная ошибка	1,096850841			
Наблюдения	19			
				Значимость F
Дисперсионный анализ				1,76863E-14
	df	SS	MS	F
Регрессия	2	988,7506917	494,3753459	410,9241442
Остаток	16	19,24930828	1,203081768	
Итого	18	1008		
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	33,19617367	1,046894106	31,70919911	7,16617E-16
X	0,323653256	0,020156781	16,05679295	2,74105E-11
X1	0,163428217	0,076801983	2,127916629	0,049237922

Все параметры уравнения регрессии статистически значимы. В обычной форме модель записывается следующим образом:

$Y = 33{,}196 + 0{,}324X + 0{,}163X1, \quad R^2 = 0{,}981, \quad \overline{R}^2 = 0{,}979, \quad F = 410{,}924.$

$(1{,}047) \ (0{,}020) \quad (0{,}077)$

Сделайте следующий шаг и введите переменную X2. После проведения процедуры регрессионного анализа должна быть получена следующая модель:

$Y = 34{,}043 + 0{,}177X + 0{,}251X1 + 0{,}162X2, \quad R^2 = 0{,}983, \quad \overline{R}^2 = 0{,}980, \quad F = 294{,}531.$

$(1{,}163) \quad (0{,}102) \quad (0{,}095) \quad (0{,}110)$

Хотя статистическая значимость уравнения в целом несомненна, а значение $\overline{R}^2$ даже слегка возросло, полученную модель нельзя признать удовлетворительной.

Параметры регрессии значительно изменились, а их стандартные ошибки резко возросли. В результате параметр при X оказался статистически незначимым.

Данные обстоятельства наводят на мысль, что включение переменной X2 привело к возникновению мультиколлинеарности – нестрогой линейной зависимости между объясняющими переменными.

Для проверки этой гипотезы воспользуйтесь процедурой корреляционного анализа. Рассчитайте коэффициенты корреляции между всеми переменными, представленными в таблице исходных данных. В меню пакета анализа выберите пункт **Корреляция**. Заполните соответствующие поля и нажмите **ОК**. Должны быть получены следующие результаты корреляционного анализа.

	Y	X	X1	X2
Y	1			
X	0,987674	1		
X1	0,820479	0,784568	1	
X2	0,971153	0,984687	0,704677	1

Высокое значение ($R = 0{,}985$) коэффициента корреляции между переменными X и X2 со всей очевидностью свидетельствуют о наличии мультиколлинеарности. Таким образом, переменные X и X2 не должны быть одновременно представлены в модели. Одну из них, имеющую наибольшее P-значение, следует исключить, а именно X2.

Выполните процедуру множественного регрессионного анализа в среде **Statgraphics Plus for Windows.**

Окно ввода данных для уравнения регрессии вызывается при нажатии кнопки «Multiple regression» или путем вывода из главного меню пункта **Relate → Multiple Regression.**

Результаты моделирования с тремя объясняющими переменными приведены ниже.

Табличные опции множественной регрессии кроме вывода итогов позволяют вывести на экран:

информацию о статистической значимости каждой из переменных при их последовательном включении в модель;

доверительные интервалы оценок параметров;

корреляционную матрицу для оценок параметров;

расчетные (прогнозные) значения переменной Y, остатки, стандартные ошибки и доверительные интервалы прогноза;

наблюдения, имеющие необычные (чрезмерно большие) остатки и наблюдения, влияющие сильнее определенного уровня на полученные оценки модели.

Графические опции дополнительно к опциям для простой регрессии позволяют показать эффект влияния каждой независимой переменной.

```
Multiple Regression Analysis
-----------------------------------------------------------------------------
Dependent variable: Y
-----------------------------------------------------------------------------
                                       Standard          T
Parameter               Estimate         Error       Statistic        P-Value
-----------------------------------------------------------------------------
CONSTANT                 34,0426        1,16343        29,2605         0,0000
X                       0,176823       0,101784        1,73724         0,1028
X1                      0,251367      0,0952893        2,63794         0,0186
X2                      0,161671       0,110004        1,46969         0,1623
-----------------------------------------------------------------------------

                           Analysis of Variance
-----------------------------------------------------------------------------
Source             Sum of Squares     Df  Mean Square    F-Ratio      P-Value
-----------------------------------------------------------------------------
Model                     991,174      3      330,391     294,53       0,0000
Residual                  16,8263     15      1,12176
-----------------------------------------------------------------------------
Total (Corr.)              1008,0     18

R-squared = 98,3307 percent
R-squared (adjusted for d.f.) = 97,9969 percent
Standard Error of Est. = 1,05913
Mean absolute error = 0,799107
Durbin-Watson statistic = 1,90226 (P=0,2842)
Lag 1 residual autocorrelation = -0,00845982
```

Вредные (ведущие к возникновению мультиколлинеарности) или статистически незначимые переменные могут быть удалены из модели вручную поочередно, в порядке убывания их Р-значений, либо автоматически. В последнем случае через контекстное меню **Analysis Options** вызываются дополнительные опции множественного регрессионного анализа **Multiple Regression Options.**.

Далее, переключатель **Fit** ставится в положение **Forward Selection** (Отбор последовательным включением) или **Backward Selection** (Отбор последовательным исключением).

Применение любой из названных процедур к табличным данным приводит к получению подобранной модели.

```
Multiple Regression Analysis
-----------------------------------------------------------------------------
Dependent variable: Y
-----------------------------------------------------------------------------
                                       Standard          T
Parameter               Estimate         Error       Statistic        P-Value
-----------------------------------------------------------------------------
CONSTANT                 33,1962       1,04689         31,7092         0,0000
X                       0,323653     0,0201568         16,0568         0,0000
X1                      0,163428      0,076802         2,12792         0,0492
-----------------------------------------------------------------------------

                           Analysis of Variance
-----------------------------------------------------------------------------
Source             Sum of Squares     Df  Mean Square    F-Ratio      P-Value
-----------------------------------------------------------------------------
Model                     988,751      2      494,375     410,92       0,0000
Residual                  19,2493     16      1,20308
-----------------------------------------------------------------------------
Total (Corr.)              1008,0     18

R-squared = 98,0903 percent
R-squared (adjusted for d.f.) = 97,8516 percent
Standard Error of Est. = 1,09685
Mean absolute error = 0,855518
Durbin-Watson statistic = 2,09216 (P=0,3050)
Lag 1 residual autocorrelation = -0,118399
```

В завершение занятия выполните следующие действия:

1.Проанализируйте результаты регрессионного анализа, используя необходимые опции, и убедитесь в адекватности модели и статистической значимости ее параметров.

2.Сделайте прогноз значений переменной Y для следующих значений факторных переменных, подставив их в соответствующие столбцы, и рассчитайте 95-процентные интервалы прогноза.

X	50	60	70	80
X1	30	30	40	40

Часть 2

МОДЕЛИ ЛИНЕЙНОЙ РЕГРЕССИИ С ГЕТЕРОСКЕДАСТИЧНЫМИ И АВТОКОРРЕЛИРОВАННЫМИ ОСТАТКАМИ

Вернемся к модели простой регрессии. Теперь, однако, усложним задание. Пусть множеству выборочных значений валовой прибыли (X) соответствуют три альтернативных множества значений инвестиций: Y, Y1 и Y2. Соответствующие числовые данные приведены в таблице.

X	Y	Y1	Y2
25	45	44	46
24	45	43	47
35	47	47	47
39	48	50	49
42	50	51	49
37	47	49	45
45	49	49	49
44	51	48	47
55	53	57	53
52	55	56	51
67	57	55	61
63	56	54	60
73	61	64	62
67	59	62	62
70	59	56	61
80	63	60	65
81	65	68	64
90	67	71	64
88	68	61	63

При проведении регрессионного анализа необходимо проверять выполнение условий Гаусса–Маркова. К наиболее частым нарушениям относятся *гетероскедастичность* (нарушение второго условия) и *автокорреляция* случайного члена (нарушение третьего условия). Автокорреляция обычно присуща данным, представленным в виде динамических рядов.

При отсутствии гетероскедастичности обычные коэффициенты регрессии имеют наиболее низкую дисперсию среди всех несмещенных оценок, являющихся линейными функциями от наблюдений y. Если имеет место гетероскедастичность, то оценки МНК, которые мы до сих пор использовали, неэффективны. Можно, по меньшей мере, в принципе найти другие оценки, которые имеют меньшую дисперсию и, тем не менее, являются несмещенными.

Гетероскедастичность может быть обнаружена визуально на основе изучения графика остатков. Более строгий вывод может быть получен при использовании специальных тестов

Тест ранговой корреляции Спирмена. При выполнении теста ранговой корреляции Спирмена предполагается, что дисперсия случайного члена будет либо увеличиваться, либо уменьшаться по мере увеличения x, и поэтому в регрессии, оцениваемой с помощью

МНК, абсолютные величины остатков и значениях будут коррелированны. Данные пол: и остатки упорядочиваются, и коэффициент ранговой корреляции определяется как

$$r_{xe} = 1 - \frac{6\sum D_i^2}{n(n^2-1)},$$

где D_i^2 – разность между рангом x_i и рангом $|e_i|$.

Тест Глейзера. Тест Глейзера позволяет несколько более тщательно рассмотреть характер гетероскедастичности. Мы снимаем предположение о том, что σ_i, пропорционально x_i, и хотим проверить, может ли быть более подходящей какая-либо другая функциональная форма, например

$$\sigma_i = \alpha + \beta x_i^\gamma$$

Чтобы использовать данный метод, следует сначала оценить регрессионную зависимость y от x с помощью обычного МНК. Затем вычислить абсолютные величины остатков $|e_i|$ по приведенной функции для данного значения γ. Можно построить несколько таких функций, изменяя значение γ. В каждом случае нулевая гипотеза об отсутствии гетероскедастичности будет отклонена, если оценка β значимо отличается от нуля. Если при оценивании более чем одной функции получается значимая оценка β, то ориентиром при определении характера гетероскедастичности может служить наилучшая из них.

Последствия *автокорреляции* в некоторой степени сходны с последствиями гетероскедастичности. Коэффициенты регрессии остаются несмещенными, но становятся неэффективными, и их стандартные ошибки оцениваются неправильно (вероятно, они смещаются вниз, т. е. занижаются).

Предположение о наличии автокорреляции также может быть сделано на основе изучения графиков остатков. Однако большинство статистических пакетов предлагают в качестве стандартной процедуры при проведении регрессионного анализа расчет коэффициента автокорреляции остатков первого порядка (лаг равен 1) r и статистики Дарбина–Уотсона d.

$$r = \frac{\sum e_{t-1}e_t}{\sum e_t^2},$$

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}.$$

Откройте рабочий лист пакета **Statgraphics Plus for Windows** и занесите туда исходные данные из последней таблицы.

Повторите регрессию X на Y и убедитесь в отсутствии признаков гетероскедастичности и автокорреляции остатков. Для этого нажмите кнопку графических опций, установите соответствующие флажки и рассмотрите графики остатков, а также проанализируйте значения r и d.

Перейдите теперь к изучению регрессионной зависимости Y1 от X.

После проведения необходимых расчетов получается следующая модель:

Y1 = 35,056 + 0,352X; R^2=0,856; $\bar{R}^2 = 0,847$; F=101,01; d=1,958; r= –0,058.

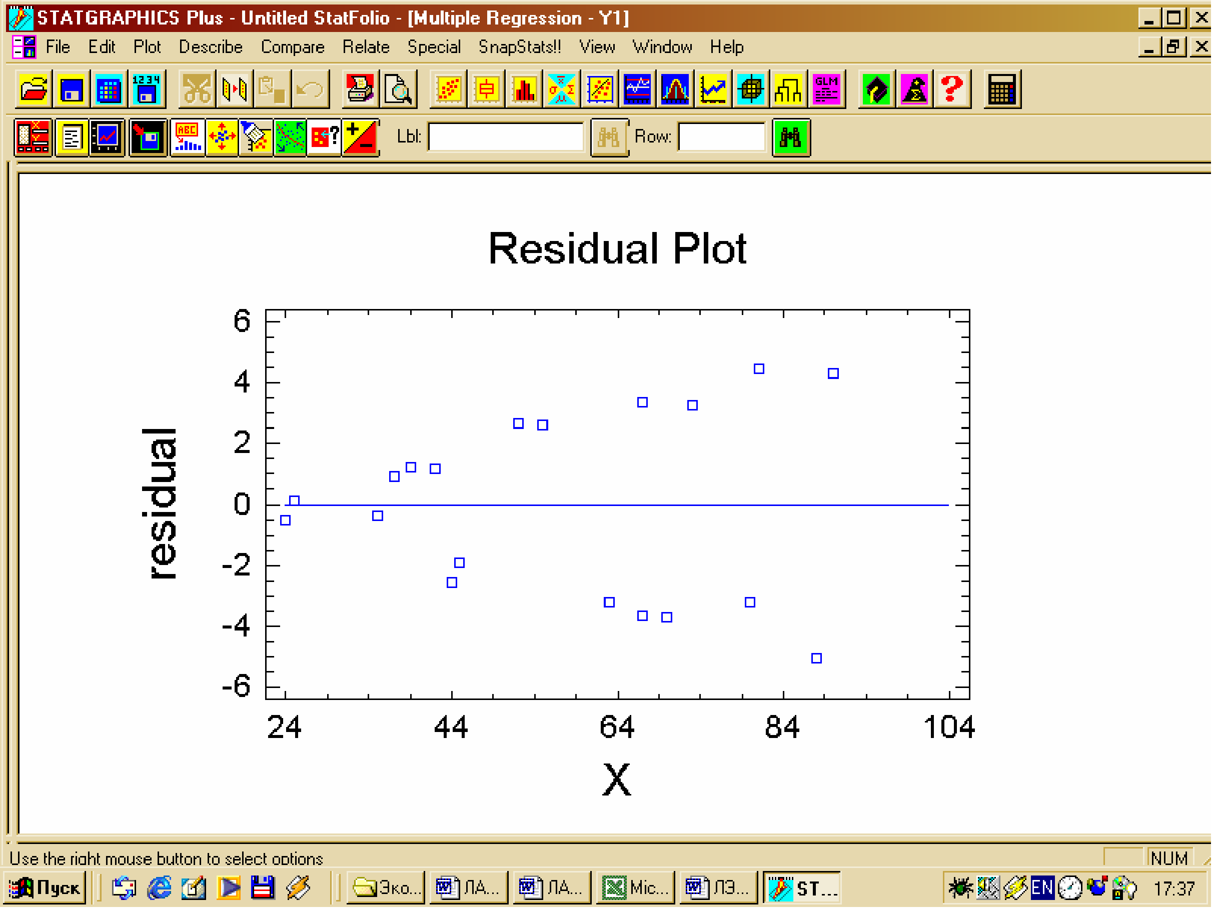
(2,106) (0,035)

Все оценки параметров модели представляются статистически значимыми. Однако, посмотрев на график **Residuals versus X** (Остатки против X), можно увидеть, что абсолютная величина остатков имеет тенденцию возрастать с ростом объясняющей переменной. Это дает основания предположить наличие гетероскедастичности.

Проверим данную гипотезу через тест ранговой корреляции Спирмена. Для этого сохраните остатки в файле с исходными данными, нажав четвертую слева кнопку сохранения результатов (**Save Results Options**), и установите флажок в окошке **Residuals**. Дайте имя новой переменной, характеризующей остатки, например, R1, и

нажмите **OK**. Эта переменная теперь будет присутствовать в окнах диалога наряду с исходными переменными.

Введите переменную, отображающую абсолютные величины полученных остатков, назвав ее, например, AbsR1. Для этого в диалоговом окне **Modify Column** выделите пункт **Formula** и нажмите кнопку **Define.**



Откроется диалоговое окно **Generate Data,** в правой части которого перечислены операторы формул, в том числе и Abs(?), т.е. абсолютная величина. Двойным щелчком переведите оператор в верхнее поле, а вместо знака вопроса подставьте в скобки название переменной, в данном случае R1.

Затем вызовите окно диалога **Describe → Numeric Data → Multiple Variable Analysis**. Введите переменные X и AbsR1 и нажмите OK. Откройте окно табличных опций и выберите опцию **Rank Correlations.** Полученное значение коэффициента Спирмена $r_{x,e} = 0{,}947$ статистически значимо, что говорит о гетероскедастичности остатков.

Для более точной проверки примените тест Глейзера. График остатков позволяет предположить линейную форму зависимости стандартного отклонения остатков σ отX. Следовательно, можно принять $\gamma = 1$.

Постройте уравнение регрессионной зависимости модуля остатков AbsR1 от X. Должна быть получена модель вида:

AbsR1 = –1,296 + 0,068X; $R^2 = 0{,}909$; $\bar{R}^2 = 0.904$; $F = 169{,}54$; $d = 2{,}092$; $r = -0{,}074$.

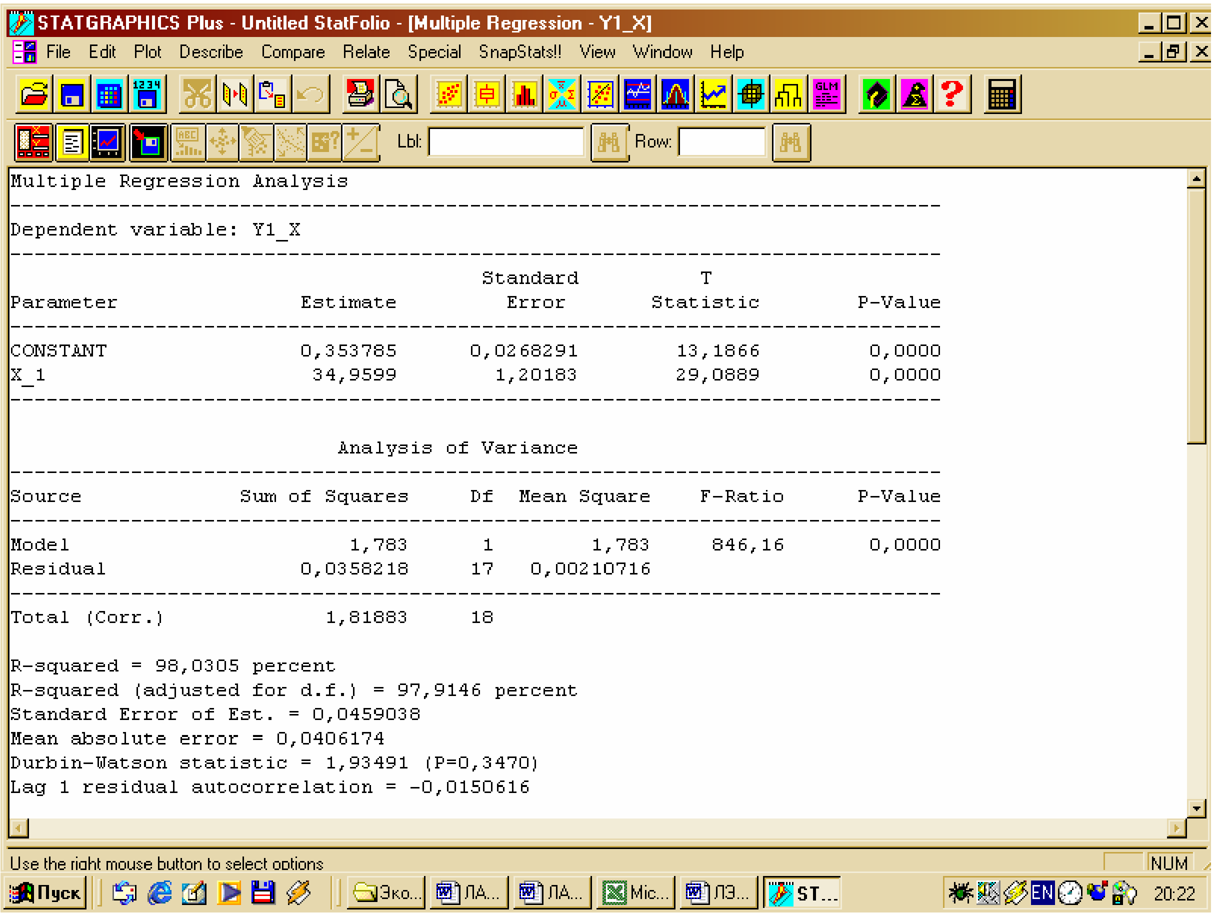
(0,312) (0,005)

Очевидно, все параметры модели статистически весьма значимы, так что гетероскедастичность можно считать доказанной. В целях устранения гетероскедастичноси применим обобщенный метод наименьших квадратов. Воспользуемся тем обстоятельством, что остатки имеют тесную линейную связь с независимой переменной X.

Разделите каждое наблюдение на соответствующее ему значение *x*, то уравнение примет вид:

$$\frac{y_i}{x_i} = \alpha \frac{1}{x_i} + \beta \frac{x_i}{x_i} + \frac{e_i}{x_i}.$$

Затем оцените регрессионную зависимость *y/x* от *1/x*, включив в уравнение постоянный член. Коэффициент при *1/x* будет эффективной оценкой *α*, а постоянный член — эффективной оценкой *β*. Проделав соответствующие расчеты в файле исходных данных, введите новые переменные Y1_X и X_1. В результате регрессионный анализ зависимости первой от второй должен дать следующие результаты.




```
Multiple Regression Analysis
-----------------------------------------------------------------------------
Dependent variable: Y1_X
-----------------------------------------------------------------------------
                                         Standard          T
Parameter                Estimate          Error       Statistic       P-Value
-----------------------------------------------------------------------------
CONSTANT                 0,353785       0,0268291       13,1866         0,0000
X_1                       34,9599         1,20183       29,0889         0,0000
-----------------------------------------------------------------------------

                           Analysis of Variance
-----------------------------------------------------------------------------
Source             Sum of Squares     Df  Mean Square     F-Ratio      P-Value
-----------------------------------------------------------------------------
Model                       1,783      1        1,783      846,16       0,0000
Residual                0,0358218     17   0,00210716
-----------------------------------------------------------------------------
Total (Corr.)             1,81883     18

R-squared = 98,0305 percent
R-squared (adjusted for d.f.) = 97,9146 percent
Standard Error of Est. = 0,0459038
Mean absolute error = 0,0406174
Durbin-Watson statistic = 1,93491 (P=0,3470)
Lag 1 residual autocorrelation = -0,0150616
```

Итоговая модель выглядит следующим образом:

$$\frac{Y}{X} = 34{,}960\frac{1}{X} + 0{,}354; \quad R^2 = 0{,}980; \ \bar{R}^2 = 0{,}979; \ F = 846{,}16; \ d = 1{,}935; \ r = -0{,}015.$$

(1,202) (0,027)

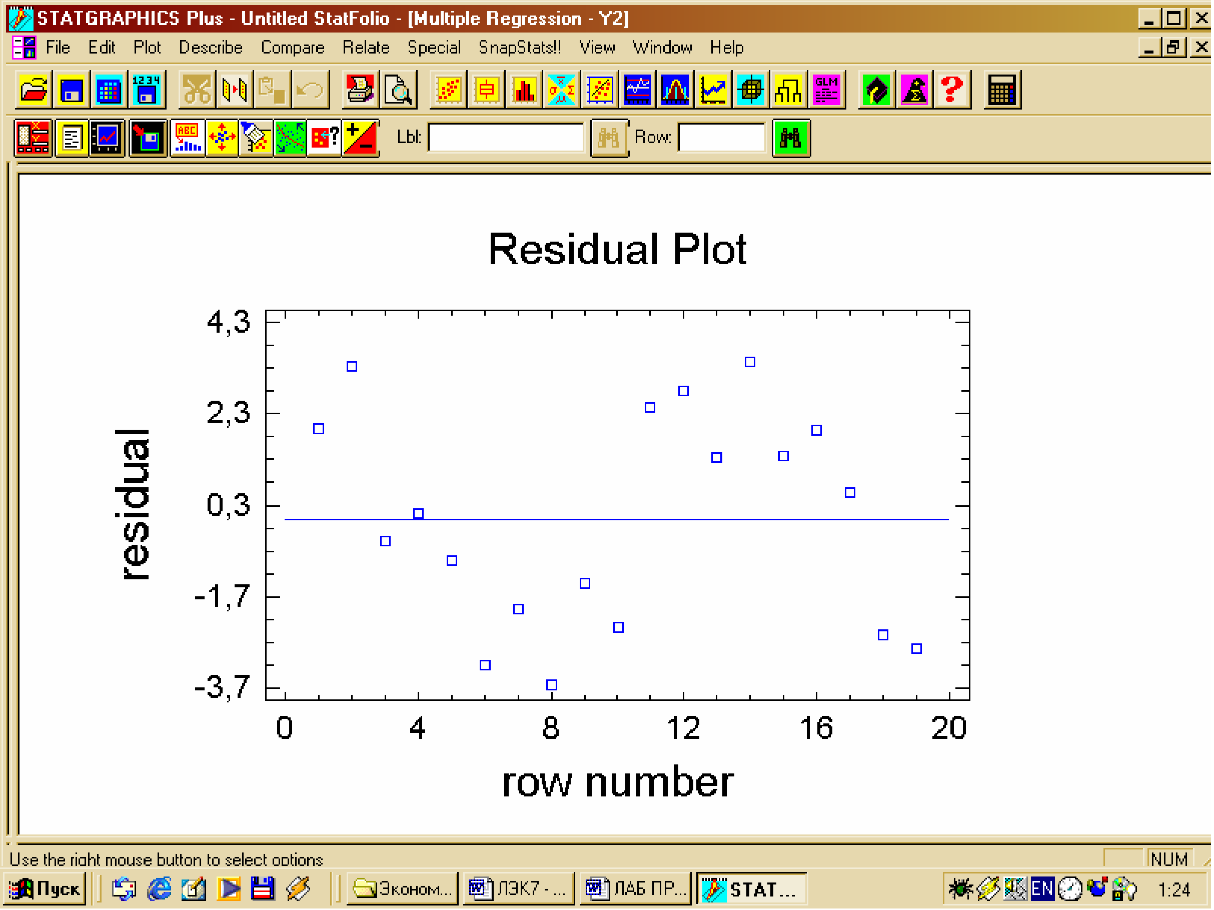
Как видите, качество регрессии повысилось.

Перейдите теперь к рассмотрению регрессионной модели зависимости Y2 от X.

Проведя соответствующие модельные расчеты, получим следующие оценки параметров:

$$Y2 = \underset{(1{,}645)}{35{,}372} + \underset{(0{,}027)}{0{,}346}X; \quad R^2 = 0{,}904; \quad \bar{R}^2 = 0{,}898; \quad F = 159{,}38; \quad d = 0{,}799; \quad r = 0{,}540.$$

Значения коэффициента автокорреляции остатков первого порядка (лаг равен 1) r и статистики Дарбина–Уотсона d убедительно свидетельствуют о наличии значительной автокорреляции остатков. Такой же вывод можно сделать после выбора графической опции **Residuals versus Row Number** (Остатки против порядковых номеров) и изучения соответствующего графика.



Примите меры по устранению автокорреляции.

Для этого следует воспользоваться преобразованием Бокса–Дженкинса:

$$y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + e_t - \rho e_{t-1},$$

где ρ – параметр авторегрессии.

Примените данную процедуру, приняв в качестве оценки величины ρ, полученное в модели значение первого коэффициента автокорреляции r. Обозначьте:

$Y2\#_t = Y2_t - \rho Y2_{t-1}$, $X\#_t = X_t - \rho X_{t-1}$, $\alpha\# = \alpha(1-\rho)$.

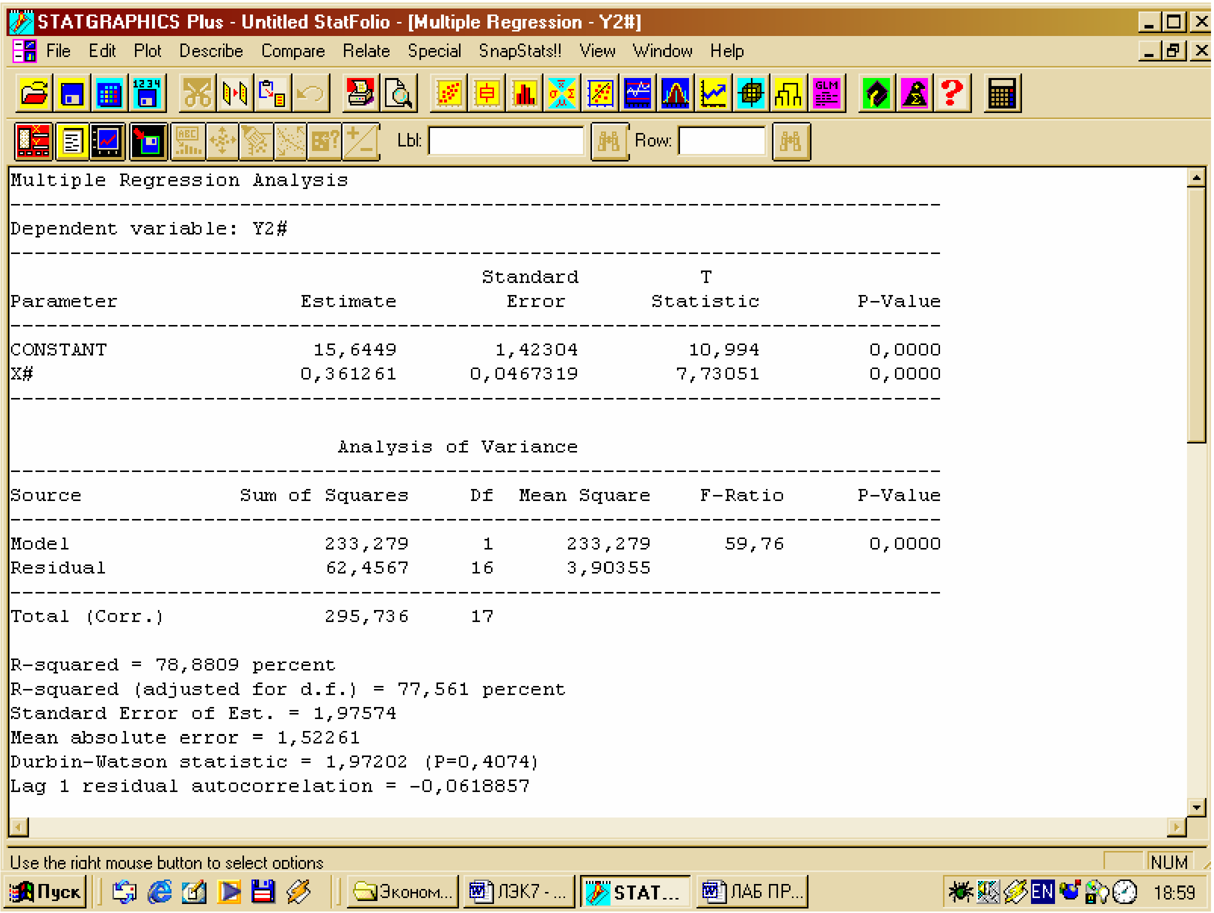
Тогда преобразованное уравнение

$Y2\#_t = \alpha\# + \beta X\#_t + u_t$, где $\alpha\# = \alpha(1-\rho)$, не содержит автокорреляцию, поскольку $u_t = e_t - \rho e_{t-1}$ независимы.

В качестве оценки параметра ρ примем полученное в модели значение первого коэффициента автокорреляции $r = 0{,}54$. Рассчитайте с использованием окон диалога **Modify Column** и **Generate Data** значения новых переменных Y2# и X# и постройте уравнение регрессии.

Для расчета X# скопируйте переменную X сначала *<u>без первого наблюдения</u>* в переменную X_, а затем *<u>без последнего наблюдения</u>* и со сдвигом на один шаг вперед в переменную X_1. Таким образом, каждому наблюдению X_t будет соответствовать наблюдение с лагом 1 X_{t-1}. Значения X# рассчитываются по формуле X# = X_-0,54*X_1.

Аналогичным образом, создав переменные У2_ и У2_1, где У2_1 является переменной У2 с лагом 1, определяем значения переменной У2# (т.е. по формуле У2# = У2_-0,54*У2_1).



Полученные статистические оценки достаточно убедительно свидетельствуют не только об очевидной статистической значимости параметров модели, но и об отсутствии автокорреляции остатков.

С учетом того, что $\alpha = \dfrac{\alpha\#}{1-r} = \dfrac{15{,}645}{1-0.54} = 34{,}011$, очищенное от автокорреляции исходное уравнение регрессии может быть записано следующим образом:

$$Y2 = 34{,}011 + 0{,}361X;\quad R^2{=}0{,}789;\quad \overline{R}^2 = 0{,}776;\quad F{=}\ 59{,}76;\quad d{=}\ 1{,}972;\quad r{=}{-0{,}062}\ .$$

(3,093) (0,047)

Сравнивая оценки регрессионной модели, нетрудно видеть, что стандартные ошибки оценок МНК меньше стандартных ошибок оценок, полученных в результате авторегрессионного преобразования. Оценки МНК оказываются заниженными и, следовательно, неэффективными.

В качестве самостоятельного задания повторите процесс моделирования на основе преобразования Бокса-Дженкинса, выбрав в качестве оценки параметра авторегрессии коэффициент авторегрессии первого порядка, получаемый при оценивании по МНК уравнения регрессии $e_t = \rho e_{t-1} + u_t$.

Эконометрика

Учебник