

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Оренбургский государственный университет

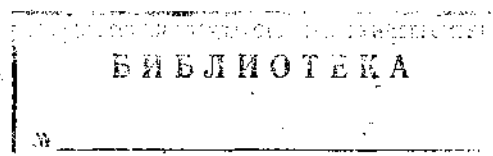
Кафедра математических методов и моделей в экономике

А.Г. РЕННЕР, Г.Г. АРАЛБАЕВА, О.А. ЗИНОВЬЕВА

РЕГРЕССИОННЫЙ АНАЛИЗ

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ
К ЛАБОРАТОРНОМУ ПРАКТИКУМУ**

Рекомендовано к изданию Редакционно-издательским советом
Оренбургского государственного университета



Оренбург 2002

1
Э
Ф
к
Х

519.2 (075)

ББК 22.172я7

Р-39

УДК 519.233.5(076.5)

Рецензент

Кандидат экономических наук Дьяконова С.В..

Реннер А.Г., Аралбаева Г.Г., Зиновьева О.А.

Р-39 Регрессионный анализ: Методические указания к лабораторному практикуму. -Оренбург:ОГУ, 2002. - 31с

Методические указания содержат описание реализации регрессионного анализа средствами диалоговой системы STADIA и статистического пакета STATISTICA с последующим анализом результатов, а также индивидуальные задания для лабораторной работы. Методические указания предназначены для выполнения лабораторной работы по курсу «Математическая статистика» для студентов всех специальностей.

Р.н. 23Р09432002

ББК22.172я7

© Реннер А.Г., 2002
© Аралбаева Г.Г., 2002
© Зиновьева О.А., 2002
© Оренбург, ОГУ, 2002

Содержание

Введение.....	4
1 Описание лабораторной работы №5.....	4
2 Постановка задачи.....	4
3 Порядок выполнения работы в системе STADIA.....	4
4 Порядок выполнения работы в системе STATISTICA.....	13
5 Интервальные оценки для параметров связи	25
6 Содержание письменного отчета.....	26
7 Вопросы к защите.....	26
Список использованных источников.....	26
Приложение А (обязательное).....	27

Введение

Под регрессионным анализом будем понимать методы оценки и анализа зависимости осредненных значений $My/x_1, x_2, \dots, x_k = f(x_1, x_2, \dots, x_k)$ от неслучайных переменных x_1, x_2, \dots, x_k , в качестве которых отбираются факторы, оказывающие существенное влияние на результирующий показатель. Как правило, помимо существенных факторов на результирующий показатель оказывают влияние множество второстепенных факторов, совместное влияние которых следует учесть в модели при ее построении и анализе.

В качестве функции регрессии берут функцию $\tilde{f}(x_1, x_2, \dots, x_k, \bar{\beta})$ из некоторого параметрического класса функций, хорошо аппроксимирующую $f(x_1, x_2, \dots, x_k)$ и, по возможности, «удобную» как в вычислительном плане, так и в плане оценки ее параметров.

1 Описание лабораторной работы №5

Лабораторная работа №5 включает следующие этапы:

- постановку задачи;
- ознакомление с порядком выполнения работы в пакетах STADIA., STATISTICA;
- выполнение расчетов индивидуальных задач на компьютере и анализ результатов;
- подготовку письменного отчета с выводами по работе;
- защиту лабораторной работы.

2 Постановка задачи

По данным Приложения А провести регрессионный анализ.

1. Подобрать и оценить функцию регрессии, наилучшую по качеству подгонки.
2. Исследовать (если это возможно) уравнение регрессии на значимость.
3. Для значимой модели регрессии исследовать значимость коэффициентов.
4. Построить доверительные интервалы для значимых параметров связи.
5. Провести экономический анализ результатов.

3 Порядок выполнения работы в системе STADIA

Запустить диалоговую систему STADIA. В меню системы **Окна** → **Установки** необходимо установить величину уровня значимости. Установим ее равной 0.05.

Подготовка данных. Ввести данные для проведения регрессионного анализа. Окно с частью исходных данных представлено на рисунке 1

Таблица данных										
	y	x1	x2	x4	x5	x6	x7	x8	x9	x10
1	1.450	6.400	13.260							
2	1.300	7.800	10.160							
3	1.370	9.760	13.720							
4	1.650	7.900	12.850							
5	1.910	5.350	10.630							
6	1.680	9.900	9.120							
7	1.940	4.500	25.840							
8	1.890	4.880	23.390							
9	1.940	3.460	14.680							
10	2.060	3.600	10.050							
11	1.960	3.560	13.990							
12	1.020	5.650	9.680							
13	1.850	4.280	10.050							
14	.880	8.850	9.130							
15	.620	8.530	5.370							
16	1.090	7.190	9.860							

Рисунок 1

На первом этапе необходимо выбрать результирующую переменную Y и одну из переменных X_i , для которых подберем и оценим функцию регрессии с наивысшим коэффициентом детерминации.

Выбор процедуры. Для проведения регрессионного анализа выберем в меню **Статистические методы** в разделе **Регрессионный анализ** пункт **Простая регрессия**. В появившемся на экране запросе **Переменные регрессии** необходимо указать в качестве Y – **переменной** объясняемую (результатирующую, зависимую) переменную, а в качестве X – **переменной** - объясняющую (независимую) переменную (рисунок 2).

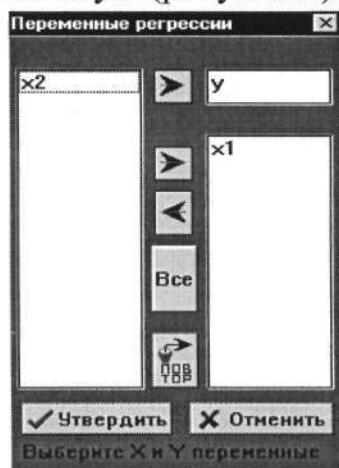


Рисунок 2

После нажатия кнопки запроса **Утвердить** программа выдаст меню моделей регрессии (рисунок 3). В правой части меню выбора приведены различные спецификации регрессионных моделей. Выбор ускоряется нажатием на быстрые клавиши, сопоставленные моделям.

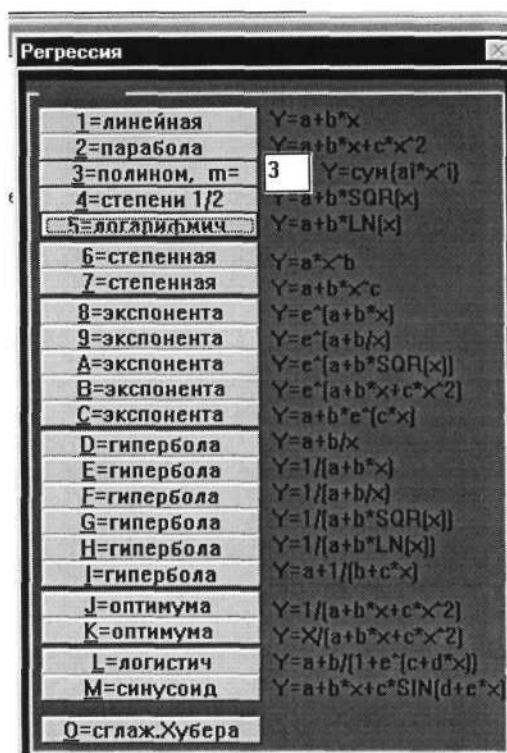


Рисунок 3

Выберем пункт 1 – линейная или просто нажмем клавишу 1.

Далее появляется бланк запроса значений независимой(мых) переменной (ных) X (рисунок 4), для которого по полученной оценке уравнения регрессии нужно прогнозировать. Запросы повторяются до нажатия кнопки **Отменить**.



Рисунок 4

Далее процедура предлагает построить график экспериментальных точек и график функции регрессии (рисунок 5). Построенный график может быть сохранен в отдельном графическом окне.

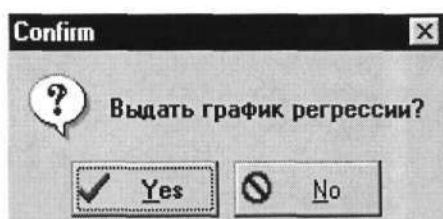


Рисунок 5

При положительном ответе нажимаем на кнопку **Yes**, выдается окно в данном примере с линейной регрессионной моделью и зоной доверительного интервала (рисунок 6).

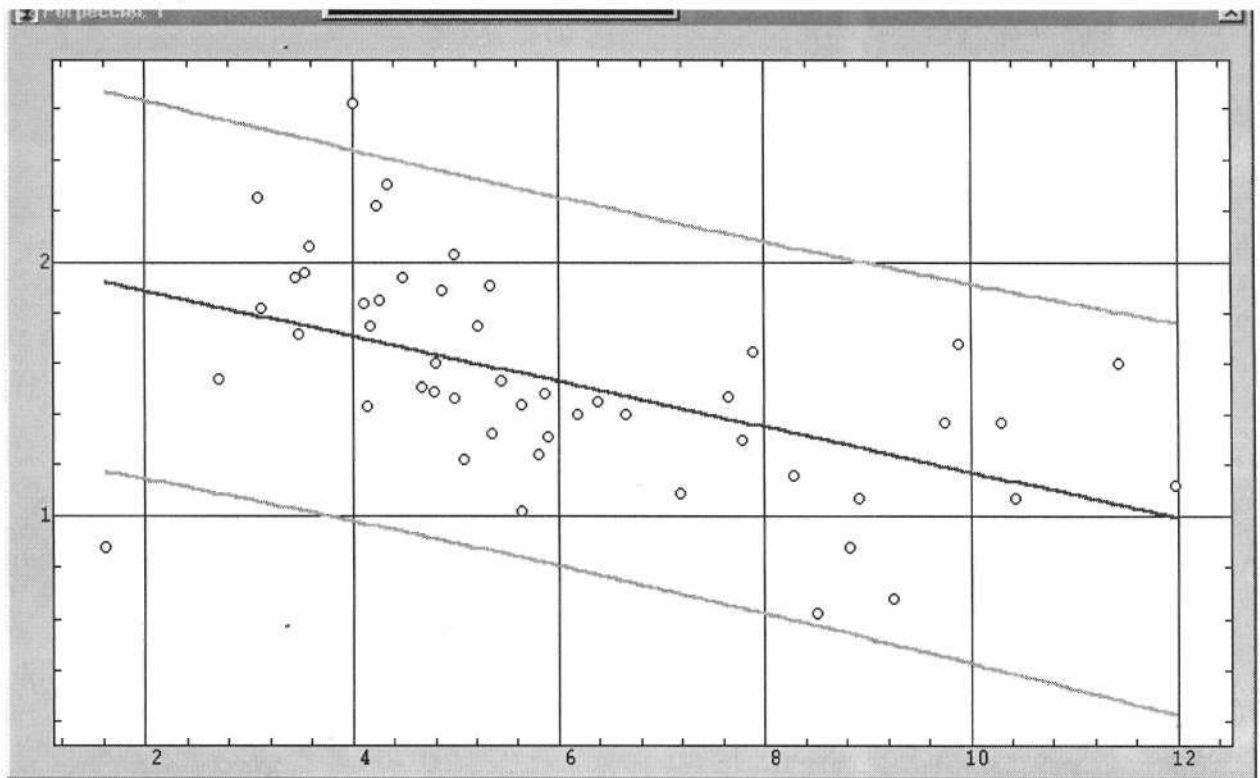


Рисунок 6

Исходные данные распределены внутри доверительного интервала.

Затем пользователю предлагается меню дополнительных возможностей процедуры (рисунок 7).



Рисунок 7

Если мы хотим сразу увидеть результаты расчетов, нажимаем на кнопку **Закончить анализ**. Результаты расчетов оценок параметров парного линейного уравнения регрессии и их анализа представлены в виде отчета на рисунке 8.

Результаты

ПРОСТАЯ РЕГРЕССИЯ. Файл:

Переменные: x1, y

Модель: линейная $Y = a_0 + a_1 * x$

Коефф.	a0	a1
Значение	2,07	-0,0895
Ст.ошиб.	0,137	0,0212
Значим.	0	0,0002

Источник	Сум.кв.др.	Степ.св	Средн.кв.др.
Регресс.	2,33	1	2,33
Остаточн	6,26	48	0,13
Вся	8,59	49	

Множество R	R^2	R^2прив	Ст.ошиб.	F	Значим
0,52079	0,27122	0,25604	0,36111	17,9	0

Гипотеза 1: <Регрессионная модель адекватна экспериментальным данным>

Рисунок 8

Стандартная выдача результатов регрессионного анализа включает следующие последовательные компоненты:

- оценки параметров уравнения регрессии или модель, записанную в общем виде;
- таблицу значений коэффициентов модели со стандартными ошибками вычисления каждого коэффициента и со значимостями коэффициентов;
- таблицу дисперсионного анализа со столбцами:

Источник Сумма квадратов Степень свободы Средняя сумма квадратов

и тремя строками для сумм квадратов: $Q_{\text{рег}}$ - регрессионная, $Q_{\text{ост}}$ - остаточная и $Q_{\text{общ}}$ -общая; таблицу оценки качества построенной модели с выводами проверки нулевой гипотезы ($H_0: b_1=0$ при альтернативной $H_1: b_1 \neq 0$), в которой приводятся: оценка множественного коэффициента корреляции (R); оценка коэффициента детерминации (R^2); приведенная или несмещенная оценка R^2 ; стандартная ошибка вычислений; значение статистики Фишера F и значимость p нулевой гипотезы. Как видно из отчета, оценка коэффициента детерминации составила 0.271. Это означает, что 27.1% доли результирующей переменной Y объясняется факторной переменной X_1 и дополняющая до 100% величина соответствует изменению неконтролируемых (случайных) показателей. Полученная модель имеет значимо отличный от нуля коэффициент $b_1 = -0.0895$, поскольку значимость коэффициента не превышает заданного уровня 0.05, а модель в целом адекватна экспериментальным данным. Оценка уравнения регрессии выглядит следующим образом:

$$\hat{Y} = 2.07 - 0.0895X_1 .$$

(0.0212)

Результаты расчетов могут также содержать остатки и их стандартизованные значения, а также стандартные ошибки остатков и доверительные интервалы для них (в виде допустимого отклонения для 95% уровня доверия).

Кoeff.		a0	a1
Значение		2,07	-0,0895
Ст. ошиб.		0,137	0,0212
Значим.		0	0,0002

Источник	Сум. квадр.	Степ. св.	Средн. квадр.
Регресс.	2,33	1	2,33
Остаточн	6,26	48	0,13
Вся	8,59	49	

Множеств R	R^2	R^2прив	Ст. ошиб.	F	Значим
0,52079	0,27122	0,25604	0,36111	17,9	0

Гипотеза 1: <Регрессионная модель адекватна экспериментальным данным>

x1=6, Y=1,54		Урегр	остаток	Ст. остат	Ст. ошиб	Довер. инт
Хэксп	Yэксп					
6,4	1,45	1,5	-0,0495	-0,139	0,365	0,724
7,8	1,3	1,37	-0,0742	-0,208	0,367	0,728
9,76	1,37	1,2	0,171	0,479	0,373	0,741
7,9	1,65	1,37	0,285	0,797	0,367	0,728
5,35	1,91	1,59	0,316	0,885	0,365	0,725
9,9	1,68	1,19	0,494	1,38	0,374	0,742
4,5	1,94	1,67	0,27	0,756	0,366	0,727
4,88	1,89	1,64	0,254	0,712	0,365	0,726
3,46	1,94	1,76	0,177	0,496	0,369	0,732
3,6	2,06	1,75	0,31	0,867	0,368	0,731
3,56	1,96	1,75	0,206	0,577	0,368	0,731
5,65	1,02	1,57	-0,547	-1,53	0,365	0,724
4,28	1,85	1,69	0,161	0,449	0,367	0,728
8,85	0,88	1,28	-0,4	-1,12	0,37	0,734

Рисунок 9

Процедура также позволяет вывести график остатков при нажатии на клавишу **1 Анализ остатков** (рисунок 10) и сохранить остатки отдельной переменной в базе данных пакета.

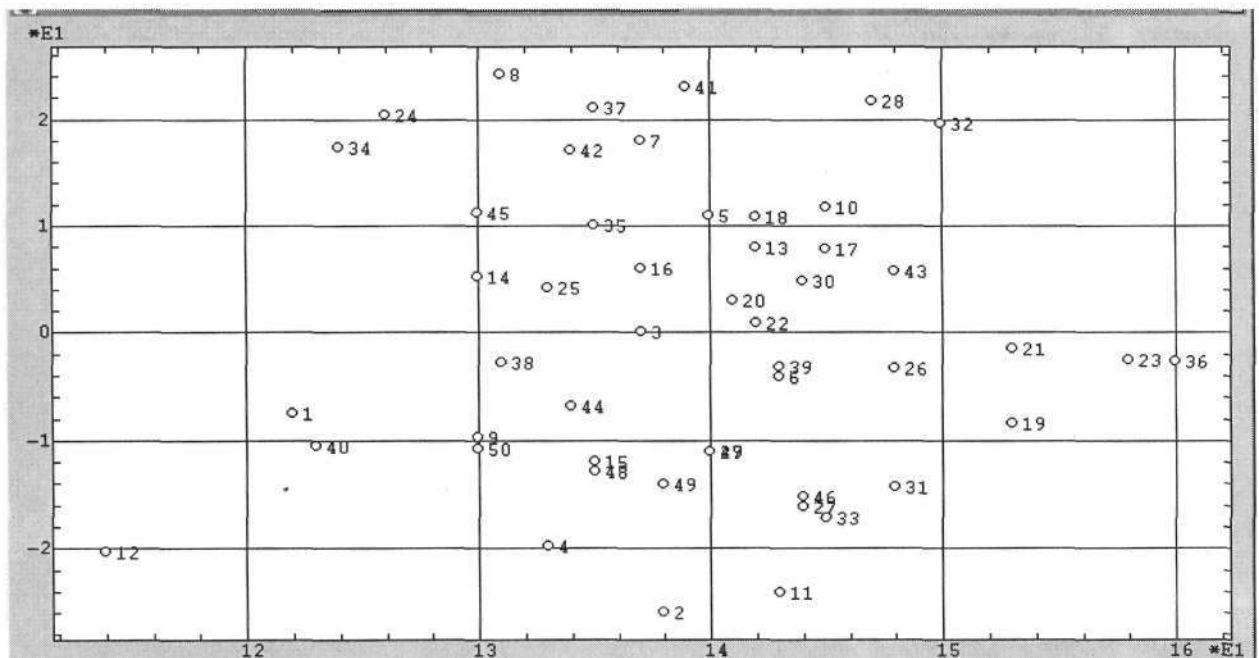


Рисунок 10

Распределение остатков на рисунке 10 достаточно однородно.

Процедура 2 = Прогнозирование позволяет получить прогноз вперед с заданным числом точек прогноза и шагом прогноза. Результаты вычисления прогноза представлены на рисунке 11. В столбце **Довер. инт.** фигурирует величина допустимого отклонения от прогноза при 95% уровне доверия

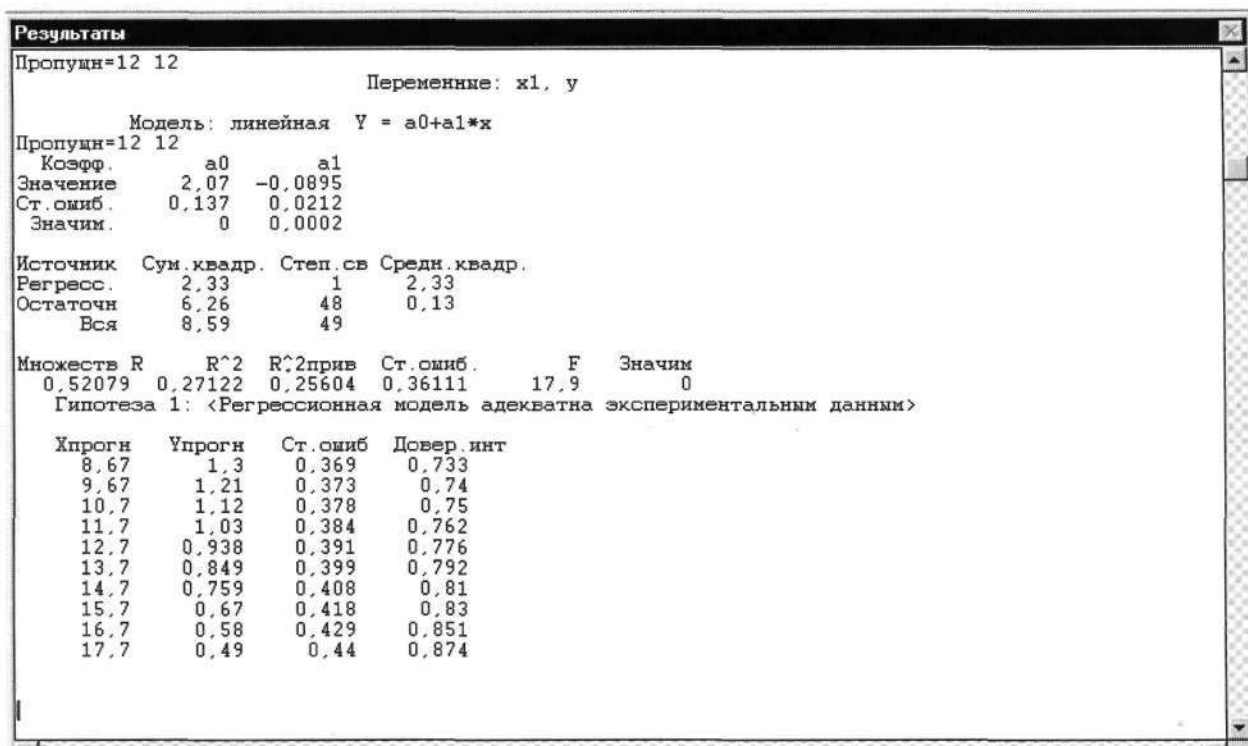


Рисунок 11

Процедура также строит график прогноза (рисунок 12).

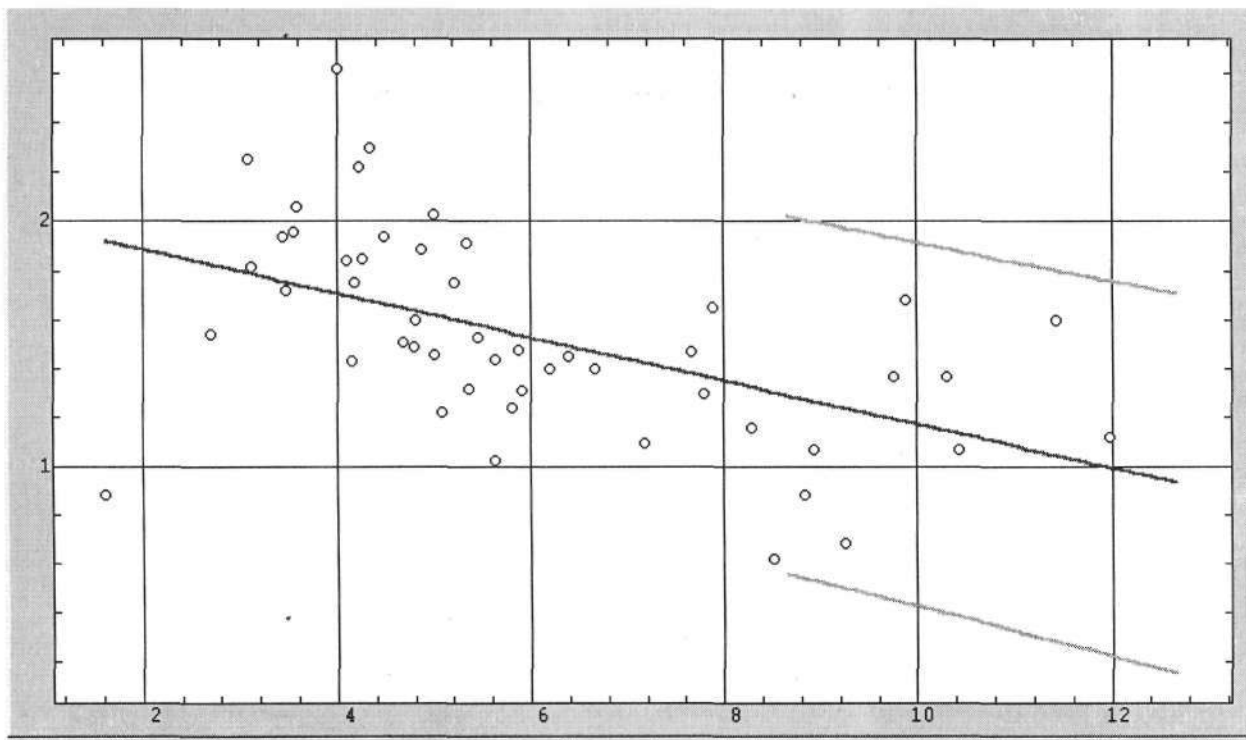


Рисунок 12

Недостаточно высокий коэффициент детерминации может быть объяснен, помимо влияния неучтенных при построении модели факторов, неправильным выбором спецификации уравнения регрессии. Поэтому попытаемся подобрать наиболее подходящий нелинейный вид уравнения. Система позво-

ляет использовать другие модели регрессии. Выбрав пункт **О = Общая/ нелинейная модель** в меню выбора статистических методов, можно задать вид зависимости соответствующей формулой из окна рисунка 7. В результате появляется меню спецификаций моделей однопараметрической регрессии (рисунок 3). Выберем (подбором), как наиболее подходящую, параболическую модель. Отчет с результатами расчетов представлен на рисунке 13.

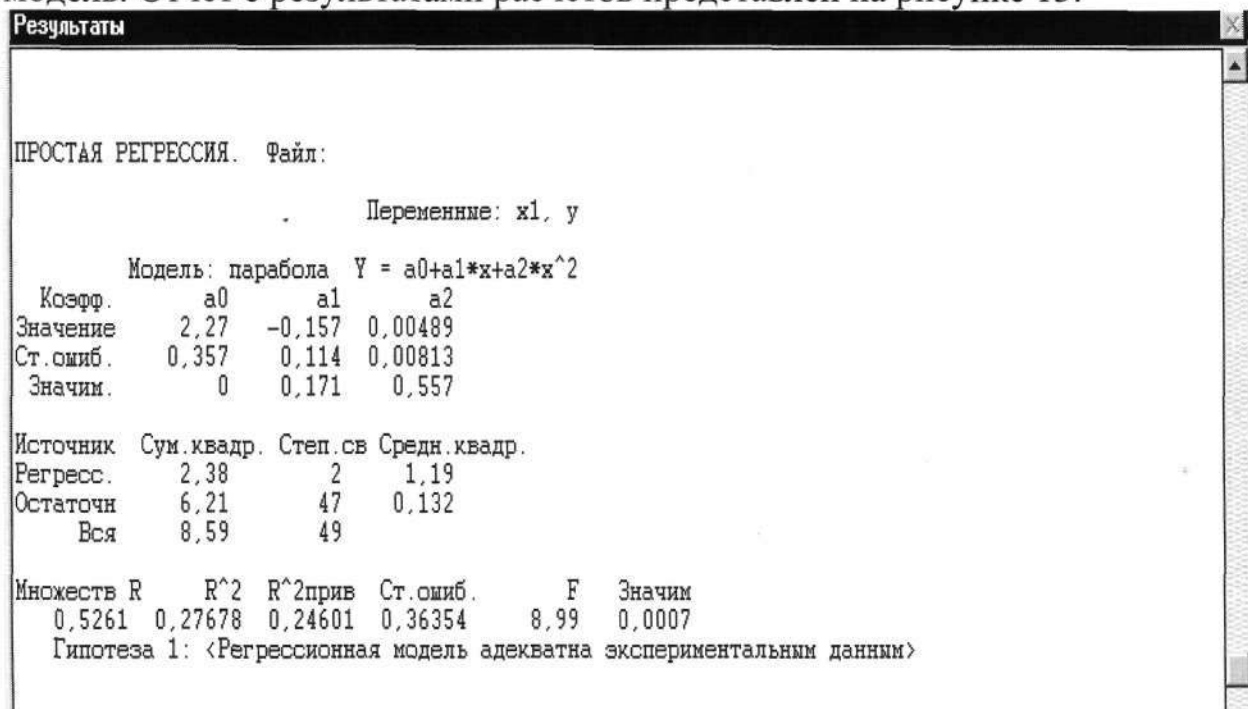


Рисунок 13

Кривая, соответствующая параболической зависимости представлена на рисунке 14.

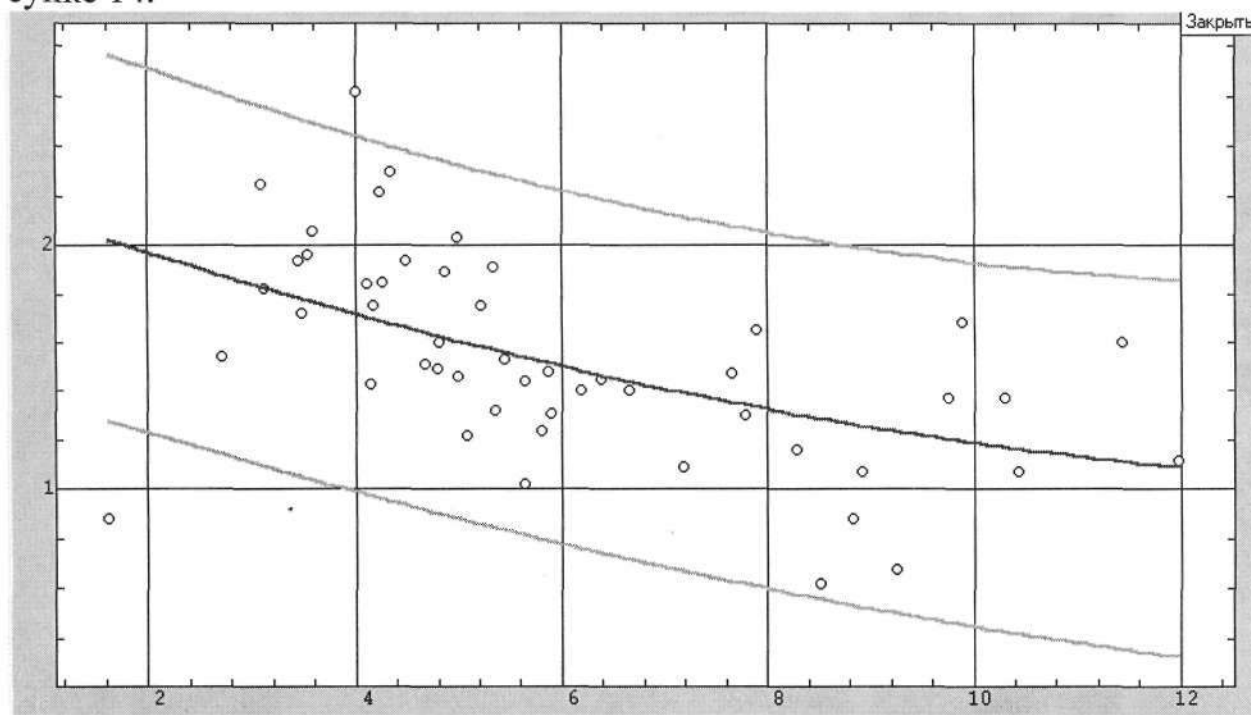


Рисунок 14

По результатам отчета видно, что коэффициент детерминации стал хоть и незначительно, но больше и составил 0.277. Модель также адекватна экспериментальным данным, но коэффициенты в модели не значимы:

$$\hat{Y} = 2.27 - 0.157X_1 + 0.00489X_1^2.$$

(0.171) (0.557)

Возможно для улучшения модели и увеличения величины коэффициента детерминации целесообразно использовать другие переменные (в нашем примере X_2) и построить множественное уравнение регрессии. Обычно начинают строить классическую линейную модель множественной регрессии. В меню **Статистические методы** в разделе **Регрессионный анализ** необходимо выбрать пункт **М – Множественная линейная**. В появившемся запросе **Переменные регрессии** выберем переменные (рисунок 15).

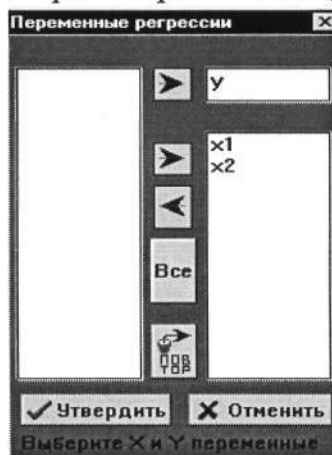


Рисунок 15

Отчет с результатами полученных оценок параметров множественного линейного уравнения регрессии представлен на рисунке 16.

РЕЗУЛЬТАТЫ						
МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ. Файл:						
Кэфф.	a0	a1	a2			
Значение	1,65	-0,0832	0,0295			
Ст. ошиб.	0,167	0,0189	0,00797			
Значим.	0	0,0001	0,0008			
Источник	Сум.кв.др.	Степ.св	Средн.кв.др.			
Регресс.	3,74	2	1,87			
Остаточн	4,85	47	0,103			
Вся	8,59	49				
Множеств R	R^2	R^2прив	Ст.ошиб.	F	Значим	
0,66019	0,43585	0,41184	0,32108	18,2	0	
Гипотеза 1: <Регрессионная модель адекватна экспериментальным данным>						
x1=1, x2=1, Y=1,6						
Хэксп	Уэксп	Урегр	остаток	Ст.остат	Ст.ошиб	Довер.инт
6,4	1,45	1,51	-0,058	-0,185	0,321	0,637
7,8	1,3	1,3	-4,64E-5	0,000148	0,323	0,641
9,76	1,37	1,24	0,128	0,407	0,328	0,652
7,9	1,65	1,37	0,279	0,887	0,323	0,641
5,35	1,91	1,52	0,392	1,25	0,321	0,638
9,9	1,68	1,09	0,585	1,86	0,329	0,653
4,5	1,94	2,04	-0,0974	-0,31	0,322	0,64
4,88	1,89	1,93	-0,0434	-0,138	0,322	0,638
3,46	1,94	1,79	0,145	0,462	0,324	0,644
3,6	2,06	1,65	0,414	1,32	0,324	0,643
3,56	1,96	1,77	0,194	0,617	0,324	0,644

Рисунок 16

Оценка уравнения регрессии выглядит следующим образом:

$$\hat{Y} = 1.65 - 0.0832X_1 + 0.0295 X_2 ,$$

(0.0) (0.0)

$R = 0.66$, $R^2 = 0.436$, стандартные ошибки для коэффициентов b_1 и b_2 составили соответственно 0.0189 и 0.00797.

Коэффициенты при переменных значимо отличны от нуля (т.к. значимость составила для каждого из них 0.0) и модель в целом адекватна экспериментальным данным (т.к. значимость критерия Фишера также оказалась равной 0). Коэффициент детерминации стал больше и составил 0.436, т.е. 43.6% доли результирующей переменной Y объясняется факторными переменными X_1 и X_2 , а 56.3% соответствует влиянию неучтенных в модели показателей.

5 Порядок выполнения расчетов в пакете STATISTICA

Методы регрессионного анализа, позволяющие моделировать статистические зависимости между двумя или несколькими переменными, можно реализовать в пакете STATISTICA модулем Множественной регрессии (**Multiple Regression**). В нем реализованы различные методы множественной, пошаговой и фиксированной нелинейной регрессии (в частности, полиномиальной, экспоненциальной, логарифмической). Программа позволяет вычислить все необходимые статистики и оценить адекватность построенной модели.

Рассмотрим процедуру реализации регрессионного анализа в системе STATISTICA. После запуска модуля на экране откроется основное окно системы, и автоматически загружается последний файл, с которым работал пользователь. Одновременно с этим появляется Стартовая панель модуля, содержащая основные операции, которые доступны в запущенном модуле, и позволяющая определить различные параметры анализа. Если стартовая панель модуля закрыта, то откроем ее с помощью меню **Analysis** – Анализ (рисунок 17)

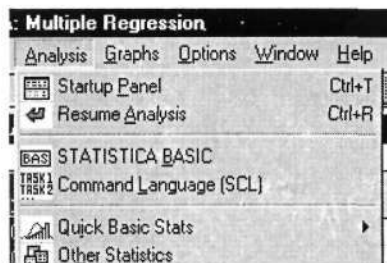


Рисунок 17

и выберем команду **Startup Panel** – Стартовая панель (рисунок 18).

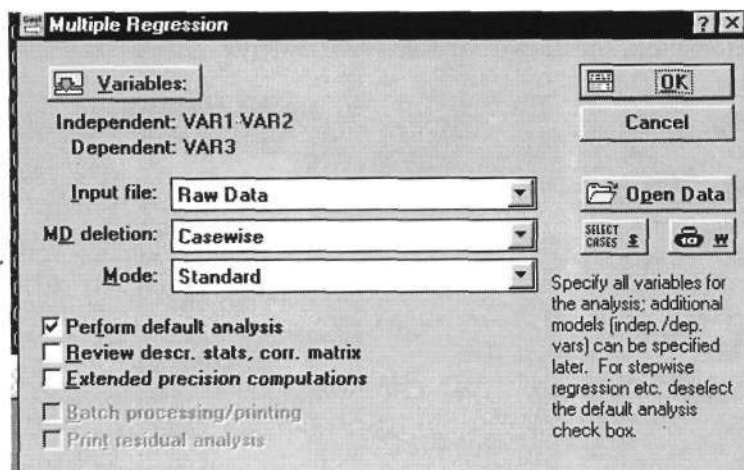


Рисунок 18

После появления стартовой панели модуля следует открыть нужный файл с исходной информацией – для этого используется кнопка **Open Data** – открыть данные. Окно с частью данных для анализа представлено на рисунке 19.

STATISTICA: Multiple Regression

File Edit View Analysis Graphs Options Window Help

45

Data: LAB5.STA 10v * 50c

NUM	1	2	3	4	5	6	7	8	9	10
VAL	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10
1	1,450	6,400	13,260							
2	1,300	7,800	10,160							
3	1,370	9,760	13,720							
4	1,650	7,900	12,850							
5	1,910	5,350	10,630							
6	1,680	9,900	9,120							
7	1,940	4,500	25,830							
8	1,890	4,880	23,390							
9	1,940	3,460	14,680							
10	2,060	3,600	10,050							
11	1,960	3,560	13,990							
12	1,020	5,650	9,680							
13	1,850	4,280	10,030							
14	,880	8,850	9,130							
15	,620	8,520	5,370							
16	1,090	7,190	9,860							

Рисунок 19

Далее необходимо выбрать зависимую (результатирующую, объясненную) и независимые (объясняющие) переменные для анализа.

Для задания переменных воспользуемся кнопкой **Variables** – Переменные из Стартовой панели (рисунок 20).

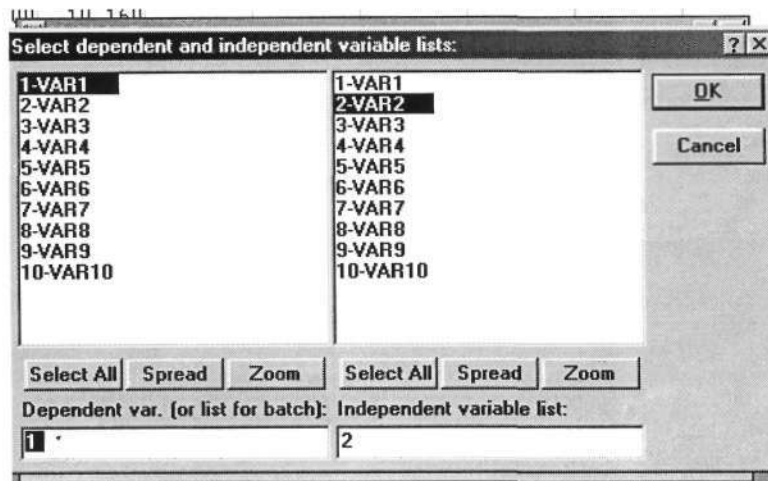


Рисунок 20

В окне **Select dependent and independent variable list** – Выбор зависимой переменной и списка независимых переменных, выделяя имя переменной в левой части окна, производится выбор зависимой переменной **Dependent**. В правой части окна выбираем независимую переменную (**Independent**). Выбор нескольких несмежных переменных производят при нажатой клавише **CTRL**. После выбора переменных необходимо щелкнуть на кнопке **OK**, вновь окажемся в Стартовой панели модуля Множественная регрессия. Здесь можно задать дополнительные опции и параметры анализа. Строка **Input file** определяет тип входной информации. Если входная информация представляет собой массив исходных данных, следует оставить **Raw Data** (необработанные данные). В следующей строке можно задать правило обработки пропущенных данных.

Поле **Mode** при отказе от параметра **Standard** (рисунок 21) и выборе параметра **Fixed non-linear**

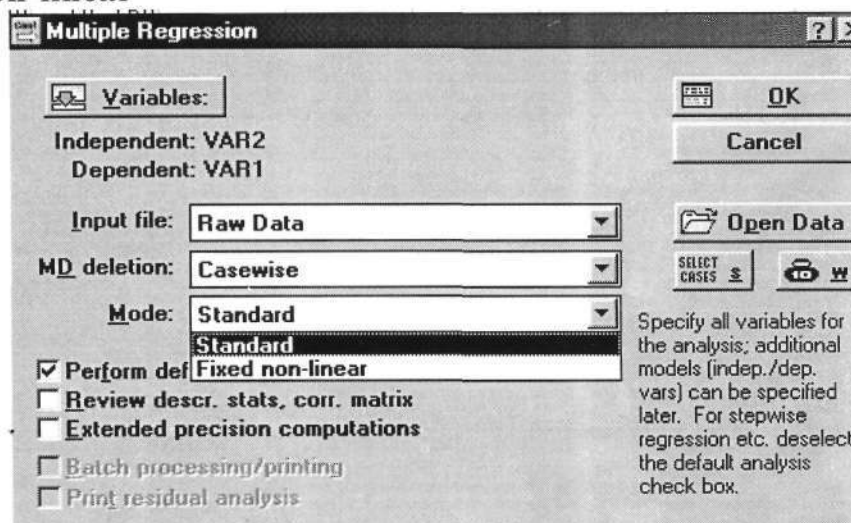


Рисунок 21

позволит перейти к построению фиксированных нелинейных моделей, среди которых полиномиальные и степенные уравнения регрессии (рисунок 22).

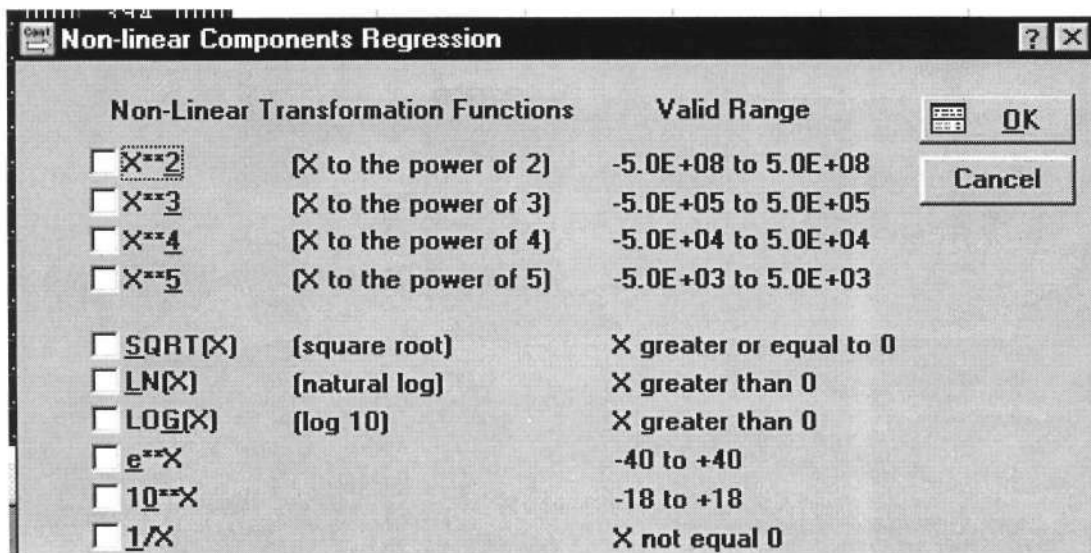


Рисунок 22

Перед выбором спецификации модели имеет смысл отобразить данные на графике. Пакет включает в себя большое количество разнообразных категорий и типов графиков в пространстве, на плоскости, научные графики в различных системах координат, деловые графики и другие. Для вызова графических возможностей используют меню **Graphics – Графика**, где выбирают тип графика, например **Scatterplots** - двумерные диаграммы рассеяния (рисунок 23).

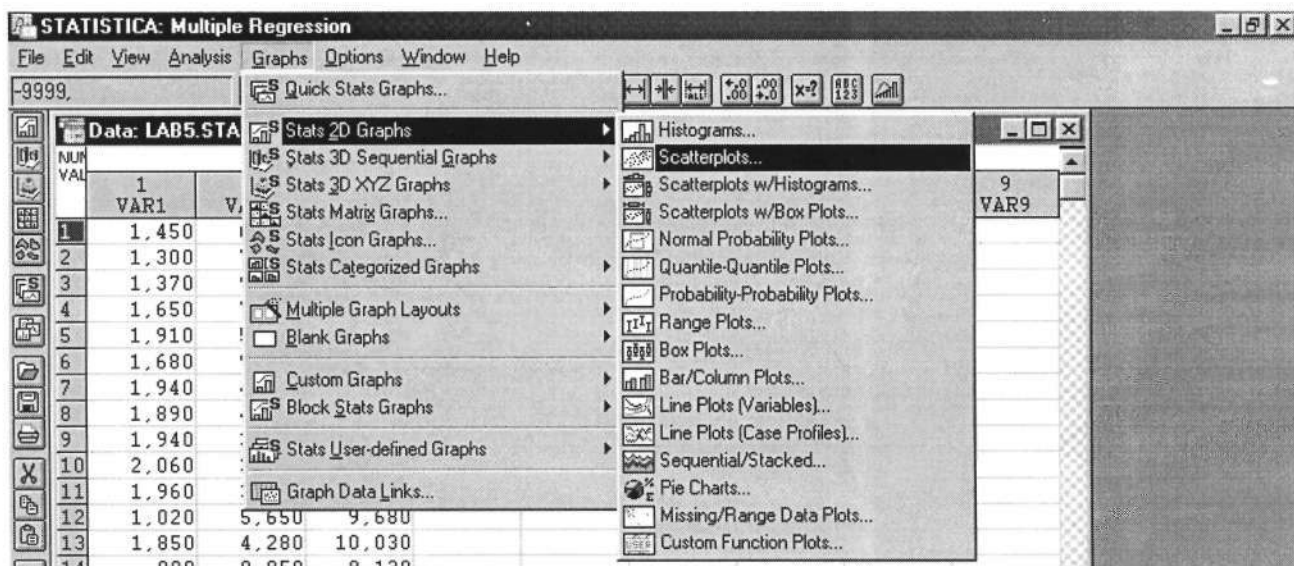


Рисунок 23

В появившемся диалоговом окне выбирают необходимые переменные (рисунок 24) и предполагаемый тип модели.

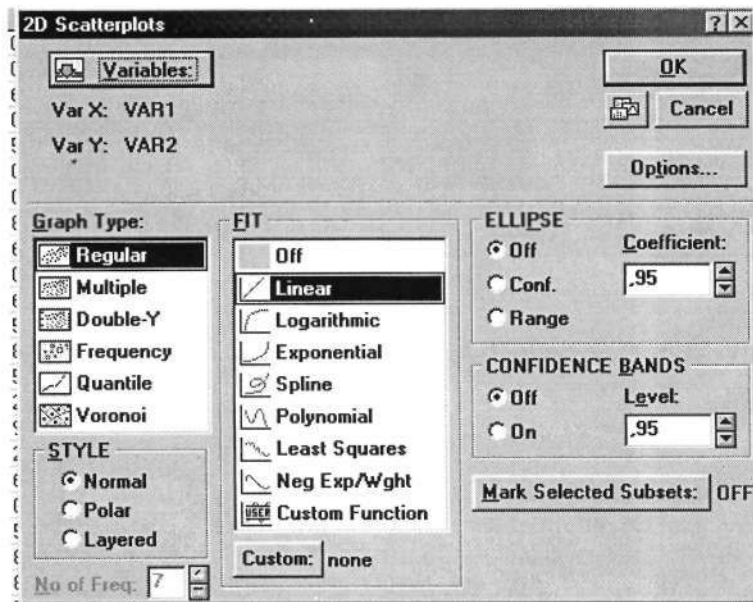


Рисунок 24

Построим линию регрессии между интересующими нас переменными (рисунок 25).

[Next](#) [Quit](#)

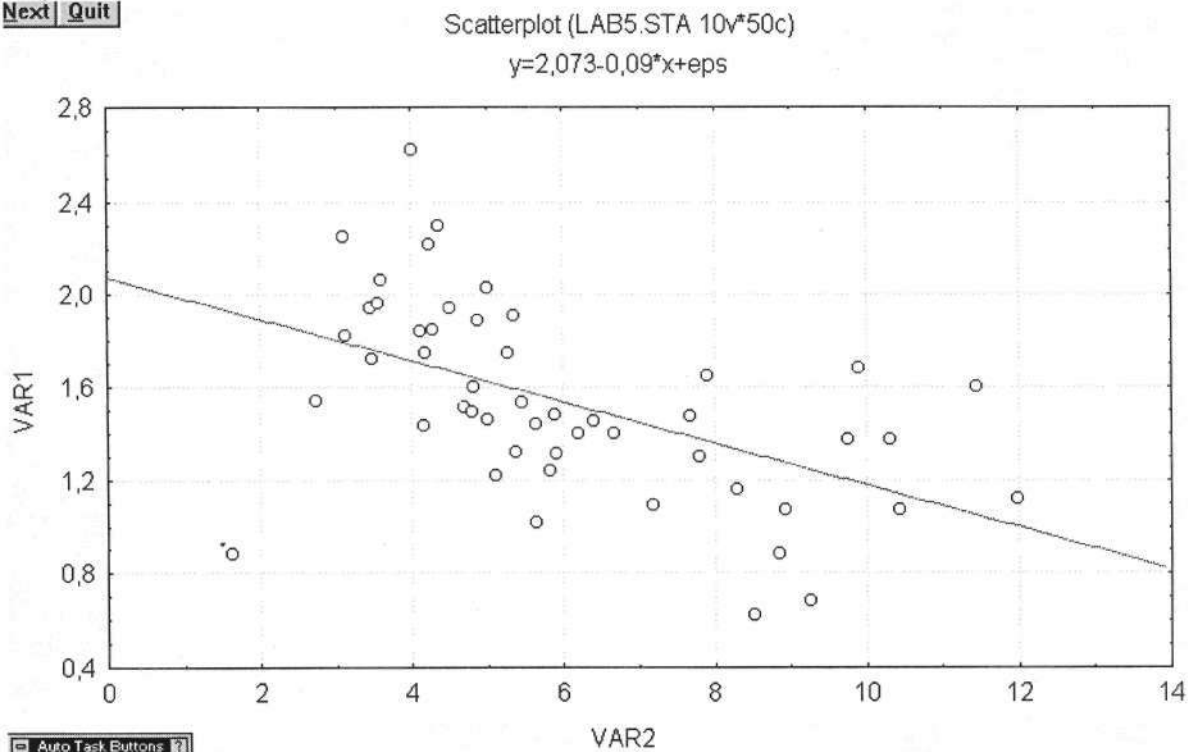


Рисунок 25

Как видно, по расположению исходных данных несколько затруднительно определить спецификацию модели. Предположим, что между рассматриваемыми переменными имеется линейная зависимость.

Для построения модели, необходимо вновь вернуться к окну рисунка 20 щелкнуть на кнопке **ОК** в правом углу окна. Система произведет расчеты и выведет окно с результатами (рисунок 26).

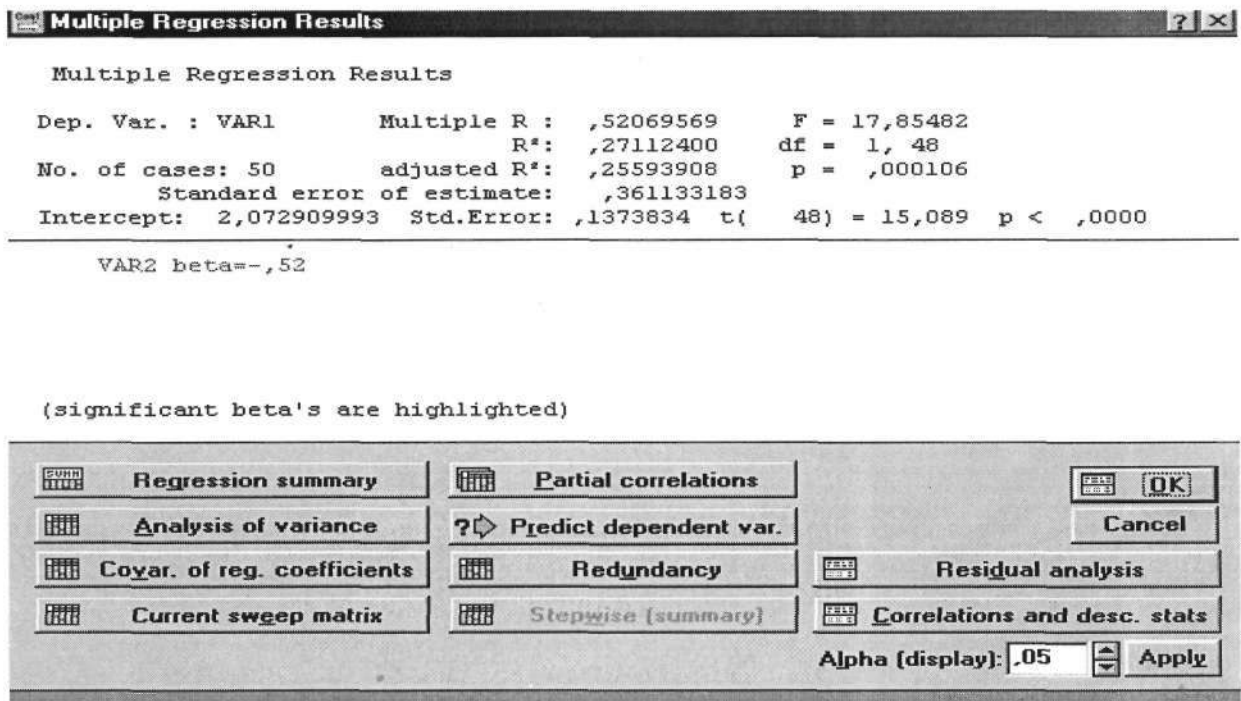


Рисунок 26

Нажатием на кнопку **Regression summary** получим информацию с рассчитанными оценками параметров уравнения регрессии (рисунок 27).

Continue...						
R= .52069569 RI= .27112400 Adjusted RI= .25593908 F(1,48)=17,855 p<.00011 Std.Error of estimate: .36113						
N=50	BETA	St. Err. of BETA	B	St. Err. of B	t(48)	p-level
Intercept			2,072910	.137383	15,08850	.000000
VAR2	-,520696	.123227	-,089571	.021198	-4,22550	.000106

Рисунок 27

В первую очередь обращаем внимание на коэффициент детерминации R^2 . Для линейной регрессионной модели $R^2 = 0.271$, т.е. вариация результирующей переменной (VAR1) лишь на 27.1% объясняется вариацией факторной переменной X_2 (VAR2). Оценка уравнения регрессии выглядит следующим образом:

$$\hat{y} = 2.073 - 0.896 X_2. \quad (0)$$

Коэффициент при переменной X_2 значимо отличается от нуля, т.к. нулевая гипотеза ($H_0: \beta_1=0$ при альтернативной $H_1: \beta_1 \neq 0$) отклонена, поскольку значимость нулевой гипотезы равна 0.000106. Регрессионная модель адекватна экспериментальным данным. Вообще говоря, следовало бы исследовать регрессионные остатки на нормальный характер распределения, но, поскольку, нас не устраивает качество модели ($R^2 = 0.271$ – мало), попробуем построить оценки уравнения регрессии в классе нелинейных. Для этого в окне рисунка 17 в поле **Mode** выберем нелинейный вид зависимости (рисунок 28).

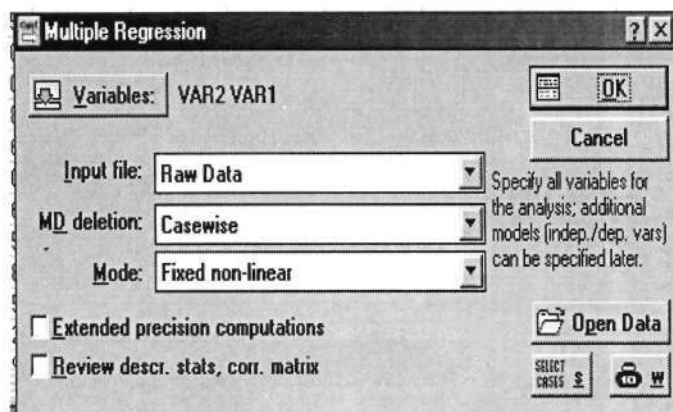


Рисунок 28

Процедура предлагает включить в модель свободный член **Intercept** (рисунок 29). Соглашаемся нажатием на кнопку **Ok**.

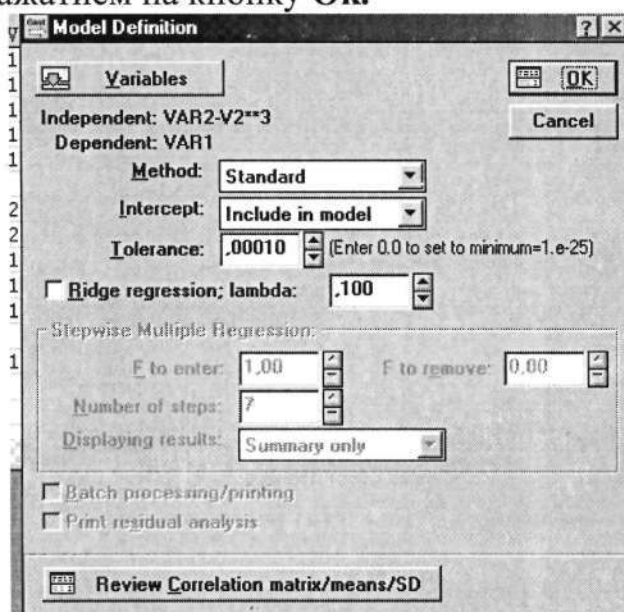


Рисунок 29

Задаем вид предполагаемой нелинейной зависимости, отмечая знаком соответствующее окошко (рисунок 30).

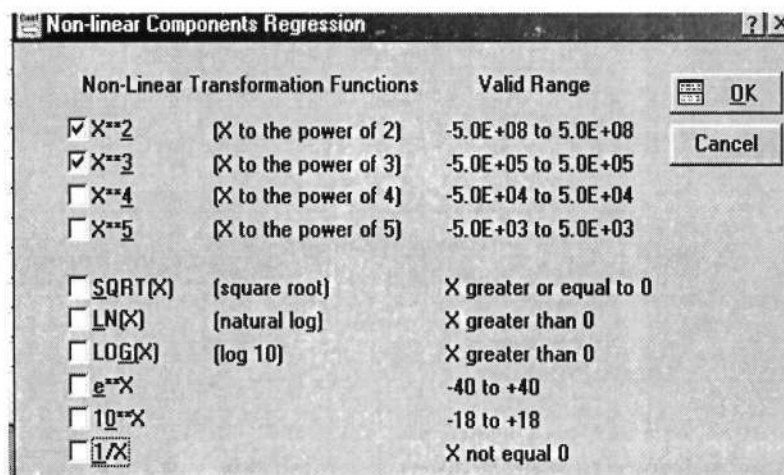


Рисунок 30

Выбираем необходимые переменные для построения модели (рисунок 31).

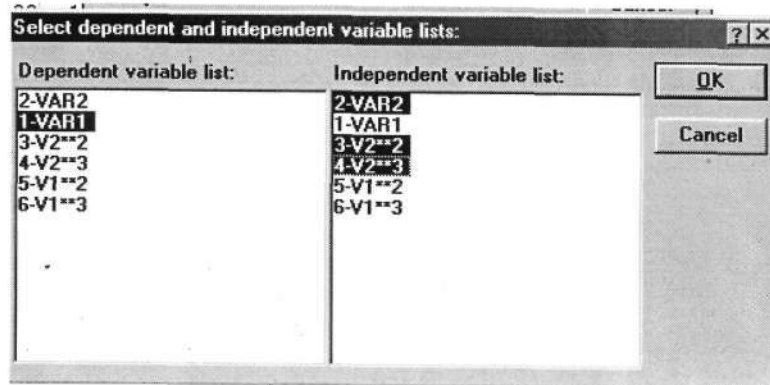


Рисунок 31

Получим следующие результаты по построению оценок коэффициентов нелинейной модели (рисунок 32).

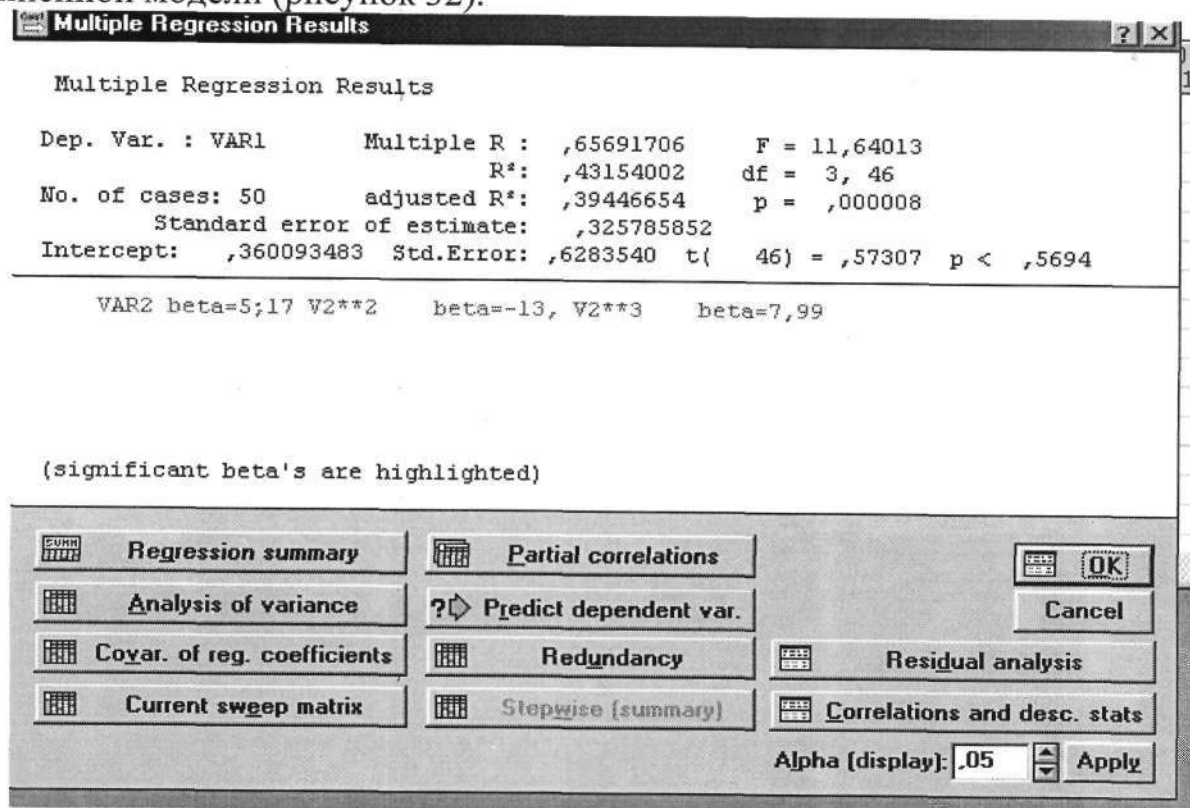


Рисунок 32

Красным высвечены значимые оценки коэффициентов уравнения регрессии. Более подробную информацию получим после нажатия на кнопку **Regression summary** (рисунок 33).

Continue...						
R= .65691706 RI= .43154002 Adjusted RI= .39446654 F(3,46)=11,640 p<.00001 Std.Error of estimate: .32579						
N=50	BETA	St. Err. of BETA	B	St. Err. of B	t(46)	p-level
Intercept			.360093	.628354	.57307	.569385
VAR2	5,1743	1,819288	.890094	.312956	2,84416	.006624
V2**2	-13,4771	3,965666	-.165950	.048831	-3,39844	.001408
V2**3	7,9894	2,257555	.008392	.002371	3,53898	.000931

Рисунок 33

Оценка уравнения регрессии выглядит следующим образом:

$$\hat{y} = 0.36 + 0.89X_1 - 0.166 X_1^2 + 0.083X_1^3,$$

(0) (0) (0)

Как видно из отчета, нулевые гипотезы о незначимости коэффициентов ($H_0: \beta_j = 0$ при альтернативной $H_1: \beta_j \neq 0$) отклоняются и оценки коэффициентов регрессии значимо отличаются от нуля ($p < 0.05$), а уравнение регрессии значимо, т.е. модель адекватна экспериментальным данным, поскольку значимость нулевой гипотезы составила 0.000008.

Далее можно приступить к исследованию остатков регрессионной модели. Значения остатков в программе получаются как разность между исходными значениями зависимой переменной и так называемыми **Predicted values** – предсказанными значениями с помощью модели. Остатки исследуются в специальном окне **Residual analysis** – Анализ остатков в окне рисунка 32 (рисунок 34).

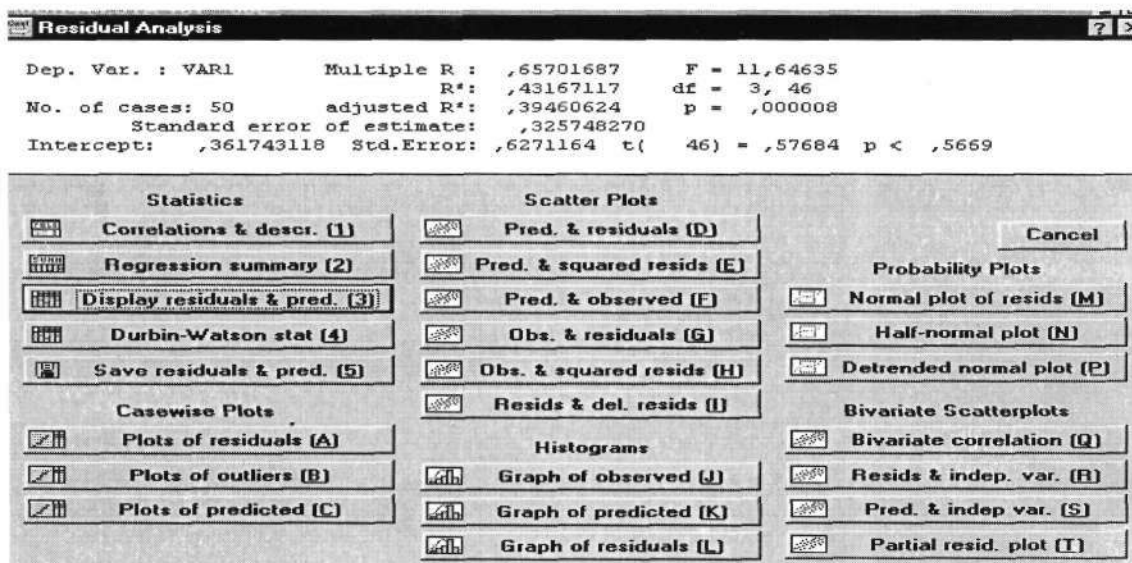


Рисунок 34

Информация о значениях остатков может быть получена нажатием на кнопку **Pred. & residuals [D]** (рисунок 35).

Predicted & Residual Values (labor444.sta)							
Dependent variable: VAR1							
Case No.	Observed Value	Predictd Value	Residual	Standard Pred. v.	Standard Residual	Std. Err. Pred. Val	M
1	1.450000	1.459003	-.009003	-.27265	-.02764	.065577	
2	1.300000	1.188622	.111378	-1.25561	.34192	.083467	
3	1.370000	1.041301	.328699	-1.79119	1.00906	.098216	
4	1.650000	1.172235	.477765	-1.31518	1.46667	.084862	
5	1.910000	1.656967	.253033	.44704	.77677	.061690	
6	1.680000	1.049730	.630270	-1.76055	1.93484	.099062	
7	1.940000	1.769502	.170498	.05616	.52340	.062292	
8	1.890000	1.726746	.163254	.70072	.50116	.061751	
9	1.940000	1.800649	.139351	.96939	.42779	.076594	
10	2.060000	1.805135	.254865	.98570	.78240	.072408	
11	1.960000	1.804160	.155840	.98216	.47841	.073500	
12	1.020000	1.604884	-.584884	.25770	-1.79551	.062023	
13	1.850000	1.787488	.062512	.92155	.19190	.063093	
14	.880000	1.056522	-.176522	-1.73586	-.54190	.094666	
15	.620000	1.086189	-.466189	-1.62800	-1.43113	.092223	

Рисунок 35

С помощью функциональных кнопок в окне рисунка 34 можно просмотреть остатки модели в графическом виде. Рассмотрим график остатков, например, на нормальной вероятностной бумаге – **Probability plots** (рисунок 36) или гистограммы для проверки справедливости предположения о нормальном распределении остатков (рисунок 37).

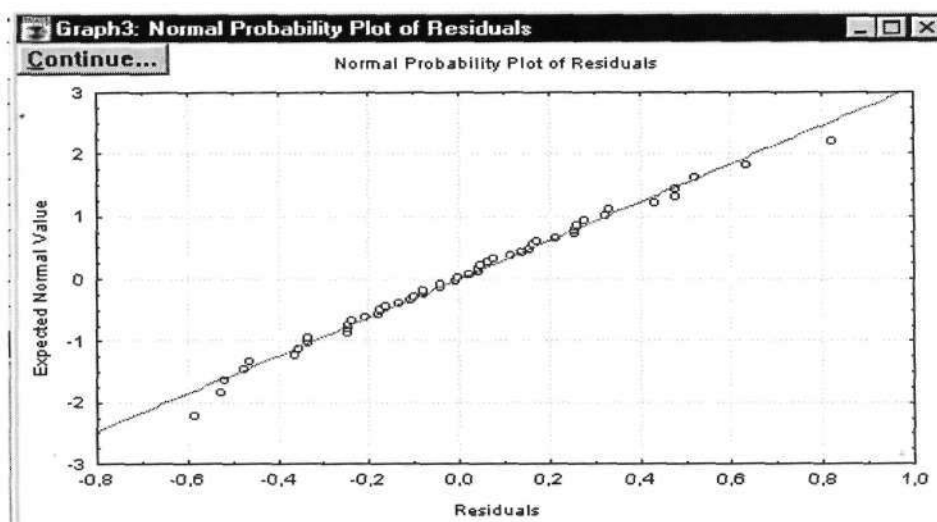


Рисунок 36

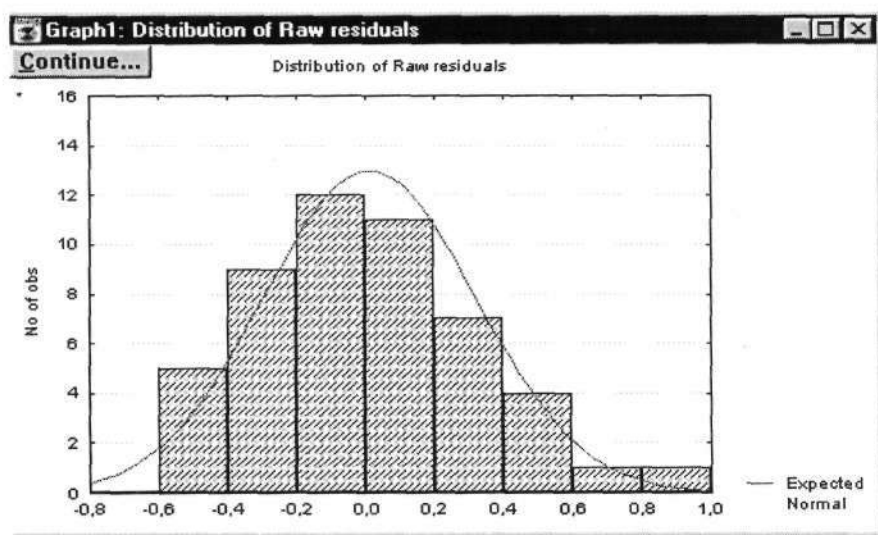


Рисунок 37

Как видно из рисунков, эмпирические (ступенчатая фигура) и теоретические (сплошная линия) частоты распределения примерно совпадают, следовательно, регрессионные остатки близки к нормальному закону распределения. Для более подробного анализа скопируем остатки в виде отдельной переменной в окно рисунка 19. Затем из меню **Basic Statistics and Tables** выбираем пункт **Frequency tables** (рисунок 38),

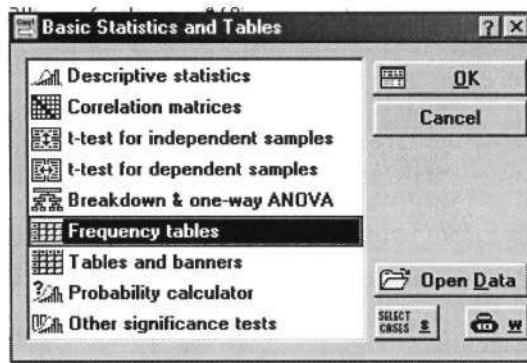


Рисунок 38

высвечивается следующее окно выбора анализа остатков (рисунок 39):

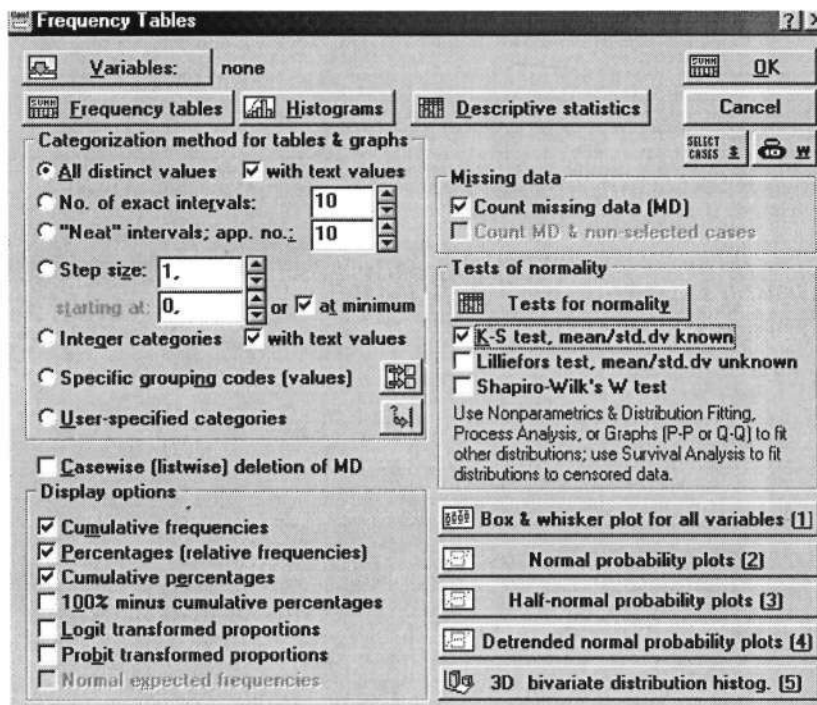


Рисунок 39

Предлагаются тесты для проверки соответствия остатков нормальному закону распределения **Tests for normality**. Выберем тест Колмогорова-Смирнова, затем после нажатия на кнопку появляются результаты анализа (рисунок 40):

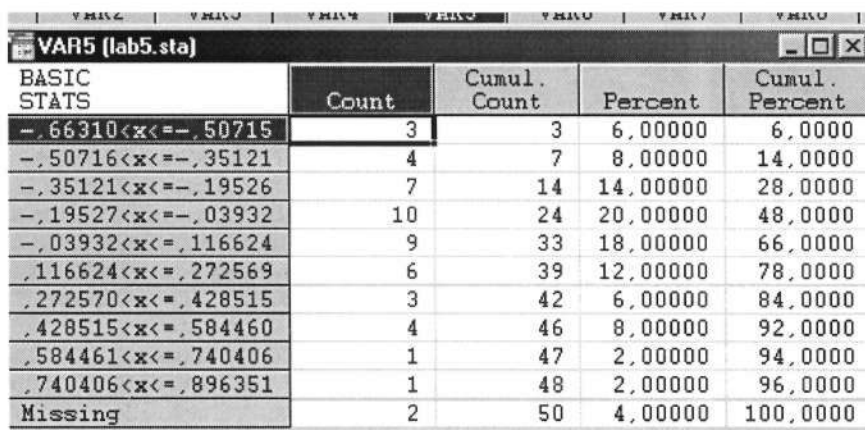
Kolmogorov-Smirnov Test [lab5.sta]			
BASIC (Mean & standard deviation known)			
STATS			
Variable	N	max D	p
VAR5	48	.056687	p > .20

Рисунок 40

Согласно теста Колмогорова-Смирнова величина $\lambda = \frac{D}{\sqrt{n}} = \frac{0.057}{\sqrt{48}} = 0.0082$

меньше критического уровня $\bar{\varepsilon} = 0.19$, выбранной из таблицы Колмогорова-Смирнова, следовательно, анализируемые остатки имеют нормальный закон распределения. Можно также вывести на экран таблицу с интервалами, содержащими частоты данных, попавших в эти интервалы (таблицы могут быть использованы для проверки по критерию Пирсона), служащие для по-

строения гистограммы распределения частот (рисунок 41). Для вывода таблицы необходимо нажать на кнопку **Descriptive statistics** из окна рисунка 36.



BASIC STATS	Count	Cumul. Count	Percent	Cumul. Percent
- .66310 < x <= - .50715	3	3	6.00000	6.0000
- .50716 < x <= - .35121	4	7	8.00000	14.0000
- .35121 < x <= - .19526	7	14	14.00000	28.0000
- .19527 < x <= - .03932	10	24	20.00000	48.0000
- .03932 < x <= .116624	9	33	18.00000	66.0000
.116624 < x <= .272569	6	39	12.00000	78.0000
.272570 < x <= .428515	3	42	6.00000	84.0000
.428515 < x <= .584460	4	46	8.00000	92.0000
.584461 < x <= .740406	1	47	2.00000	94.0000
.740406 < x <= .896351	1	48	2.00000	96.0000
Missing	2	50	4.00000	100.0000

Рисунок 41

Таким образом, исследуемая модель адекватна экспериментальным данным, регрессионные остатки имеют нормальный закон распределения.

Коэффициент детерминации для модели равен 0.432. Это несколько лучше предыдущей линейной модели, но, скорее всего, целесообразно более лучшую модель искать в классе множественных.

Для построения уравнения линейной множественной регрессии выберем зависимую (объясненную) и независимые (объясняющие) переменные для анализа. Для их задания воспользуемся кнопкой **Variables** – Переменные из Стартовой панели (рисунок 42).

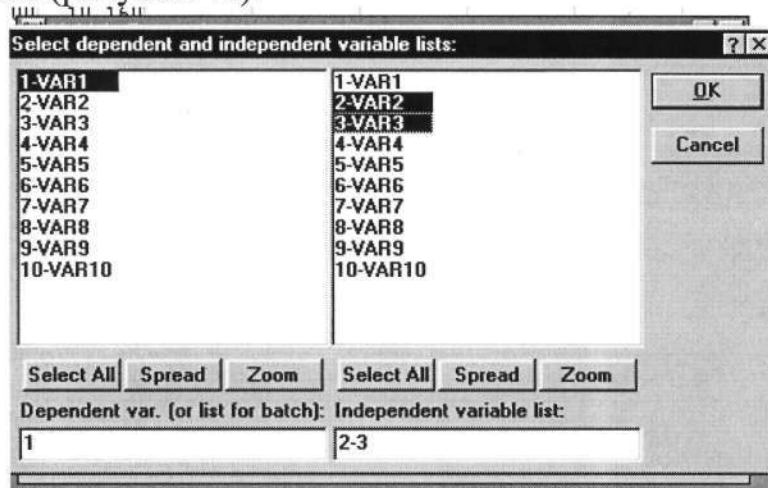


Рисунок 42

Отчет с результатами расчетов представлен на рисунке 43.

7 Интервальные оценки для параметров регрессии

Интервальное оценивание параметров регрессии в рассмотренных программных средствах не производится. Поэтому получим интервальные оценки в других приложениях, например в Excel. Интервальная оценка с доверительной вероятностью γ для параметров β_j имеет вид:

$$b_j - t_\alpha \hat{S}_{b_j} \leq \beta_j \leq b_j + t_\alpha \hat{S}_{b_j},$$

где t_α находят по таблице t – распределения при вероятности $\alpha = 1 - \gamma$ и числе степеней свободы $v = n - k - 1$.

Определим интервальные оценки коэффициентов уравнения с доверительной вероятностью $\gamma = 0.95$ (рисунок 45). Величину t_α определим в окне Excel, используя **Мастер функций**, категорию **Статистические**, функцию **СТЮДРАСПОБР**.

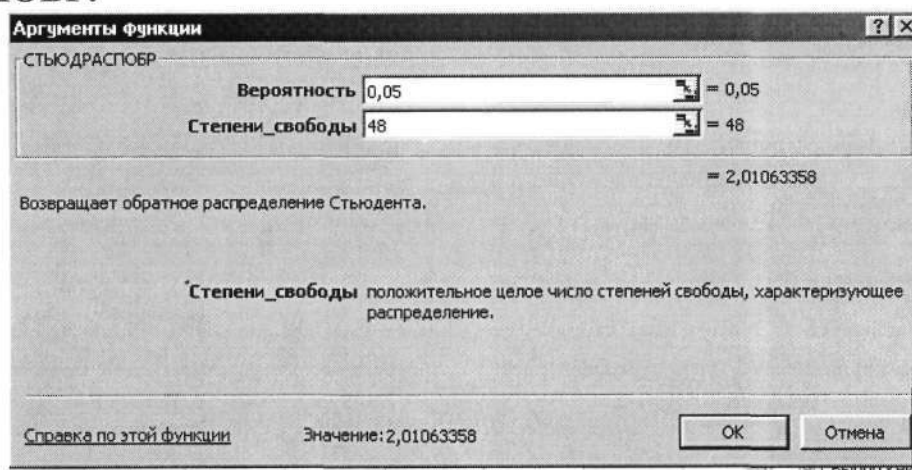


Рисунок 45

Пример расчета приведен на рисунке 46.

E5		f Для множественной регрессии						
	A	B	C	D	E	F	G	H
1								
2								
3								
4								
5		Для парной регрессии			Для множественной регрессии			
6								
7		n=	50		k=	2		
8		k=	1					
9		α =	0,05					
10		ν =	48		ν =	47		
11		t=	2,010634		t=	2,011739		
12								
13		b0=	2,072574		b0=	1,649522		
14		b1=	-0,08954		b1=	-0,08324		
15					b2=	0,29505		
16		Δ_{80}	4,167187		Δ_{80}	3,318407		
17		Δ_{81}	-0,18002		Δ_{81}	-0,16745		
18					Δ_{82}	0,593563		
19								
20								

Рисунок 46

6 Содержание письменного отчета

Отчет должен быть оформлен на листах формата А4 с титульным листом, оформленным соответствующим образом и содержать следующее:

- 1) постановку задачи с вариантом выборок;
- 2) краткое изложение теории регрессионного анализа;
- 3) результаты компьютерной обработки данных;
- 4) анализ полученных результатов;
- 5) выводы по полученным результатам.

7 Вопросы к защите

- 7.1 Сформулируйте основные задачи регрессионного анализа.
- 7.2 Перечислите основные предпосылки регрессионного анализа.
- 7.3 В чем заключается основная идея метода наименьших квадратов?
- 7.4 Что характеризует и как оценивается остаточная дисперсия?
- 7.5 На какие составляющие разлагается общая дисперсия?
- 7.6 Как объясняются параметры уравнения регрессии?
- 7.7 Как проверяется значимость уравнения регрессии и значимость коэффициента регрессии?
- 7.8 Как строятся интервальные оценки параметров уравнения регрессии?
- 7.9 Как строится доверительная оценка для интервала прогнозирования?
- 7.10 Какие показатели оценивают адекватность построенной модели?

Список использованных источников

1. Айвазян А.М., Мхитарян В.С. Прикладная статистика и основы эконометрики: Учебник для вузов. –М.: ЮНИТИ, 1998. –1022с.
2. Боровиков В.П., Боровиков И.П. STATISTICA – Статистический анализ и обработка данных в среде Windows. –М.: Инф.изд. Дом "Филин", 1998. -608с.
3. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы: Учебник. –М.: Финансы и статистика, 1998, -352с

Приложение А1
(обязательное)

Таблица 1 - Варианты индивидуальных заданий

Номера вариантов	Переменные Y, X ₁ , X ₂	Номера вариантов	Переменные Y, X ₁ , X ₂
1	1,2,3	16	2,3,4
2	1,2,4	17	2,3,5
3	1,2,5	18	2,3,6
4	1,2,6	19	2,3,7
5	1,2,7	20	2,4,5
6	1,3,4	21	2,4,6
7	1,3,5	22	2,4,7
8	1,3,6	23	2,5,6
9	1,3,7	24	2,5,7
10	1,4,5	25	2,6,7
11	1,4,6	26	3,4,5
12	1,4,7	27	3,4,6
13	1,5,6	28	3,4,7
14	1,5,7	29	3,5,6
15	1,6,7	30	3,5,7

Таблица 2 - Значения переменных

№п/п	1	2	3	4	5	6	7
1	73	48	99	31	284	68	279
2	69	40	83	28	265	47	245
3	72	52	106	30	298	40	354
4	72	50	107	23	264	35	212
5	65	39	79	25	232	60	323
6	67	49	100	25	272	37	285
7	56	38	80	25	292	41	240
8	70	47	96	27	245	46	361
9	63	41	98	31	274	46	236
10	64	50	97	25	256	53	246
11	70	52	92	28	291	43	357
12	67	36	90	25	290	53	361
13	60	55	108	30	257	54	343
14	63	43	107	28	258	50	390
15	80	45	96	28	309	45	208
16	71	56	86	31	257	50	301
17	74	45	98	23	316	25	306
18	68	55	97	28	251	44	334

Продолжение таблицы 2

1	2	3	4	5	6	7	8
19	65	63	128	26	278	54	342
20	73	47	88	26	268	56	274
21	57	48	117	28	273	48	336
22	71	55	110	27	280	44	277
23	66	37	94	25	244	49	320
24	76	38	82	26	243	50	278
25	70	46	97	27	269	46	266
26	68	41	91	28	255	46	294
27	74	42	118	27	229	47	308
28	69	53	87	29	272	39	271
29	68	57	104	24	263	49	228
30	71	39	89	23	249	44	414
31	60	46	71	27	252	49	324
32	56	56	86	27	229	53	348
33	71	47	97	28	284	44	211
34	68	51	89	26	265	47	356
35	66	52	101	29	298	51	350
36	60	50	93	26	264	56	350
37	70	47	112	23	232	39	301
38	69	48	67	27	249	46	338
39	72	54	97	29	292	57	297
40	70	52	82	24	245	45	268
41	61	47	81	27	274	55	294
42	62	57	92	25	256	46	309
43	63	53	90	29	291	52	284
44	71	48	110	29	290	45	276
45	65	46	96	25	257	43	289
46	70	43	110	24	258	48	234
47	70	49	97	30	309	50	298
48	68	31	102	25	257	44	264
49	78	40	101	31	231	38	302
50	66	49	112	27	288	68	323

Таблица 3 – Варианты индивидуальных заданий

№ варианта	Номера признаков Y, X ₁ , X ₂	№ варианта	Номера признаков Y, X ₁ , X ₂
1	11,5,1	17	11,7,2
2	11,5,2	18	11,7,3
3	11,5,3	19	11,7,4
4	11,5,4	20	11,7,8
5	11,5,6	21	11,7,9
6	11,5,7	22	11,7,10
7	11,5,8	23	11,10,1
8	11,5,9	24	11,10,2
9	11,5,10	25	11,10,3
10	11,6,1	26	11,10,4
11	11,6,2	27	11,10,6
12	11,6,3	28	11,10,8
13	11,6,4	29	11,10,9
14	11,6,7	30	11,10,7
15	11,6,10	31	11,6,4
16	11,7,1	32	11,4,6

Таблица 4 - Показатели уровня жизни населения в 1994г.

Страны	Потребление продуктов на душу населения							Число врачей на 100000ч.	Смертность на 100000 насел.	ВВП по ППС в % к США	Расходы на здравоох. в % к ВВП	Урожайность зернов. ц/га
	Мясо	Масло животн.	Сахар	Алко-голь	Фрукты	Хлебо-продукты	Кг					
	кг	кг	кг	л	Кг	кг	кг					
	x1	x2	x3	x4	x8	x9	x5	x11	x6	x7	x10	
Россия	55	3.9	30	5	28	124	44.5	84.98	20.4	3.2	14.4	
Австралия	100	2.6	47	8.2	121	87	32.5	30.58	71.4	8.5	11.6	
Австрия	93	5.3	37	12	146	74	33.9	38.42	78.7	9.2	56.1	
Азербайджан	20	4.1	12.4	7.9	52	141	38.8	60.34	12.1	3.3	16.4	
Армения	20	3.7	4.3	6.5	72	134	34.4	60.22	10.9	3.2	13.5	
Белоруссия	72	3.6	28	5.4	38	120	43.6	60.79	20.4	5.4	22.4	
Бельгия	85	6.9	48	11	83	72	41	29.82	79.7	8.3	65.5	
Болгария	65	3	18	9.5	92	156	36.4	70.57	17.3	5.4	27.8	
Великобритания	67	3.5	39	8.8	91	91	17.9	34.51	69.7	7.1	62.3	
Венгрия	73	1.7	40	10.9	73	106	32.1	64.73	24.5	6	39.8	
Германия	88	6.8	35	8.1	138	73	38.1	36.63	76.2	8.6	56.9	
Греция	83	1	24	8.8	99	108	41.5	32.84	44.4	5.7	37.4	
Грузия	21	3.8	36	9.8	55	140	55	62.64	11.3	3.5	18.6	
Дания	98	5	38	10.3	89	77	36.7	34.07	79.2	6.7	54.4	
Ирландия	99	3.3	31	9.6	87	102	15.8	39.27	57	6.7	64.2	
Испания	89	0.4	26	8.95	103	72	40.9	28.46	54.8	7.3	22.6	
Италия	84	2.2	27	9.6	169	118	49.4	30.27	72.1	8.5	46	
Казахстан	61	4.2	19.2	7.2	10	191	38.1	69.04	13.4	3.3	7.9	
Канада	98	3.1	44	7.4	123	77	27.6	25.42	79.9	10.2	25.4	
Киргизия	46	4.1	23.5	6.7	20	134	33.2	53.13	11.2	3.4	17	