

КОРРЕЛЯЦИЯ И СТАТИСТИЧЕСКАЯ ЗАВИСИМОСТЬ

Регрессионный анализ

Стукач Олег Владимирович

Каф. КИСМ, пр. Ленина, д. 2, оф. 204

☎ (3822)-701777*2754

tomsk@ieee.org

<http://ieee.tpu.ru/pages/stukach.htm>



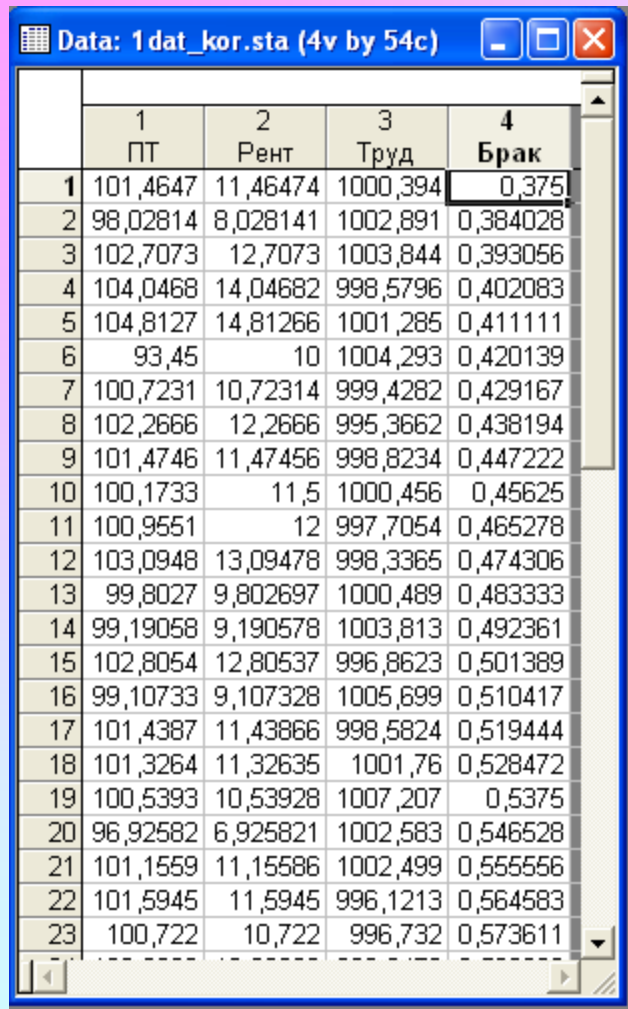
Основная задача регрессионного и корреляционного анализа состоит в выявлении связи между случайными переменными

Стукач О.В. Программный комплекс Statistica в решении задач управления качеством: учебное пособие / О. В. Стукач ; Национальный исследовательский Томский политехнический университет (ТПУ). - 1 компьютерный файл (pdf; 2.4 MB). - Томск : Изд-во ТПУ, 2011. - ТПУ. - Adobe Reader. - <[URL:http://www.lib.tpu.ru/fulltext2/m/2011/m426.pdf](http://www.lib.tpu.ru/fulltext2/m/2011/m426.pdf)>.

Содержание

- 1 Общие сведения о корреляционном анализе
- 2 Пример корреляционного анализа данных в программе STATISTICA
- 3 Общие сведения о регрессионном анализе
- 4 Пример регрессионного анализа данных в программе STATISTICA

Корреляция



	1 ПТ	2 Рент	3 Труд	4 Брак
1	101,4647	11,46474	1000,394	0,375
2	98,02814	8,028141	1002,891	0,384028
3	102,7073	12,7073	1003,844	0,393056
4	104,0468	14,04682	998,5796	0,402083
5	104,8127	14,81266	1001,285	0,411111
6	93,45	10	1004,293	0,420139
7	100,7231	10,72314	999,4282	0,429167
8	102,2666	12,2666	995,3662	0,438194
9	101,4746	11,47456	998,8234	0,447222
10	100,1733	11,5	1000,456	0,45625
11	100,9551	12	997,7054	0,465278
12	103,0948	13,09478	998,3365	0,474306
13	99,8027	9,802697	1000,489	0,483333
14	99,19058	9,190578	1003,813	0,492361
15	102,8054	12,80537	996,8623	0,501389
16	99,10733	9,107328	1005,699	0,510417
17	101,4387	11,43866	998,5824	0,519444
18	101,3264	11,32635	1001,76	0,528472
19	100,5393	10,53928	1007,207	0,5375
20	96,92582	6,925821	1002,583	0,546528
21	101,1559	11,15586	1002,499	0,555556
22	101,5945	11,5945	996,1213	0,564583
23	100,722	10,722	996,732	0,573611

Очень важно установить логическую связь между двумя рядами явлений или двумя совпадающими во времени явлениями, либо же дать им разумное объяснение

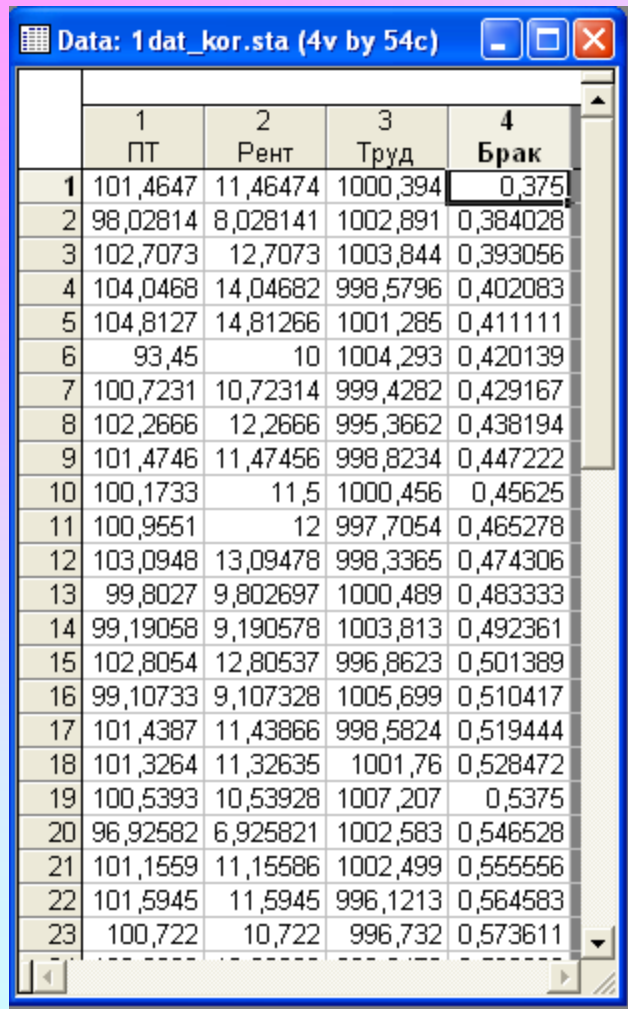
ПТ – производительность труда

Рент – рентабельность

Труд – трудоемкость единицы продукции

Брак – удельный вес потерь от брака

Корреляция



	1 ПТ	2 Рент	3 Труд	4 Брак
1	101,4647	11,46474	1000,394	0,375
2	98,02814	8,028141	1002,891	0,384028
3	102,7073	12,7073	1003,844	0,393056
4	104,0468	14,04682	998,5796	0,402083
5	104,8127	14,81266	1001,285	0,411111
6	93,45	10	1004,293	0,420139
7	100,7231	10,72314	999,4282	0,429167
8	102,2666	12,2666	995,3662	0,438194
9	101,4746	11,47456	998,8234	0,447222
10	100,1733	11,5	1000,456	0,45625
11	100,9551	12	997,7054	0,465278
12	103,0948	13,09478	998,3365	0,474306
13	99,8027	9,802697	1000,489	0,483333
14	99,19058	9,190578	1003,813	0,492361
15	102,8054	12,80537	996,8623	0,501389
16	99,10733	9,107328	1005,699	0,510417
17	101,4387	11,43866	998,5824	0,519444
18	101,3264	11,32635	1001,76	0,528472
19	100,5393	10,53928	1007,207	0,5375
20	96,92582	6,925821	1002,583	0,546528
21	101,1559	11,15586	1002,499	0,555556
22	101,5945	11,5945	996,1213	0,564583
23	100,722	10,722	996,732	0,573611

Во многих случаях первый шаг анализа состоит в вычислении корреляционной матрицы всех переменных и проверке значимых зависимостей (ожидаемых и неожиданных).

После того как это сделано, следует понять общую природу обнаруженной статистической значимости.

ПТ – производительность труда

Рент – рентабельность

Труд – трудоемкость единицы продукции

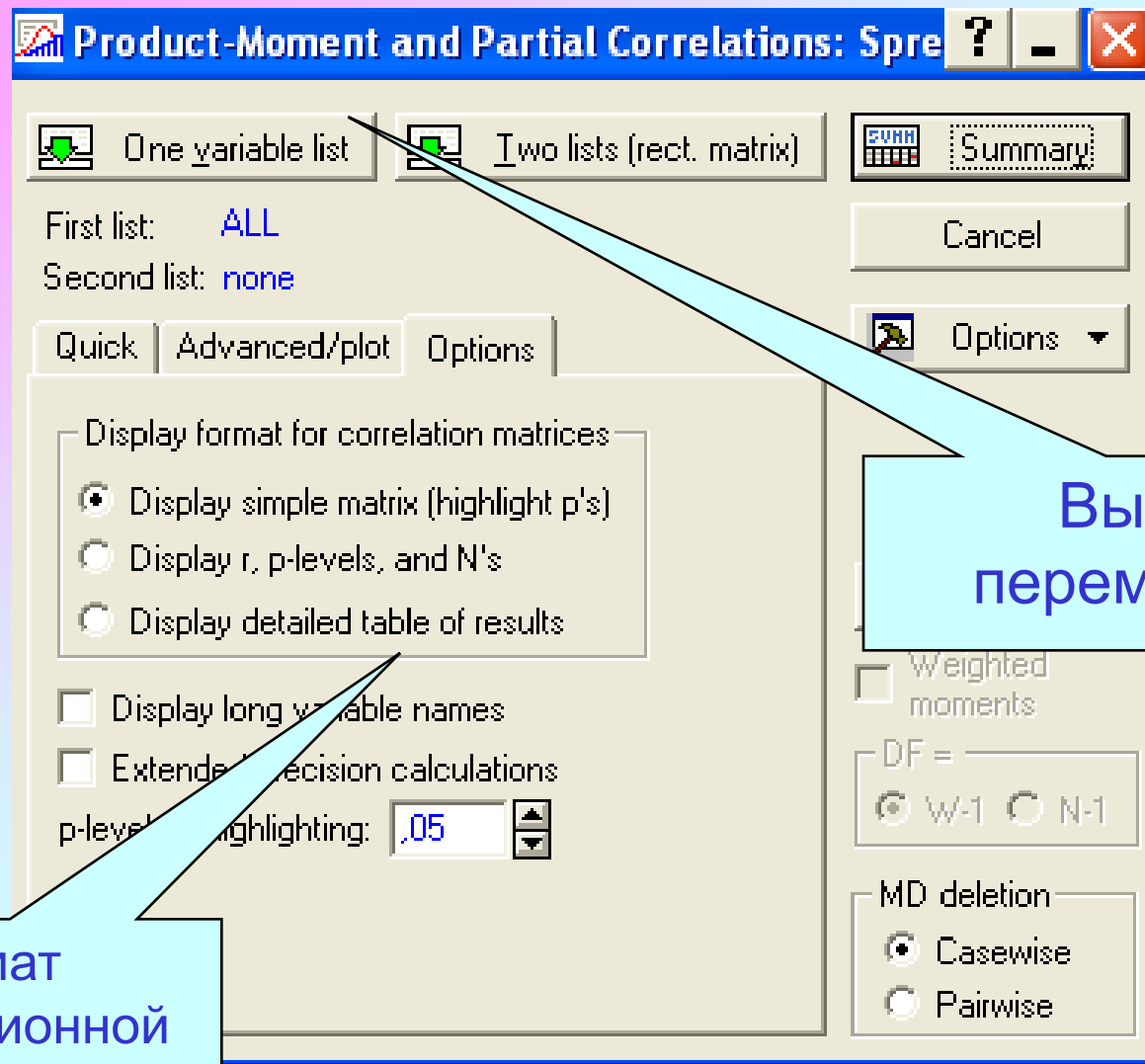
Брак – удельный вес потерь от брака

Корреляция

Коэффициент корреляции

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

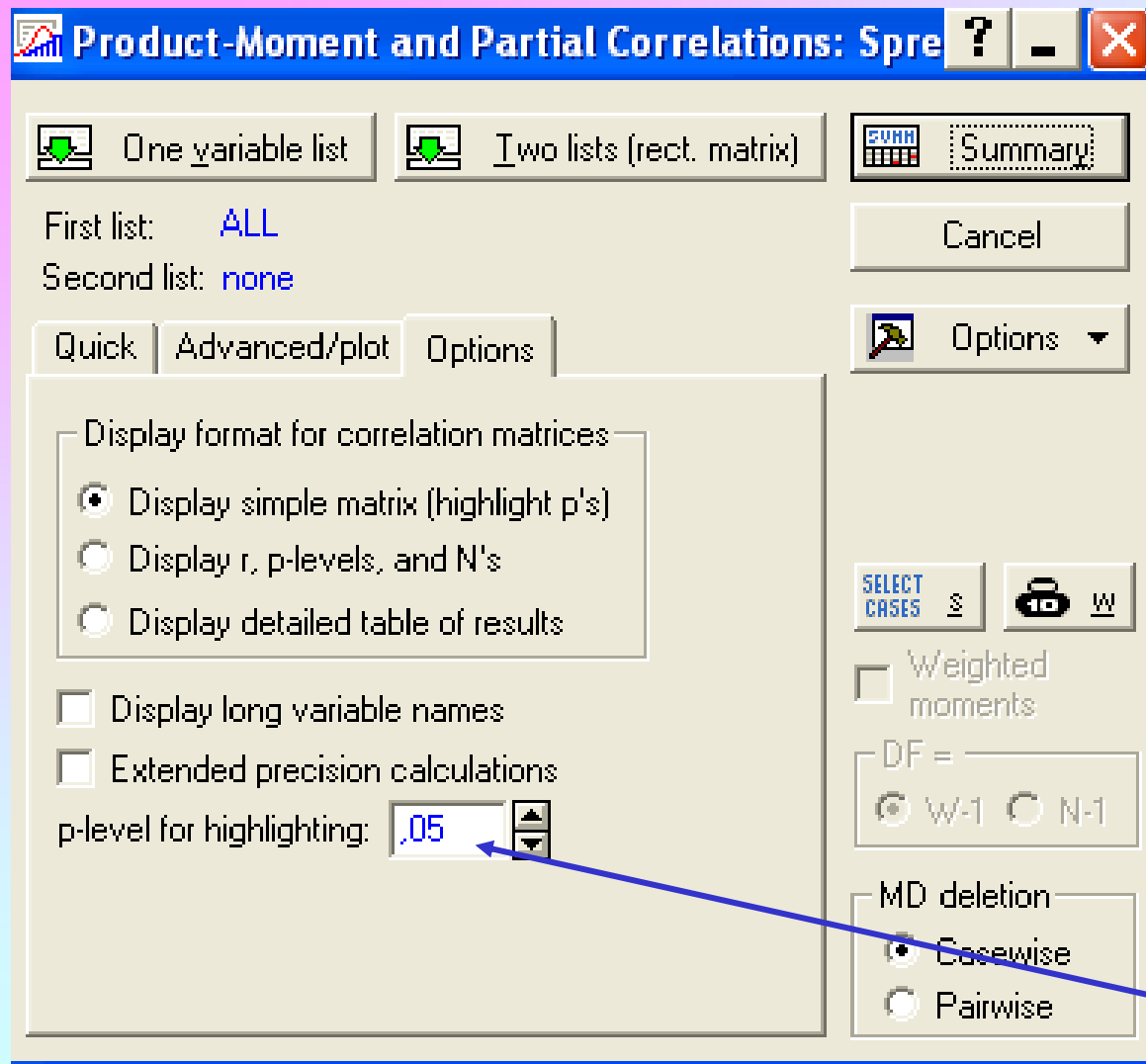
Корреляция



Выбор
переменных

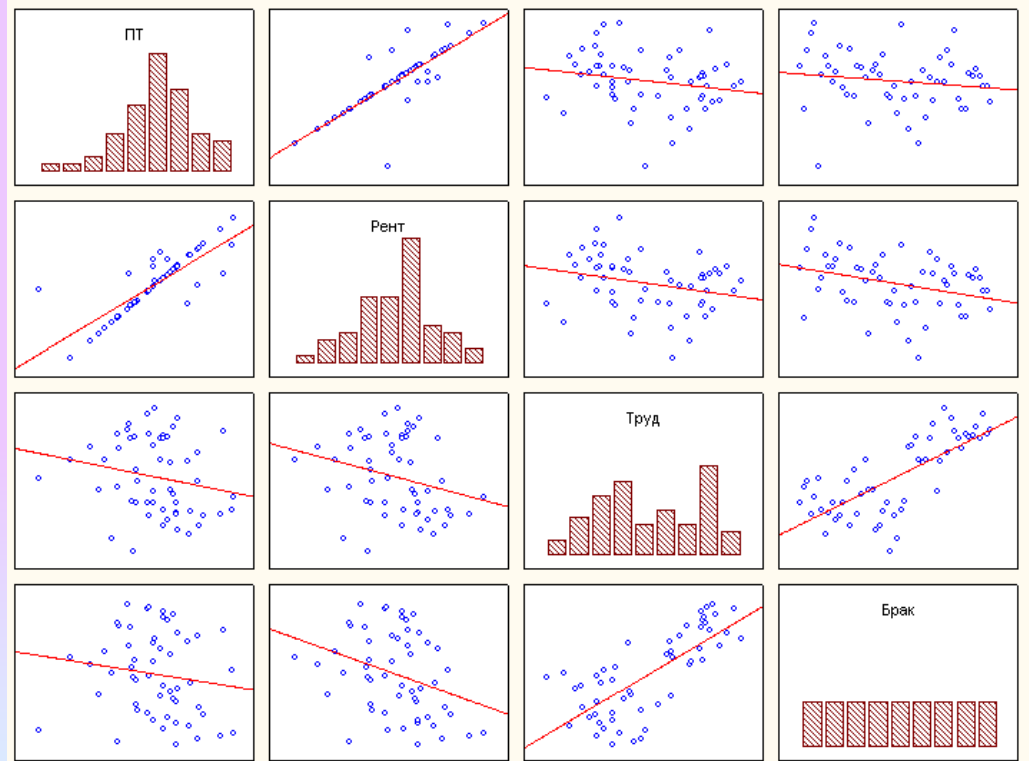
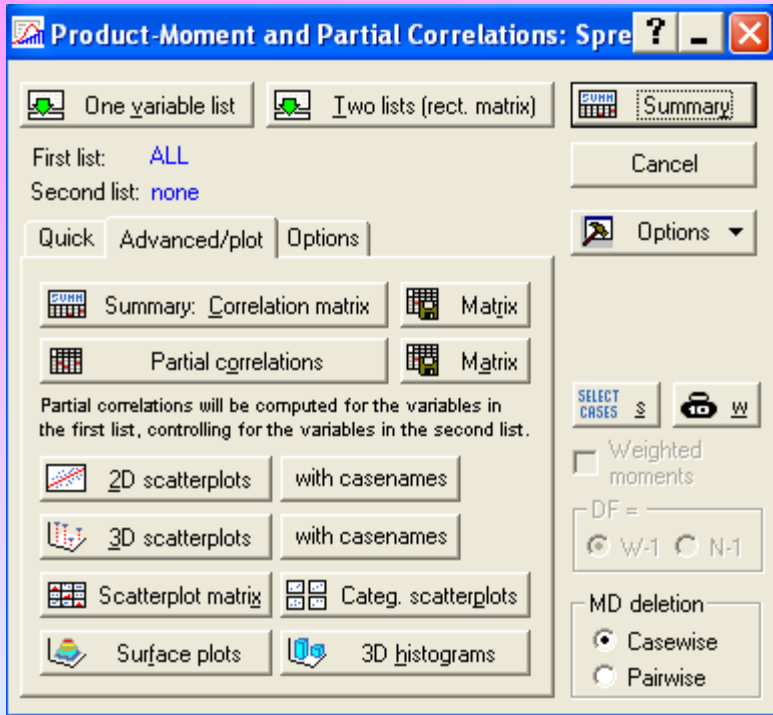
Формат
корреляционной
матрицы

Корреляции: изменение уровня значимости



Вы можете указать уровень значимости (альфа) (0.05, по умолчанию) для выделения значимых коэффициентов корреляции в таблице результатов.

Корреляция



Workbook6* - Correlations (Spreadsheet15)

Correlations (Spreadsheet15)
Marked correlations are significant at $p < .05000$
N=54 (Casewise deletion of missing data)

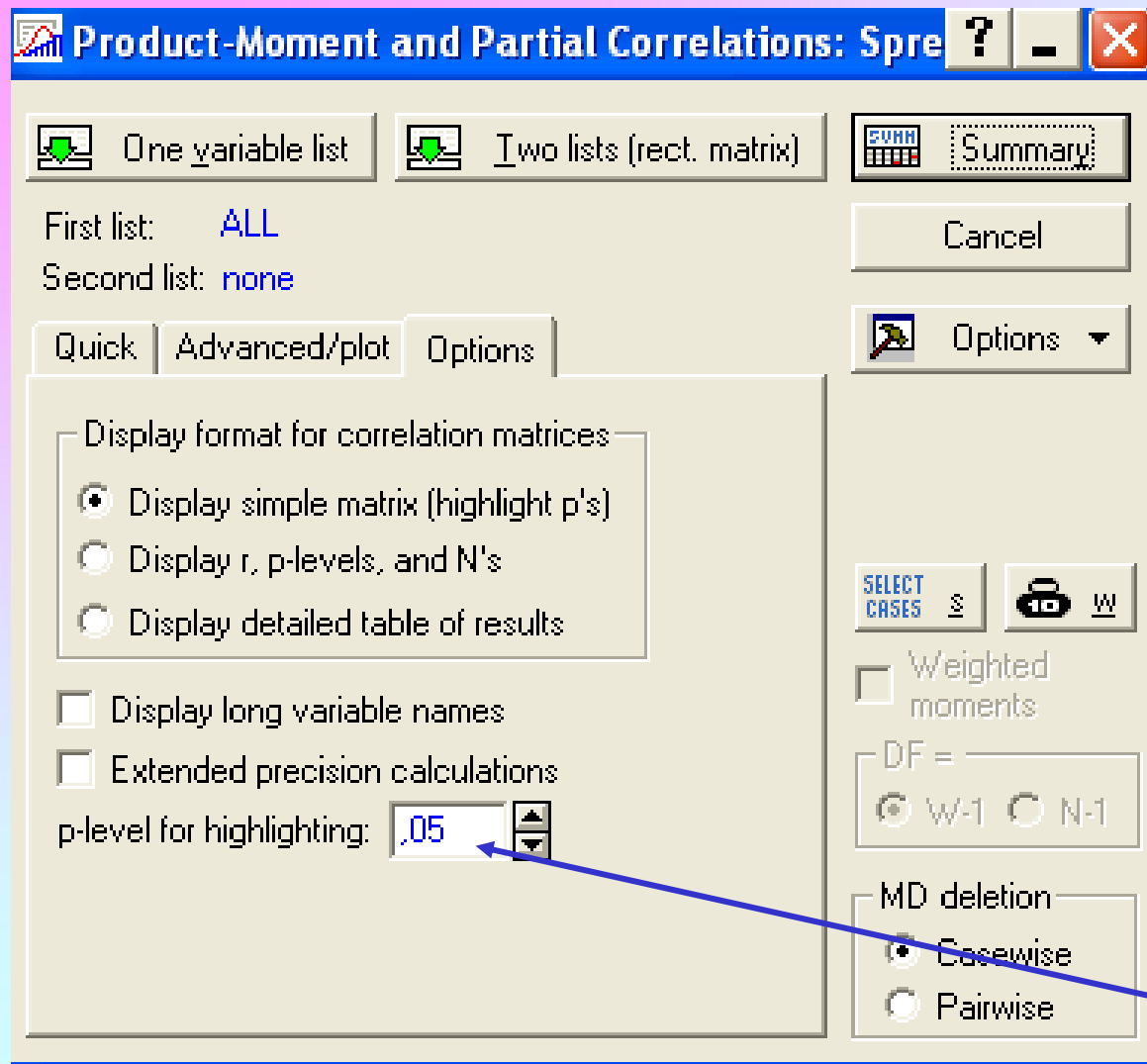
Variable	Рент	Труд	Брак
ПТ	0,82	-0,20	-0,15

«Как только я собираюсь в баню с друзьями, так в стране очередное ЧП случается»
министр МЧС С.К. Шойгу

Корреляция

Correlations (Spreadsheet1)							
Marked correlations are significant at $p < ,05000$							
N=10 (Casewise deletion of missing data)							
Variable	Var1	Var2	Var3	Var4			
Var1	1,00	0,68	-0,38	0,27			
Var2	0,68	1,00	-0,06	0,11			
Var3	-0,38	-0,06	1,00	-0,56			
Var4	0,27	0,11	-0,56	1,00			

Корреляции: изменение уровня значимости



Вы можете указать уровень значимости (альфа) (0.05, по умолчанию) для выделения значимых коэффициентов корреляции в таблице результатов.

Что вам угрожает

- ☹ Ложные корреляции
- ☹ Незначимость коррелированных переменных
- ☹ Несостоятельность
- ☹ Случайность

Регрессионный анализ

Стукач Олег Владимирович

Каф. КИСМ, пр. Ленина, д. 2, оф. 204

 (3822)-701777*2754

tomsk@ieee.org

<http://ieee.tpu.ru/pages/stukach.htm>



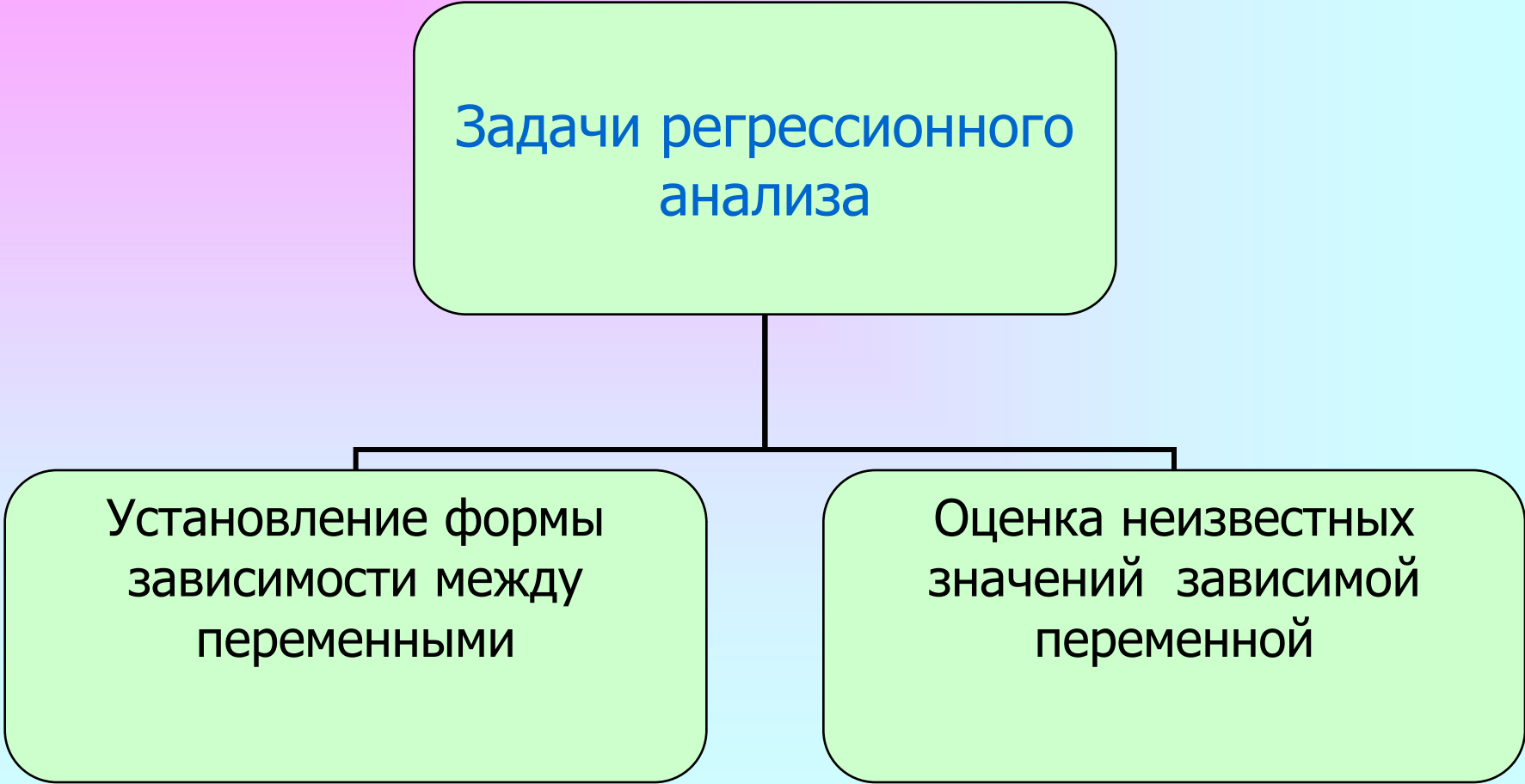
Регрессионный анализ

Во многих случаях исследуются объекты, характеризующиеся несколькими признаками.

Регрессионный анализ – определение характеристик распределения каждого признака и того, насколько эти признаки связаны между собой, можно ли по значению одного признака сделать какие-либо выводы о предполагаемом значении другого.

Задачи регрессионного анализа

Задачи регрессионного
анализа



```
graph TD; A[Задачи регрессионного анализа] --> B[Установление формы зависимости между переменными]; A --> C[Оценка неизвестных значений зависимой переменной];
```

Установление формы
зависимости между
переменными

Оценка неизвестных
значений зависимой
переменной

Пример регрессионного анализа данных

(Дорожный Поверхностный Монитор)

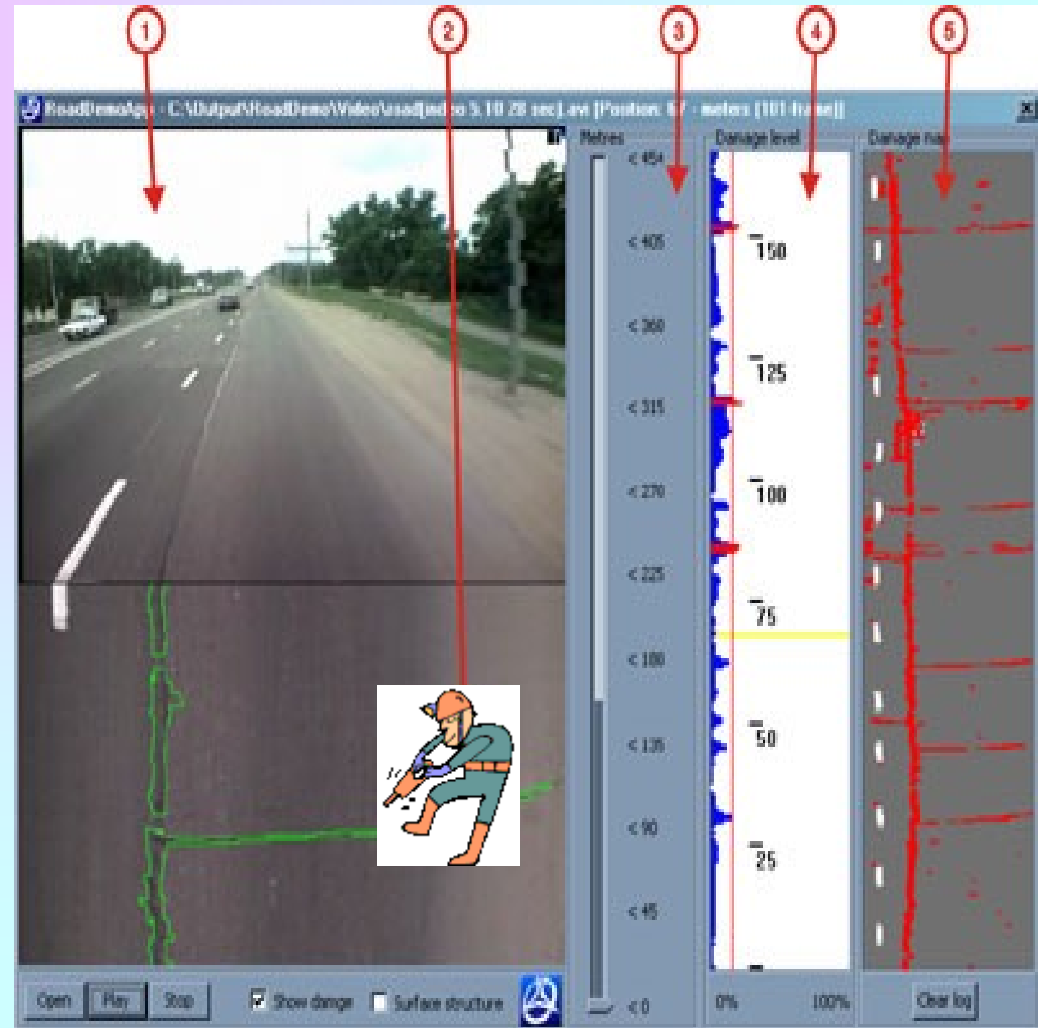
1- исходное видео, снятое видеокamerой, установленной на движущейся машине;

2 - видео в удобном для анализа поверхности ракурсе. В этом окне также могут отображаться зеленым цветом контуры повреждений.

3 - километраж

4 - статистика повреждений дорожного покрытия.

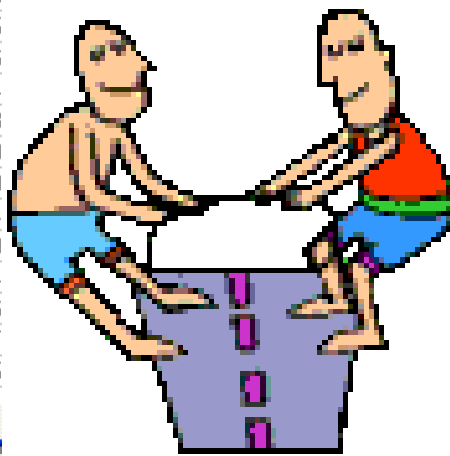
5 - окно, в котором отображается карта повреждений дорожного полотна. Белым цветом обозначается разметка, а красным - контуры повреждений.



Пример регрессионного анализа данных

Data: 3_doroga.sta (4v by 70c)

	1 Повр.	2 T1	3 T2	4 T3
1	108,6499	109,75	6,7	53,94
2	111,9129	94,45	5,1	4,83
3	96,23004	92,19	7,4	23,33
4	95,68411	105,89	6,5	9,13
5	111,5913	109,51	7,8	24,20
6	94,20351	80,12	5,8	30,46
7	109,416	106,56	5,7	4,88
8	108,9486	106,21	3,7	28,66
9	107,1998	105		
10	113,5048	106		
11	94,77887	84		
12	100,0957	103		
13	110,7379	90		
14	96,55905	100		
15	107,925	94		
16	112,379	110		
17	103,6192	94		
18	112,7804	116		
19	90,07166	67		
20	93,00613	85		
21	112,3285	101		

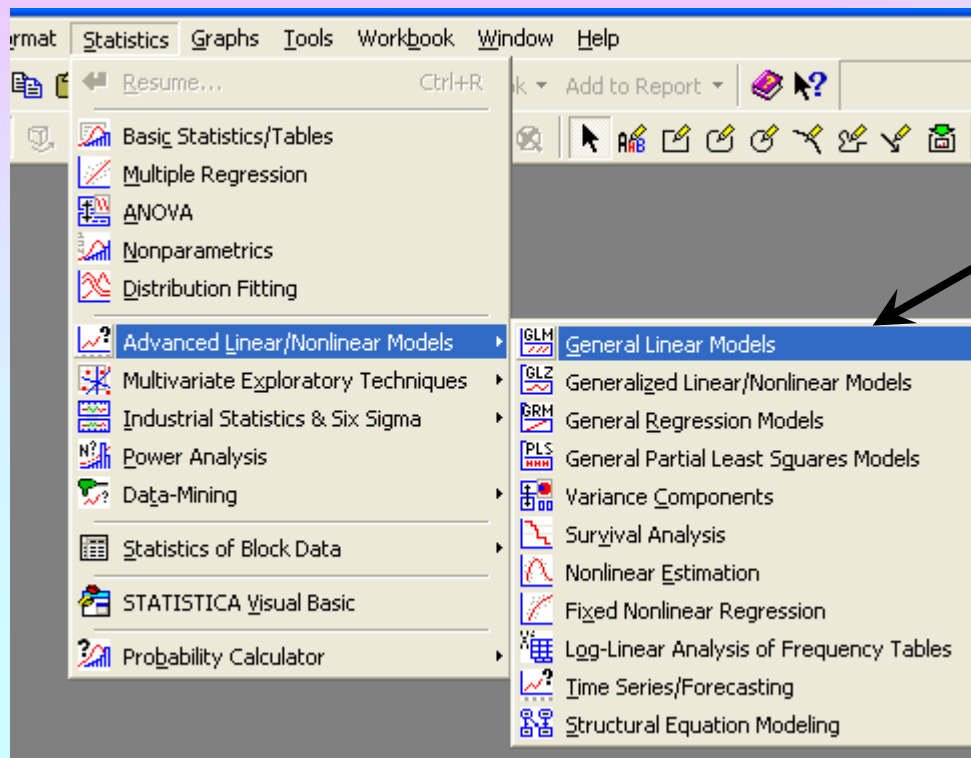


В таблице приведена статистика степени повреждений (Повр.) в зависимости от технологических параметров дорожных работ (T1, T2, T3). Необходимо выразить зависимость между степенью повреждений и этими параметрами.

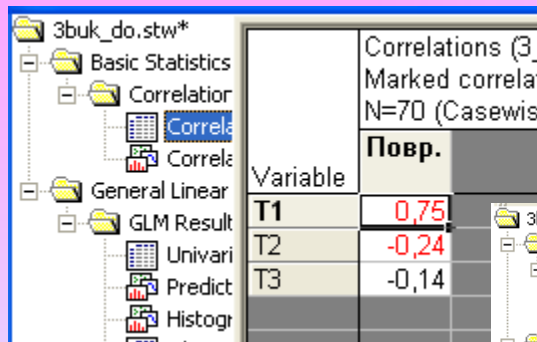
Имея такую формулу, можно прогнозировать качество дорожных работ и сроки очередного ремонта, если технологические параметры изменятся так, что данных для этого случая не окажется в таблице.

Этапы проведения регрессионного анализа в программе STATISTICA

1. Исследование корреляций
2. Графический анализ
3. Использование модуля «Множественная регрессия» (Advanced Linear/Nonlinear Models – General Linear Models)



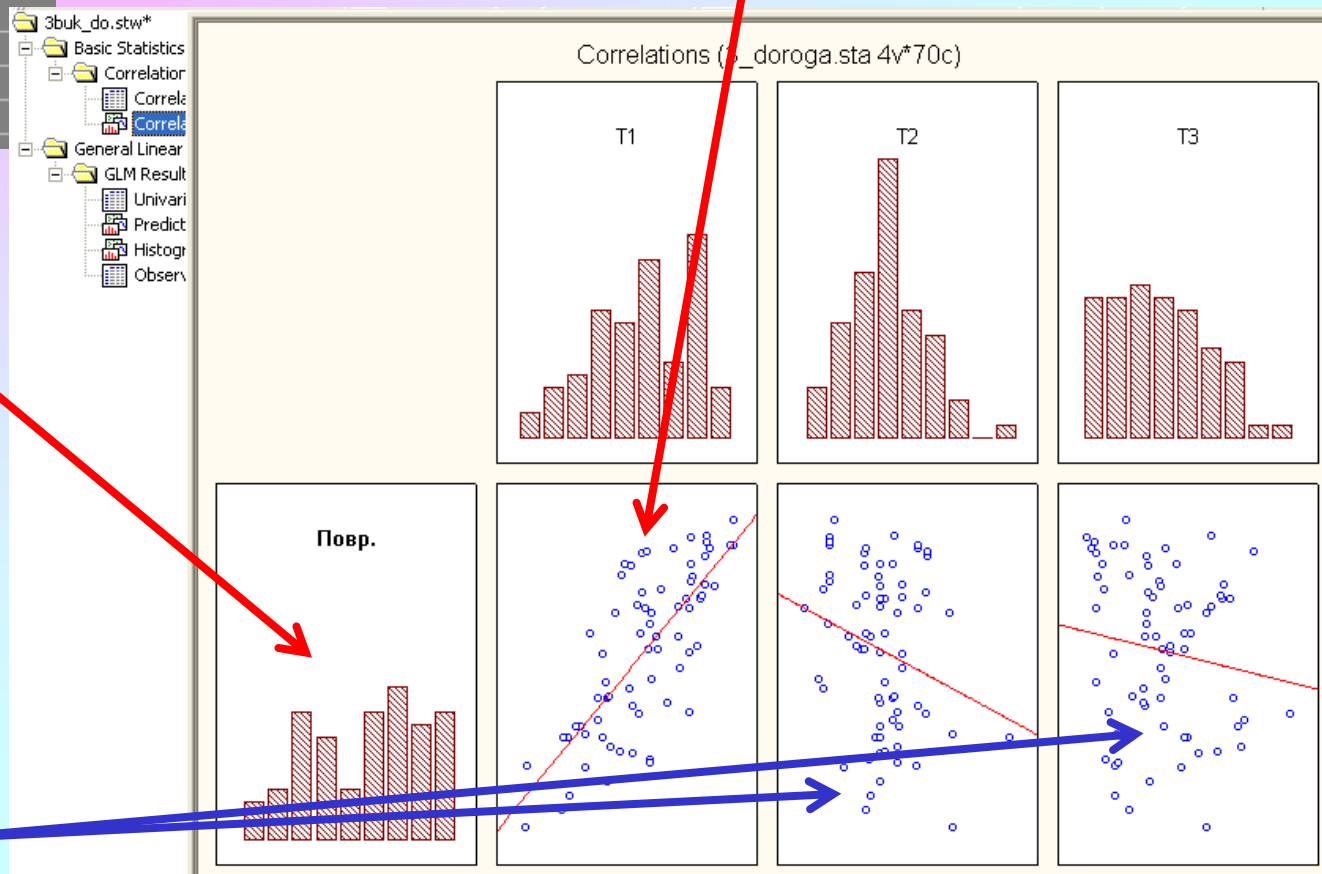
Шаг 1: Исследование корреляционных зависимостей



Сильная корреляция

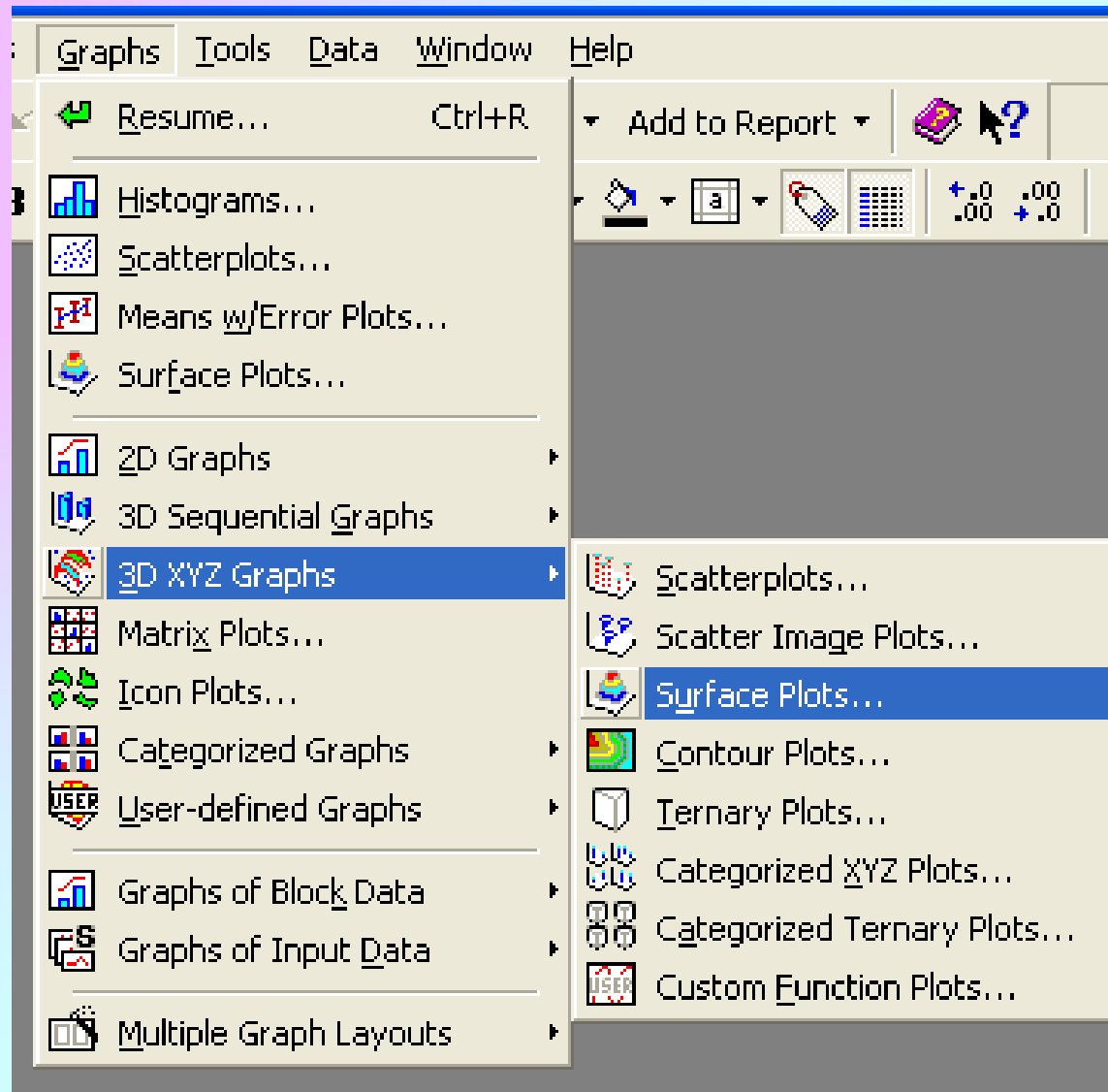
Бимодальное распределение

Слабая корреляция

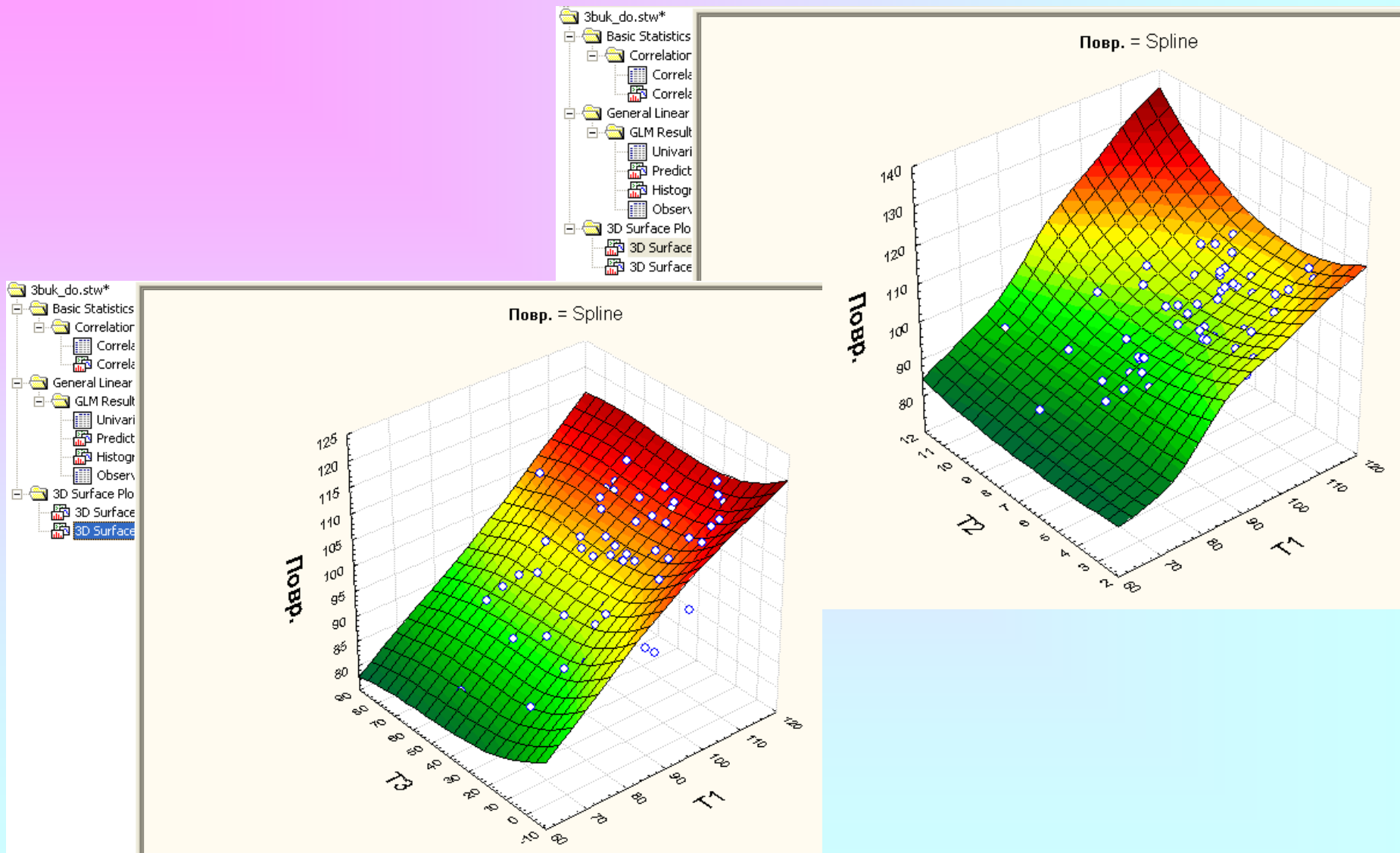


Шаг 2: Графический анализ

Теперь имеет смысл отобразить данные на графике. Для этого необходимо выбрать модуль *Graph – 3D XYZ Graphs – Surface Plots*

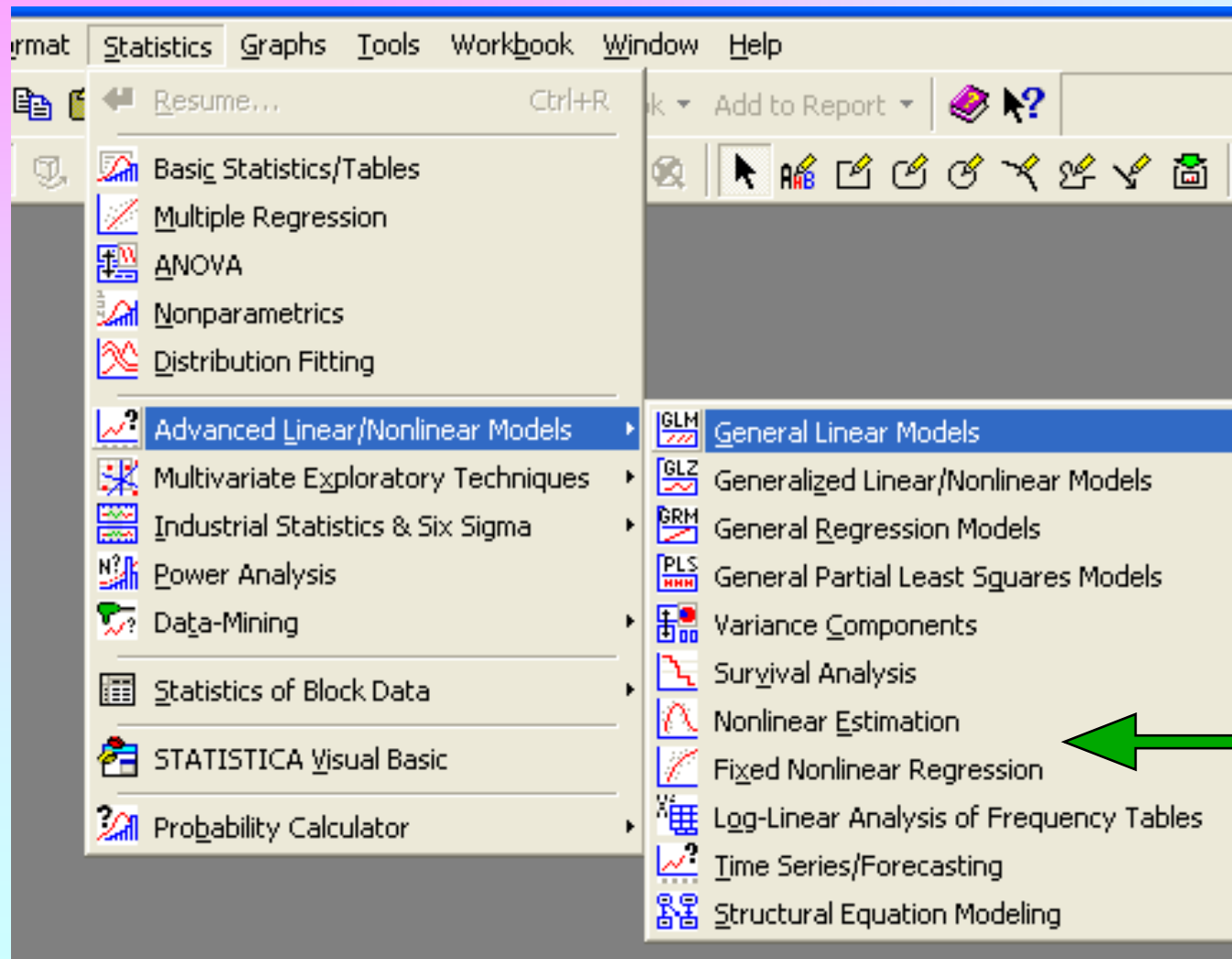


Шаг 2: Графический анализ



У нас 4-мерное пространство, поэтому исследуем 2 трёхмерных графика

Шаг 3: собственно регрессионный анализ



Шаг 3: собственно регрессионный анализ

Математическая постановка задачи

$$\text{Повр.} = \alpha + \beta X + \varepsilon$$

Будем строить линейную модель влияния характеристик техпроцесса на степень повреждений дорог

$$X' = \begin{pmatrix} T1 \\ T2 \\ T3 \end{pmatrix}$$

Вектор-строка независимых переменных

Наша задача – методом наименьших квадратов определить коэффициенты модели α и β .

Целевая функция

Суть метода наименьших квадратов

Рассмотрим модель:

$$Y = \alpha + \beta X + \varepsilon$$

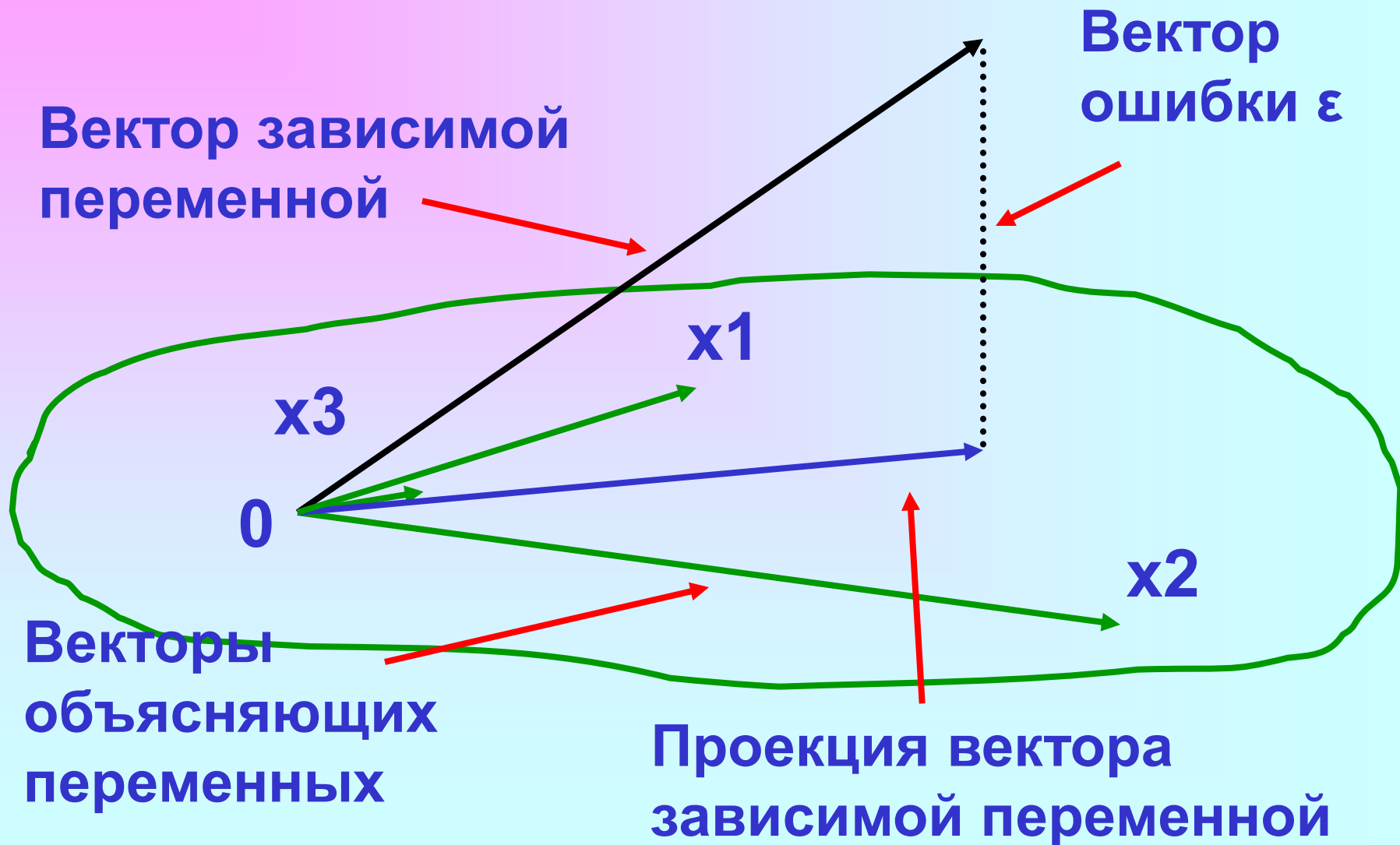
**Ошибка
“белый шум”**



Выбираем коэффициенты α и β такими, чтобы минимизировать следующую функцию:

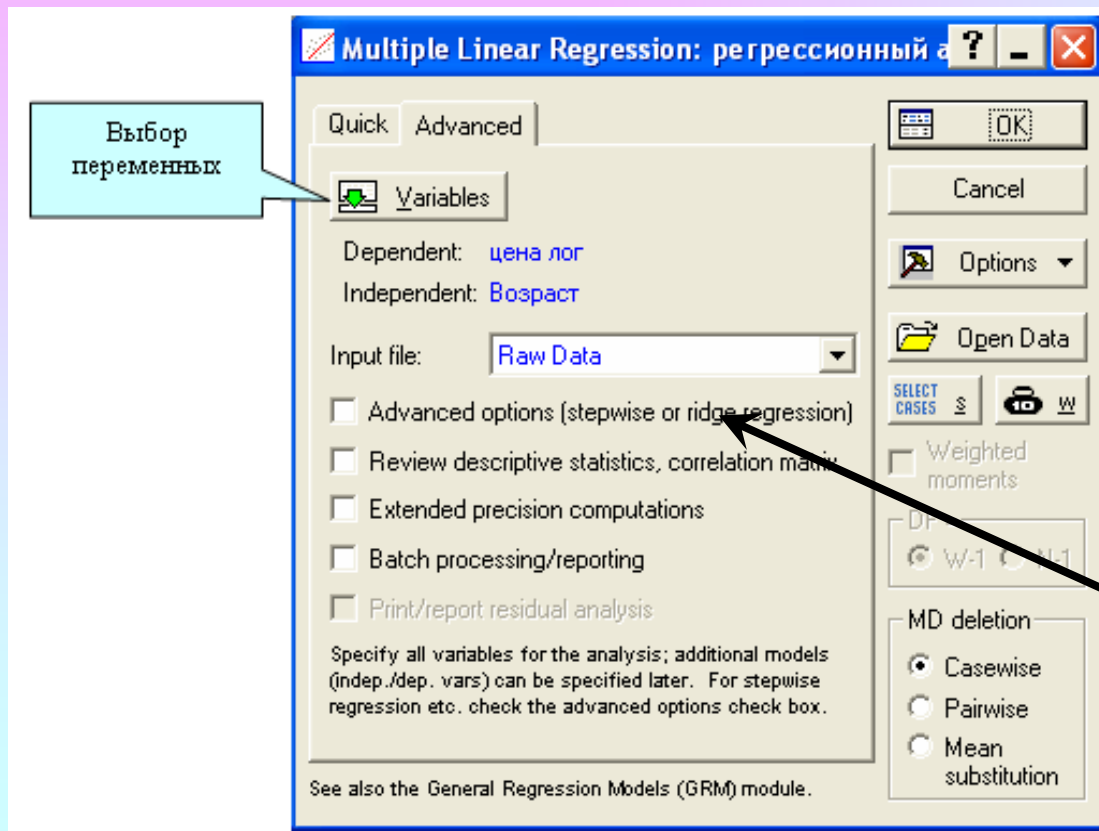
$$e^2 = (Y - \alpha - \beta X)^2 \rightarrow \min$$

Графическая иллюстрация



Собственно регрессионный анализ

Модуль *Statistics/Multiple Regression*



Начинаем анализ

Диалоговое окно
выбора переменных
для анализа

Собственно регрессионный анализ

The screenshot shows the 'Multiple Regression Results' dialog box in SPSS. The window title is 'Multiple Regression Results: 3regres-.sta'. The main area displays regression statistics: Multiple R = .99402816, F = 373,3964, R² = .98809197, df = 2,9, adjusted R² = .98544575, p = .000000, Standard error of estimate = 355,78790992, Intercept = -705,6802151, Std. Error = 221,8025, t(9) = -3,182, p = .0112. Below this, 'Месяц beta = ,063' and 'Работа beta = ,940' are shown, with the latter being highlighted in red. The 'Alpha for highlighting effects' is set to .05. The bottom section contains a grid of buttons for various analyses: Summary: Regression results, ANOVA (Overall goodness of fit), Covariance of coefficients, Current sweep matrix, Partial correlations, Redundancy, Stepwise regression summary, and ANOVA adjusted for mean. There is also an 'Options' dropdown menu.

Multiple Regression Results

Dependent: **Простой** Multiple R = ,99402816 F = 373,3964
R² = ,98809197 df = 2,9
No. of cases: 12 adjusted R² = ,98544575 p = ,000000
Standard error of estimate: 355,78790992
Intercept: -705,6802151 Std. Error: 221,8025 t(9) = -3,182 p = ,0112

Месяц beta = ,063 Работа beta = ,940

(significant betas are highlighted)

Alpha for highlighting effects: .05

Quick | Advanced | Residuals/assumptions/prediction

Summary: Regression results | Partial correlations | Options

ANOVA (Overall goodness of fit) | Redundancy

Covariance of coefficients | Stepwise regression summary

Current sweep matrix | ANOVA adjusted for mean

Информационное поле

Заданный уровень значимости

Исследование остатков

Поле просмотра результатов

Коэффициенты модели

Окно результатов регрессионного анализа

Верхняя часть окна – информационная; нижняя содержит функциональные кнопки, позволяющие всесторонне просмотреть результаты анализа

Коэффициент детерминации

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

$$\hat{Y} = aX + b$$

$$\sum_i (Y_i - \bar{Y})^2 = \sum_i (Y_i - \hat{Y}_i)^2 + \sum_i (\hat{Y}_i - \bar{Y})^2,$$

$$1 = \frac{\sum_i (Y_i - \hat{Y}_i)^2}{\sum_i (Y_i - \bar{Y})^2} + \frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2}.$$

$$Y_i - \hat{Y}_i$$

$$\begin{aligned} r(Y, \hat{Y}) &= \frac{\sum_i (Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})}{\sqrt{\sum_i (Y_i - \bar{Y})^2} \cdot \sqrt{\sum_i (\hat{Y}_i - \bar{Y})^2}} \\ &= \frac{\sum_i (Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})(\hat{Y}_i - \bar{Y})}{\sqrt{\sum_i (Y_i - \bar{Y})^2} \cdot \sqrt{\sum_i (\hat{Y}_i - \bar{Y})^2}} \\ &= \frac{\sum_i [(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) + (\hat{Y}_i - \bar{Y})^2]}{\sqrt{\sum_i (Y_i - \bar{Y})^2} \cdot \sqrt{\sum_i (\hat{Y}_i - \bar{Y})^2}} \\ &= \frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sqrt{\sum_i (Y_i - \bar{Y})^2} \cdot \sqrt{\sum_i (\hat{Y}_i - \bar{Y})^2}} \\ &= \sqrt{\frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2}}. \end{aligned}$$

$$r(Y, \hat{Y})^2 = \frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2}$$

Итоги регрессии

3buk_do.stw* - Univariate Results for Each DV (3_doroga.sta)

Univariate Results for Each DV (3_doroga.sta)						
Sigma-restricted parameterization						
Effective hypothesis decomposition						
GENERAL Effect	Degr. of Freedom	Повр. SS	Повр. MS	Повр. F	Повр. p	
Intercept	1	2011,793	2011,793	66,90085	0,000000	
"T1"	1	2392,255	2392,255	79,55285	0,000000	
"T2"	1	3,579	3,579	0,11900	0,731216	
"T3"	1	29,092	29,092	0,96744	0,328912	
Error	66	1984,703	30,071			
Total	69	4689,409				

Ошибки

Коэффициенты

Предсказательная сила модели

Наша модель: Повр. = 2011,793 +

$$\left(\begin{array}{l} 2392,255 * T1 \\ 3,579 * T2 \\ 29,092 * T3 \end{array} \right)$$

Исследование остатков

Остатки (Residuals) – разность между исходными (наблюдаемыми) значениями зависимой переменной и предсказанными значениями.

Предсказанные значения (модельные) – значения, предсказанные с помощью модели.

Исследуя остатки, можно оценить степень адекватности модели. Вначале для оценки адекватности модели лучше всего использовать визуальные методы и затем, если потребуются, перейти к статистическим методам.

Окно анализа остатков

Residual Analysis: регрессионный анализ.sta

Dependent: ценалог Multiple R : ,96385398
R?: ,92901449
No. of cases: 14 adjusted R?: ,9230999
Standard error of estimate: ,201482
Intercept: 1,143890598 Std. Error: ,115805 t(12) = 10,040 p < ,0000

Quick | Advanced | Residuals | Predicted | **Scatterplots** | Probability plots | Outliers | Save

Summary: Residuals & predicted

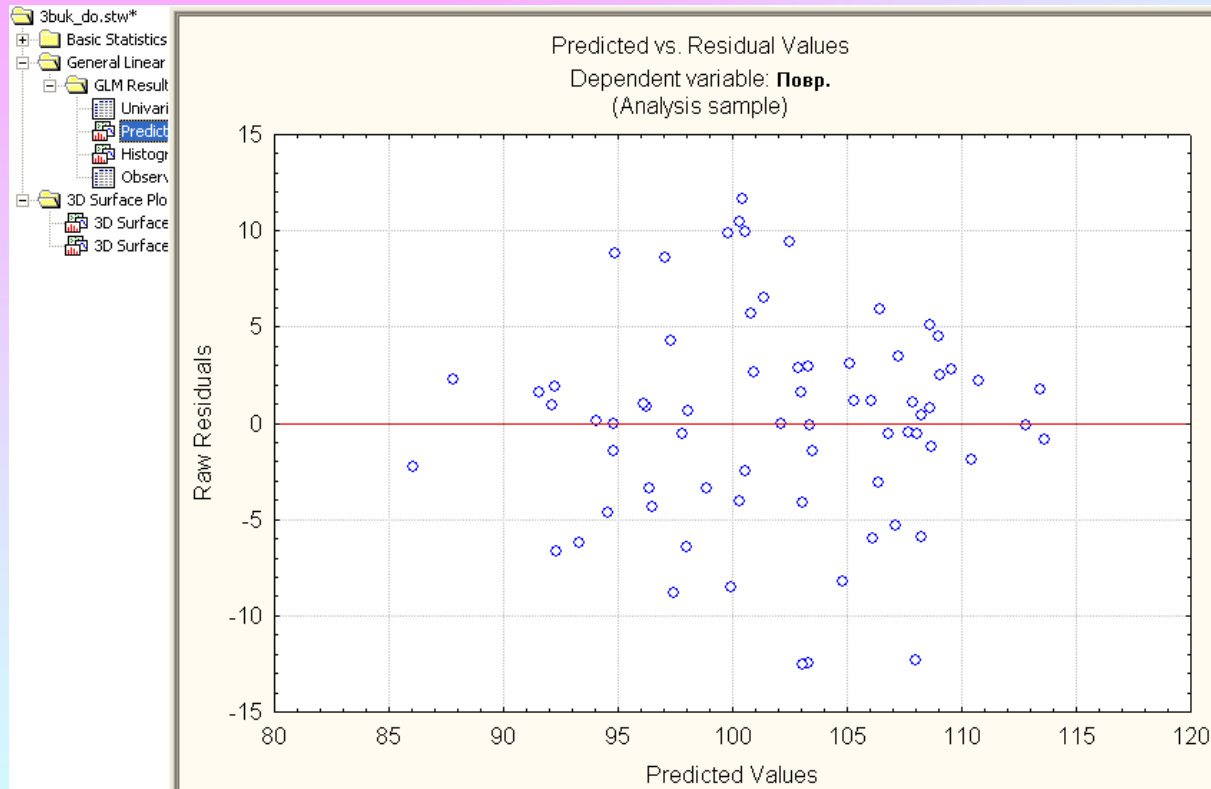
Normal plot of residuals

Summary | Cancel | Options

Построить график остатков в линейной модели зависимости логарифма цены от предсказанных значений

Построить график остатков на нормальной вероятностной бумаге

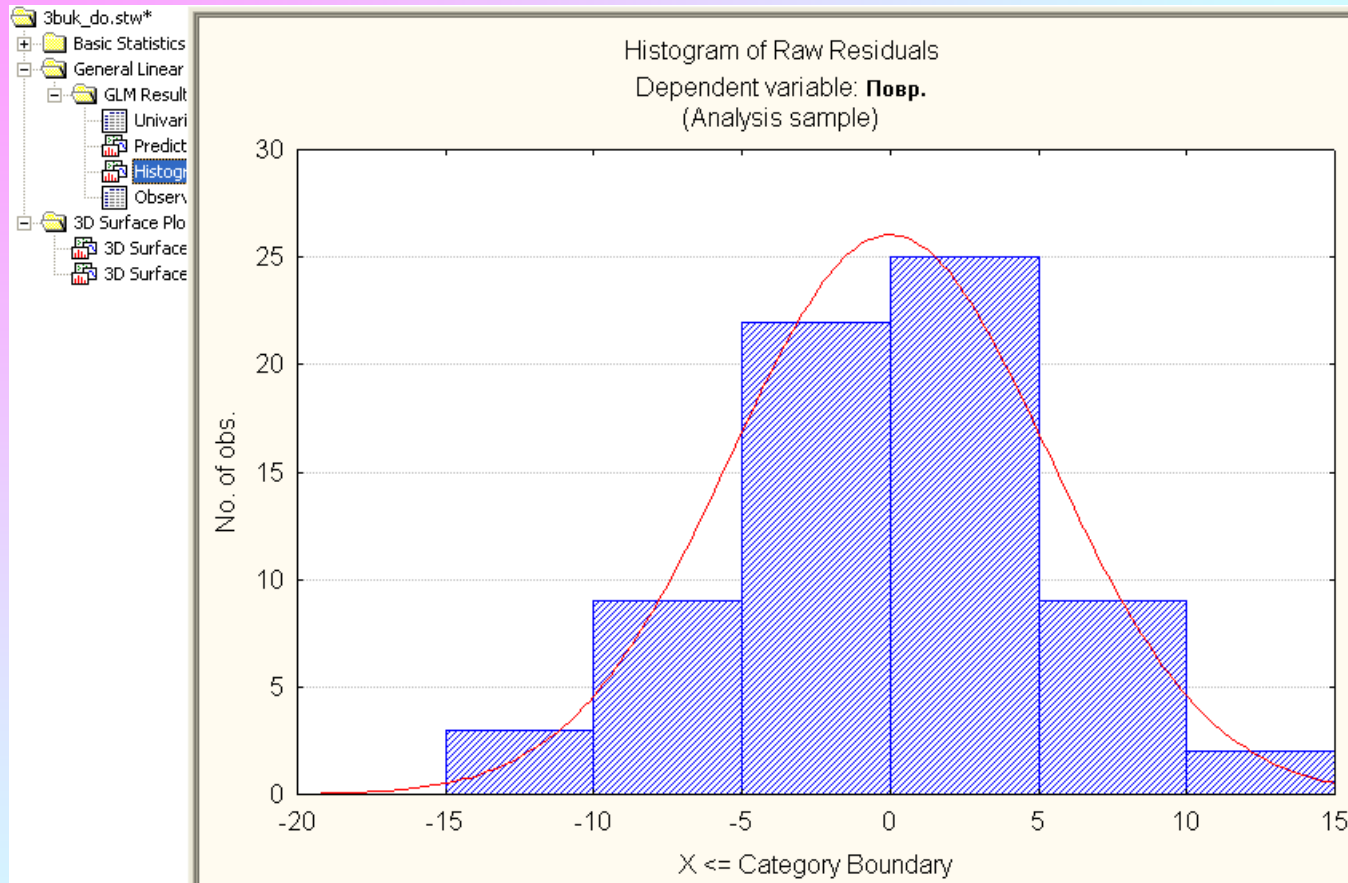
График остатков в линейной модели зависимости степени повреждений от предсказанных значений



Из графика видно, что остатки хаотично разбросаны относительно прямой, в их поведении нет закономерностей. Нет оснований говорить, что остатки связаны между собой, нет также резко выделяющихся остатков.

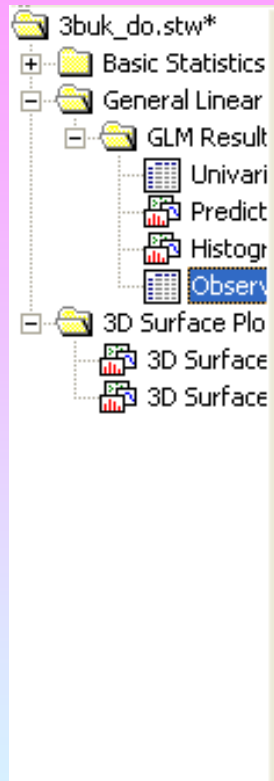
Отсюда можно заключить, что **модель достаточно адекватно описывает данные.**

Гистограмма остатков



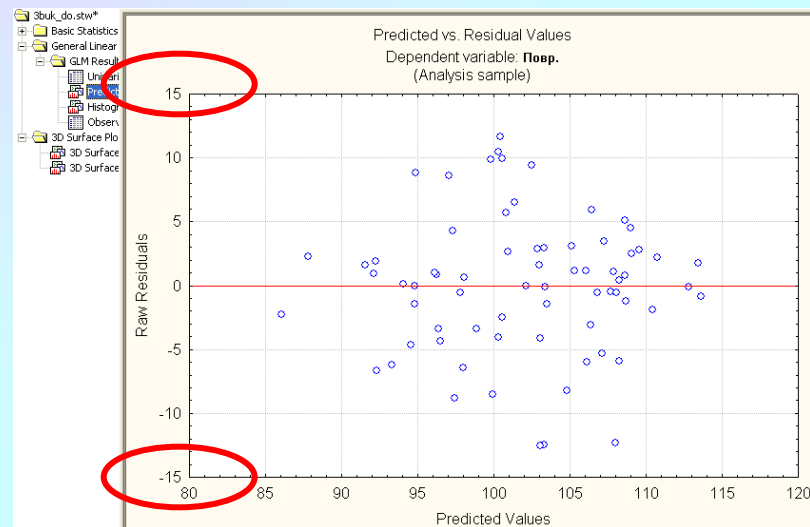
Из гистограммы видно, что остатки достаточно хорошо соответствуют подгоночной кривой нормального закона. Поэтому предположение о нормальном распределении ошибок выполнено.

Таблица реальных значений, модельных значений и остатков



Observed, Predicted, and Residual Values Sigma-restricted parameterization (Analysis sample)			
GENERAL Case number	Повр. Observed	Повр. Predictd	Повр. Resids
1	108,6499	108,2591	0,3908
2	111,9129	102,5055	9,4074
3	96,2300	100,2834	-4,0533
4	95,6841	107,9909	-12,3068
5	111,5913	109,0752	2,5161
6	94,2035	94,0778	0,1257
7	109,4160	108,6229	0,7931
8	108,9486	107,8697	1,0789
9	107,1998	106,0655	1,1343
10	113,5048	109,0239	4,4809
11	94,7789	94,8274	-0,0485
12	100,0957	106,1041	-6,0084
13	110,7379	100,2887	10,4492
14	96,5590	104,8178	-8,2588
15	107,0250	104,2012	2,8238

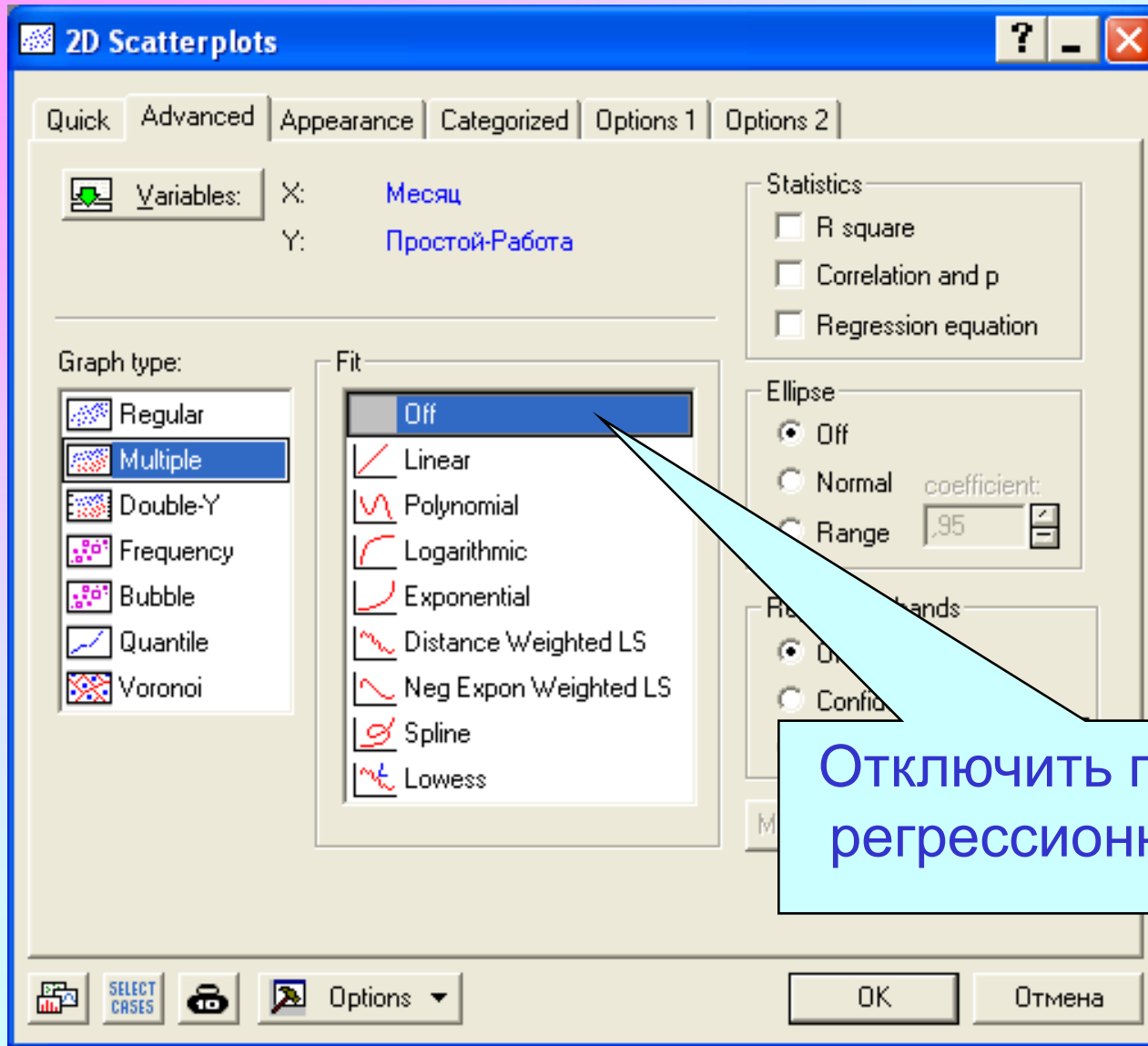
Из таблицы видно, что некоторые остатки все-таки велики. Поэтому линейная модель довольно грубая и требуется построение нелинейной модели.



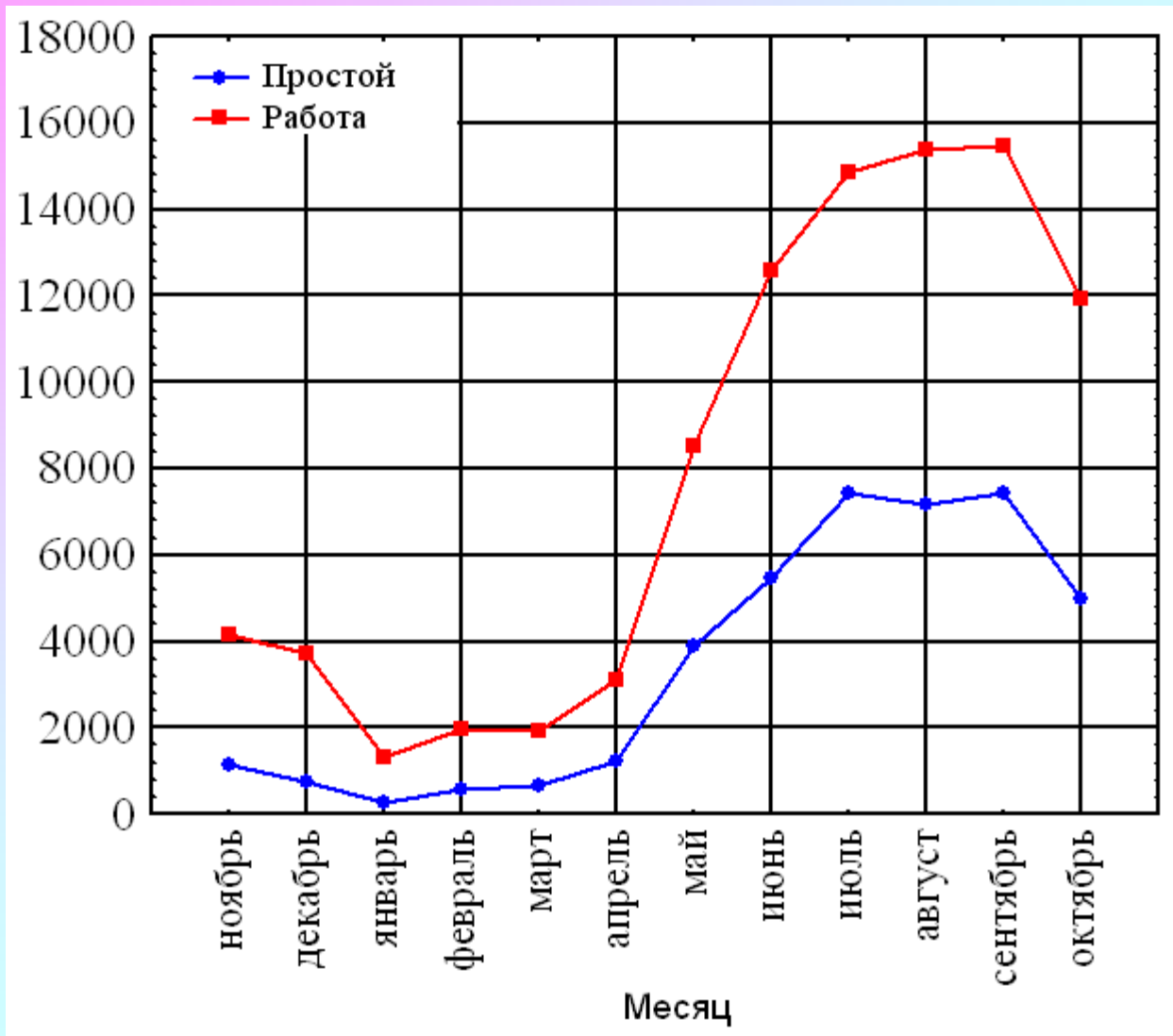
Пример

Месяц	Простой	Работа
ноябрь	1130,01	4137,63
декабрь	734,42	3704,00
январь	265,40	1328,40
февраль	586,60	1961,60
март	666,70	1939,70
апрель	1232,00	3116,00
май	3888,35	8509,35
июнь	5465,39	12588,89
июль	7412,33	14875,50
август	7168,66	15388,08
сентябрь	7416,68	15450,67
октябрь	5001,41	11944,82

Пример



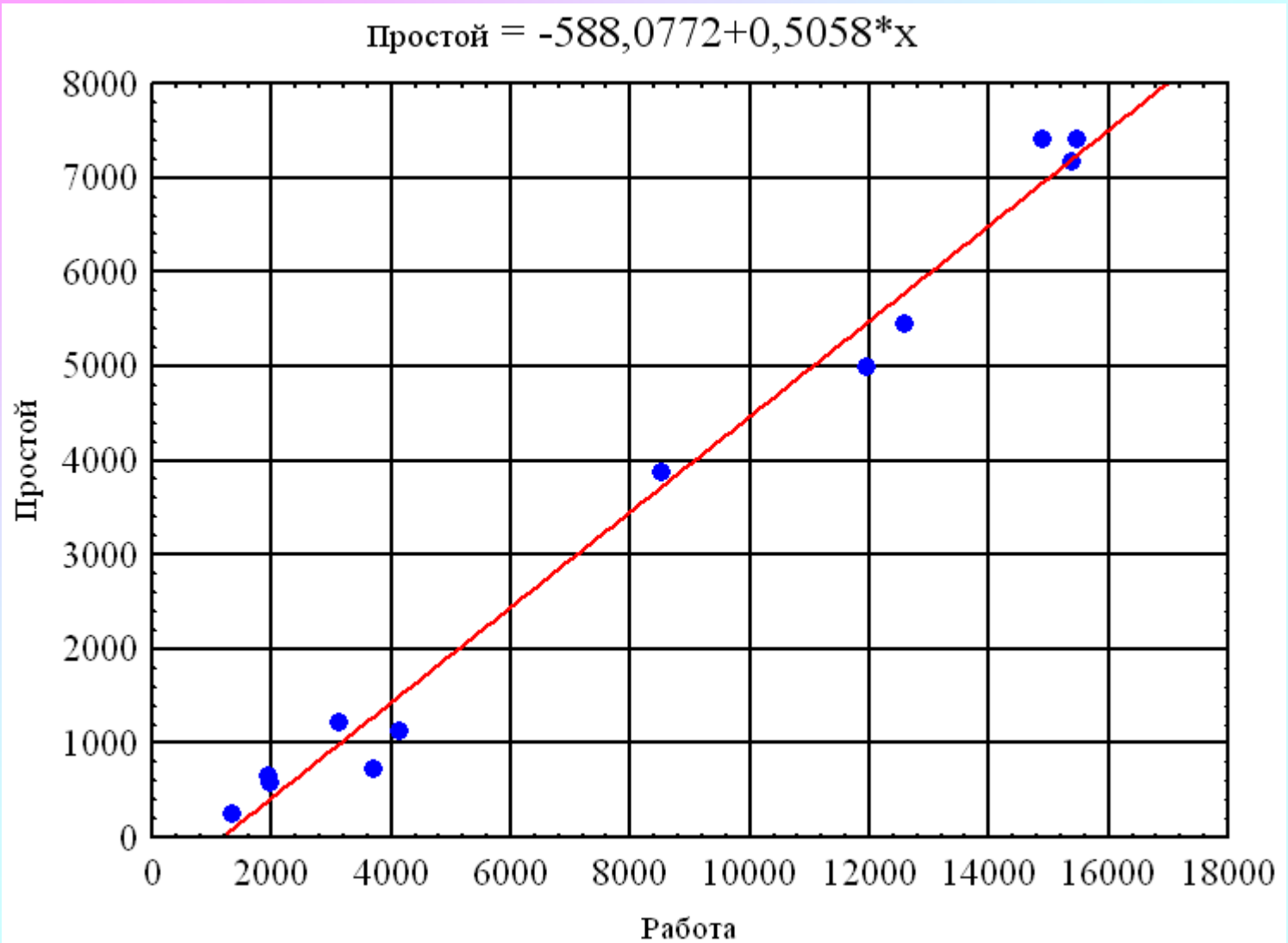
Пример



Пример

	Месяц	Простой	Работа
Месяц	1,00	0,86*	0,85*
Простой	0,86*	1,00	0,99*
Работа	0,85*	0,99*	1,00

Пример



Пример

Multiple Regression Results: 3regres-.sta

Multiple Regression Results

Dependent: **Простой** Multiple R = ,99402816 F = 373,3964
R² = ,98809197 df = 2,9
No. of cases: 12 adjusted R² = ,98544575 p = ,000000
Standard error of estimate: 355,78790992

Intercept: -705,6802151 Std. Error: 221,8025 t(9) = -3,182 p = ,0112

Месяц beta=,063 **Работа beta=,940**

(significant betas are highlighted)

Alpha for highlighting effects: .05

Quick Advanced Residuals/assumptions/prediction

Summary: Regression results Partial correlations
ANOVA (Overall goodness of fit) Redundancy
Covariance of coefficients Stepwise regression summary
Current sweep matrix ANOVA adjusted for mean

OK
Cancel
Options

Пример

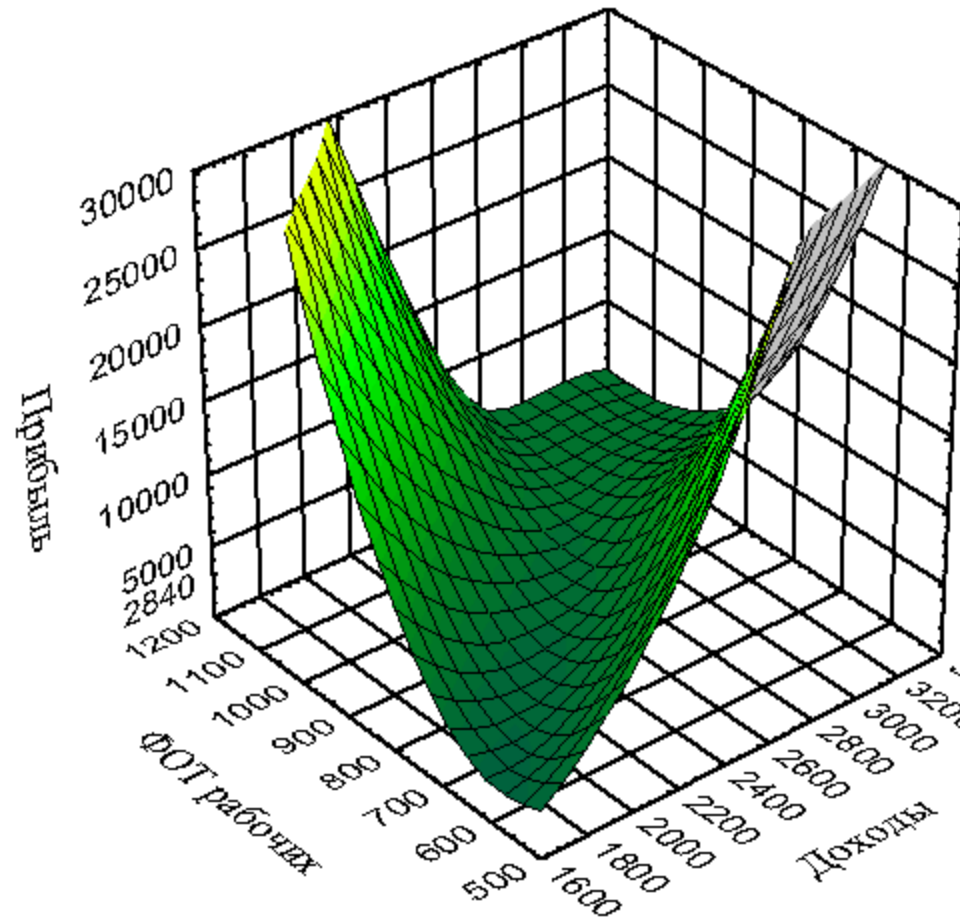
Regression Summary for Dependent Variable: Простой (3regres-.sta)							
R= ,99402816 R ² = ,98809197 Adjusted R ² = ,98544575							
F(2,9)=373,40 p<,00000 Std.Error of estimate: 355,79							
N=12	Beta	Std.Err. of Beta	B	Std.Err. of B	t(9)	p-level	
Intercept			-705,680	221,8025	-3,18157	0,011157	
Месяц	0,062537	0,069694	51,152	57,0059	0,89731	0,392925	
Работа	0,940149	0,069694	0,479	0,0355	13,48967	0,000000	

Фиксированная нелинейная регрессия

Экономические показатели некоторого предприятия

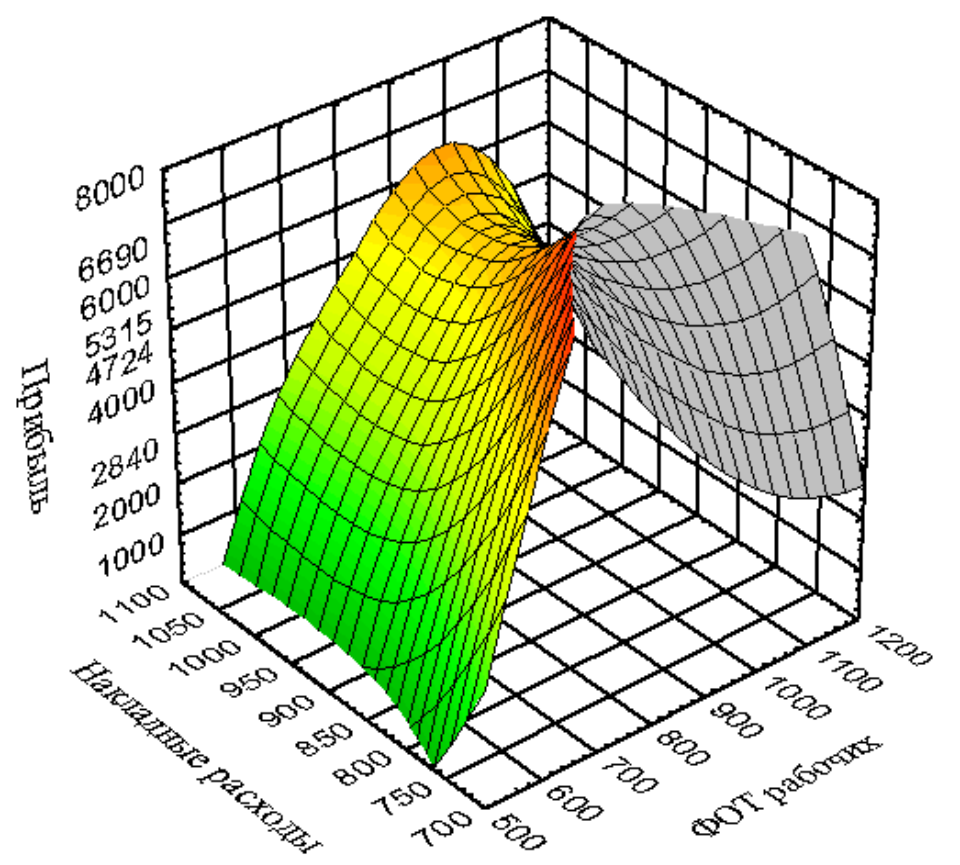
Месяц	Прибыль	Доходы	ФОТ рабочих	Накладные расходы
	y	x_1	x_2	x_3
Январь	2839,6	1675,9	567,9	757,6
Февраль	3354,9	2050,2	696,0	900,9
Март	4302,6	2382,1	795,6	954,7
Апрель	6690,5	2798,3	880,6	896,1
Май	5414,0	2735,3	1012,4	974,6
Июнь	4805,3	2552,7	843,1	1017,4
Июль	5680,1	2987,6	1092,5	1072,3
Август	5315,2	3171,8	1046,9	1054,0
Сентябрь	4724,5	2902,7	1099,3	1070,6

Другие показатели не учитываются, чтобы не усложнять модель



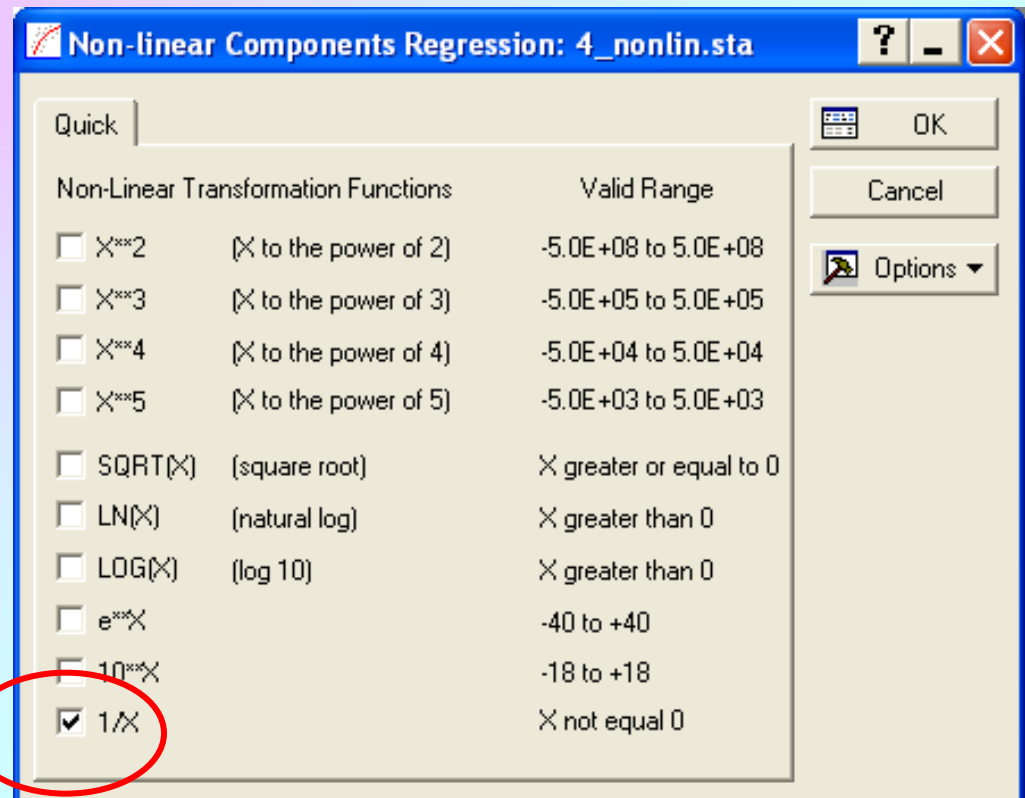
Характерно, что экономисты предприятия планировали прибыль исходя из линейной модели, неадекватность которой была очевидна для них же самих

$$y = b_1 / x_1 + b_2 / x_2 + b_3 / x_3 + \varepsilon$$

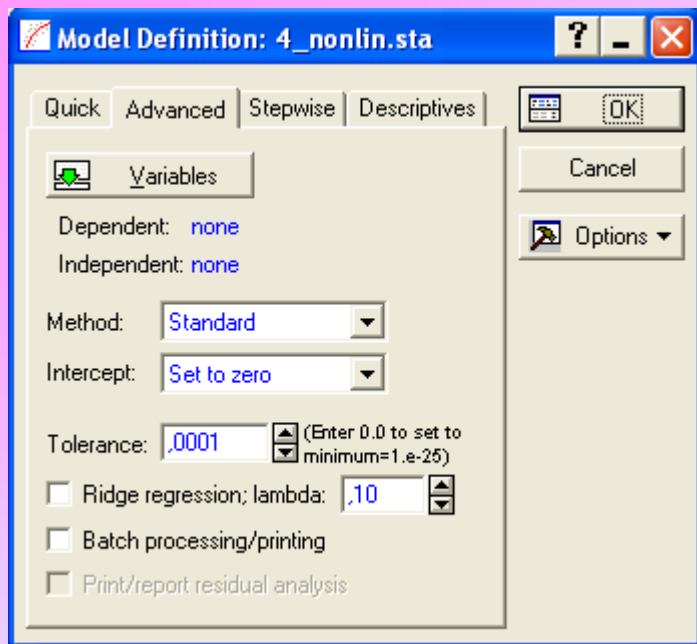


запуск модуля фиксированной регрессии Statistics/ Advanced Linear/Nonlinear Models/Fixed Nonlinear Regression и выбора переменных после нажатия на кнопку ОК в диалоговом окне Non-linear Components Regression

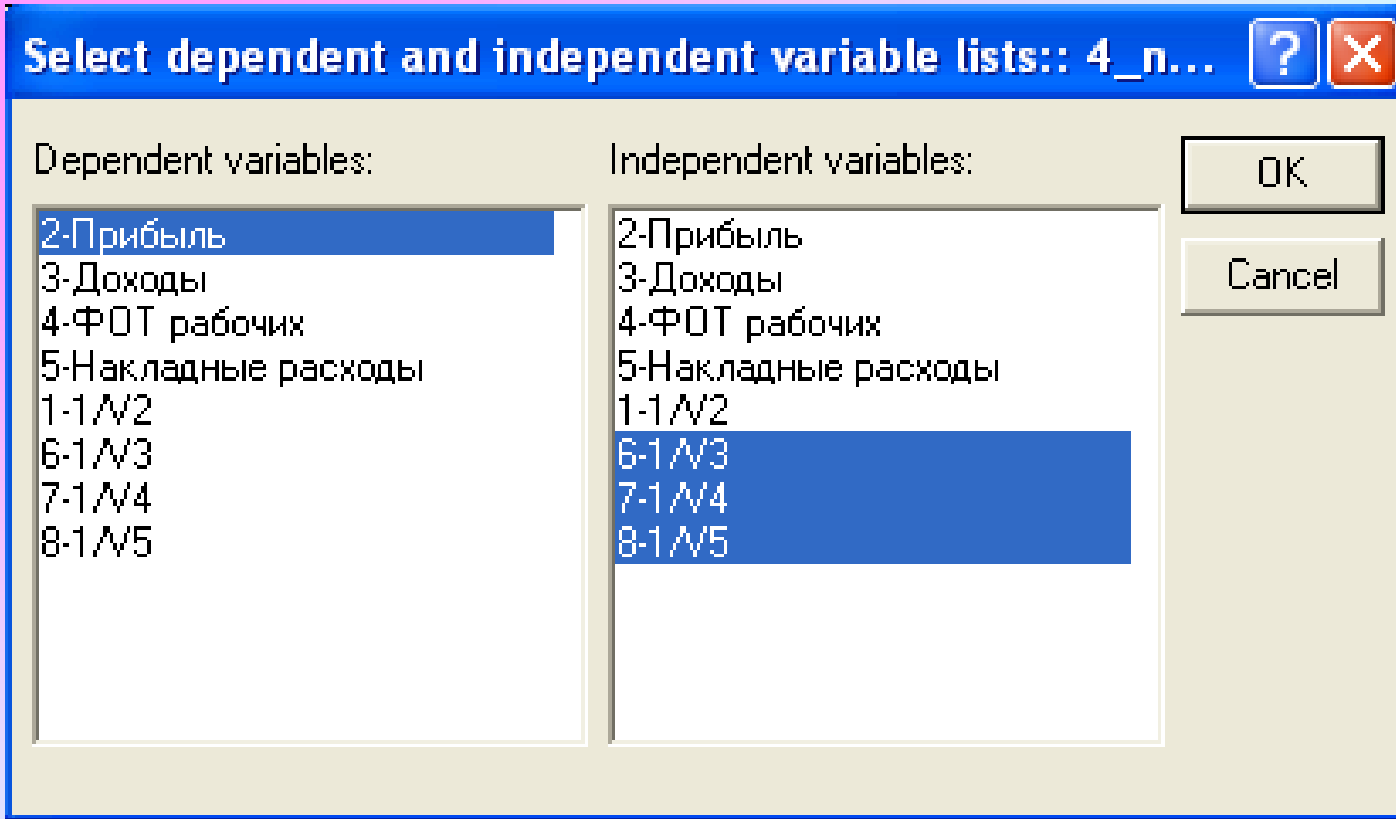
можно выбрать типы преобразования переменных в виде широко распространённых математических функций. Для нашего примера это $1/x$



$$y = b_1 / x_1 + b_2 / x_2 + b_3 / x_3 + \varepsilon$$



необходимо уточнение зависимой и независимых переменных фиксированной нелинейной регрессионной модели. Оно производится на следующем шаге при помощи кнопки **Variables** диалогового окна **Model Definition**. В этом окне установим значение поля **Intercept** на **Set to zero**, что позволит получить регрессионную модель без свободного члена уравнения, то есть $b_0=0$



Зависимой переменной (Dependent variables) в нашем случае будет **Прибыль** у; независимыми (Independent variables) - обратные величины 3, 4 и 5 переменной по списку, то есть $1/x_1$, $1/x_2$ и $1/x_3$

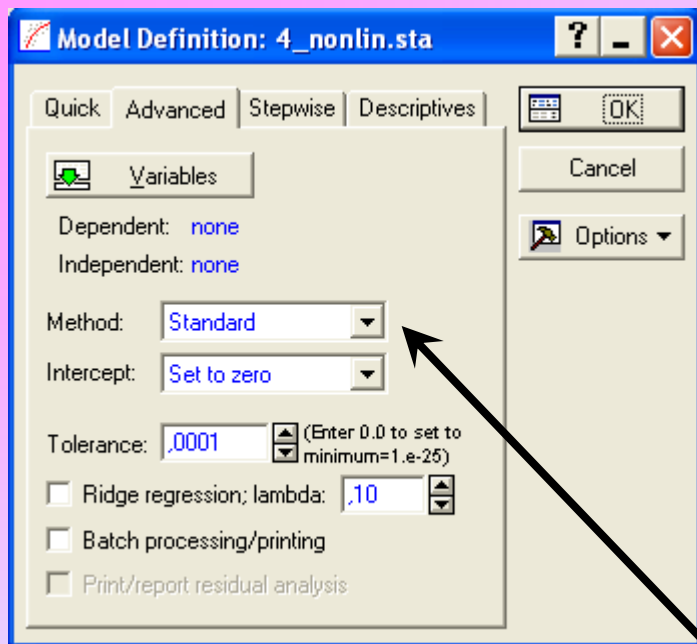
Результат

Ошибка уравнения

Regression Summary for Dependent Variable: Прибыль (4_.sta)						
R= ,99619175 R?= ,99239801 Adjusted R?= ,98859702						
F(3,6)=261,09 p<,00000 Std.Error of estimate: 525,35						
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(6)	p-level
N=9						
1/V3	-1,80743	0,775576	-21671821	9299465	-2,33044	0,058606
1/V4	-0,59693	0,713163	-2444052	2919948	-0,83702	0,434647
1/V5	3,34061	0,300167	15608932	1402524	11,12917	0,000031

$$y = -21671821,4 / x_1 - 2444052,3 / x_2 + 15608932,1 / x_3 + \varepsilon$$

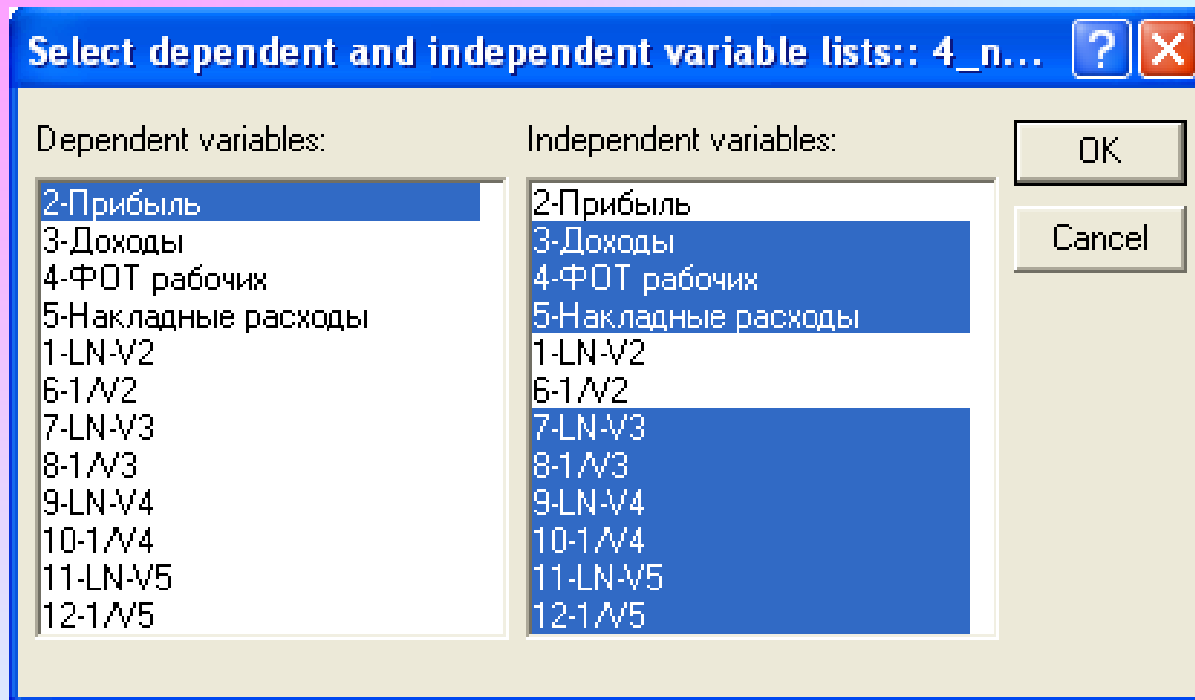
Все коэффициенты уравнения значимы по уровню 0,05. Уравнение объясняет 99,24 % ($R^2 = 0,9924$) вариации зависимой переменной. По анализу остатков можно убедиться в адекватности полученной модели



Пошаговая регрессия

При помощи опции Method пользователь может отказаться от стандартного проведения регрессионного анализа (**Standard**) и воспользоваться методами пошагового включения переменных в регрессионную модель (**Forward stepwise**) или пошагового исключения переменных (**Backward step wise**) из регрессионной модели

Не все переменные из числа указанных останутся в модели



Некоторые закономерности можно найти чисто математическим путём, между тем как непосредственное наблюдение не позволяет установить даже их присутствия

Regression Summary for Dependent Variable: Прибыль						
R= ,96521693 R ² = ,93164372 Adjusted R ² = ,90885829						
F(2,6)=40,888 p<,00032 Std.Error of estimate: 356,81						
N=9	Beta	Std.Err. of Beta	B	Std.Err. of B	t(6)	p-level
Intercept			85001	17684	4,80674	0,002980
1/V3	-1,72306	0,230057	-22705835	3031601	-7,48972	0,000293
LN-V5	-0,98550	0,230057	-10351	2416	-4,28372	0,005184

$$y = 85001,1 - 22705835,2/x_1 - 10350,98 * \ln(x_3) + \varepsilon$$

Согласно модели, прибыль тем больше, чем больше фонд оплаты труда. В уточнённой модели прибыль зависит от дохода и накладных расходов, но не зависит от фонда оплаты труда. Компьютер исключил как незначимую эту переменную, поступив куда умнее руководителя предприятия, вечно экономящего на зарплате

С одной стороны, для получения надёжных прогнозов значений отклика y в модель нужно включать как можно больше независимых переменных. С другой стороны, с увеличением их числа возрастает дисперсия прогноза и увеличиваются затраты, связанные с получением информации о дополнительных переменных, поэтому желательно включать в уравнение как можно меньше переменных